Technische Universität München
Wissenschaftszentrum Weihenstephan für
Ernährung, Landnutzung und Umwelt
Professur Protein Modelling

Structure-based Prediction and Modeling of the Inter-Domain Association Angles of T-Cell Receptors

Thomas Hoffmann

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzende:	Prof. Dr. Aphrodite Kapurniotu
Prüfer der Dissertation:	1. Prof. Dr. Iris Antes
	2. Prof. Dr. Angela Krackhardt

Die Dissertation wurde am 06.06.2016 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 12.07.2016 angenommen.

In Erinnerung an
Papa
und an
Artur

Acknowledgements

I would like to thank Prof. Aphrodite Kapurniotu for accepting the chair position of my defense committee as well as Prof. Iris Antes and Prof. Angela Krackhardt for assessing my thesis.

In Prof. Antes I have found a supervisor, who is at least as fascinated of T-cell receptors as me. She made it possible for me to continue my work on these molecules, was open for new ideas, and supported me with valuable advice, constructive criticism, and all her expert knowledge. Furthermore, she gave me the opportunity to expand my horizon by involving me in several other projects of different biological topics. I am also grateful for her confidence in my skills in planing and managing our high performance compute cluster system, which allowed me to upgrade my expertise in this field.

Prof. Krackhardt enabled me joining the *Graduiertenzentrum Weihenstephan* (GZW) by immediately agreeing to become my mentor. As an expert for T-cells she recommended a lot of beneficial literature and her interest in my work was extremely motivating. Our successful collaboration resulted in a publication embedded in this thesis.

A further fruitful joint work was together with Prof. Heinrich Leonhardt, Carina Frauer and Sebastian Bultmann in the field of epigenetics. I am very happy about the broad resonance on the publication to which this pleasant cooperation led. Additionaly, I thank Hanieh Mirzaei and Dima Kozakov for providing their rigid body optimization code, Stefan Bietz for the provision of a Protoss license, and Antonia Stank for preprocessing MHC structures.

I am very grateful to Antoine Marion, since he contributed with his expertise to my method and he finalized the corresponding manuscript for publication. Especially, Antoine made every effort reading my thesis and suggested many corrections and improvements. Merci.

I would like to thank Guy Georges for the invitation to the *Mini-Symposium on Computational Biology @ Roche Innovation Center Penzberg*, which was a great experience. In this context I am grateful to Christiane Diekmann and Jan Werrstein for hosting me. The GZW funded my journey to the ISMB in Boston, and thus I thank Michael Beisswenger for advising me and for assisting me to complete the application form. I also gratefully look back to the nice time I spent in New York

with Nomsa Buchholz, Marc Tobias Kunisch, and Max. In the same way I thank Jule Veit and Sebastian Häfele for hosting me in Berkeley and for the unforgettable trip through California.

My PhD project was supported from the beginning by the diligent helpers Olli Avieny, Klaus Thelen, Christoph Baur, Manuel Glaser, and Dominik Moser involved in my move to or within Freising.

My work was accompanied by many nice colleagues. I am thankful to our administrative assistents Bianca Riedl and Iwona Kul-Federspiel for their efforts in organizing things. Bianca additionally introduced us to the Munich area and various interesting board games. Our student Sabrina Hecht spent a lot of time in organizing our attendence at the *Weihenstephaner Berglauf*. I thank Manuel Glaser, Matthias Bierschenk, Atanas Patronov, Martin Zachmann, and Katharina Kopp for interesting discussions during coffee breaks. I also appreciate Sabine Schweizer for many useful advices and Dominik Moser for the initial administration of our computers.

I am grateful to the following people for revealing the typos and cryptic sentences I produced for this thesis and in previous publications: Prof. Iris Antes, Matthias Bierschenk, Manuel Glaser, my sisters Sabine and Birgit, Prof. Angela Krackhardt, Iwona Kul-Federspiel, Josy Kübler, Katharina Kopp, Jörg Lippold, Antoine Marion, Atanas Patronov, Katja Schulenburg, Ilke Ugur, Daniel Weiller, and Martin Zachmann. In Boston I was awarded with the *ISMB PDB Poster Prize* for a poster, which I optimized based on many suggestions by Prof. Antes, Ewa Chudyk, Manuel Glaser, Atanas Patronov, and Harpreet Shah *et al.*.

Many people helped me after I broke my leg in the mountains this year during the completion of this thesis and thus I thank the *Bergwacht Schliersee*, Janosch Kunz, Lisa Schubert, Katja Schulenburg, Fabian Härtl, Rebekka, Josy Kübler, Daniel Weiller, Sophie Allendorf, and Resi Endriß. But I am also appreciative of the restful trips to the mountains, to islands, lakes, or to amazing events with Andrea Rütschle, Anselm Rohland, Antoine, Bianca, Christiane, Christoph, Daniel, Fred Gunner, Helene, Ilke, Jan, Janosch, Josy, Johannes Riedl, Katja, Klaus, Manuel, Markus, Max Herrmann, Marcel Baudy, Michael Betz, Olli, Paula, Thomas Schenk, and Volker Menrad. Daniel for selecting nice hiking spots and events, Helene and Markus for teaching me in climbing, and Manuel for a lot of activities. I thank Max H., Volker, my cousin Nicole, and Michael Betz for remote motivation.

Finally I am very grateful to my family and especially to my parents for all the support over many years. The days spent together in Südtirol were motivating and will keep in a very nice memory. I thank my sister Birgit for sending the established *Geistesblitze* and for proofreading the manuscript as well as my sister Sabine, Heiko, and Klara for the relaxing time in Berlin and for many suggestions for my thesis. Special acknowledgements are dedicated to Klara for the skillful programming of the autopilot.

Abstract

T-cell receptors (TCRs) are highly diverse, heterodimeric, and membrane anchored proteins, which allow the vertebrate immune system to distinguish between *self-* and *non-self peptides* presented by *major histocompatibility complex* (MHC) molecules on the surfaces of different cells. Knowledge of the 3D structure of TCR:antigen complexes allows the theoretical investigation of the T-cell signal transduction, the development of vaccines, and the rational optimization of these receptors in the context of adoptive T-cell therapy. However, due to the high variability of TCRs, for the majority of these complexes there exists no experimental structure and thus homology modeling must be applied to build a theoretical 3D model. Two TCR-specific aspects not considered in general homology modeling approaches are the rotation of the TCR variable domains and the varying orientations in which TCRs bind to different MHC:peptide complexes (pMHC).

In the first part of this thesis an analysis method for an efficient investigation of TCR inter-domain angles is presented. The method uses a simplified cuboid-based description of the individual TCR domains and allows for a precise Euler angle measurement of the orientation of the antigen recognizing TCR variable α- and β-domains, Vα and Vβ, despite their high structural and sequential diversity. Within a diverse set of experimental structures of TCR:peptide:MHC (TCRpMHC) complexes differing in the represented TCR clonotypes and pMHC ligands, the domain orientations were compared using a cluster analysis based on the Euler angle distance metric. This analysis revealed that the differences in the TCR inter-domain angles are specific for the different clonotypes. Through a grid-based analysis of all structures a common *center of rotation* (CoR) was localized, which is stabilized by conserved glutamine-glutamine interactions between both chains.

To obtain the necessary experience and background for the development of a force field based prediction method for TCR inter-domain angles on the basis of the above analysis results, a molecular dynamics-based application project was performed in the second part of the thesis. This project consisted of several molecular dynamics simulations of the DNA binding protein domain *Uhrf1:SRA* bound to different DNA substrates. These substrates contained either methylated or hydroxymethylated cytosine bases (5mC or 5hmC). The simulations allowed a molecular explanation of the experimentally observed differences in the binding affinities of 5hmC and 5mC. For this purpose different analyses were performed based on the obtained molecular dynamics trajectories, such as the computation of binding energies, the calculation of conserved hydrogen bond networks, and the identification of conserved water molecules.

In the third part of this thesis a force field based method was developed for the prediction of the orientation of TCR Vα/Vβ domains. The approach was extended to an arbitrary number of domains in order to concurrently predict the relative

orientation of the two variable TCR domains as well as the orientation of the whole TCR with respect to the pMHC complex. The method is implemented in a modular manner based on a rigid body energy minimization approach, allowing for a rigid body optimization of the angle between individual domains. Additionally, due to its modularity, the optimization approach can be easily extended towards the inclusion of additional properties, such as side chain flexibility. The latter is demonstrated by the inclusion of local side chain flexibility for the glutamine residues localized at the CoR into the TCR inter-domain optimization procedure.

The performance of the method was evaluated by its application on two test sets containing 75 TCR structures and 53 TCRpMHC complexes, and yields a rate of correct predictions of 89% and 72%, respectively. Examining single parts of the prediction pipeline reveals two main results: First, the correction and optimization of the glutamine residues at the CoR improves the predictions emphasizing the importance of this conserved position in the TCR. Second, the slightly lower performance of the method applied on complexes containing both the TCR and the pMHC can be explained by the preplacement of the ligand and carries potential for improvement. The remarkable high prediction rates obtained by the new approach demonstrate the suitability and also the relevance of the proposed methodology in the field of TCRpMHC structure prediction. Therefore, future approaches will most likely need to take the association properties of the TCR itself and its docking angle onto the pMHC complexes into account, in order to obtain accurate theoretical models of such complex assemblies. Due to its modular implementation the cuboid method presented in this thesis can readily be combined with currently available homology modeling methods to create such theoretical models. In addition, it can be easily extended for the concurrent optimization of the domain orientations as well as side chain- and backbone conformations in arbitrary proteins.

Zusammenfassung

T-Zell-Rezeptoren (TCRs) sind hochdiverse, heterodimere, membranverankerte Oberflächenmoleküle. Sie erlauben dem Immunsystem der Vertebraten, zwischen *Selbst-* und *Nicht-Selbst-Peptiden* zu unterscheiden, die auf der Oberfläche verschiedener Zellen durch Moleküle des *Haupthistokompatibilitätskomplexes* (MHC) präsentiert werden. Die Kenntnis der 3D-Struktur von TCR:Antigen-Komplexen erlaubt die theoretische Untersuchung der T-Zell-Signaltransduktion, die Impfstoffentwicklung und die gezielte Optimierung des Rezeptors für die T-Zell-Therapie. Da aber aufgrund der hohen Variabilität der TCRs nur für einen geringen Teil dieser Komplexe experimentell aufgeklärte Strukturen existieren, ist die Anwendung von Homologiemodellierung zur Erzeugung theoretischer 3D-Modelle unabdingbar. Zwei TCR-spezifische Aspekte werden in allgemeinen Homologiemodellierungsverfahren nicht berücksichtigt, nämlich die Rotation der variablen TCR-Domänen und die verschiedenen Ausrichtungen, mit denen ein TCR einen MHC:Peptid-Komplex (pMHC) bindet.
Im ersten Teil dieser Dissertation wird eine Analysemethode zur effizienten Untersuchung der TCR-Interdomänenwinkel präsentiert. Die simplifizierte Repräsentation der einzelnen TCR-Domänen durch Kuboide erlaubt eine genaue Eulerwinkelmessung der variablen, antigenerkennenden TCR-Domänen (Vα und Vβ), trotz deren hoher struktureller und sequentieller Diversität. Innerhalb einer Menge verschiedener experimenteller TCR:Peptid:MHC-Strukturen (TCRpMHC-), die sich in den darin enthaltenen TCR-Klonotypen und den pMHC-Liganden unterscheiden, wurden die Domänenaurichtungen durch ein Clusteranalyseverfahren anhand der Eulerwinkeldistanzmetrik verglichen. Die Analyse zeigt, dass unterschiedliche Interdomänenwinkel spezifisch für verschiedene TCR-Klonotypen sind. Mit Hilfe einer Gitteranalyse aller Strukturen wurde ein gemeinsamer Drehpunkt (CoR) der variablen Domänen lokalisiert, der durch Wechselwirkungen aus beiden Ketten stammender, konservierter Glutaminresiduen stabilisiert wird.

Um das notwendige Hintergrundwissen für die Entwicklung einer kraftfeldbasierten Vorhersagemethode für TCR-Interdomänenwinkel auf Grundlage der obigen Analyse zu sammeln, wurde für den zweiten Teil der Dissertation ein Anwendungsprojekt aus dem Bereich der Moleküldynamik durchgeführt. Dieses Projekt bestand aus mehreren Moleküldynamiksimulationen der DNS-bindenden Proteindomäne *Uhrf1:SRA* im Komplex mit verschiedenen DNS-Substraten. Diese Substrate enthielten entweder methyliertes oder hydroxymethyliertes Cytosin (5mC oder 5hmC). Die Simulationen erlaubten, die experimentell beobachteten Unterschiede der Bindeaffinitäten von 5mC und 5hmC in molekularer Hinsicht zu erklären. Hierfür wurden verschiedene Analysen mit den aus den Moleküldynamiksimulationen gewonnen Trajektorien durchgeführt, wie etwa die Berechnung von Bindeenergien,

die Bestimmung konservierter Wasserstoffbrückennetzwerke und die Identifikation von konservierten Wassermolekülen.

Im dritten Teil dieser Dissertation wurde eine kraftfeldbasierte Methode zur Vorhersage der Orientierung von TCR Vα/Vβ-Domänen entwickelt. Das Verfahren wurde dann für eine beliebige Anzahl von Domänen erweitert, um gleichzeitig die Stellung der beiden variablen TCR Domänen bezüglich zum pMHC-Liganden im gesamten TCRpMHC-Komplex vorherzusagen. Kern der Implementierung ist ein modularer Energieminimierer, der eine Festkörperoptimierung der Domänen durchführt. Aufgrund seiner Modularität kann das Optimierungsverfahren leicht um zusätzliche Eigenschaften des Komplexes erweitert werden, wie etwa Seitenkettenflexibilität. Letzteres wird anhand der Hinzunahme lokaler Seitenkettenflexibilität der am Drehpunkt lokalisierten Glutaminresiduen zur Optimierung demonstriert.

Die Leistungsfähigkeit der Methode wurde zum einen für einen Testdatensatz von 75 TCR-Komplexen und zum anderen für 53 TCRpMHC-Komplexe ermittelt, wobei 89% der TCR- und 72% der TCRpMHC-Orientierungen korrekt vorhergesagt wurden. Eine Analyse einzelner Teile der Methode zeigte auf, dass erstens die Korrektur der Glutaminausrichtungen am Drehpunkt die Vorhersageergebnisse verbessert und unterstreicht damit auch die Wichtigkeit dieser in der Analyse aufgedeckten Eigenschaft der TCRs. Zweitens lässt sich die etwas geringere Leistung der Methode bei der Anwendung auf die TCRpMHC-Komplexe gegenüber den ungebundenen TCRs mit der Vorplatzierung des pMHCs erklären, bei der noch Verbesserungsmöglichkeiten bestehen. Die bemerkenswert hohen, mit der Methode erzielten Vorhersageraten zeigen die Eignung und auch die Bedeutung des hier vorgeschlagenen neuen Verfahrens für die Vorhersage von TCRpMHC-Strukturen. Daher werden zukünftige Modellierungsansätze, um akkuratere theoretische Modelle zu erhalten, wahrscheinlich die Assoziationseigenschaften der TCR-Domänen sowie den Bindewinkel des pMHC-Liganden berücksichtigen müssen. Die für diese Arbeit implementierte kuboidbasierte und modulare Vorhersagemethode kann mit vorhandnen Homologiemodellierungsverfahren kombiniert werden, um damit derartige theoretische Modelle zu erzeugen. Außerdem kann die Methode leicht derart erweitert werden, dass eine Optimierung von Domänenorientierungen gleichzeitig mit der Optimierung von Seitenketten- oder Rückgratkonformationen für beliebige Proteine erfolgt.

Contents

List of Publications

Carina Frauer, Thomas Hoffmann, Sebastian Bultmann, Valentina Casa, M. Cristina Cardoso, Iris Antes, Heinrich Leonhardt, (2011) *Recognition of 5-Hydroxymethylcytosine by the Uhrf1 SRA Domain.* PLoS ONE 6(6): e21306. doi:10.1371/journal.pone.0021306.

Thomas Hoffmann, Angela M. Krackhardt, Iris Antes, (2015) *Quantitative Analysis of the Association Angle between T-cell Receptor Va/Vβ Domains Reveals Important Features for Epitope Recognition.* PLoS Comput Biol 11(7): e1004244. doi:10.1371/journal.pcbi.1004244

Thomas Hoffmann, Antoine Marion, Iris Antes, (2016) *DynaDom: Structure-based prediction of T-cell receptor inter-domain and T-cell receptor-peptide-MHC association angles.* BMC Struct Biol (submitted).

List of Figures

List of Tables

List of Abbreviations

$\mathcal{C}_\mathcal{E}$	energy and RMSD based success criterion.
$\mathcal{C}_\mathcal{R}$	RMSD based success criterion.
Å	Ångström.
β2m	*beta-2-microglobulin*.
5caC	*5-carboxylcytosine*.
5fC	*5-Formyllcytosine*.
5hmC	*5-hydroxymethylcytosine*.
5mC	*5-methylcytosine*.
AA	all atom.
AB	*antibody*.
AG	*antigen*.
AMBER	*Assisted Model Building with Energy Refinement*.
APC	*antigen presenting cell*.
BFGS	*Broyden-Fletcher-Goldfarb-Shanno*.
BLAST	*basic local alignment search tool*.
BU	*biological unit*.
C	constant (domain).
CASP	*Critical Assessment of Techniques for Protein Structure Prediction*.
CD	*cluster of differentiation*.
CD4	T-Cell co-receptor molecule specific for MHC class II molecules.
CD8	T-Cell co-receptor molecule specific for MHC class I molecules.
CDR	*complementary determining region*.
CG	*conjugate gradient*.
CHARMM	*Chemistry at HARvard Macromolecular Mechanics*.
CLIP	*class II-associated Ii (invariant) chain peptide*.
CoR	*Center of Rotation*.
CoR_β	*Center of Rotation* of the variable β domain.
CoR_μ	*Center of Rotation* of the pMHC ligand.
CPU	*central processing unit*.
CTL	*cytotoxic T lymphocyte*.

d	diversity (gene segment).
D-A	donor-acceptor.
DNA	*deoxyribonucleic acid*.
EAD	*Euler angle distance*.
EMSA	*electrophoretic mobility shift assay*.
ER	*endoplasmic reticulum*.
ES	*evaluation setting*.
FF	*force field*.
GFP	*green fluorescent protein*.
GSL	*GNU scientific library*.
HLA	*human leukocyte antigen*.
IG	*immunoglobulin*.
IMGT	*international ImMunoGeneTics project*.
ITAM	*immunoreceptor tryrosine-based activation motif*.
j	joining (gene segment).
LAT	*linker of activation in T-cells*.
LIE	*linear interaction energy*.
LJ	*Lenard-Jones*.
LRA	*linear-response approximation*.
MBD	*methyl-CpG binding domain*.
MBP	*5-methylcytosine binding protein*.
MD	*molecular dynamics*.
MHC	*major histocompatibility complex*.
MSA	*multiple sequence alignment*.
NKT	*natural killer T-cell*.
NMR	*nuclear magnetic resonance*.
NP	*nondeterministic polynomial time*.
OPLS	*Optimized Potentials for Liquid Simulations*.
OPMD	*optimized potential molecular dynamics*.
PBC	*periodic boundary conditions*.
PDB	*Protein Data Bank*.
PDBA	*pull-down DNA binding assay*.
pdf	*probability density function*.
PME	*particle mesh Ewald*.
pMHC	*peptide-MHC molecule complex*.
PR	*positions restraints*.

Q	glutamine.
Q-Q	glutamine-glutamine.
QO	*glutamine carboxamide group optimization*.
RB	*rigid body*.
RMSD	*root mean square deviation*.
RMSF	*root mean square fluctuation*.
SAG	*superantigen*.
SCR	*structural conserved region*.
SD	*steepest descent*.
SO	*subprocess operator*.
SPC	*simple point charge*.
SRA	*SET and RING finger Associated*.
SVR	*structural variable region*.
TAP	*transporter associated with antigen processing*.
TBM	*template based modeling*.
TCR	*T-cell receptor*.
TCRpMHC	*T-cell receptor-peptide-major histocompatibility complex molecule*.
TDG	*thymine DNA glycosylase*.
TeT	*ten-eleven translocation*.
TIP3P	*transferable intermolecular potential 3 points*.
TLR-4	*toll-like receptor 4*.
TRAV	*T-cell receptor alpha variable*.
Uhrf	*Ubiquitin-like, containing PHD and RING finger domains*.
UPGMA	*Unweighted Pair Group Method with Arithmetic mean*.
V	variable (domain).
v	variable (gene segment).
Vα	variable α.
Vβ	variable β.
vdW	van der Waals.
YFP	*yellow fluorescent protein*.
ZAP70	*zeta-chain associated protein*.

Chapter 1

Introduction

T-cell receptors (TCRs) are highly diverse heterodimeric molecules, which allow T-cells to detect pathogenic peptides on the surfaces of our cells. The current chapter first introduces the role of T-cells in the immune response (Sec. 1.1) by highlighting (i) the mechanisms of *antigen* (AG) presentation, (ii) explaining the genetic reasons for the TCR diversity, (iii) giving structural insights into the molecular mechanisms of the AG recognition, and (iv) reviewing models of the so far not fully understood signal transduction of T-cells. The vast variety of possible TCRs does not allow for a comprehensive experimental structural solvation of this signaling molecule. Therefore, alternative methods are required to predict the molecular structure of the TCR recognizing a pathogen, such as homology modeling. Sec. 1.2 describes the basic concept of homology modeling, which allows for the construction of a protein structure based on its sequence and related structural templates. However, even though all TCRs share a high homology, the straight forward homology modeling of the TCR and the TCR:AG complex is hampered by structural features, such as inter-domain orientations. The aim of the present thesis is to analyze such features and to develop a method, which makes these features accounted during the modeling processes. The aim of this thesis is detailed in Sec. 1.3, which also outlines this cumulative work.

1.1 The Role of T-Cells in the Adaptive Immune Response

The immune defense of vertebrates is a complex system made of evolutionary older and younger components. The following sections briefly introduce this wide topic and are based on the book *Immunobiology: the immune system in health and disease* by Jeneway [137], if not cited otherwise. For a more comprehensive introduction, the reader is referfed to the latter and other text books.

We distinguish between the old innate and the younger adaptive immune sys-

tem. The innate immune response includes inherited specific mechanisms, such as the recognition of Gram-negative pathogens by macrophages. Lipopolysaccharides (LPS) occur on these bacteria and trigger a signaling cascade in which the specific *toll-like receptor 4* (TLR-4) is involved [265]. This and other innate mechanisms are immediately available, thus efficient and fast, but are altogether restricted to a limited number of targets. In contrast, the adaptive immune response is more universal but delayed. The adaptive immunity is based on B and T lymphocytes (also further referred to as B-cells and T-cells), which both arise from hematopoietic stem cells in the bone marrow. B-cells are capable of producing pathogen-neutralizing *antibodies* (ABs) with high specificity against certain AGs but can also act as specialized *antigen presenting cell*s (APCs) (see below).

The focus of the present thesis is placed on ($\alpha\beta$) T-cells, which distinguish between self and non-self (pathogenic) peptides presented on *major histocompatibility complex* (MHC) molecules on cell surfaces. Sec. 1.1.1 describes the function of different classes of MHC molecules and the origin of their presented peptides. Different T-cell clones express distinct TCRs and a vast variety of these clones allows the immune system to recognize a tremendous number of different peptides. T-cell diversity is further introduced in Sec. 1.1.2. The signal transduction upon the TCR binding a cognate *peptide-MHC molecule complex* (pMHC) is still not fully understood, leading to different suggested signaling models (Sec. 1.1.4). In the past decades several structures of TCRpMHC complexes and their single components were resolved using X-ray crystallography. Sec. 1.1.3 describes some general structural properties of the TCR, the pMHC, and the TCRpMHC complex and provides several other structural observations, which might be important for the antigen recognition and the signaling.

1.1.1 Major Histocompatibility Complex Molecules Present Antigens

The major histocompatibility gene complex is a highly polymorphic genomic region that encodes several proteins involved in immunity. MHC molecules are cell surface proteins grouped into class I or class II. These two classes of MHC molecules and a further MHC-like protein called *CD1d* participate in the presentation of AGs to certain types of T-cells.

MHC class I molecules present peptides originating from cytosolic proteins. The presentation pathway is illustrated in Fig. 1.1 and is briefly described in what follows. The cytosolic proteins are first cleaved into peptides by the multiprotease complex called *proteasome* and are then translocated by the *transporter associated with antigen processing* (TAP) into the *endoplasmic reticulum* (ER). In the ER several chaperones are involved in the loading of empty MHC class I molecules with

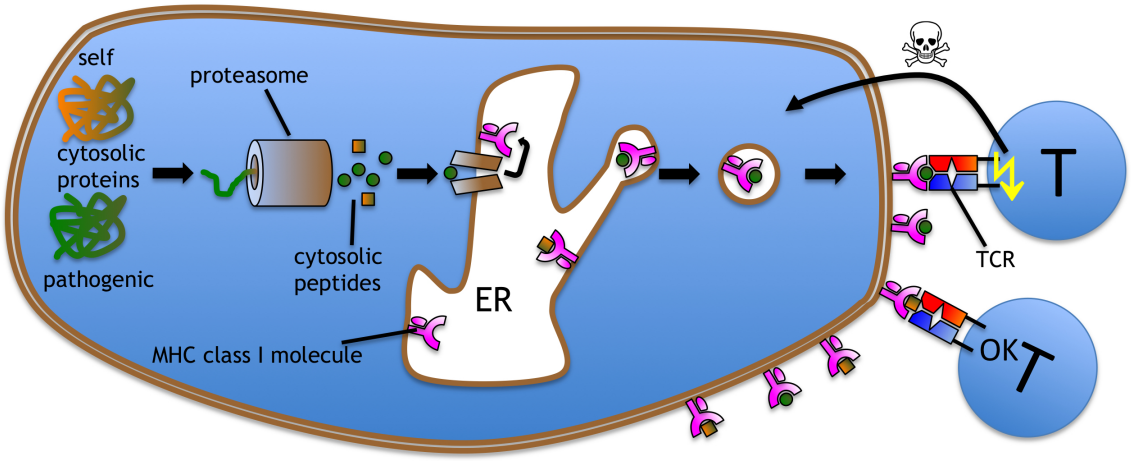


Figure 1.1: **The MHC Class I Pathway and the Role of CD8 T-Cells.** Cytosolic proteins are cleaved by the proteasome and the resulting peptides (self: square; non-self: green circle) are transported to the endoplasmic reticulum (ER) and loaded to MHC class I molecules. CD8 T-Cells (TCR: red/blue; co-receptor not shown) recognize peptide:MHC complexes on the cell surface and distinguish between self and non-self peptides. Peptides originating from tumor or virus proteins activate the T-cell, which induces apoptosis in the unhealthy cell.

peptides arriving through the TAP. After the loading process, vesicles containing the membrane anchored pMHC complexes are transported to the cell surface. The total of all pMHC class I complexes on the cell surface is a condensed image of a cell's proteome and thus indicates the health state of a cell [120]. In a healthy cell exclusively, self-peptides are presented on the surface, whereas on the surface of viral infected or degenerated (tumor) cells, non-self peptides can be found presented by the MHC class I molecules. Pathogens presented by MHC molecules of class I are detected by so called CD8+ T-cells, which carry a CD8 coreceptor. If a naïve CD8 T-cell meets the appropriate AG for the first time, it is activated, becomes a cytotoxic T effector cell (*cytotoxic T lymphocyte*, CTL), and then proliferates. The function of CTLs is to initiate apoptosis of virus infected or tumor cells.

The MHC class II presentation pathway is illustrated in Fig. 1.2 and is briefly described in what follows. MHC molecules of class II appear on the surface of so called APCs and display peptides of exogenous origin. APCs are for the most part macrophages, dendritic cells and B-cells, which are able to uptake pathogens or their products via endocytosis from the extracellular space into endosomes. In theses endosomes a digestion takes place resulting in peptides, which replace the placeholder molecule *CLIP* in MHC class II molecules. The pMHC class II complexes migrate to the surface of the APC to activate T helper cells (T_H-cells, CD4+ cells). T_H-cells carry the CD4 co-receptor and are involved in the triggering of AB production by B-cells.

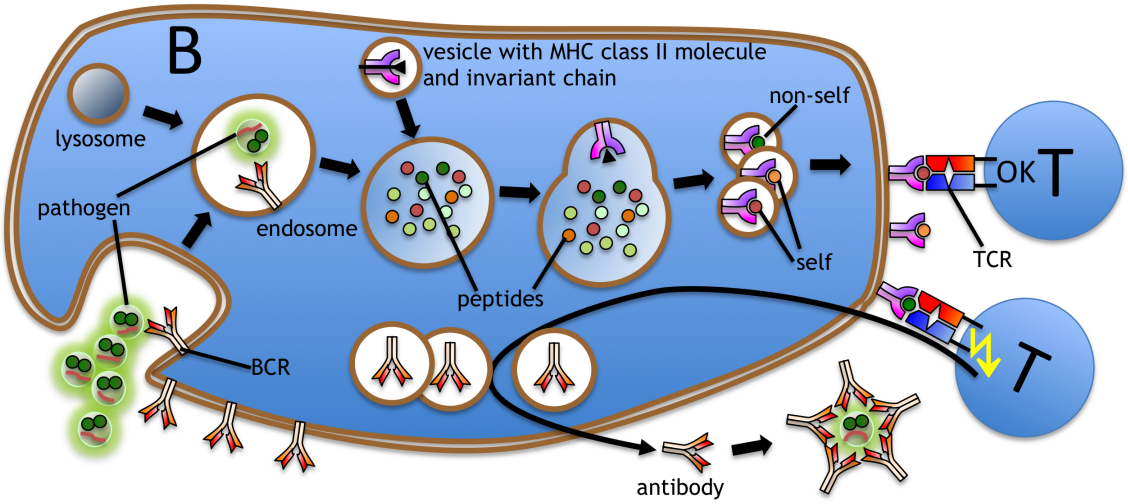


Figure 1.2: **The MHC Class II Pathway and the Role of CD4 T-Cells.** Exogenous material as pathogens are taken up by an antigen presenting cell (APC, here a B cell). Endosomes merge first with lysosomes containing digestive enzymes and second with vesicles containing MHC class II molecules (the CLIP contained in the invariant chain is shown as a black triangle). The MHC class II molecules are loaded with digestive products (peptides) and exposed to CD4 T-cells (helper T_H-cells; co-receptor not shown). Activated T-cells stimulate the B cell for antibody production. For simplicity of illustration the APC and the B cell are the same. Abbreviations: BCR = B-cell receptor, MHC = major histocompatibility complex, TCR = T-cell receptor.

Unlike the MHC class I and class II molecules, the MHC-like surface molecule CD1d does not present peptidic but lipidic ligands and is not polymorphic [18, 222]. The lipid:CD1d complex indicates a tumor cell, viral infection, or an infection by intracellular bacteria. *Natural Killer T-cells* (NKTs) detect such complexes and initiate apoptosis. Such non-peptidic AGs are, however, out of the scope of this thesis.

1.1.2 Diversity of T-Cell Receptors and MHC Molecules

T-cells develop from hematopoeitic stem cells and mature in the thymus. Along this maturing process the cells differentiate to either MHC class I restricted CD8+ T-cells or to MHC class II restricted CD4+ T-cells and acquire their antigen specificity. The difference in the AG specificity of two distinct T-cell clonotypes depends on their expressed TCR. The sequence of the expressed TCR is determined by somatic recombination, which facilitates the same enzymatic mechanism as responsible for the diversification of ABs in B-cells. In contrast, a further diversity is achieved by somatic hypermutation in B-cells.

As described in more detail in Section 1.1.3.1, the pMHC binding part of the TCR consists of two associated *immunoglobulin* (IG) domains, the variable α (Vα) and the variable β (Vβ) domains. During the maturation process of the T-cell the

gene sequence coding a variable domain is assembled from different gene segments, for which multiple gene loci exist. The combination of different loci is one reason for the diversity. In the case of the TCR Vα domain, two segments are connected, the variable (v) and the joining (j) segment. For the TCR Vβ domain an additional short diversity (d) segment is inserted between the v and the j segments. In this process, also referred to as v(d)j recombination, *non-template nucleotides* can be randomly added at the junction sites between two segments. This highly diverse area is located at the CDR3 loop (see Section 1.1.3.1 for structural details), which mainly contacts the presented peptide in the TCRpMHC complex. The CDR1 und CDR2 loops of the TCR mainly contact the MHC molecule and are exclusively germline encoded. Thus, a co-evolution between the CDR1/2-coding v segments and the MHC alleles can be assumed. To summarize, the diversity of T-cell receptors is achieved by (i) the combination of two chains, (ii) the somatic recombination of gene segments, and (iii) the addition of non-template nucleotides. The repertoire of functional TCRs per individual human being is estimated to $2.5 \cdot 10^7$ and for mice to $2 \cdot 10^6$ [12, 44, 284].

MHC class I molecules bind peptides using a pocket formed by one single protein chain. For *homo sapiens* this chain is encoded at three different loci (HLA A, -B, and -C), *i. e.* an individual expresses six different MHC class I molecules. The three loci are highly polymorphic - for each locus, several hundreds of alleles are reported.

In the case of the MHC class II molecule the binding pocket is assembled by two independent protein chains. The loci encoding the two chains are also polygenic and, for all except one, highly polymorphic (HLA DRα, -DRβ, -DPα, -DPβ, -DQα, and -DQβ).

Annoted sequences of *human leukocyte antigen* (HLA) and TCR segment alleles are collected in the *international ImMunoGeneTics project* (IMGT) databases[1] [95, 233]. In the present thesis the IMGT naming scheme for TCR alleles and the unique residue numbers are used [163, 164].

1.1.3 Structural Basis of the Peptide-MHC Recognition by T-cell Receptors

X-ray structures of human or murine TCR heterodimers and TCRpMHC complexes appeared for the first time in the *Protein Data Bank* (PDB) in the late 1990s [26, 27, 89, 90]. Until today more than 200 elucidated structures containing a TCR are available. The structures contain different TCRs in the bound and unbound state with different ligands. In the present thesis, subsets of these structural sets are used as described in more detail in Chapters 2 and 4. The following sections describe general structural features of TCRs (Sec. 1.1.3.1), MHC molecules (Sec. 1.1.3.2), and

[1] IMGT/HLA database: `http://www.ebi.ac.uk/ipd/imgt/hla/`; IMGT/GENE-DB: `http://www.imgt.org/genedb/`

finally features of the complex assembled by both molecule types (Sec. 1.1.3.3). An X-ray structure of a TCRpMHC complex is illustrated in Fig. C.1, p. 218.

1.1.3.1 General Structural Properties of TCRs

The T-cell signalosome involves several different proteins (see Sec. 1.1.4) and can be partitioned into an extracellular and an intracellular area. The TCR plays a central role in the interaction of these molecules, since TCRs specifically recognize pMHCs complexes.

The TCR is a heterodimeric complex consisting of an α chain and a β chain and is similar to the general structure of Fab-Fragments of ABs [90]. Both chains consist of five regions: The first region is the cytoplasmic tail consisting only of a few amino acids and is followed by the second region, a transmembrane domain. This transmembrane domain is connected to the hinge (H) region in the extracellular space. The H regions of both chains are covalently bound to each other *via* a disulphide bridge. The H region connects two subsequent IG-like domains to the transmembrane domain of the TCR, which bear both glycosylation sites. The first IG-like domain is referred to as constant (C) domain and is highly conserved for each of the two chain types. The second IG-like domain is called the variable (V) domain and is responsible to the diversity of the TCR.

As typical for IG-like domains, a V domain (see Fig. 1.3) consists of two layers of antiparallel β-strands, and the two layers are covalently bound to each other by a conserved disulphide bridge. Three distal loops form the pMHC specific binding site of the TCR and are named as *complementary determining region* 1 - 3 (CDR1, CDR2, and CDR3; see Fig. 1.3). The CDR1 and CDR2 loops are germline encoded in the v segment, whereas the CDR3 loop sequence is determined during T-cell maturation by the junction of v and j (respectively v, d, and j) segments (see Sec. 1.1.2). Depending on the used segments and on the v(d)j recombination process the CDR loops of two TCRs can differ in their length, thus complicating direct sequence and structure comparisons of multiple TCRs. This problem has to be considered in the methodology to be developed in this thesis, as introduced in Sec. 1.3.

In the TCR complex the two chains associate such that the Vβ domain contacts the Vβ domain and the Cα domain contacts the Cβ domain. Comparisons of structures of the A6 TCR bound to different ligands showed differences in the relative orientation of the Vβ domain with respect to the Vβ domain [70, 71, 88, 89]. This additional flexibility of TCRs was proposed by Li *et al.* in analogy to the rearrangement of V_H and V_L domains in the Fab fragments of antibodies [169]. Such alterations could be observed in several antibody X-ray structures [15, 28, 29, 54, 117, 158, 160, 172, 203, 231, 255, 257, 266]. Furthermore, alterations of the relative orientations of the TCR constant domains were reported for several X-ray structures and the effect

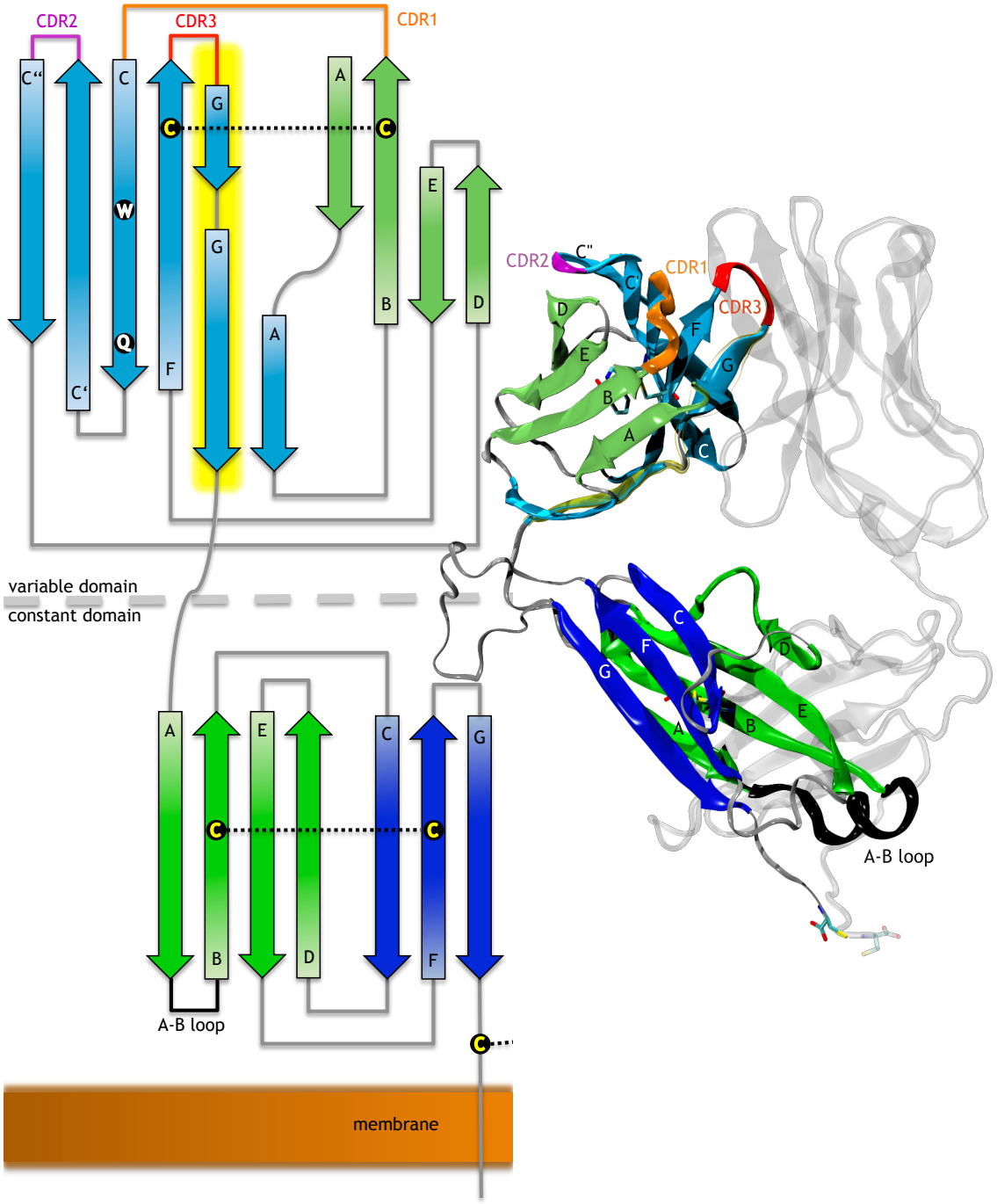


Figure 1.3: **Structural Localization of the T-Cell Receptor Subunits.** The secondary structure topology is schematically depicted on the left and a structural example (PDB ID 1tcr) focusing the β chain (colorized) in the context of the α chain (transparent gray) is provided on the right. The two stacked sheets of the *immunoglobulin-like* domains are indicated in green and blue in both domains of the chain and the strands are labeled according to the *IMGT naming scheme*. The *complementary determining region* (CDR) in the variable domain and the *A-B loop* in the constant domain are highlighted in orange, magenta, red, and black. Special conserved residues, as the central tryptophan (W), disulphide bridge (black dotted line) forming cysteines (C), and a glutamine (Q) interacting with the opposite chain, are shown as circles or in licorice representation in the left or right picture, respectively. The part of the structure encoded by the j-segment is marked yellow.

is often referred to as *scissoring* [10, 67, 87, 88, 116, 134, 147, 188]. A first direct comparison of the Vα/Vβ association angle for several TCR colontypes was performed by McBeth *et al.* in 2008 [188]. Dunbar *et al.* developed a methodology (ABangle) to measure the orientation of the V_H and V_L in ABs [73]. The authors then applied the ABangle method also to a non-redundant set of TCRs [74]. Both methods will be revisited in Chapter 2, which focuses on the development and application of a new domain angle measuring method relevant for the TCR modeling (Chapter 4).

Early comparisons of TCR sequences allowed to identify 40 conserved sites in the V domain framework regions, including glutamine residues at the Vα:Vβ binding interface [49]. In the first ever resolved TCR X-ray structure, a bifurcated hydrogen bond interaction at the Vα:Vβ binding interface between glutamine (Q) residues of both domains was observed and both residues belong to the set of the previously mentioned conserved sites [90]. The interaction of these conserved Q residues could also be observed for antibodies [43, 49]. This conserved interaction is revisited in Chapter 2, where it is shown to be an important pivot point for the Vα/Vβ domain arrangement. The glutamine-glutamine (Q-Q) interaction is explicitly taken into account in the modeling process, which is developed in Chapter 4.

1.1.3.2 General Structural Properties of MHC Molecules

The major histocompatibility gene complex encodes two different classes of antigen presenting molecules displaying peptides of either cytosolic or extracellular origin to different types of T-cells (see Sec. 1.1.1 for details). MHC class I and class II molecules are very similar to each other in their general shape, since both form a peptide binding groove utilizing two domains. This binding groove consists of a vaulted plane of β-strands, which is spatially limited by α-helices.

MHC class I molecules are heterodimeric membrane surface proteins, consisting of a heavy and a light chain (see Fig. C.1, p. 218 for illustration). The light chain is the highly conserved soluble *beta-2-microglobulin* (β2m) , which is not encoded in the MHC gene complex. The heavy chain consists of three domains, namely α_1, α_2, and α_3. The IG-like α_3 domain, is anchored to the cell membrane, and offers a binding site to the CD8 coreceptor molecule of T-cells [243]. The binding groove is constituted by the α_1 and the α_2 domain and binds peptides of 8 to 12 amino acids length.

MHC class II molecules are also heterodimeric, but in contrast to class I, both chains (α and β) consist of two domains. The IG-like α_2 and β_2 domains are each anchored to the membrane and the latter bears a binding site for the CD4 co-receptor molecule of T-cells [155]. A further difference to class I molecules is, that the binding groove is constituted by domains of both chains, the α_1 and the β_1 domains. In contrast to class I molecules, the class II binding groove has an open shape and does

not restrict the size of the bound peptide.

1.1.3.3 Structural Properties of TCR-peptide-MHC Complexes

In the TCRpMHC complex the peptide presenting groove faces to the variable domains of the TCR. The two TCR variable domains Vβ and Vβ together constitute the pMHC binding site with a contribution of the six *complementary determining region* (CDR) loops. The germline encoded CDR1α, CDR2α, CDR1β, and CDR2β loops mainly contact the helices of MHC peptide binding domains, whereas the more diverse CDR3α and CDR3β loops predominantly interact with the presented peptide. Such an interaction pattern was already suggested before any X-ray structure of a TCRpMHC complex was available [260]. The structure solvation of the murine 2C TCR and the human A6 TCR in the MHC bound state showed this so called diagonal binding mode, which was then assumed to be general [70, 89, 90]. Some years later, when more structural data became available, Rudolph *et al.* investigated a broader set of complexes [237]. This study confirmed the diagonal binding mode and also showed the pMHC docking angle on the TCR to be in a certain range. Rudolph *et al.* determined the docking angle by computing the dot product of two vectors. The first vector is embedded into conserved residues of a helix in the MHC binding groove and the second vector connects the conserved disulphide bridges of the Vα:Vβ domains. In the context of the present thesis two limitations of the measure used in the above study has to be mentioned: First, the orientation of the line between the two disulphide bridges is not independent from the Vα:Vβ domain rearrangement. I And second, with this method it is not decidable whether the pMHC is flipped by 180°. A structure was indeed recently published with a TCR binding the pMHC in an exceptional reverse diagonal mode [25].

1.1.4 T-Cell Signaling Models

A variety of proteins constitute the T-cell signalosome, which spreads from the intracellular to the extracellular space and includes the TCR [289]. When the TCR binds the pMHC AG a signal transduction cascade is initiated. At different steps of this cascade specialized kinases phosphorylate so called *immunoreceptor tyrosine-based activation motif*s (ITAMs). The TCR molecule itself does not carry such ITAMs, but is associated to the intracellular membrane-bound ζ-chain homodimer, offering several of these phosphorylation sites. The initial phosphorylation in the cascade takes place at ITAMs of the cytoplasmic tail of the CD3 molecule, which is present in two different heterodimeric forms and interacts with the extracellular constant domains of the TCR. The further downstream of the signal continues with the phosphorylation of *linker of activation in T-cells* (LAT) by the activated form of the kinase *zeta-chain associated protein* (ZAP70). The activation of this kinase

depends on two events. In the first event the ZAP70 binds the phosphorylated ITAMs of the ζ-chains. The second event of activation is the phosphorylation of ZAP70 by the co-receptor associated kinase *Lck*. Besides the co-receptors CD4 and CD8, other costimulatory proteins were discovered to intervene into the signaling cascade, such as the CD28 molecule [296].

The intersubunit communication in the signalosome is still not fully understood and it is not known which events affect the initial phosphorylation of the CD3 heterodimer [94]. Several (not mutably exclusive) signaling models were proposed in the past years and mostly (i) *segregation*-, (ii) *aggregation*-, (iii) *conformational change*-models were discussed (reviewed in Refs. [45, 50, 273]).

The kinetic segregation model assumes a spontaneous and random phosphorylation of the CD3 while the half-life of the phosphorylated state depends on the presence of a CD45-associated kinase [63, 64, 274]. In this model the formation of the TCRpMHC complex causes a tight and narrow cell-cell contact, squeezing away the large CD45 molecule from the contact gap.

In aggregation models multimers of subunits of the signalosome constitute aggregates bringing the kinases close to their substrates, such as the pseudodimer model [133]. In the pseudodimer model two TCRpMHC complexes are assumed to be close to each other while TCR1 is exposed to a self peptide (pMHC1) and TCR2 binds a non-self peptide. The co-receptor bound to pMHC1 leads to a phosphorylation of TCR2 signalosome.

The third type of model considers conformational changes to be involved in the signal transduction. Conformational changes in the CDR loops can indeed be observed in several structures (*e. g.* the BM3.3 TCR [229]). However, these loop alterations are rather believed to play a role in the TCRs plasticity to adapt to different ligands instead of being communicated to the conserved parts of the molecule [94]. Such conformational changes can even enable some TCR to recognize different classes of MHC molecules [92]. Gil *et al.* found a conformational change of the CD3ε chain induced by an unknown process in the TCRs upon ligand binding and speculated about a movement of the TCR to be responsible for the alteration of the associated CD3 molecule [94]. A conformational change in the *A-B loop* of the TCR constant α (Cα) domain observed for two different TCRs was proposed as a possible key event in the communication between the TCR and the associated CD3 molecule (see Fig. 1.3 for loop location) [19, 148]. Conformational changes on domain level were introduced in Section 1.1.3.1 and will be further investigated in Chapter 2. The rearrangement of the TCR Vα and Vβ is rather seen as a further degree of freedom for the plasticity of the TCR and major changes are not correlated to the signal strength in the antigen recognition [71, 134, 188]. However, the change of Vα/Vβ inter-domain angles could be involved in the scissoring effects influencing the constant domain orientations and induce the conformational change of the

A-B loop.

1.2 Homology Modeling

Homology modeling (also referred to as *comparative modeling* or *template based modeling* (TBM)) techniques are based on the observation that proteins of evolutionary related sequences share a similar three-dimensional structure and that for a target protein sequence the structure is predicted using already resolved related proteins as templates [48, 247]. The first application of this approach was the "*ex silico*" modeling of bovine *α-lactalbumin* on the structure of the homologue hen's egg-white *lysozyme* using wire and plastic tubes in the year 1969 [39]. Then several computational methods found their way in this field and have been regularly compared in the *Critical Assessment of Techniques for Protein Structure Prediction* (CASP) experiment since 1995 until present [205, 206]. A leading method in this competition over several years is the I-TASSER server [298].

Generally, homology modeling methods consist of four simultaneously or consequtively executed steps, which are (i) the identification of (related) template structures, (ii) the alignment of the target structure with the templates, (iii) the model building, and (iv) the evaluation of the model, where step (iii) includes the modeling of the core, the loop regions, and the side chain placement [14]. Many of such approaches are based on the partitioning of the system into *structural variable region*s (SVRs) and *structural conserved region*s (SCRs) [105, 277].

The software tool *MODELLER*, which uses the satisfaction of spatial restraints, has evolved during the past two decades and is used in more than 2,000 published homology modeling studies [241]: The *MODELLER* program is based on a *conjugate gradient* (CG) algorithm (see Sec. 4.1.2.1), optimizing several spatial features represented as a so called *probability density function* (pdf), which depend on the used template(s). In the initial publication of the program, 21 features are considered in the modeling process, such as alpha carbon distances, dihedral angles, or solvent accessibility. The features are obtained from the aligned template structures and are then transfered to the target sequence. The optimization process starts from a random conformation of the model and iteratively alter the coordinates of the atoms such that the restraints are satisfied as good as possible. Furthermore, energy terms of the CHARMM force field (see also Chapter 3) are included in the objective function of the optimization to avoid unphysical conformations.

High quality models with an *root mean square deviation* (RMSD) below 2 Å can be expected for sequence identities of more than 40% between the target and the template [277]. Furthermore, the success of homology modeling also depends on the (multiple) sequence alignment, as well as the number and quality of the used templates [14]. A comparison of several homology modeling tools showed that models

can be improved by placing side chains using specialized tools, such as *SCWRL* or *IRECS* [110, 157, 277]. Other specialized methods focus on the modeling of SVR or loops and are classified into *ab initio* and template-based methods (see ref. [193] for review). Different attempts and methods of specially modeling TCRs or TCRpMHC complexes are reviewed in the introduction of Chapter 4 in Sec. 4.1.1.

1.3 Aim and Outline of this Thesis

T-cell receptor (TCR) molecules are highly diverse thus enabling the adaptive immune system to distinguish between a tremendous amount of *antigens* (AGs) (see Sec. 1.1). Structural models of the TCR in its unbound state and bound to an AG can open doors to new exciting biomedical applications: First, virtual screening of T-cell epitopes allows for the development of new vaccines and a straightforward identification of organ or stem cell donors [7, 85, 238]. Second, adoptive T-cell therapy can be improved by the rational design of optimized TCRs and by the prediction of TCR mispairing [21, 204, 301]. Third, theories of the so far not fully understood T-cell signalling can be verified and improved (see Sec. 1.1.4). For these applications the structural models need to be very accurate on the atomistic level as alterations of one or a few atoms by *e. g.* mutation of one amino acid in the TCR or *peptide-MHC molecule complex* (pMHC) ligand may lead to drastic changes in the resulting signal. Due to the very high diversity of TCRs (see Sec. 1.1.2) it is only possible to obtain experimental structures for a small fraction of the total T-cell repertoire and thus accurate computational models urgently needed [197]. As high sequence similarity is an essential characteristic of MHC molecules, the TCR constant domains, and the TCR variable domain framework regions, intuitively high quality comparative models of TCRpMHC complexes with a deviation below 1 Å could be expected (see Sec. 1.2). However, this assumption is only valid treating all domains of the complex separately. Considering the TCR or the TCRpMHC complex as a whole unit, the model accuracy can be much lower due to the flexibility of the domain linking regions and thus the relative orientation of the single domains.

The topic of the present thesis is the computational prediction of the above discussed domain-domain orientations. This work led to two main studies, which build up on each other: First, analysis of the relative orientation of individual complex components with respect to each other (*i. e.* the two TCR variable domains and the pMHC ligand with respect to the TCR) and investigation whether potential differences in the orientation are an artifact of the method for structure elucidation or are an intrinsic feature of the TCRs. Second, based on the results of the analysis, development of a method allowing the prediction of the domain orientations.

For the structural analysis, which will be described in Chapter 2, a cuboid-based method was developed, which allows the measurement of the relative orientation

angle of the TCR variable (V) domains and also of the relative TCR/pMHC angle in a unique manner. The angles determined for a broad set of structures were clustered and then correlated to different TCR clonotypes and pMHC ligands. Furthermore, common *Centers of Rotation* (CoRs) were determined for both cases, *i. e.* TCR variable domain and TCR/pMHC orientations, and the CoR sequence and structure-based features were characterized in detail. The results of this analysis were published in Publication 2 (see Appendix B, pp. 121) of the present thesis.

In the second results-based chapter of the thesis (Chapter 3) an application study will be discussed, which served as an introduction to molecular mechanics *force field* (FF) *molecular dynamics* (MD) techniques, which are a requirement for the development of the predictive method in Chapter 4. That work covers DNA binding proteins bound to differently modified DNA substrates. The results were published in Publication 1 (see Appendix A, pp. 103) embedded in the present thesis.

In Chapter 4 a new method is presented, which allows for the prediction of the angular orientation of TCR Vα/Vβ domains and which was subsequently generalized for the prediction of multimeric protein complexes. The final method was evaluated by modeling the orientation of the pMHC ligand with respect to the TCR simultaneously with the adjustment of the Vα/Vβ orientation. The method uses a multidimensional *rigid body* (RB) energy minimizer based on the work of Mirzaei *et al.* [200]. The underlying program was designed in a modular manner and can be extended easily to the optimization of additional features such as local side chain conformations, which was implemented for conserved residues at the *Center of Rotation* of the variable β domain (CoR_β) identified in the structural analysis (see above and Chapter 2). Fundamental for the implementation of the optimization algorithm is the usage of molecular mechanics FFs, as introduced in Chapter 3. This part of the thesis led to Publication 3 (see Appendix C, pp. 171) embedded in this dissertation.

In Chapter 5 the results are summarized and ideas for further improvements, extensions, and applications of the new method are discussed.

Chapter 2

Structural Analysis of TCRpMHC complexes

The high diversity of *T-cell receptors* (TCRs) allows the adaptive immune system to recognize a vast variety of pathogens (see Chapter 1 for details). The present chapter is about the analysis of a broad set of different TCR and TCRpMHC complex structures, in order to derive general properties of the receptor. The results of the analysis will serve as a basis for a modeling method presented in Chapter 4, as the goal of the present thesis is the development of a method predicting the association geometries in TCRpMHC complexes.

Sec. 2.1 introduces previous analyses of TCR association angles and provides the theoretical background for the measurement used in the analysis presented in the current chapter. In Sec. 2.2 the cuboid-method is developed, which is used for the determination of variable α (Vα)/variable β (Vβ) domain association angles and for the localization of a *Center of Rotation* of the variable β domain (CoR_β). Furthermore, the CoR_β is correlated to conserved residues using a sequence based analysis. Sec. 2.3 first summarizes Publication 2 of the present cumulative thesis. This publication provides the results of the TCR Vα/Vβ association angle analysis, as it shows (i) that the relative orientation of Vβ domain with respect to the Vα domain can vary depending on the clonotype of the TCR, (ii) that the orientation can differ between the bound and the unbound state of the TCR, suggesting a two step mechanism for the antigen recognition, and (iii) that a *Center of Rotation* (CoR) exists at the interface between the Vα and the Vβ domain stabilized by a conserved glutamine-glutamine (Q-Q) interaction. Second, beyond the scope of Publication 2, the docking orientation of *major histocompatibility complex* (MHC) molecules were examined.

2.1 Introduction

The part of a TCR that recognizes the *peptide-MHC molecule complex* (pMHC) is assembled by the association of the two V domains, Vα and Vβ (see Sec. 1.1.3.1, p. 6). Plasticity regarding the orientation of the two domains was early suggested and shown for particular TCR clonotypes [70, 71, 89, 169]. Sec. 2.1.1 describes two studies, which are based on different methodologies and systematically compared broad set of TCR clonotypes regarding their Vα/Vβ orientation angles. The latter section also sketches out a method, which was suggested for the unified measuring of pMHCs docking angles in TCRpMHC complexes and had been applied in a further study on a set of several structures. Sec. 2.1.2 motivates for the new analysis provided in the current chapter using a broader data set and the new related *cuboid-method*, which was specially designed to be compatible with the modeling approach described in Chapter 4. The background of approaches incorporated in the cuboid-method are depicted in the two subsequent sections: In Sec. 2.1.3 the concept of hierarchical clustering and significance analysis is briefly described (for a more comprehensive introduction see Refs. [78, 144]), as well as the measure based on the *Euler angle distance* (EAD). Sec. 2.1.4 deals with structural alignment methods with a particular focus on the requirements for the cuboid based analysis used in this chapter and the modeling procedure in Chapter 4.

2.1.1 Current Studies Comparing TCR and TCRpMHC Domain Geometries

In the following, three different studies dealing with association geometries in TCRpMHC complexes are presented. The first one measures and compares the docking angle of pMHC with respect to TCRs while the two other studies treat the Vα/Vβ inter-domain geometries of different TCRs.

The work by Rudolph *et al.* was the first comparative investigation of pMHCs docking angles based on a broad set of X-ray structures and showed the application of a unified methodology for the measurement of the angular range adopted by the pMHCs [237]: The examined data set contained 24 different structures of TCRpMHC complexes, while 17 of them contained an MHC class I molecule and in the remaining 7, a TCR was exposed to an MHC class II molecule. The data available at that time allowed to investigate the structures of 16 different TCR clonotypes, and some cases in which different peptides are presented to the same TCR type. The suggested unified approach for the determination of the pMHCs docking angle is based on the computation of the inner product of two vectors, which both represent the orientation of the Vα:Vβ complex of the TCR and of the pMHC, respectively. The first vector is defined as the connection between the two conserved

disulphide bridges in the Vα and in the Vβ domains (Sγ-atom centroids). The second vector is embedded in the MHC binding groove by fitting it in the alpha carbon atom coordinates of both MHC helices (see Sec. 1.1.3.2 for a structural description of MHC molecules). In Sec. 2.1.2 the particular definition of the first vector is taken up again in the context of Vα/Vβ flexibility.

McBeth *et al.* were the first to examine and compare the relative Vα/Vβ orientation for an, at that time, comprehensive set of TCR X-ray structures [188]: In this study, a data set of 37 X-ray structures of 21 different TCR clonotypes were taken into account, while the set contained bound (to MHC class I as well as MHC class II) and unbound TCRs and also contained three examples of the *natural killer T-cell* (NKT) TCR subgroup. The authors noticed differences in the Vα/Vβ orientation by comparing structures of different clonotypes as well as comparing the same clonotype of TCRs in the bound and the unbound state. The approach deployed for the determination of the Vα/Vβ angles was the *pseudo-dyad-method*, which had been established for the calculation of inter-domain angles in structures of *antibodies* (ABs). This method uses two angles to define a pseudo-dyad axis and a third angle to define a rotation around this axis, such that the Vβ domain is superposed well onto the Vα domain. A limitation of this method, given by the structural dissimilarity between Vα and the Vβ domains, will be revisited in Sec. 2.1.2.

Dunbar *et al.* developed the *ABangle* methodology initially to describe the V_H/V_L inter-domain geometry of ABs in an absolute manner [73]: The orientation is described by five angles and a distance vector, which are determined by embedding two planes into a set of conserved residues found in both domains. Later, the method was ported from ABs to the related TCRs in the context of rational design of TCR-like ABs [74]. In the latter study, the method was applied to a broad and non-redundant set of X-ray structures of different TCR clonotypes, but also complexes containing NKT TCRs or structures containing *superantigens* (SAGs) were included into the set. The design of this data set will be taken up again in Sec. 2.1.2. In another study, the ABangle method was applied on trajectories of a series of *molecular dynamics* (MD) simulations (see also Chapter 3 for an introduction into MD) [152]: The series consists of 172 different MD simulations of a TCRpMHC complexes differing only in the bound peptide to examine the impact of the ligand immunogenicity on the Vα/Vβ geometry.

2.1.2 Motivation for a New, Cuboid-Method-Based Analysis

An aim of the present work is to develop a prediction method for the Vα/Vβ associtation geometry and the pMHCs orientation based on structural properties of the TCRpMHC complexes (see Sec. 1.3). In the previous section, several studies dealing with the systematic analysis of the inter-domain geometries in TCRpMHC

complexes were introduced. These studies differ in their *foci* and thus in the applied methods and the used data sets. The focus of the present thesis is the development of a modeling approach for the prediction of inter-domain angles, as presented in Chapter 4, and is not covered in particular by the previous analyses. Thus, a new approach for the angular measurement is required.

The method for the determination of the docking angle of the pMHCs introduced by Rudolph *et al.* (see Sec. 2.1.1) is facing three limitations with respect to the present work: First, an assumed flexibility in the Vα/Vβ orientation influences the calculation of the pMHCs docking angle by translating the cysteine residues, thus altering the orientation of one of the two vectors. Second, the relative, three-dimensional transformation of the pMHCs is projected into a plane and is represented loosing the information about two of the three angles of rotation. Third, the information about the directionality of the pMHCs is lost, *i. e.* it cannot be determined whether the ligand rotated by 180°. Additionally, nowadays much more structural data is available requiring to revisit the measurement of the TCRpMHC docking angles.

In the study by Dunbar *et al.* (see Sec. 2.1.1) the measuring of the Vα/Vβ orientation was performed on a *non-redundant* set of TCR structures, which is sufficient and desired for goal of this experiment to compare TCRs and ABs. In the present thesis, it is however desired to compare TCRs of the same clonotype bound to different ligands, in order to investigate the influence of the latter. Thus, redundancy is a required feature of the data set used to derive structural data for the modeling process. Additionally, the data set of Dunbar's study contains structures of NKT TCRs (see also Sec. 1.1.1) as well as TCRs bound to SAGs. General observations about TCRpMHC complexes cannot be derived from complexes of the latter two types, since there the TCRs are not (exclusively) bound to pMHC. Besides the used data set, the methodology of Dunbar *et al.* is not suitable for the goal of the present thesis. A rotation in the three-dimensional space can be represented by three parameters. The ABangle method, however, uses a more complex but absolute representation, which is less appropriate for a direct comparison based on a *metric* (see also Sec. 2.1.3). For the assembly of the complex in the modeling process, it is beneficial to represent the rotation in a relative manner requiring only three angular parameters to keep the approach simple.

The study by McBeth *et al.* (see Sec. 2.1.1) allows for a description of the Vα/Vβ orientation in a relative manner using only three angular parameters, but faces two limitations: First, this pioneering study was based on a limited data set while newer structural data allows for more comparisons of the same TCR clonotypes in the unbound state and bound to far more different ligands, endorsing a new attempt of the analysis. Second, the pseudo-dyad method used in this study entails the risk of imprecision, since the underlying superposition of the Vβ onto the Vα domain is

biased by the (dis)similarity between the two domains. In the current chapter the cuboid-method is introduced, which gets along without cross-chain superimpositions in order to increase the precision of the measurement.

2.1.3 Clustering and Bootstrapping

Cluster data analysis is an early bioinformatics method, which was first applied in the field of phylogeny in the late 1950s [252, 253]. Since then, in the past decades, much effort was put on the advancement of this approach, which then spread to the realm of many other disciplines [144]. Generally, such cluster analysis methods group objects according to their similarity, which therefore requires to be quantified by a metric. The various methods are either agglomerative or divisive and mainly differ in the manner in which the similarity between two sets of already clustered objects is computed (also referred to as *linkage*). The most commonly used methods are (i) *average-*, (ii) *single-*, (iii) *complete-*, (iv) *centroid-linkage*, and (v) *Ward's minimum variance method*.

The *average-linkage* method, also referred to as the *Unweighted Pair Group Method with Arithmetic mean* (UPGMA), computes the distance $D(\mathcal{A}, \mathcal{B})$ between two sets $\mathcal{A}$ and $\mathcal{B}$ of already clustered objects as [253]:

$$D(\mathcal{A}, \mathcal{B}) = \frac{1}{|\mathcal{A}| \cdot |\mathcal{B}|} \sum_{a \in \mathcal{A}} \sum_{b \in \mathcal{B}} \delta(a, b), \tag{2.1}$$

where the distance $\delta(a, b)$ between the two objects a and b is required to be an *ultra-metric*, and $|\mathcal{X}|$ denotes the cardinality of a cluster $\mathcal{X}$. The *single linkage* method, which is also referred to as the *minimum linkage-* or *nearest neighbor* method, defines the distance $D(\mathcal{A}, \mathcal{B})$ between two sets $\mathcal{A}$ and $\mathcal{B}$ of clustered objects as [191, 252]:

$$D(\mathcal{A}, \mathcal{B}) = \min_{a \in \mathcal{A}, b \in \mathcal{B}} \delta(a, b), \tag{2.2}$$

where $\delta(a, b)$ is the distance between the two objects a and b. A phenomenon, not desired for some applications, is the tendency of the single-linkage method to merge weakly related clusters, which are only similar at a single point (*chaining*). This chaining phenomenon is avoided by the *complete-linkage* approach also referred to as *maximum linkage-* or *farthest neighbor* method, which uses the maximum instead of the minimum in Eq. 2.2 [66]. The *centroid* based methods, are also referfed to as the *k-means-* or *k-median-algorithm*. The algorithm partitions a data set into a number of k clusters of a similar sizes and minimizes the sum of squares (*i. e.* the variance) within a cluster [181]. Another approach resulting in clusters of minimized sums of squares is *Ward's (minimum variance) method*, which in contrast to the centroid methods does not require knowledge about the number of clusters to be found and

additionally allows clusters of different cardinalities [283]. In other words, the latter method preferably finds an arbitrary number of compact clusters, which can differ in their size. Thus, Ward's method was selected to be used for the analysis of the TCR Vα/Vβ inter-domain geometries, as further described in Sec. 2.2. To assess the significance of a clustering, resampling methods are often used, such as the *bootstrapping* approach [76].

2.1.3.1 Euler Angle Distance Measure

Euler Angles, which consist of the three angular components Φ, Θ, and Ψ, vividly describe the relative orientation of two Cartesian coordinate systems of the same origin. The sequentially executed rotations around three different axes according to the three angle components are illustrated in Fig. 2.1. For a detailed definition the reader is referred to textbooks in theoretical physics, such as *Classical Mechanics* by Goldstein *et al.* [98]. Here, three properties of Euler angles shall be summarized: First, the order of execution regarding the single components is important. Second, the same rotation can be described by several different Euler angles (*e. g.* the two different Euler angles $(\pi, 0, 0)$ and $(0, \pi, \pi)$ result in the same rotation). Third, a rotation by 90° around one axis leads to a superimposition of two rotation axes and thus to a lost degree of freedom. This phenomenon is also referred to as *gimbal lock*.

In the current chapter, a method is introduced, which measures the relative orientation of protein domains and evaluates the results using a clustering approach requiring a metric. The extent of the torsions of two rigid bodies A and B can be compared utilizing the *Euler angle metric*, which is defined as follows [69]:

$$d_E(A, B) = \sqrt{(\Phi_A - \Phi_B)^2 + (\Theta_A - \Theta_B)^2 + (\Psi_A - \Psi_B)^2} \tag{2.3}$$

It must be considered that Euler angles are only unambiguous within the intervals $-\pi \leq \Phi \leq \pi$, $-\pi/2 \leq \Theta \leq \pi/2$, and $-\pi \leq \Psi \leq \pi$ and therefore the properties of a metric are fulfilled only within these ranges [130]. In the analysis presented in the current chapter (see Sec. 2.3) these requirements are sufficiently served, making the normalization of the angular range or the use of an alternative metric (*e. g.* the *quaternion metric*) obsolete [130].

2.1.4 Structural Alignment

Structural alignment is an important method, which is applied in many fields, such as homology modeling (see Sec. 1.2), for the comparison of protein structures, for the elucidation of protein structures, or for the evaluation of trajectories obtained from MD simulations (see Chapter 3). The *Kabsch-algorithm* is often applied for the superimposition of different conformations of the same protein, as it is the case

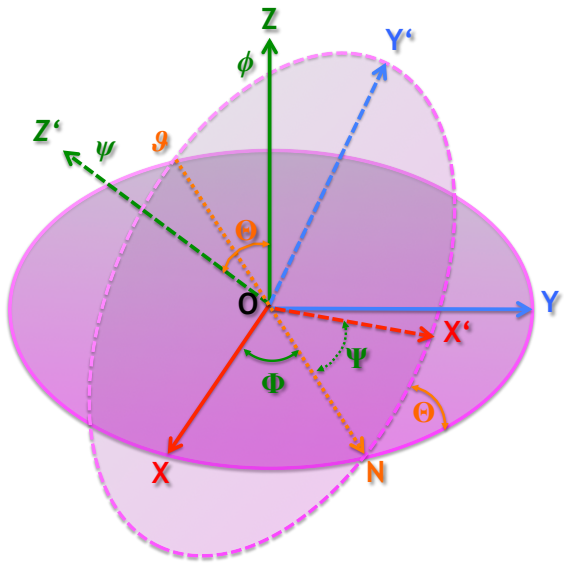


Figure 2.1: **Coordinate System Transformation with Eulerian Angles.** The initial coordinate system XYZ is shown as solid red, green, and blue axes, while the axes of the new coordinate system transformed at the origin O is depicted with dashed lines. First, the X axis is rotated within the XY-plane (purple, solid border) around the Z-axis by the Euler angle component Φ leading to the *line of nodes* N (orange, dotted). Then, the XY-plane is tilted about N and the Euler angle component Θ. Finally, the axes X' and Y' a formed by a rotation within the tilted plane (purple, dashed border) by the angle Ψ around the new axis Z'. Greek lower case letters ϕ, ϑ, and ψ denote the axes, around which is rotated by the angles Φ, Θ, and Ψ.

for MD trajectory analysis. This approach determines analytically a rotation matrix that maps two equally sized point clouds onto each other with minimal *root mean square deviation* (RMSD) [142]. A method using quaternions is available, which was shown to be equivalent to the Kabsch-algorithm [56].

However, focusing on the alignment of similar structural partitions in two *different* proteins, the superpositioning problem becomes – in terms of computational complexity theory – *NP-hard* and therefore, it is practically only solvable heuristically [153]. As reviewed in Ref. [177], several approaches have been developed, which mainly differ in the internal representation of the proteins, in the scoring function, and in their heuristic search.

The pairwise/multiple structural alignment tool *DALI* (*Distance Matrix Alignment*) makes use of a *C-alpha based distance map* (C-map) for protein representation and of a Monte Carlo approach in order to search for *similar fragment pairs*, which are based on a structural alphabet [126, 177]. In the context of the present work it is worth noting that due to local sequence similarities or yet identities of the TCRs, sequence information should not be considered in order to avoid a bias of the structural alignments. The use of the structural alphabet in the case of DALI ensures the exclusively structure based superpositioning. Thus, and because DALI is available as a web service, which can be queried using a command line tool, this algorithm was chosen for the methodology of the present work [111]. Besides the resulting superposed structures, the DALI algorithm provides information about the mapped residues which were used for the structural alignments. This mapping is a helpful feature used for the analysis method presented in the remaining part of this chapter.

2.2 Methods

In the following, first the cuboid-method is introduced in a general manner and the data set is described (Sec. 2.2.1), which were both used for the analysis of the TCR V domain orientations. Then several analyses based on this cuboid approach are described: Secs. 2.2.2 and 2.2.3 introduce the methods used in Publication 3 in order to analyze the relative TCR Vα/Vβ orientation and in order to detect and describe the CoR, respectively. Sec. 2.2.4 deals with methods used in order to analyze the pMHCs orientations and is separated into two parts. The first part covers the adaption of the cuboid method to measure the relative orientation of the pMHCs with respect to the TCR Vα domain, which was not performed in the analysis presented in Publication 2, but was needed for the modeling of the TCRpMHC complex in Publication 3 (Chapter 4). The second part deals with the question of the directionality of the pMHCs docking angle (here also referred to as the *pMHC flipping problem*). The latter study remained unpublished so far.

2.2.1 The Cuboid-Method for Domain Rotation Analysis

The following two sections describe data sets used for the analysis and the steps necessary to represent the TCR domains as cuboids, further referred to as the *cuboid-method*.

2.2.1.1 Data Sets

For the analysis of the TCR Vα/Vβ inter-domain geometries the *Protein Data Bank* (PDB) was queried for structures containing TCR α chains as well as TCR β chains, but δγ TCRs were excluded [27]. With this database search structures of free TCRs were found, as well as structures of TCRs bound to other molecules. The latter were of interest for the analysis, only if the TCRs were solely bound to pMHCs ligands. Thus, structures containing *superantigens* were discarded, such as a 2C T7 TCR with the PDB Id 2icw or the JM22 TCR structures with the PDB Ids 2xn9 and 2xna [242, 281]. For the same reason NKT TCRs were identified by a sequence analysis (see Sec. 2.2.2.1) and excluded from the data set, since these receptors do not bind pMHCs ligands but the MHC-like, lipid-presenting CD1d1 molecules (see Sec. 1.1.1). The structures, which were found at the date of the data base search and which where actually used, are listed in Tab. 2.1. In this table the particular clonotype and the ligand type (MHC class I or -class II) is assigned to each structure, based on the sequence analysis (see *ibid.*). More details about the structures are provided in Tabs. B.1 and B.S5 (pp. 128 and 162). Structures containing multiple *biological unit* (BU)s also referred to as *crystallographically independent molecules in the asymmetric unit*) were split and each BU was treated as a separate and

independent structure in the data set. The TCR constant domains as well as the MHC α_3 domain, the *beta-2-microglobulin* (β2m) molecule, and MHC β_2 domain were identified based on sequence comparison (see *ibid.*) and were removed from all structures. This choice was motivated by two reasons: First, many of the available structures only contain the V domains of the TCRs and the binding domains α_1 and β_1 of MHC class II molecules or α_1 and α_2 of the MHC class I molecules. Second, a superpositioning on a whole TCR chain consisting of the constant and the V domain could induce a bias due to a possible relative twist between the two domains (see *scissoring* in Sec. 1.1.3.1). Furthermore, all solvent molecules, salt ions, and all auxiliary molecules for crystallization were removed from the structures. Then, all structures were superposed onto the TCR Vα domain according to the procedure described in Sec. 2.2.1.2 below.

2.2.1.2 Superpositioning and Cuboid Placement

In the following a methodology is described, which is used for two purposes. The first use is the superpositioning of the TCR- or TCRpMHC structures onto their Vα domains. The second use is to represent the diverse domains in an unified manner, allowing for the measurements of the relative domain orientations (for illustration see Fig. B.1 B+C, p. 124). The structural alignment according to the Vα domain consists of two steps, which are first, the determination of a superpositioning template and second, the actual alignment of all structures according to this template.

For the determination of the template, the structure with the PDB identifier 2bnu was chosen as a reference molecule, having a high resolution of 1.4 Å [46]. The Vα domain was separated from the reference structure and is further referred to as the *initial Va template* $T_{\alpha,0}$. The Vα domains of all structures in the data set were separated and then superposed on the $T_{\alpha,i}$ template, using DaliLite command line tool (see Sec. 2.1.4) in an iterative process [111, 126]. After each iteration step i, the intersection of the mapped residues between the structures and the current template $T_{\alpha,i}$ were determined. A template $T_{\alpha,i+1}$ then contained the residues of such an intersection and the process was repeated until no alterations were observable in the final template T_α (the residues of the template are listed in Appendix D).

For actual superpositioning, all complexes of the data set were aligned on the T_α template according to their Vα domains. The MHC-bound and free TCR-structures, which were superposed to their Vα domains, are contained in two different sets further referred to as $\mathcal{S}_{bound}$ and $\mathcal{S}_{unbound}$, respectively. The union of both sets containing all these superposed structures is further referred to as $\mathcal{S}_{all} = \mathcal{S}_{bound} \cup \mathcal{S}_{unbound}$. As a result of the superpositioning, the Vα domains of all structures were fixed according to structurally conserved areas, whereas the Vβ domains (and the pMHCs) were displaced (see Fig. B.1 B, p. 124).

In order to allow the quantification of the domain twists, a cuboid based represen-

Table 2.1: **Structures used in the Analysis Data Set**

TCR Type	PDB	MC[a]	Lit.	TCR Type	PDB	MC[a]	Lit.
1G4	2bnu	U	[46]	LC13	3kps	I	[178]
1G4	2bnq	I	[46]	MEL5	3hg1	I	[52]
1G4	2bnr	I	[46]	OB.1A12	2wbj	II	[109]
1G4 AV-wt	2f54	I	[75]	OB.1A12	1ymm	II	[107]
1G4 c5c1	2pyf	U	[244]	RA14	3gsn	I	[102]
1G4 c5c1	2pye	I	[244]	SB27	2ak4	I	[267]
1G4 c49c50	2f53	I	[75]	SB27[K16Dα]	3kxf	I	[41]
1G4 c58c62	2p5w	I	[244]	TCR MS2-3C8	3o6f	II	[293]
1G4 c58c61	2p5e	I	[244]	TK3 wt	3mv7	I	[103]
3A6	1zgl	I	[171]	TK3 Q55H	3mv8	I	[103]
A6	2gj6	I	[88]	TK3 Q55A	3mv9	I	[103]
A6	1qsf	I	[71]	1934,4	2pxy	II	[81]
A6	1qse	I	[71]	1F1E8	3mff	U	[272]
A6	3d3v	I	[216]	226 TCR	3qiu	II	[208]
A6	3d39	I	[216]	226 TCR	3qiw	II	[208]
A6	1qrn	I	[71]	2B4	3qib	I	[208]
A6	1ao7	I	[89]	2B4	3qjf	U	[208]
A6	3h9s	I	[33]	2C	1tcr	U	[90]
A6	3pwp	I	[34]	2C	1g6r	I	[67]
AS01	3o4l	I	[198]	2C	1mwa	I	[176]
B7	1bd2	I	[70]	2C	2ckb	I	[91]
cf34	3ffc	I	[101]	2C [T7-wt-s]	2oi9	I	[53]
DM1	3dxa	I	[10]	2C m13 [T7-s]	3e3q	I	[138]
DM1	3dx9	U	[10]	2C m6 [T7-s]	2e7l	I	[53]
E8	2ian	II	[68]	2C m67 [T7-s]	3e2h	I	[138]
E8	2iam	II	[68]	2W20	3c6l	II	[60]
E8	2ial	U	[68]	5c.c7	3qjh	U	[208]
ELS4	2nx5	I	[268]	AHIII12.2	2uwe	I	[199]
ELS4	2nw2	U	[268]	AHIII12.2	2jcc	I	[199]
HA1.7	1fyt	II	[116]	AHIII12.2	1lp9	I	[42]
HA1.7	1j8h	II	[115]	B3K506	3c5z	II	[60]
Hy.1B1	3pl6	II	[248]	BM3.3	1nam	I	[230]
JM22	2vlj	I	[134]	BM3.3	1fo0	I	[228]
JM22	2vlk	I	[134]	BM3.3	2ol3	I	[187]
JM22	1oga	I	[258]	cl19	2z31	II	[81]
JM22	2vlm	U	[134]	D10	1d9k	II	[227]
JM22 [S99βA]	2vlr	I	[134]	KB5-C20	1kj2	I	[229]
KK50.4	2esv	I	[123]	N15	1nfd	U	[280]
LC13	1mi5	I	[149]	TCR 21.30	3mbe	II	[295]
LC13	1kgc	U	[147]	TCR172.10	1u3h	II	[186]
LC13	3kpr	I	[178]	YAe62	3c60	II	[60]

[a] Bound state / bound MHC class (I or II); U=unbound.

tation was used. For this purpose, templates were prepared as described for the T_α template for the corresponding domains (Vβ or pMHC). The templates differ from the T_α template as they contained both the conserved residues of the domain and the points defining the cuboid (the according residues are defined in Appendix D). The Vβ cuboid was defined to be the maximum extent of the represented domain of the reference structure and to have edges parallel to the coordinate axes of the reference structure coordinate system. In contrast to the superpositioning procedure described above, the structure is not superposed on the template but the reverse (*i. e.* aligning the template and the template-coupled cuboid according to the domain). This procedure was applied in order to subsequently superpose the T_β template onto all Vβ domains contained in the set $\mathcal{S}_{all}$, using the tool *DaliLite* [111, 126]. The orientation of the β-cuboid relative to the 2bnu coordinate system represents the orientation of the Vβ domain within a complex. The sets of orientations are further referred to as $\mathcal{S}^O_{bound}$, $\mathcal{S}^O_{unbound}$, and $\mathcal{S}^O_{all}$. Note: Additionally, cuboids were placed on and merged with the separated domains (Vα and Vβ) by the same procedure leading to the sets $\mathcal{S}_\alpha$ and $\mathcal{S}_\beta$, which will be used in the modeling procedure introduced in Chapter 4.

2.2.2 Clustering of the TCR V domain Association Angle

This section presents how the orientations of the Vβ domains were measured and clustered. The calculation of the rotational angles is based on the unique cuboid representation as obtained from the superpositioning presented in Sec. 2.2.1.2. The clustering was carried out for two different sets of cuboid orientations, $\mathcal{S}^O_{unbound}$ and $\mathcal{S}^O_{all}$. The measuring was performed by use of the program *TCRanalyze*, which was implemented for the special purpose of the present thesis utilizing the BALL library [121]. The program reads a set of β-cuboids and computes the Euler angles with respect to the reference coordinate system represented by an α-cuboid. In the cases in which multiple BUs were available for one PDB identifier, the data set contained multiple structures and thus multiple cuboid representation for this identifier. In such cases, the Euler angles were determined for each of the multiple cuboid representations and were then merged to one median Euler angle. In other words, to each PDB identifier a single Euler angle triplet was assigned, even if it contained several independent BUs. For all PDB identifiers found in one set ($\mathcal{S}^O_{unbound}$ or $\mathcal{S}^O_{all}$) the pairwise EAD according to Eq. 2.3 was computed and a distance matrix was assembled (see Sec. 2.1.3.1) [69]. The distances were clustered according to Ward's method (see Sec. 2.1.3) and the significance of the clustering was examined using a bootstrapping approach [76, 283]. The clustering and bootstrapping was performed with the program *R* and the use of the packages *pvclust*, while the number of bootstrapping replica was set to 10^6 [225, 261]. For illustration of the results, the

clustered distance matrix was presented as a heat map together with a dendrogram containing the significance values obtained by the bootstrapping (see Sec. 2.3). The lines and columns as well as the found significant main clusters were annotated with the TCR clonotype names obtained by the sequence analysis, as described in the following section.

2.2.2.1 Sequence Based Annotation of TCR and MHC Alleles

All sequences occurring in the used structures were annotated for the MHC alleles and the segment alleles of the TCRs. The goal was to assign the different TCR clonotypes to the found clusters (see Fig. B.3, p. 129) and to identify differences between TCRs (or MHC molecules), sharing the same name in the literature.

For the MHC molecules, all known human sequences were downloaded from the IMGT/HLA database (see Sec. 1.1.2) and for the murine MHC molecule alleles a BLAST search was performed [4, 233]. All allele names and mutations were compared to those reported in the PDB or stated in the corresponding literature of the structures.

For the annotation of the TCR segment alleles, the web-based tool *IMGT/DomainGapAlign* was used, which utilizes the IMGT/Gene-DB (see Sec. 1.1.2) [77, 95]. The tool provides several results for a TCR α- or β-sequence query, such as sequence alignments to the most similar v- and the used j-segment alleles, an identification of the CDR loops, and a list of mutations with respect to the assigned segment alleles. The latter list allows for a localization of mutations in the secondary structure elements of the *immunoglobulin* (IG) like domains according to the naming scheme presented in Fig. 1.3 [163]. The TCR names found in the literature were double checked and extended names were assigned to some structures, if the names were ambiguous or mutations had not been considered. These names were assigned to the single lines and columns of the clustering heat map (*i. e.* to the different PDB Ids) in order to find correlations between the cluster affiliation and the corresponding TCR clonotypes. In particular, the analysis was performed to clarify, whether the differences in TCR inter-domain angles are an artifact of the crystallization process or are specific to an TCR type. Especially, several TCRs assigned to the type 2C by the literature were investigated in more detail for sequence differences, since this TCR type was the only one found in different clusters (see Text B.S1).

2.2.3 Grid Based Determination of the Center of Rotation

In this section, a method is described, which is used to determine and analyze a CoR_β and is illustrated in Fig. B.1, p. 124. The approach consists of two steps: In the first step a grid based method is applied in order to localize the CoR spatially.

In the second step, the previously localized CoR is assigned to residues found at this position in the TCR structures and then these residues are analyzed regarding their sequence position. The conservation at this position is investigated by applying sequence comparison methods to all known TCR alleles.

The grid analysis approach is a further feature of the program *TCRanalyze*, which was introduced in Sec. 2.2.2. For this analysis all Vβ cuboids in the set $\mathcal{S}_{all}^{O}$ were considered, including all different BUs per PDB identifier. The basic idea behind the approach was to find the most invariant point of the cuboids with respect to the rotation. For this purpose, grids were placed in all the Vβ cuboids in same manner, such that (i) each grid line, column, and plane is parallel to the cuboid edges, and (ii) the grid points are indexed in each grid in the same order. For each grid index, the pairwise distances with respect to all structures (*i. e.* BUs) were computed and then for each index the variance of the occurring distances was determined as:

$$\mathrm{var}(g_i) = \frac{1}{n^2-1}\sum_{k=1}^{n}\sum_{l=1}^{n}\left(\delta_S(g_{i,k},g_{i,l}) - \frac{\sum_{k=1}^{n}\sum_{l=1}^{n}\delta_S(g_{i,r},g_{i,s})}{n^2}\right), \tag{2.4}$$

where $\delta_S(g_{i,a},g_{i,b})$ is the Euclidean distance between the grid point $g_{i,a}$ with the index i in the grid of structure a and the grid point $g_{i,b}$ with the same index i in the grid of structure b, and n is the cardinality $|\mathcal{S}_{all}^{O}|$. The desired CoR is then the coordinate of the grid point with the lowest distance variance.

Then, the residues found closest to the CoR were determined for both chains of all structures and were located in an *multiple sequence alignment* (MSA) of all TCR alleles. For this purpose, all functional TCR variable segment alleles were obtained from the IMGT/Gene-DB (see Sec. 1.1.2) [95]. Then the sequences were aligned (separately for Vα and Vβ), using the program *MAFFT* in the localpair mode, a maximum number of refinement iterations of 1,000, and the Blosum62 substitution matrix [114, 143] . The residues located around the CoR were all aligned in the same columns of the two MSAs and thus they can be considered to be conserved. In the MSA not only sequences occurring in the structures but all known variable segments were used. This extension of data allows a statistical analysis of the distribution of amino acid types at the CoR, which is generally occupied by glutamine (Q) residues for both chains. However, the analysis also showed exceptional residues of a lower frequency, as shown in Tab. B.2, p. 130. The available structures were scanned for these exceptions for further structural analysis (see Fig. B.4, p. 131).

2.2.4 Analysis of the orientation between the TCR and its pMHC Ligand

In Chapter 4, a modeling approach to predict the orientation of the Vα and Vβ domains of TCRs will be developed based on the analysis results of Publication 2,

which is presented in the current chapter. Then the method will be extended to additionally predict the orientation of the pMHC ligands. The analysis of the pMHC orientations, which is required for the modeling, is not contained in Publication 2, but is part of Publication 3. The pMHCs analysis is, however, described in the current chapter in the following Sec. 2.2.4.1. Furthermore, in Sec. 2.2.4.2, the general directionality of the pMHCs docking angle is addressed. A method is provided, which systematically determines for a set of structures, whether the pMHCs ligand is flipped by 180° with respect to the other structures.

2.2.4.1 Adaption of the Cuboid-Method to the pMHC docking angle

A superpositioning template $T_\mu{}'$ of the MHC class α_1 and α_2 domains was prepared[1] as described in Sec. 2.2.1.2. Structures containing MHC class I molecules were selected from the data set $\mathcal{S}_{bound}$ (superposed to 2bnu Vα, see Sec. 2.2.1.2) and cuboids were placed by aligning the template $T_\mu{}'$. The method was not applied for MHC class II molecules, even though, in principle, there should not be any limitation. The tool *TCRAnalyze* was used in order to (i) apply a grid analysis for the determination of the MHC class I specific *Center of Rotation* of the pMHC ligand (CoR_μ) and in order to (ii) compute all pairwise EADs (compare Sec. 2.2.1.2). In this case unlike for the TCR domains, no clustering was performed and only a single average conformation of all structures was required for the pMHC. Those structure M was determined, for which the sum of all pairwise angle distances regarding all other structures was a minimum. Then the final superpositioning template T_μ was created in two steps. First, $T_\mu{}'$ was superposed to the structure M and second, the conserved residues of $T_\mu{}'$ were enclosed in a new cuboid, which had edges parallel to the 2bnu coordinate system and an extent of the MHC α_1 and α_2 domains. Finally, cuboids were placed around all separated MHC class I molecules by superpositioning T_μ onto them and were used later for the modeling in Chapter 4.

2.2.4.2 The MHC Flipping Problem

To answer the question whether, the MHC (-like) molecules (MHC class I, class II and CD1d1; further referred to as ligands) all bind to the TCR in the same orientation or if these ligands may contact the TCR in a flipped manner, a comprehensive set of 183 X-ray crystal structures of ligand bound TCRs was systematically analyzed. The method by Rudolph *et al.* for measuring MHC docking angles does not consider TCR inter-domain flexibility and the axis defining the MHC orientation is fitted into the binding groove without considering a directionality, as mentioned in Sec. 2.1.2 [237]. Here, a method is presented, which is similar to that by Rudolph *et al.*, but in contrast (i) uses cuboid centroids instead of the conserved cysteine

[1] This step was carried out by Antonia Stank within her Master's Thesis, which I supervised.

residues of the TCR and (ii) places the MHC axis similar to the cuboid placement. A focus of the presented method is the used data set, which was desired to be comprehensive and correct in the assignment of the different chain types. The latter criterion is often not fulfilled in the PDB due to mis-assignments. For the sake of completeness, it has to be noted that the analysis was performed before the structures 4y19 and 4y1a were published, showing the first structural proof of a flipped MHC docking state [25].

The investigated data set contains 111 MHC class I bound αβ TCRs, 33 MHC class II bound αβ TCRs, 33 NKT TCRs bound to a CD1d1 molecule, 4 αβ TCRs bound to the MHC class I related molecule MR1, and 2 structures of an MHC class I bound γ/αβ TCR (PDB Ids: 4qrr, 4wo4; [214]). A MSA of all murine and human, functional *T-cell receptor alpha variable* (TRAV) sequences obtained from the IMGT/Gene-DB was computed with the *MAFFT* tool, using the same parameters as described in Sec. 2.2.3 [95, 114, 143]. Based on this MSA, a consensus sequence (see Sec. D.2) was determined with *Ugene* [211]. The consensus sequence was used to perform a BLAST search in the PDB [4, 26, 27]. This search results in all TCR structures bound to MHC (like) molecules, but also unbound TCRs as well as other IG like molecules (*e. g.* ABs). The structures were split into their BUs and all BUs containing less than 2 or more than 5 protein chains were removed from the dataset. For the remaining structures the conserved domains were assigned to each protein chain of a BU [183]. The conserved domain information was used to filter the dataset for αβ TCRs bound to MHC class I- and MHC class II molecules, or MHC-like CD1d1 molecules. Structures containing SAGs were neglected (compare Sec. 2.2.1.1). Then for each chain the locus and allele of the V domain was determined with the IMGT/DomainGapAlign web tool (see also Sec. 2.2.2.1) [77].

All structures were superposed onto T_α and cuboids were placed around the TCR V domains with the same method as described in Sec. 2.2.1.2. To uniquely mark the direction of the ligand, a superpositioning template $T_{\mu\alpha 2}$ was prepared from the structure 1ao7, containing only the two helices of the MHC α_2 domain (the template residues are listed in Appendix D) [89]. This helix template, which is coupled to a direction vector, was superposed onto each Vα-aligned structure (MHC I and CD1d1: α_2 domain; MHC II: β_2 domain).

Then the two vectors $\vec{T}$ and $\vec{M}$ were determined (with respect to the 2bnu coordinate system), where $\vec{T}$ is the connecting vector from the Vα-cuboid center to the Vβ-cuboid center and $\vec{M}$ is the vector of the (transformed) helix templates C-terminal carbon atom to the N-terminal nitrogen backbone atom. Both vectors were projected into the xy-plane of the Vα-cuboid ($\vec{T}', \vec{M}'$) and then the cross product $\vec{C} = \vec{T}' \times \vec{M}'$ was computed. For the case that on the one hand all MHC (-like) molecules and on the other hand all the TCRs are found in the same main orientation one can expect: $(\forall_{i\in\mathcal{D}}(\vec{C}_i \cdot \overrightarrow{(0,0,1)} < 0) \oplus (\forall_{i\in\mathcal{D}}(\vec{C}_i \cdot \overrightarrow{(0,0,1)} > 0)$ where $\mathcal{D}$ is the

set of superposed structures; *i. e.* the sign of the cross product vectors z-component is expected to be either positive or negative for all investigated structures.

2.3 Results and Discussion

The results are presented in two parts. The first part covers the analysis of the TCR Vα/Vβ orientations as a summary of Publication 2, which is embedded into Appendix B of the present cumulative dissertation. The second part treats the results of the cuboid-method based analysis of the pMHCs docking angle. This analysis is contained in Publication 3, which is presented in Chapter 4 covering the modeling topic rather than the preliminary analyses. Furthermore, the result of an unpublished analysis of the pMHCs docking directionality is presented briefly, while the corresponding data is made available in the Appendix D.

2.3.1 Publication 2: Quantitative Analysis of the Association Angle between T-cell Receptor Vα/Vβ Domains Reveals Important Features for Epitope Recognition

The results of the cuboid-method applied on T-cell receptors in order to anlayze the orientations of their V domains was published in the peer reviewed open access journal PLoS Computational Biology:
Thomas Hoffmann, Angela M. Krackhardt, Iris Antes, (2015) *Quantitative Analysis of the Association Angle between T-cell Receptor Va/Vβ Domains Reveals Important Features for Epitope Recognition.* PLoS Comput Biol 11(7): e1004244. doi:10.1371/journal.pcbi.1004244
Submitted: November 6, 2014; Accepted: March 17, 2015; Published: July 17, 2015

In the following two sections the article is summarized and my contribution to this work is declared. The full article and the supplementary material are included in Appendix B (pp. 121) of this thesis.

2.3.1.1 Summary

The recognition of *major histocompatibility complex* (MHC) molecule presented peptide antigens by *T-cell receptors* (TCRs) is an initial process in the adaptive immune response (see Sec. 1.1 and Fig. B.1 A, p. 124). It is not elucidated yet, whether conformational changes of the TCR play a role in the not fully understood signal transduction pathway (see Sec. 1.1.4). A flexibility in the relative TCR variable α (Vα)/variable β (Vβ) inter-domain orientation could have an impact on the *antigen* (AG) recognition, which is indicated by the superpositioning of different TCR

structures and illustrated in Fig. B.1 B. A quantitative analysis of such shown TCR domain transformations is hampered by the high structural und sequential diversity of these receptors, requiring the development of a method for a precise measurement.

The focus of Publication 2 was the development of such an analysis method and the application to a broad set of TCR structures. The underlying concept of this analysis was to represent the highly diverse TCR V domains by simple geometric bodies, namely cuboids, in order to obtain a comparable form of the different TCR structures (see Fig. B.1 A-C, p. 124). In the analysis of different structures containing various unbound TCRs and receptors bound to several peptide:MHC combinations (see Tabs.B.S1, B.S5, and B.S4, p. 154 and pp. 160), the cuboid based representation allowed to determine the rotation of the TCR Vβ domain relative to the Vα domain expressed by Euler angles (see Sec. 2.1.3.1).

The pairwise *Euler angle distance* (EAD) is used as a metric in order to cluster the TCR inter-domain twist according to the extent of the rotations. Six significant clusters were found and verified using a bootstrapping approach (see Sec. 2.1.3 and Figs. B.3 and B.S1, pp. 129+167). An assignment of TCR types to the used structures based on detailed sequence- and mutation analyses (Tabs. B.S1 and B.S2, pp. 154) showed that TCRs of the same clonotype adopt similar relative orientations of their V domains. This correlation supports the assumption of a type specific feature rather than an observation of a crystallographic artifact. The comparison of bound and unbound TCR structures using this method (Figs. B.S2 and B.S3, pp. 168) reveals differences in the conformations of both states and leads to the conclusion that certain angles are stabilized by the bound ligand. This effect is illustrated in Fig. B.2 A, p. 125 and furthermore the Tabs. B.3 and B.4, pp. 133 provide angular differences in the bound and unbound states for selected TCR clonotypes.

A further analysis method based on cuboids was developed, which makes use of a grid approach and was used in order to detect a CoR of the relative TCR Vα/Vβ orientation (see Fig. B.1 D). Examining the environment of the found CoR, which is located at the interface between the Vα and the Vβ domains, revealed interactions between both TCR chains involving glutamine (Q) residues at this position (Tab. B.S3, p. 158). A MSA based analysis beyond the sequences of the used structures but gathering all available TCR alleles, quantified the conservation of Q residues to 89% and 98% for the Vα and the Vβ domain, respectively (Tab. B.2, p. 130). *I. e.* in few cases exceptions for the "Q-Q hinge" are possible and the obviously important interaction at this position is most likely compensated by the replacing residues. The current available structures give insight into some, however, not all possible substitutions, as illustrated in Fig. B.4, p. 124.

2.3.1.2 Contribution

I am first author of this publication having designed and performed all programming and analyzed the resulting data. I created all figures and tables used in all parts of the publication.

2.3.2 Analysis of the MHC orientation

Analyses of the pMHCs orientation in TCRpMHC complexes were performed beyond the scope of Publication 2, which only treated TCR Vα/Vβ domains. Sec. 2.3.2.1 provides the results of the Euler angle analysis of the pMHCs class I orientations needed for the TCRpMHC modeling in Chapter 4. Sec. 2.3.2.2 presents the analysis of the MHC docking angle directionality.

2.3.2.1 The MHC Class I Average Conformation and -CoR$_\mu$

In order to model the pMHC orientation in Chapter 4 the cuboid-method was adapted to this type of molecules (see Sec. 2.2.4.1). The structure exhibiting the pMHCs with the lowest Euler angle distance with respect to all other structures was determined to be the BU containing the chain identifiers R,S, T, and P of the structure with the PDB identifier 3e3q [138]. The center of rotation for the MHC class I molecules is located at the center of the bound peptide at position 31.2, -48.2, 34.5 Å according to the 2bnu coordinate system (see Fig. C.1, p. 218). Unlike the CoR$_\beta$ of the TCRs, which is located at conserved Q residues, the CoR$_\mu$ cannot be assigned to conserved residues of the binding groove formed by the MHC α_1 or α_2 domains.

2.3.2.2 The MHC Flipping Problem

Whether TCRs can bind MHC molecules in both directionalities of the diagonal binding mode is especially of interest for the modeling of the TCRpMHC complexes, because it might be necessary to consider also the flipping. As described in Sec. 2.2.4.2, a broad set of bound TCR structures was analyzed regarding the directionality of the pMHC (-like) ligand. Care was taken to extract as many structures as possible from the PDB and to assign the different chain types not by the possibly incorrect entries in the PDB files but instead by a comparison with annotated sequences. The precise reassignment was necessary, since a mis-assignment (*i. e.* swapping of α and β) of the TCR chains would have lead to a reverse result in the measurement. The results are provided in Tab. D.2, p. 242, listing the PDB identifier, the PDB chain identifiers assigned to the different chain types of the different molecules, the IMGT allele name of the TCR chains, the ligand type, the coordinates of both vectors, and their in-plane cross product z-component. The sign of all

entries in the last column does not differ for the 183 used structures, meaning that all ligands are bound with the same directionality. The recently published structures 4y19/4y9a were not available at the time of the presented analysis and were not considered in the data set. These structures, however, exhibit an exception from the findings as they provide evidence of a flipped pMHCs ligand [25].

Chapter 3

Molecular Dynamics – An Application Study

T-cell receptor structures were analyzed in Chapter 2 and it was shown that these heterodimeric molecules feature a diversity in their association angles. In Chapter 4, an approach for the prediction of this association angle will be developed, based on molecular mechanics *force field*s (FFs) and on the previously described features of *T-cell receptors* (TCRs). The present chapter introduces the theoretical aspect of these empirical FFs, using a practical application (*i. e.* DNA binding proteins) to illustrate these concepts. Sec. 3.1 introduces briefly DNA binding proteins in the context of epigenetics and then describes the molecular FFs and their application in the *molecular dynamics* (MD) approach. Sec. 3.2 describes the *in vitro* and *in silico* methodologies used in this collaboration project, which resulted in a publication included into the cumulative dissertation at hand and being the basis of the current chapter. A summary and my contribution to this publication is provided in Sec. 3.3.

3.1 Introduction

The context of the project described in this chapter is epigenetics. The first part of this introduction, Sec. 3.1.1, gives a glance on epigenetic DNA modifications and the related enzymatic mechanisms. The goal of the project is then defined in Sec. 3.1.2. In Secs. 3.1.3 and 3.1.4 the theoretical background for the computational methods used in this chapter are provided, namely molecular mechanics FFs and MD simulations.

3.1.1 Biological Background: 5-Methylcytosine and 5-Hydroxymethylcytosine are Epigenetic Markers

When the *International Human Genome Sequencing Consortium* released the results of the *Human Genome Project* in the year 2004, an amount of 20,000 to 25,000 protein-coding genes was estimated for *homo sapiens* [131]. These genes carry the information for all types of human cells. The regulation of the gene expression is achieved on the one hand by transcription factors and on the other hand by mechanisms silencing genes for a long term or even permanently. Gene silencing plays an important role in the differentiation of cells but also generally in the phenotypic character of a cell as a response to its environment. The field of *epigenetics* is defined by Russo *et al.* as:

> *The study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence.* [239]

Epigenetic modifications were found to be correlated with several diseases, such as various forms of cancer, Alzheimer's disease, and diabetes [79, 173, 223, 232, 269]. Therefore, epigenetics is a growing topic in the field of personalized medicine [226].

Methylation and demethylation of the DNA base cytosine in CpG island regions are key mechanisms of the heritable changes and multiple enzymes are involved in these processes [31, 235]. The subsequent sections briefly introduce the proteins taking part in DNA methylation (Sec. 3.1.1.1), the recognition of epigenetic signals for gene silencing (Sec. 3.1.1.2), and introduce the two modifications relevant for this work, *5-methylcytosine* (5mC) and *5-hydroxymethylcytosine* (5hmC) as well as other so far known modified bases (Sec. 3.1.1.3).

3.1.1.1 Mechanisms of Gene Silencing by DNA Methylation

The methylation of cytosine is catalyzed by methyltransferases [99]. The different enzymes are involved in two different so far known paths. The first path is the *de novo* methylation and is supported by the proteins *Dnmt3a* and *Dnmt3b* [165, 210]. In this process, new epigenetic information is written on the DNA as a response of the cell to the environmental conditions. The second path is the maintenance methylation and is supported by the protein *Dnmt1* [51, 167, 168]. In this process, epigenetic information is transferred from one DNA strand to the complementary strand after DNA replication. The maintenance methylation allows for an epigenetic memory.

3.1.1.2 Recognition of Silenced Genes

In a minor case only, the silencing of genes is achieved by methylation of promoter regions to prevent the binding of transcriptional factors [17]. Instead, the modified

DNA regions are actively recognized by specialized proteins, the *5-methylcytosine binding proteins* (MBPs), which stabilize silent chromatin states [245]. Currently three different families of MBPs are known, namely the *methyl-CpG binding domain* (MBD) family, the *Ubiquitin-like, containing PHD and RING finger domains* (Uhrf) protein family, and the Kaiso protein family. The Uhrf and the Kaiso families share the preference of binding hemimethylated DNA, in which only one strand of a CpG island is modified [11, 13, 35, 112]. This chapter focuses on the two proteins *Uhrf1* (also known as *NP95* or *ICBP90*) and *MeCP2* as representatives of the Uhrf family and the MBD family, respectively, which are both highly expressed in brain tissue [96, 156, 263]. It was reported that *MeCP2*, in contrast to *Uhrf1*, discriminates between methylated and hydroxymethylated DNA [270, 271]. Crystal structures revealed that *Uhrf1*, unlike *MeCP2*, makes use of a base flipping mechanism to bind the methylated DNA [11, 13, 112, 122]. *Uhrf1* was proposed to act as a cofactor in the *Dnmt1* maintenance methylation pathway (see Sec. 3.1.1.1) [1, 35, 250]. Several other interactions of *Uhrf1* were found attributing this molecule a central role, such as (i) histone tail modifications, (ii) the histone lysine transferase *G9a*, and (iii) the *de novo* DNA methyltransferases *Dnmt3a* and *Dnmt3b* (see Sec. 3.1.1.1) [146, 192, 236].

3.1.1.3 Different DNA Modifications and their Possible Role in Demethylation

5mC is often referred to as the fifth DNA base. In addition to 5mC, several other modifications of cytosine have been discovered. The oxidation of 5mC results in the sixth DNA base 5hmC. These two additional bases (*i. e.* 5mC and 5hmC) are in the focus of the current chapter. The oxidation of 5mC was shown to be catalyzed by *ten-eleven translocation* (TeT) enzymes during a demethylation process [135, 156, 264]. The role of 5hmC in the demethylation mechanism was discussed to be either passive or active [106, 132, 270, 292]. In these considerations, the oxidation of 5mC hides the epigenetic signal from the *Dnmt1* in the maintenance methylation process (Sec. 3.1.1.1). Several active demethylation pathways were suggested, which are either direct or indirect [194]. As such is the excision of the 5mC nucleotide making use of the DNA repair mechanisms an active indirect mechanism [93]. A further conceivable active demethylation pathway could be the enzymatic direct and iterative elimination of the methyl group, which is supported by the discovery of TeT proteins being able to convert 5mC to *5-Formyllcytosine* (5fC) and to *5-carboxylcytosine* (5caC) [136]. However, the identification of a decarboxylase directly converting 5caC to cytosine is still lacking [136]. Alternatively the oxidation products of 5mC are suggested to be involved in an indirect active pathway, in which the 5caC base is excised by the *thymine DNA glycosylase* (TDG) [113, 297].

3.1.2 Aim of the Project

The collaboration project resulting in Publication 1 (Sec. 3.3 and Appendix A, p. 103) combines *in vitro* and *in silico* methods in order to investigate the binding of methylated and hydroxymethylated DNA substrate to representatives of the MBD and Uhrf proteins (see Sec. 3.1.1). The goal of the *in vitro* experiments was to elucidate the preference of both proteins of binding methylated versus hydroxymethylated DNA. These experiments were carried out with fully modified and with hemi modified DNA substrates. The application of *in silico* approaches were intended to rationalize the experimental results of *Uhrf1* binding hemi modified DNA, containing either a 5mC or a 5hmC base (see Sec. 3.1.1.3) and to give structural insights about the binding pocket for the discovery of potential catalytic sites.

3.1.3 Empirical Molecular Mechanics Force Fields

A molecular mechanics FF describes the potential energy of a multi particle system as a function of the particle coordinates. The forces affecting each particle can be calculated as the derivative of the energy. A FF is defined by its functional form and by a set of empirical parameters, as further described in Sec. 3.1.3.1. In the context of molecular mechanics, the topology of a molecule (*i. e.* a particle system) is a set of descriptors for tuples of particles in the system. Such a descriptor includes a functional term and a set of parameters, describing a property of the particle tuple. In other words, the topology describes the connectivity of a system and it defines how the FF terms are applied on parts of the molecule. In the past decades many FFs for different purposes have been developed, differing in their functional form and parametrization. Sec. 3.1.3.2 briefly introduces a selection of them. For a more detailed insight about this broad topic I refer the reader to one of the many text books, such as *Molecular Modeling – Principles and Application* (Chapter 4) by Leach [161].

3.1.3.1 Functional Form

As described above, a molecular FF allows to calculate the interaction energy (and the resulting forces) between a set of atoms based on their coordinates. Generally, two main types of interactions are considered in molecular mechanics: To the first type of interactions belong the non-bonded interactions including van der Waals (vdW) interaction and Coulomb interaction (electrostatic). The second type describes bonded interactions and includes bond-stretch, bond-angle, and torsion energies. A general functional form of a FF describing a system can be defined as follows:

$$E_{pot} = E_{bond} + E_{angle} + E_{torsion} + E_{vdW} + E_{Coulomb}, \tag{3.1}$$

where E_{pot} is the total potential energy of the system and the other terms are detailed in what follows.

3.1.3.1.1 Bond Stretch Term. The bond stretch term is often modeled as a harmonic potential (Eq. 3.2), which is a good compromise between computational efficiency and accuracy. The use of the *Morse* potential would provide a more precise mathematical description of a bond, but is computationally to expensive for most applications. Some FF implementations find a trade-off between both cases. The harmonic description of the bond stretch term of Eq. 3.1 is formulated as follows:

$$E_{bond} = \sum_{b \in \mathcal{B}} \frac{k_b}{2}(d_b - d_{b,0})^2 \tag{3.2}$$

The $\mathcal{B}$ list contains all the bound atom pairs (A, B) and d_b is the distance between A and B. The constants k_b and $d_{b,0}$ are atom type dependent and respectively correspond to the potential force constant and to the reference distance for the pair (A, B).

3.1.3.1.2 Bend Angle Term. For a triple of three atoms A, B, and C in which B is covalently bound to both atoms A and C, the bend angle θ can be computed between the two bonds. The bend energy is usually computed as a function of the deviation between the actual angle and a theoretical or empirical optimum using a harmonic potential similar to the case of the bond stretching term (Sec. 3.1.3.1.1):

$$E_{angle} = \sum_{a \in \mathcal{A}} \frac{k_a}{2}(\theta_a - \theta_{a,0})^2 \tag{3.3}$$

The $\mathcal{A}$ list contains all the atom triples (A, B, C) for which the two covalent bonds $A-B$ and $B-C$ exist and θ_a is the angle between the two bonds. The constants k_a and $\theta_{a,0}$ are atom type dependent and respectively correspond to the force constant and to the reference angle for the triple (A, B, C).

3.1.3.1.3 Torsion Angle Term. For a tuple of four atoms A, B, C, and D and the three covalent bonds $A-B$, $B-C$, and $C-D$, the rotation angle ω of the bond $A-B$ relative to the bond $C-D$ according to the rotation axis $B-C$ is determined as the dihedral angle between the two planes ABC and BCD. The torsion term in Eq. 3.1 is generally based on the Pitzer potential of the form [220]:

$$E_{torsion} = \sum_{t \in \mathcal{T}} \sum_{m=1}^{M_t} \frac{k_{t,m}}{2}(1 + \cos(m\omega_t - \omega_{t,m,0})) \tag{3.4}$$

The $\mathcal{T}$ list contains bound atom triples (A, B, C, D) as described above and t is the torsion tuple index. The integer m denotes the multiplicity of the term and runs from 1 to M_t, which depends on the atom types in t. The multiplicity m determines the periodicity and $\omega_{t,m,0}$ is the phase shift angle. The constant $k_{t,m}$ is the energy barrier from one minimum to another.

3.1.3.1.4 Electrostatic Interactions, which belong to the non-bonded type, are computed according to Coulomb's law:

$$E_{Coulomb} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \tag{3.5}$$

For a system composed by N particles, the electrostatic potential is computed for each particle pair (i, j). The charges of both particles of a pair are given by the constant parameters q_i and q_j, the distance between the particles is denoted as r_{ij}, and ε_0 is the dielectric constant. The computational effort grows with N^2 and only minor energy contributions of pairs with a large distance are accounted. Thus, for large systems (*e. g.* containing thousands of solvent molecules) cut-off distances are used very often.

3.1.3.1.5 *Van der Waals* Interactions are modeled by the empirical *Lenard-Jones* (LJ) potential, which allows a rapid computation of these interactions for large systems:

$$E_{vdW} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(\frac{n}{n-m} \left(\frac{n}{m} \right)^{m/(n-m)} \varepsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^n - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^m \right) \right) \tag{3.6}$$

Setting $n = 12$ and $m = 6$ leads to the widely used LJ 12-6 potential:

$$E_{vdW} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right) \right) \tag{3.7}$$

For a system composed by N particles, the LJ potential is computed for each particle pair (i, j). The distance between i and j is given by r_{ij}. Here the parameters ε_{ij} and σ_{ij} depend on the types of both atoms/particles i and j. The parameter ε_{ij} determines the well depth (*i. e.* the minimum at $r_{ij,min}$) of the function term for one pair. The parameter σ_{ij} defines the distance between the two atoms, for which the energy contribution is zero. The r^{-12} term models repulsive interactions and contributes with a high energy for the overlap cases ($0 \leq r_{ij} < r_{ij,min}$), whereas the r^{-6} term models attractive interactions and asymptotically approaches zero. Thus, since long-range interactions only provide small energy contributions, cut-off

distances are often used to reduce the computational efforts (compare Sec. 3.1.3.1.4).

3.1.3.2 Different Force Fields

The first molecular FF appeared after the introduction of the molecular mechanics methods in the late 1950s [2]. Over the decades several FFs were developed. To the most popular biomolecular FFs count the AMBER FF [55], the CHARMM FF [179, 180], the OPLS FF [139, 141], and the GROMOS FF [212]. The development of these FFs is continued: *e. g.* in contrast to the CHARMM22 FF the CHARMM27 knows additional atom types and parameters to handle lipids and nucleic acids. The FFs share the general functional form provided in Sec. 3.1.3.1, but some differ in details of the functional form. *E. g.* in the CHARMM FFs the additional Urey-Bradley cross term for angle bending and a further cross term treating protein backbone dihedral angles are present. Furthermore, the FFs or even the different versions within one FF family differ in their parameters, which are not interchangeable. Force fields can be classified into (i) all atom FFs, which treat each atom as a separate particle, (ii) united atom FFs, which summarize carbon atoms and their bound uncharged hydrogen atoms to one particle, and (iii) coarse grained FFs, which treat groups of several atoms as beads. All atom FFs allow the highest precision to the cost of compute time, whereas coarse graining facilitates the application of molecular mechanics methods to very large systems. The Gromos FF is of the united atom type, whereas the OPLS FF and the AMBER FF are available in the all atom (AA) and the united atom variant, and recent versions of the CHARMM FF are only available in the all atom variant. The Martini FF is a representative a of coarse grained FF [184].

3.1.4 Molecular Dynamics Simulations

The X-ray crystallography technique allows detailed insight into the structure and the function of proteins. A limitation of this method is that the resolved molecule is shown in a quasi frozen crystal state snapshot, instead of a solvated state at room or body temperature. Furthermore, the protein conformation is not necessarily the same at each temperature and the protein functionalities are often affected by conformational changes. To assess a better understanding of proteins, it is thus crucial to consider and study their dynamics. Such studies are possible *in vitro* using *nuclear magnetic resonance* (NMR) spectroscopy for small proteins in a short time scale. For *in silico* applications, Alder and Wainwright introduced the MD simulation technique based on a simple potential function to study the motion of particles in the late 1950s [2]. A milestone was the first MD simulation of a folded protein by McCammon *et al.* [189]. Over the past 40 years, the MD technique was further improved together with the computational power to allow the simulation of

large protein complexes for long time scales. Nevertheless, the basic principle of recent simulations remains the same, namely solving Newtons equations of motion for each particle i of a system as:

$$\vec{F_i} = m_i\vec{a_i} = m_i\frac{\mathrm{d}\vec{v_i}}{\mathrm{d}t} = m_i\frac{\mathrm{d}^2\vec{r_i}}{\mathrm{d}t^2}, \tag{3.8}$$

where $\vec{a_i}$, $\vec{v_i}$, and $\vec{r_i}$ are are respectively the acceleration, velocity, and position vectors of the particle i, $\vec{F_i}$ is the force applied on i, and m_i is the mass of this particle. Furthermore, the force is the spatial derivative of the potential energy U of a system, which can be obtained using a molecular mechanics FF (Sec. 3.1.3):

$$\vec{F_i} = -\frac{\mathrm{d}U_i}{\mathrm{d}\vec{r_i}} \tag{3.9}$$

Starting from an initial set of coordinates and velocities, the position of each particle can be determined as a function of time. Practically, finite difference methods are required to integrate the equations of motion, as briefly described in Sec. 3.1.4.1. Further aspects of MD simulations, such as heat- and pressure control, as well as periodic boundary conditions are described in the Secs. 3.1.4.2 and 3.1.4.2. Finally, a selection of implementations is presented in Sec. 3.1.4.3. For deeper insight in the theory the reader is referred to text books about this topic, *e. g.* Chapter 7 of *Molecular Modeling - Principles and Applications* by Leach [161].

3.1.4.1 Integrators

According to Equations 3.8 and 3.9, the coordinates and velocities of multiple coupled particles are deterministically defined by a given set of initial conditions for an arbitrary time T in the future. However, although analytical solutions for n-body problems are theoretically available for particle systems interacting only with $1/r$-terms (*e. g.* Eq. 3.5), practical general analytical approaches of solving this problem are not available and thus numerical methods are preferred for the integration [224].

Several integration methods for MD simulations were developed, sharing a basic concept of dividing the total time T into discrete time steps. The particle positions and velocities at a time step $t + \Delta t$ are computed from the acceleration of prior time t, using Taylor expansion. The approaches make use of the approximation that the forces, which are required to determine the acceleration of a particle, are approximately constant along a small time period Δt.

The *Verlet* algorithm computes the position $\vec{r_i}(t+\Delta t)$ of a particle i for the time step $t+\Delta t$ as a function of the coordinates at two prior time steps t and $t-\Delta t$ and in dependency of the acceleration $\vec{a_i}(t)$ at time step t as follows [276]:

$$\vec{r_i}(t + \Delta t) = 2\vec{r_i}(t) - \vec{r_i}(t - \Delta t) + \Delta t^2\vec{a_i}(t) \tag{3.10}$$

The Verlet algorithm bears two limitations: First, velocities are not explicitly computed and are estimated from two time steps when needed for the calculation of the kinetic energy. Second, the algorithm is numerically imprecise due to the addition of small and large numbers.

A later variant of this algorithm is the *velocity Verlet* method, which computes for time step $t+\Delta t$ both the coordinates and the velocities, without the numerically problematic addition of big and small numbers [262]:

$$\vec{r_i}(t+\Delta t) = \vec{r_i}(t) + \Delta t\vec{v_i}(t) + 1/2\Delta t^2\vec{a_i}(t) \tag{3.11}$$

$$\begin{aligned} \vec{v_i}(t+\Delta t) &= [\vec{v_i}(t) + 1/2\Delta t\vec{a_i}(t)] + 1/2\Delta t\vec{a_i}(t+\Delta t) && (3.12)\\ &= \vec{v_i}(t+1/2\Delta t) + 1/2\Delta t\vec{a_i}(t+\Delta t) && (3.13) \end{aligned}$$

Due to dependencies, the evaluation requires three steps: In the first step, $\vec{r_i}(t+\Delta t)$ is computed ($\vec{v_i}(t)$ and $\vec{a_i}(t)$ are already known from the previous integration cycle) according to Eq. 3.11. In the second step, the two terms in Eq. 3.13 are determined by (i) computing $\vec{v_i}(t+1/2\Delta t) = \vec{v_i}(t) + 1/2\Delta t\vec{a_i}(t)$ and (ii) updating the forces according to the new coordinates, leading to $\vec{a_i}(t+\Delta t)$. Finally, Eq. 3.13 is evaluated.

Other variants of the Verlet integration scheme are (i) the *Leap-Frog* algorithm [124], which computes the velocities and the coordinates in an asynchronous manner avoiding the numerically problematic operations, and (ii) the computationally more expensive, but more precise *Beeman* integration [20]. Recently, Michels and Desbrun proposed a semi-analytical integrator [195]. The authors claim that their algorithm is in average 30 times faster than the Verlet scheme and allows for larger Δt.

3.1.4.2 Further Aspects

Technically, the FF and the integrator are the core of a MD simulation. Modern simulations consider many other aspects, which are briefly outlined in this section. A detailed description is out of the scope of this thesis, but can be found *e. g.* in the book *Molecular Modeling and Simulation: An Interdisciplinary Guide* (Chapter 13) by Schlick [246].

MD simulations are usually carried out in a simulation box with *periodic boundary conditions* (PBC), *i. e.* particles leaving the box at one border re-enter the box on the opposite side. PBC allow the particles to migrate in a pseudo infinite space. The size of the periodic box is chosen according to the non-bonded cutoff values (see Secs. 3.1.3.1.4 and 3.1.3.1.5) preventing interactions between the simulated molecule and its virtual copy. The use of a cut-off leads to an abrupt change in the potential function. One technique to keep the potential continuously differentiable is to

apply a switching function. The *particle mesh Ewald* (PME) method allows a fast and accurate summation of point charge potentials especially for simulations using periodic boundary conditions [61].

The pioneering MD simulations of biological macromolecules were carried out *in vacuo* totally neglecting solvent effects (*e. g.* Ref. [189]). To allow more realistic conditions, nowadays counter ions and solvent molecules are artificially added in simulation setups or implicit solvent models are used [150]. In explicit solvent simulations, the thousands of added water molecules significantly contribute to the total compute time of the simulation. To reduce the computational effort, the internal degrees of freedom of water molecules are rigidified and the intermolecular interactions are accounted between charged sites, as introduced with the simulation of liquid water by Stillinger and Rahman [259]. The *simple point charge* (SPC) and the *transferable intermolecular potential 3 points* (TIP3P) water models are representatives of the three-interaction-site models located at the three atoms [23, 140]. Computationally more expensive is the TIP5P model, which extends the TIP3P model by two interaction sites representing the oxygen lone pairs [182]. The rigidification of the bond geometries in the water molecules can be achieved by the application of constraints using the algorithms SHAKE-, the SETTLE-, or the LINCS [118, 202, 240, 294].

Isobaric-isothermal (also referred to as NPT ensemble) simulations require the control of the pressure and the temperature. Simple approaches scale the velocities of the particles in defined time intervals to adjust the temperature [290]. A more sophisticated method to maintain the temperature is the *Berendsen thermostat*, which couples the temperature to an external heat bath [24]. Another approach is the *stochastic randomization* method by Andersen, which simulates random collisions with particles of a virtual heat bath [5]. Furthermore, the *extended system* approach (also known as the *Nose-Hoover-thermostat*) incorporates a reservoir of kinetic energy by an additional friction term into the Newtonian equations of motion [128, 209]. The *Langevin thermostat*, as implemented in the program *NAMD*, extends the equations of motion by a friction term and a random force [215]. Barostats analogously rescale the coordinates to adjust the volume of a system using a pressure bath or utilize the extended system method, such as the *Langevin piston* method [5, 24, 80, 185].

3.1.4.3 Implementations

Various software packages for MD simulations of biomolecules are available using different FFs (see Sec. 3.1.3) and a comprehensive description of all the tools is beyond the scope of this thesis. The most widely used programs are: *CHARMM* (Chemistry at HARvard Molecular Mechanics) [37], *AMBER* (Assisted Model Building with Energy Refinement) [288], *GROMACS* (GROningen MAchine for Chemical Simulations) [22], and *NAMD* (Nanoscale MD) [207]. In the context of this thesis

the DynaCell program [6] is used, which is introduced in Sec. 4.1.4, p. 58.

The *CHARMM* program was initially released in 1983 for MD simulations with the same-named FF, but more recent versions also include many other FFs [38]. In 1984, MD code based on the *GROMOS* program was included into the *AMBER* package [213]. The focus of AMBER is mainly the implementation of the same-named FF, but recent versions also allow the usage of other FFs, such as CHARMM [57]. The *GROMACS* software was introduced in 1994 covering the functionality of the *GROMOS* software [22]. The slogan *fast, flexible, and free* accounts for clever code optimizations and a bunch of implemented FFs in this open source software [275]. The *NAMD* tool, which was first released in 1996, was from the beginning designed as a massively parallel implementation using the CHARMM FF [207]. Beside the high performance and massive scalability (*i. e.* up to several hundreds or even thousands of CPUs) the modularity of the code is another design goal of the project [207]. The modularity allows the code to be easily extended and this is achieved by an object oriented design and the use of the popular programming language C++, as well as an interface to a scripting language [215].

3.1.5 Complex Formation Energy Calculation

It is often required to compute binding affinities of a protein to a ligand if experimental data is not available. MD (Sec. 3.1.4) simulations allow the conformational sampling of complex components in the free and in the bound state to provide the energies of these sample frames. Several methods use an average over multiple samples, *e. g.* the *linear-response approximation* (LRA), *molecular mechanics with Poisson-Boltzmann* (or *general Born*) *and surface area solvation* (MM/PBSA or MM/GBSA), and the *linear interaction energy* (LIE) method [8, 9, 154, 162, 254].

The LIE method is widely used, as this approximation only takes the intermolecular interactions into account and requires two MD simulations, one for the complex in the bound state and one for the free ligand. In the following, this method is described in more detail, since it was used to estimate the binding affinities of modified DNA to *Uhrf1* (Sec. 3.2.3.1):

$$\Delta G_{bind} = \alpha \left(\left\langle E_{ele}^{L-S} \right\rangle_{PL} - \left\langle E_{ele}^{L-S} \right\rangle_{L} \right) + \beta \left(\left\langle E_{vdW}^{L-S} \right\rangle_{PL} - \left\langle E_{vdW}^{L-S} \right\rangle_{L} \right) \tag{3.14}$$

Here, $\left\langle E_{ele}^{L-S} \right\rangle_{PL}$ is the average over multiple trajectory frames of the electrostatic interaction energy of the whole complex, whereas $\left\langle E_{ele}^{L-S} \right\rangle_{L}$ accounts for the averaged electrostatic energy of the free ligand (see Sec. 3.1.3.1.4). Analogously, $\left\langle E_{vdW}^{L-S} \right\rangle_{PL}$ and $\left\langle E_{vdW}^{L-S} \right\rangle_{L}$ take the vdW interactions of the complex or of the free ligand into account (see Sec. 3.1.3.1.5). The weighting factors α and β may require a problem specific adaption [8, 282].

3.2 Methods

The described project combines *in vitro* and *in silico* approaches in order to investigate the binding behavior observed in DNA:protein complexes containing methylated or hydroxymethylated DNA. The used assays are briefly outlined in Sec. 3.2.1, but the main focus of this Chapter targets the used MD simulations of these biomolecules (Sec. 3.2.2) and on the computational evaluation (Sec. 3.2.3).

3.2.1 *In Vitro* Binding Assays

Two different assays were used to measure binding affinities of DNA binding proteins towards modified or unmodified substrates, namely an *electrophoretic mobility shift assay* (EMSA) and *pull-down DNA binding assay* (PDBA) [86, 236]. For the experiments unmodified, hemimodified, and fully modified fluorescence labeled DNA substrates were produced as well as a CpG-free DNA double strand (see Tab. A.S1, p. 112). The modified substrates either carried 5mC or 5hmC bases (see Sec. 3.1.1.3) at the palindromic CpG site and were tagged with different fluorescent labels. Constructs of *Uhrf1*, the *Uhrf1*-SRA domain, and the *MeCP2* MBD (see Sec. 3.1.1.2) domain either labeled with *green fluorescent protein* (GFP) or with *yellow fluorescent protein* (YFP) were expressed in transfected human embryonic kidney cells (HEK293T) [36, 72, 86, 236].

Using the EMSA method, the binding specificity and binding affinity were measured for all three constructs and the following substrate pairs in direct competition: (i) hemimodified methylated and hydroxymethylated DNA and (ii) fully modified methylated and hydroxymethylated DNA. The wells of a polyacrylamide gel were filled with the same amount of each of the two competing substrates and an increasing amount of the G/YFP-protein construct. In an electric field, the DNA:protein complexes show a lower mobility in the gel than the free DNA and thus, shifted bands for the protein, the complex, and the free DNA substrate can be expected. The two different fluorophores allow a distinction between the two substrates in the band of the complex and of the free DNA. The ratios of both substrates were determined by detection with a laser scanner.

Based on the PDBA method, the binding specificity of *Uhrf1* (GFP construct) towards two different substrate pairs was measured in direct competition: (i) no GpG *vs.* hemimethylated DNA and (ii) unmethylated DNA *vs.* hemimethylated DNA. The protein-GFP construct was immobilized on beads and incubated with an equimolar mixture of both differently labeled substrates. After the incubation, the immobilized GFP-protein construct was washed to remove unbound substrate. To calculate the ratio of the substrates relative to the protein-GFP concentration, the fluorescence intensities were measured for GFP and for the two different fluorophore

tags bound to the DNA. To avoid possible bias induced by the labels, the experiments were performed multiple times with substrates tagged by alternative fluorophores.

3.2.2 Simulation Setup

To investigate the dynamics and to compute binding affinities of the *Uhrf1:SRA*-domain bound to methylated or hydroxymethylated DNA (see Sec. 3.1.1), I carried out MD simulations (see Sec. 3.1.4) for the protein bound to both variants of the DNA as well as simulations of unbound DNA. The latter simulations were required for binding energy calculations, as described in Sec. 3.2.3.1. The trajectories were further evaluated regarding hydrogen bond patterns for both substrates, as described in Sec. 3.2.3.2. All simulations were based on the X-ray crystal structure with the *Protein Data Bank* (PDB) Id 3fde, which contains the *SRA*-domain bound to double stranded hemimethylated DNA [26, 112]. The transformation of the structure into the hydroxymethylated form, the separation of the DNA, as well as other preparatory steps are described in Sec. 3.2.2.1. For all simulations the *NAMD* software (Version 2.7b1) with the CHARMM22/27 parameter set was used, since this combination allows for highly scaled simulations of proteins and DNA (see Secs. 3.1.4.3 and 3.1.3.2) [179, 180, 215]. The models were carefully heated up before the MD simulations were performed, as described in Sec. 3.2.2.1.

3.2.2.1 Preparation

The preparation of the model consists of several steps: (i) removing of facultative molecules, (ii) modeling of missing atoms or modifications, (iii) solvation and ionization, (iv) creation of the systems topology, and (v) energy minimization.

All auxiliary molecules (such as EDO) required for the crystallization of the complex, as well as conserved water molecules and ions were deleted from the PDB file 3fde [112]. The *biological unit* (BU), also referred to as *crystallographically independent molecule in the asymmetric unit*, containing the protein chain A and its ligand was selected and all the other BUs were discarded. Additionally, the DNA double strand was separated for the simulation of the unbound DNA.

Using X-ray crystallography, single or multiple atoms can often not be resolved due to blurred electron density maps. Furthermore, hydrogen atoms cannot be resolved by this method. Missing protein residues can be modeled and even mutated using side-chain-placement tools, as *IRECS*, *SCWRL* (see also Sec. 1.2) [110, 157]. A further tool for this purpose is *PSFgen*, which belongs to the *NAMD* software package [215]. This tool was used to model missing hydrogen atoms, incomplete protein residues, and to transform the methylated DNA into hydroxymethylated DNA according to the patches presented in Tab. A.S3, p. 113, and to create the topology of the complex. This table also contains the FF parameters for the modified

residue 5hmC, which were adopted from the values of a serine residue.

Before the complex was solvated, an energy minimization was carried out for 50,000 steps using the *steepest descent* (SD) algorithm *in vacuo*. To solve the problem of poorly placed atoms, this method translates the atom positions according to the interaction forces (more details about energy minimization are provided in Sec. 4.1.2.1).

Subsequently, using the *VMD solvate* module, water molecules were placed around the target (*i. e.* either the protein:DNA complex or the unbound DNA) with a minimum distance of 2.4 Å [129]. The target was rotated to keep the simulation box as small as possible and the box size was chosen to cover the target with a water layer of at least 7.5 Å diameter, leading to a final system of approximately 24,000 atoms. In the topology file, the added molecules were recorded according to the TIP3P water model (see Sec. 3.1.4.2) [140]. Then charges were neutralized by replacing water molecules with sodium or chloride ions using the *VMD* command *ionize* in two alternative ways: In the first setup, only neutralization took place, and in the second setup[1] additionally the NaCl concentration was set to 0.5 M [129]. In this process, a minimum distance between an ion and the target of 5.0 Å was adhered.

Finally, the systems were again energy minimized as described above. The simulations were partitioned into two phases, a heat-up phase and a subsequent production phase. In the starting structure tension forces might occur due to the artificial placement of atoms and due to uncertainties in the original crystal structure. A simulation of the energy minimized systems directly at the target temperature of 300 K might lead to artifacts. Thus, the systems were slowly heated up to the target temperature, as described below. In the following two paragraphs, first, the general simulation conditions and then the heat-up conditions are provided (see also Sec. 3.1.4.2):

The simulations were executed using periodic boundary conditions with PME summation for long range electrostatic interactions. The non-bonded cut-off and the switching distance were set to 14 Å and 12 Å, respectively. The bond lengths were constrained using the SHAKE algorithm allowing the used integration step size of 1 fs, for which the *velocity Verlet* scheme was used (see Sec. 3.1.4.1) [202, 240, 262]. The production run was performed in the NPT ensemble using the *Langevin piston* method (target pressure: 1.01325 bar, oscillation time scale: 100 fs, damping time scale: 50 fs) for the pressure- and *Langevin* dynamics for temperature maintenance (damping coefficient 5/ps), while hydrogen atoms were not coupled to the heat bath [80, 185]. Additionally, for the simulations of the unbound DNA, the backbone was restrained, to keep the modified bases in the flipped state.

The heat-up protocol consists of ten subsequent simulations with different condi-

[1] In this setup a larger simulation box was used with a minimum water layer of 15.0 Å. The larger setups approximately contain 43,000 atoms.

Table 3.1: **Heat-Up Protocol**

T[a]	Thermostat[b]	Barostat[c]	Force[d]	Steps[e]
0	R	n/a	1000	10,000
5	R	n/a	1000	10,000
10	R	n/a	1000	10,000
20	R	n/a	1000	10,000
50	R	n/a	1000	20,000
100	R	n/a	1000	20,000
100	R	n/a	100	30,000
200	R	n/a	100	50,000
200	LD	LP	n/a	30,000
300	LD	LP	n/a	50,000

[a] Target temperature [K].
[b] Thermostat: rescaling (R) or *Langevin* dynamics (LD).
[c] Barostat: none (n/a) or *Langevin* piston (LP).
[d] Restraining force constant in [kcal mol^{-1} Å^{-2}].
[e] Number of time steps.

tions, as summarized in Tab. 3.1. The first heat-up simulation was performed under constant volume (NVT), harmonic restraints were applied to all atoms of the complex, while the velocities were rescaled every single step. In the transitions from one to the other heat-up simulation the restraints were loosened and the target temperature was increased. At a temperature of 200 K, restraints were completely switched off (exception: in the simulations of the unbound DNA the restraints were kept on the backbone). At the same temperature the ensemble was changed to NPT conditions by switching the *Langevin* piston barostat and *Langevin* dynamics on, as described above. The temperature was further increased to 300 K.

3.2.2.2 Simulation

The simulations of the complexes and the unbound DNA systems were performed for 57 ns and 15 ns, respectively. For all production runs of the complexes, the *root mean square deviation* (RMSD) and the *root mean square fluctuation* (RMSF) were computed to monitor the stability of the simulations using the *VMD* tool (see Figs. A.S3 and A.S4 in Appendix A) [129]. The RMSF was computed after superpositioning of the trajectories onto the protein in the heated starting structure for two time windows ($30-40$ ns and $50-57$ ns).

3.2.3 Trajectory Evaluation

The MD simulations were performed to theoretically confirm and to rationalize the *in vitro* measurements of the binding affinities between the *Uhrf1:SRA* domain and DNA substrates containing alternatively 5mC or 5hmC. The computation of the

binding free energy is described in Sec. 3.2.3.1. Furthermore, Sec. 3.2.3.2 outlines the hydrogen bond network analysis, which was applied in order to find structural explanations for the affinities.

3.2.3.1 Binding Energy Difference Estimation

The binding free energies $\Delta G_{bind,5mC}$ and $\Delta G_{bind,5hmC}$ of the *SRA* domain bound to either hemimethylated or hemihydroxymethylated DNA was computed using the LIE approach (see Sec. 3.1.5) [9]. To this end, for each substrate two simulations were carried out containing either the DNA:protein complex or the isolated DNA, as described in Sec. 3.2.2.1. The trajectory analysis feature of *NAMD* to compute interaction energies between groups was used on groups defined as follows [215]: The first group contained the modified nucleotide, the 3'-, and the 5'-neighbor. The second group contained a shell of residues forming the binding pocket. The pocket was defined as all the residues containing at least one atom with a distance less than 15.0 Å to the modified nucleotide. These residues were determined directly after the heat-up for both the 5mC and the 5hmC simulation, and then both lists were combined. The selection of atoms close to the binding site reduces fluctuations of the energies induced by flexible and solvent exposed residues. Neglecting these parts of the molecules is valid, since we were only interested in the energy difference $\Delta\Delta G_{bind} = \Delta G_{bind,5mC} - \Delta G_{bind,5hmC}$ between the two complexes. The computed energies were averaged over the final 10.0 ns of the trajectories and inserted into Eq. 3.14, using 1.0 and 0.5 as values for the scaling parameters α and β, respectively.

3.2.3.2 Hydrogen Bond Network Pattern Analyses

The formation of hydrogen bonds was monitored over the simulation time for several donor-acceptor (D-A) pairs in the binding pocket. For this purpose, the *hbond* plugin of *VMD* was used with thresholds of 20.0° and 3.0 Å [129]. The occurrences of hydrogen bonds over the simulation time were plotted coincidently for all D-A pairs. This type of plot allows for an easy comparison of the hydrogen bond patterns for the two substrates and facilitated the identification of a conserved water molecule.

3.3 Results and Discussion

3.3.1 Publication 1: Recognition of 5-Hydroxymethylcytosine by the Uhrf1 SRA Domain

The results of the so far described study were published in the peer reviewed open access journal PLoS ONE:

Carina Frauer*, Thomas Hoffmann*, Sebastian Bultmann, Valentina Casa, M.

Cristina Cardoso, Iris Antes, Heinrich Leonhardt, (2011) *Recognition of 5-Hydroxymethylcytosine by the Uhrf1 SRA Domain.* PLoS ONE 6(6): e21306. doi:10.1371/journal.pone.0021306
Submitted: March 18, 2011; Accepted: May 25, 2011; Published: June 22, 2011
The authors marked with an asterisk contributed equally.

In the following two sections the article is summarized and my contribution to this work is declared. The full article and the supplementary material are included in Appendix A of this thesis.

3.3.1.1 Summary

Besides the four bases adenine, cytosine, guanine, and thymine, several modified variants of these components occur in eucaryotic DNA. Methylated cytosine (*5-methylcytosine*, 5mC) is known to play a role in the control of gene expression in an epigenetic context and is often denoted as the "fifth base". Oxidation of 5mC leads to the sixth base *5-hydroxymethylcytosine* (5hmC), which was recently proven to be contained in genomic DNA. However, the role of the sixth base is still unclear and is discussed to be an independent epigenetic signal or a metabolite in the (de)methylation pathway. So far three different protein families of 5mC binding proteins are known for the recognition of methylated DNA, namely the MBP-, the Uhrf-, and the Kaiso-protein families. Only the Uhrf family seems to preferably bind hemimethylated DNA substrate in contrast to the two other families (see Sec. 3.1.1 for details). This collaborate study combines *in vitro* and *in silico* methods to investigate the binding behavior to methylated and hydroxymethylated DNA of an Uhrf representative and an MBD representative.

The *in vitro* experiments (Sec. 3.2.1) show that the SRA domain of *Uhrf1* binds both, hemimethylated and hemihydroxymethylated, DNA substrates have a similar affinity, whereas the MBD domain of *MeCP2* has a strong preference in favor of methylated substrates (see Fig. A.1, p. 105; Fig. A.S1, p. 115; Fig. A.S2, p. 116).

For the *in silico* part of the publication I modified an X-ray crystal structure of hemimethylated DNA bound to the *Uhrf1* SRA domain, by adding a hydroxy group to the cytosine (Sec. 3.2.2.1), with the result that the modified residue fits into the binding pocket (Fig. A.2, p. 106). I carried out several *molecular dynamics* (MD) simulations based on the original crystal structure and on the hydroxy-modified model (Sec. 3.2.2), in order to (i) compute the SRA:5mC and SRA:5hmC binding affinity using the *linear interaction energy* (LIE) approach (Sec. 3.2.3.1) and (ii) to investigate the dynamics of the hydrogen bond network pattern in the binding pocket for both variants: (i) The binding energy calculations are consistent with the *in vitro* experiments and show that hemihydroxymethylated DNA binds to the SRA domain with a similar or even slightly better affinity compared to hemimethylated

DNA. (ii) The hydrogen bond network of the SRA:5hmC complex compared to the SRA:5mC complex is less fluctuating over the simulation time (Fig. A.3, p. 107). This stabilization of the complex is achieved by the formation of additional hydrogen bonds between the hydroxy group of the 5hmC base and a water molecule at a conserved position (Fig. A.4, p. 108), which is also indicated in the original crystal structure (Fig. A.S5, p. 119). Since migration events of water molecules into protein cavities occur rather rarely in MD simulations and since ions influence water dynamics and the DNA conformation, I repeated the simulations with different ion concentrations. These replicas of the computations revealed reproducibility of the hydrogen bond patterns and the migration of a water molecule to the conserved position (Fig. A.S6, p. 120).

3.3.1.2 Contribution

I am co-first author of this publication having designed and performed all *in silico* experiments and analyzed the resulting data (molecular modeling of the 5hmC modification, MD simulations, hydrogen bond network analysis, binding energy calculations). I contributed the following figures and tables: Figs. 2, 3, 4, S3, S4, S5, and S6 and Tab. S3. I was not involved in the *in vitro* experiments (EMSA and PDBA).

Chapter 4

Prediction of the TCR Inter-Domain Angle and the TCRpMHC Complex Domains Orientations.

The complex formation between *T-cell receptors* (TCRs) and *peptide-MHC molecule complex* (pMHC) ligands was described in Chapter 1 as an initial event in the recognition of pathogenic peptides by the adaptive immune system. A broad structural analysis of these highly diverse complexes was provided in Chapter 2 and showed that the complex components specifically adopt different relative orientations. Furthermore, this analysis localized a *Center of Rotation* (CoR) for the TCR variable domains at the bifurcated interaction of two conserved glutamine (Q) residues. As a further basis of the current chapter, the methodology of molecular mechanics *force field*s (FFs) was introduced for the computation of intermolecular forces and of binding energies (Chapter 3). The focus of the present chapter is the development of the prediction method *DynaDom* based on the so far gained insights. This method predicts the inter-domain association angles and other features of TCRs and TCRpMHC complexes based on an energy minimizing approach and on a *rigid body* (RB) optimization method. The DynaDom method is required, since so far, these properties have not been explicitly taken into account in special homology modeling approaches targeting this immunologically relevant complex.

The introductory Sec. 4.1 reviews the state of the art in TCRpMHC modeling, provides the technical background of the DynaDom method, and describes a further problem complicating the modeling. In Sec. 4.2, the DynaDom method is first exclusively developed for association of TCR variable domains in absence of the pMHC ligand. Then the modular implementation is extended for the prediction of TCRpMHC complexes. Test calculations show that accurate models are predicted in 96% and 74% of the cases for TCRs and TCRpMHC complexes, respectively. Further tests revealed that improvements of the pMHC preplacement procedure

could increase the prediction rate. Sec. 4.3 presents the results as a summary of the Publication 3.

4.1 Introduction

In Sec. 4.1.1, the currently available methods for the modeling of TCRs and TCRpMHC complexes are reviewed, among which none explicitly takes the TCR inter-domain orientation flexibility into account. Then, in Sec. 4.1.2, the concepts of energy minimization are provided in general and for the special case of the RB optimization approach, which is implemented in DynaDom. In Sec. 4.1.3, the *NQ-flipping* problem is defined, which had to be taken into account in the context of the conserved Q residues at the CoR. In Sec. 4.1.4, finally the program *DynaCell* is briefly introduced, which hosts the new DynaDom prediction method.

4.1.1 TCRpMHC Modeling Approaches

The first structural model of a TCRpMHC complex was built by Almagro *et al.* in 1995, before any X-ray structure of a TCR or a TCRpMHC complex was elucidated[1] [3]. In this work a combination of homology modeling (Sec. 1.2) and *molecular dynamics* (MD) (Sec. 3) techniques was used to propose an atomistic model of a TCRpMHC complex of a 5c.c7 TCR exposed to the MCC peptide presented by the murine I-E^k *major histocompatibility complex* (MHC) molecule[2].

Later, several experimental structures were studied using molecular dynamics simulations [58, 59, 65, 151, 152, 196, 234, 256, 278, 279, 299, 300]. In two of these studies, 172 different peptides of known immunogenicity were modeled into a complex of a LC13 TCR and HLA-B*08:01 MHC class I molecule by side chain substitution [151, 152].

Michielin *et al.* used the first TCRpMHC complex crystal structure as template for homology modeling [197]. In this study they applied the *MODELLER* program (Sec. 1.2) as well as simulated annealing techniques to model the structure of the T1 TCR bound to the photoreactive PbSC peptide and to the murine K^d MHC class I molecule [197, 241]. The photoreactive feature of the peptide allows covalent crosslinking of the complex, which was used to directly determine the association constant of fifty different mutants of this system [145]. The properties of nearly all of these mutations were rationalized by the homology model [197]. The

[1] The structures with the *Protein Data Bank* (PDB) IDs 1tcr and 1ao7 were resolved in 1996 and 1998 [89, 90].

[2] Interestingly, two structures (PDB IDs: 4p2q and 4p2r) of this complex containing two different peptides were recently published and could give insight into the quality of this pioneering modeling [32].

same group released the tool *TCRep 3D* and applied it in the field of rational TCR design [166, 301]

Pierce and Weng released the method *TCRflexDock*, which uses the Monte Carlo based *RosettaDock* protocol in order to model TCRpMHC complexes and reached a prediction rate of 80% on a benchmark of 20 structures [104, 217]. This method was recently applied to predict models of TCRs bound to MHC-like ligands, such as CD1 and MR1 [219]. In this study the use of multiple docking starting positions improved the performance of the prediction. The same group used the scoring function *ZAFFI* and the *Rosetta interface mutagenesis tool* for the rational 100- and 400-fold enhancement of TCR binding affinities of the A6 TCR to the TAX:HLA-A2 and the DM5 TCR to the ELAGIGILTV:HLA-A2 complex, respectively [108, 218, 251].

4.1.2 Rigid Body Energy Minimization

The DynaDom method for the prediction of inter-domain association angles adopted in TCRpMHC complexes, which is developed later in this chapter (Sec. 4.2), is for the most part based on a *RB optimization* technique. In the following sections first (Sec. 4.1.2.1) the *energy minimization* methodology is introduced in a general manner emphasizing the most commonly used minimization algorithms in the field of molecular mechanics (see also Chapter 3). After that, the fast RB optimization approach used for the prediction of the inter-domain angles is presented (Sec. 4.1.2.2).

4.1.2.1 Energy Minimization Approaches

The following introduction of energy minimization algorithms used in the field of molecular mechanics is mainly based on the two following books: *Molecular modelling: principles and applications* by Leach and *Practical methods of optimization* by Fletcher [84, 161]. Generally a multi-dimensional nonlinear minimization is the problem of finding the parametrization $\vec{X}_{min}$ for a minimum solution from all possible solutions of an n-dimensional objective function $f(\vec{X})$:

$$\vec{X}_{min} = \underset{\vec{X}\in\mathbb{R}^n}{\arg\min}\, f(\vec{X}) \Leftrightarrow f(\vec{X}_{min}) = \min_{\vec{X}\in\mathbb{R}^n} f(\vec{X}) \tag{4.1}$$

A distinction is made between zero$^{\text{th}}$, first, and second order approaches for solving the problem with respect to the degree of derivative used. The Newton-Raphson method is a second order approach and the *steepest descent* (SD) algorithm, the *conjugate gradient* (CG) algorithm, and the *BFGS* alogorithm belong the first-order derivative methods and are described below. Zero$^{\text{th}}$ order approaches, as the *simplex* method, are out of the scope of this thesis. In the context of energy minimization of molecular systems, the objective function to be minimized is the energy of the

system and the first-order derivative is the negative net force (see Sec. 3.1.3) in dependency to Cartesian coordinates. Furthermore the various approaches differ in their convergence behavior and in their robustness against numerical inaccuracy of the objective function and the choice of starting parameters.

Generally, the Newton-Raphson method determines a solution for $f(\vec{X}) = 0$ by the iterative application of:

$$\vec{X}_{n+1} = \vec{X}_n - \frac{f(\vec{X}_n)}{f'(\vec{X}_n)} \tag{4.2}$$

An optimization problem of dimension d is solved by the Newton-Raphson method by solving $f'(\vec{X} = (x_1, \ldots, x_d)) = 0$ and thus the second-order derivative is required, which is for a n-dimensional function the computationally expensive Hessian matrix:

$$\mathcal{H}_{k,l} = \frac{\partial^2 f}{\partial x_k \partial x_l} \tag{4.3}$$

In the SD algorithm the parameter set $\vec{X}_{i+1}$ of the next iteration step is determined by moving along the gradient:

$$\vec{X}_{i+1} = \vec{X}_i - \lambda_i \frac{\vec{g}_i}{|\vec{g}_i|}, \tag{4.4}$$

where $|\vec{g}_i|$ and $\frac{\vec{g}_i}{|\vec{g}_i|}$ denote the norm and the unit vector of the gradient $\vec{g}_i$, and λ_i defines the local step size, which is determined by a line search. The SD algorithm features a fast convergence behavior but tends to oscillate close to the minimum.

To avoid this oscillatory behavior, the CG algorithm is often used close to the minimum after application of the SD algorithm. The search direction of the next step $\vec{v}_{i+1}$ is determined by taking the direction of the previous step $\vec{v}_{i-1}$ into account and the parameters of step $i+1$ are computed as:

$$\vec{X}_{i+1} = \vec{X}_i + \vec{v}_i, \tag{4.5}$$

where $\vec{v}_i = -\vec{g}_i + \gamma_i \vec{v}_{i-1}$. For molecular systems the computation of the scaling factor γ_i is usually computed as defined by Polak and Ribiere [221].

The algorithm denoted as *BFGS* was independently developed by C.G. Broyden, R. Fletcher, D. Goldfarb, and D.F. Shanno [40, 83, 97, 249]. This quasi-Newton approach avoids the direct expensive computation of the inverse *Hessian* matrix approximating its inverse ($\tilde{\mathcal{H}}_i^{-1}$) and features a fast convergence behavior of theoretically d iteration steps for convex functions but is less robust than other approaches. The set of parameters for the next iteration step is computed as:

$$\vec{X}_{i+1} = \vec{X}_i - \lambda_i \tilde{\mathcal{H}}_i^{-1} \vec{g}_i, \tag{4.6}$$

where λ_i is the local step size.

Various programming libraries exist in which these algorithms are implemented focusing on different aspects, such as general or special applicability, speed, or the dimension of the problem. The present thesis did not focus on a new efficient implementation of a minimizer, but the interest was rendered on the application of minimization approaches for the modeling of TCRpMHC structures. Thus, the *GNU scientific library* (GSL) was chosen as it provides a general interface for the easy switching between various implemented algorithms such as SD, CG, and BFGS [100].

4.1.2.2 The Method of Mirzaei

In the context of the present thesis, a RB is defined as an arbitrary set of atoms, on which a transformation can be applied conserving all distances between all atoms of the RB, *e. g.* a domain of a protein. A RB optimization minimizes the energy of a system in which parts of the system are transformed as an RB. Mirzaei *et al.* proposed the *RB energy minimization on manifolds for molecular docking*, which outperforms most other available approaches [200]: The method is based on the mapping of the search space onto a six-dimensional manifold by exponential coordinates. Then, a local only six-dimensional Euclidean search space can be treated by standard optimization methods as described in Sec. 4.1.2.1, which is the (limited memory variant of the) (L)BFGS [175]. The six-dimensional objective energy function is formulated in dependency of a three-dimensional translational parameter $\vec{\vartheta}$ and a three-dimensional rotational parameter $\vec{\omega} = (\omega_1, \omega_2, \omega_3)$:

$$E(\vec{\omega}, \vec{\vartheta}) = E\left(\exp([\vec{\omega}])(\vec{q}_1 - \vec{p}) + \vec{p} + \vec{\vartheta}, \ldots, \exp([\vec{\omega}])(\vec{q}_m - \vec{p}) + \vec{p} + \vec{\vartheta}\right), \tag{4.7}$$

where

$$[\vec{\omega}] = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}$$

is the cross product matrix, m is the number of atoms of the RB, $\vec{p}$ is the initial center of rotation, and $\vec{Q} = (\vec{q}_1, \ldots, \vec{q}_m)$ are the initial positions of the atoms. The gradient computation is based on the transformation according to the Rodrigues formula, which is used for the computation of the exponential function of an antisymmetric matrix:

$$\frac{\partial \exp([\vec{\omega}])}{\partial \omega_i} = \frac{\partial}{\partial \omega_i}\left(\frac{\sin|\vec{\omega}|}{|\vec{\omega}|}[\vec{\omega}] + \frac{1 - \cos|\vec{\omega}|}{|\vec{\omega}|^2}[\vec{\omega}]^2\right) \tag{4.8}$$

4.1.3 The NQ-Flipping Problem

X-ray crystallography is a method for the structure elucidation of molecules and is based on electron density maps into which the coordinates of known atoms are

placed. A limitation of this method is that oxygen and nitrogen atoms are very similar in their electron density leading to possible misinterpretation. This ambiguity particularly affects the placement of asparagine and glutamine (N and Q, naming the problem) side chains, since they contain rotatable carboxamide groups in which oxygen and nitrogen are neighbors. The position of both atoms can be swapped by a rotation of the group by 180°. A correct determination of the flipping state is often only possible by taking the hydrogen bond network into account. Four different studies agreed in the estimate that these residues were resolved in the wrong flip state in approximately 20% of all cases [127, 190, 285, 291]. Such misplacements are assumed to influence the results of MD simulations (see also Chapter 3), protein engineering, and molecular docking [127]. Due to the significant impact of the NQ-flipping problem, its correction has been addressed by the development of several software tools [16, 30, 47, 62, 127, 159, 170, 174, 190, 285–287, 291]. Since the *Center of Rotation* of the variable β domain (CoR_β) of the TCR is formed by the bifurcated interaction between two Q residues of each chain, Q misplacements at this position are likely to influence the modeling results (see also Chapter 2 and Fig. C.3 B+C, p. 220) [125]. Thus, in the context of the present thesis two of the available tools were used, namely *Reduce* and *Protoss* [30, 47, 62, 174, 291].

The Reduce tool is part of the structure analysis suite *MolProbity* and serves for the placement of hydrogen atoms to X-ray structures [47, 62, 291]: Besides the optimization of the hydrogen bond network the included NQ-flip correction additionally takes van der Waals (vdW) clashes into account by *contact dot surfaces*. Technically the identities of the oxygen and the nitrogen atom are swapped to preserve the correct bond lengths.

Protoss was used in the present thesis, since it is the newest available tool for structure protonation and NQ-flip correction [30, 174]: The tool is very fast, since it benefits from a dynamic programming approach and a docking scoring function.

Unfortunately, none of the tools can be meaningfully applied on the single components of the TCR, because in the modeling process the context of the Q residues at the CoR_β may change due to the association or transformations of the variable domains. Both methods are however used as a reference in order to allow the development of an approach, which corrects the NQ-flips simultaneously in the modeling process of the domain association.

4.1.4 The DynaCell Program

The program suite *DynaCell* by Antes is a molecular modeling and molecular dynamics tool, which hosts several methods, such as the side chain placement and mutagenesis approach *IRECS* as well as the molecular dynamics based approach for flexible docking *DynaDock* [6, 110]. A key feature of the program is the *optimized*

potential molecular dynamics (OPMD) method. The OPMD allows to perform molecular dynamics simulations with initial overlap by optimizing (see Sec. 4.1.2.1) scaling factors, which smoothly switch the non-bonded terms (see Sec. 3.1.3) of the molecular FF on. A docking run consists of the broad sampling of the ligand in the binding pocket of the receptor and of a subsequent OPMD based refinement of the poses.

The software provides the implementation of several FFs and served as a framework for the development of the new DynaDom method for the prediction of TCRpMHC inter-domain association angles.

4.2 Methods

In Chapter 2, TCR and TCRpMHC structures were analyzed and classified in six different significant clusters of variable α (Vα)/variable β (Vβ) orientations. Furthermore, the analysis pointed out that the variable TCR domains share a center of rotation stabilized by conserved glutamine residues. In the following Sec. 4.2.1 a modular method is introduced, which takes the results of the analysis into account in order to predict the orientations of only the TCR Vα and Vβ domains. The method is named as DynaDom. This method is based on RB optimization and makes use of a molecular force filed (see Chapter 3). In Sec. 4.2.2 the method is evaluated by applying it on a broad set of TCR structures for remodeling. Building on this, Sec. 4.2.3 shows the extension of the modular method in order to additionally take the pMHC orientation into account. Several aspects of the extended approach had been evaluated on a broad set of TCRpMHC structures.

4.2.1 The Prediction Pipeline for the Vα/Vβ Orientations

The modeling method for the prediction of the TCR variable domain orientations follows two main steps. First, different models of the TCRs are assembled with the different orientations found in the clustering analysis. Second, the different preliminary models are refined using a FF based RB optimization. The optimization approach was implemented in a modular manner, which derives several advantages:

1. The modular implementation allows for the sole treatment of specific parts of the protein for transformation operations as well as FF based calculations.

2. A multistage pipeline can be assembled, which takes different subsequent processes of the modeling into account, such as the subsequent execution of a skillful preplacement of domains, a partial optimization of the target complex and finally the full optimization.

3. The single pipeline stages (further referred to as pipeline steps) can be built by the arrangement of an arbitrary number of modules at the runtime. For instance, a pipeline step could consist of a concurrent optimization of a RB transformation together with a placement of side chains.

4. Different optimization algorithms can be applied in the different pipeline steps, such as simple sampling approaches or various multi-dimensional energy minimizers.

In the following Sec. 4.2.1.1, the pipeline used for the prediction of the TCR Vα/Vβ orientation is outlined. Sec. 4.2.1.2 briefly introduces the modular components used for the pipeline.

4.2.1.1 Pipeline Overview

The pipeline for the prediction of the TCR Vα/Vβ inter-domain orientations consisted of six steps. This pipeline was in turn the basis of the extended seven-stage prediction method for the orientations in ternary TCRpMHC complexes, which is described in detail in Sec. 4.2.3.1. The six-stage TCR modeling pipeline computed one model for each of the eleven starting conformations, which were derived from the six significant clusters found in the analysis provided in Chapter 2 (For more details about the starting conformations see Text C.S1, p. 222). Finally, the models were ranked according to their interaction energy, as illustrated in Fig. C.S1, p. 235. The OPLS-all atom (AA) FF (see Sec. 3.1.3) was used in the different pipeline steps for all energy and forces calculations [139, 141]. Besides an RB optimization of the TCR Vβ domain, an additional correction/adaption of the Q residues found at the CoR_{β} (see Sec. 2.3) was performed. This correction was necessary, since these important interacting residues bridging the two variable TCR domains were incorrectly resolved in a flipped state in many structures.

The seven steps of the modeling pipeline are briefly described in the following and are illustrated in Fig. 4.1. More details about the pipeline can be found in the Methods part of the embedded Publication 3, pp. 180.

Step 1 of the pipeline started with the assembly of the complex from the two variable domains in the so called *zero orientation* (see Sec. 4.2.2.1 below for details).

Step 2 In the second step, the flip states of the Q residues were corrected separately, *i. e.* interactions with the opposed chain were not taken into account. The rotation was corrected by a binary sampling, rotating the carboxamide groups by 180°, and a subsequent refinement of the rotation with an energy minimizer. After the Q-optimization, the FF energies were calculated independently for each domain for later use in Step 6.

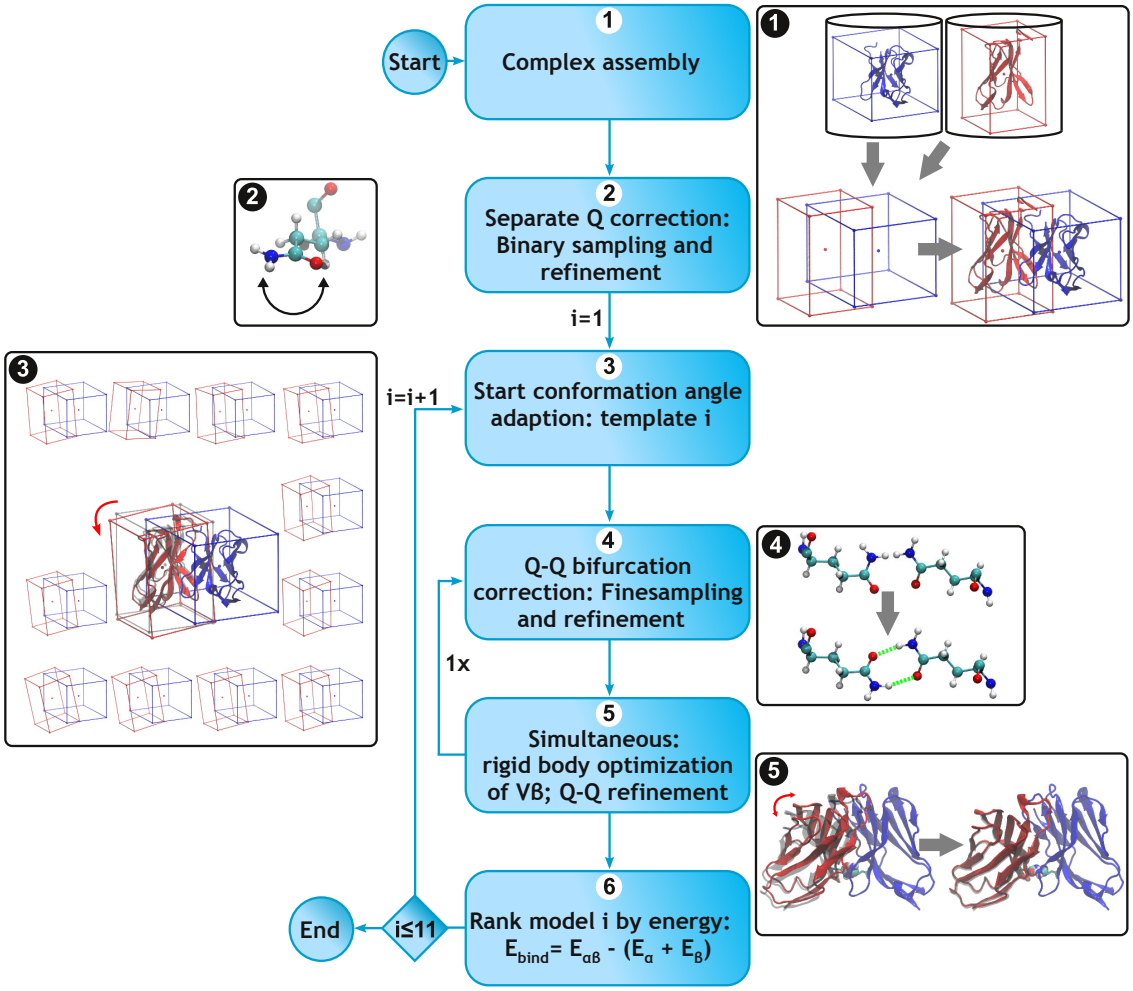


Figure 4.1: **Prediction Pipeline for the Vα/Vβ Orientations.** (1) The complex was pre-assembled from separated variable domains Vα and Vβ (blue and red) in the *zero orientation* and then (3) transformed to one of eleven starting orientations (*cluster representatives*). (5) The variable domains were treated by a *rigid body* (RB) optimization while the conserved glutamine (Q) residues were (2,4, and 5) optimized in their interaction and restrained in their distance. (6) For each cluster representative a model was computed and the best model was selected according to the energy rank.

Step 3 The initial orientation of the TCR Vβ domain was set according to one of the template orientations defined by the clustering.

Step 4 was similar to Step 2, as the Q side chains were sampled and refined again. In contrast to the previous separated Q-correction, the sampling was performed using a smaller discrete angular step size and interactions between both domains were taken into account instead of only intra-chain interactions. This step allows for the formation of a bifurcated interaction between the two carboxamide groups of both TCR chains.

Step 5 was the actual RB optimization of the TCR Vβ domain. The optimization took place while the glutamine-glutamine (Q-Q) interaction at the CoR_β was kept rotatable and additionally restraints were applied to avoid the variable domains to drift apart. The RB optimization is interrupted after few iterations and Step 4 is invoked again, in order to allow the Q-Q interaction to adapt to larger initial translations or rotations of the RB Vβ.

Step 6 follows after the continuation of the interrupted Step 5. The precomputed energies of the single domains (Step 2) were subtracted from the FF energy of the full complex in order to obtain the interaction energy.

Loop The Steps 3 to 6 were repeated for each of the eleven starting conformations obtained by the clustering. Finally the top ranking model was selected according to the interaction energy.

4.2.1.2 Algorithmic Modules

The *DynaDom* method introduced in the present thesis extends the *DynaCell* program (see Sec. 4.1.4) by a modular approach for protein structure prediction based on a multi-dimensional optimizer [6]. The modularity of the implementation allows to tailor problem specific modeling pipelines, which take particular properties of a protein class into account for the structure prediction. Such pipelines are realized by using different modules, which are assembled at the runtime. The modules are classified into five different categories, namely (i) *assembly* modules, (ii) *flow control* modules, (iii) coordinate altering *subprocess operator* (SO) modules, (iv) coordinate retaining coupled *auxiliary* SO, and (v) modules for the *evaluation*. The following paragraphs in the current section intend to provide an overview on the modules used for the modeling of TCR Vα/Vβ orientations. Technical details on the different modules can be obtained in Text C.S2, p. 223. The additionally required modules for the treatment of pMHC ligands in the ternary TCRpMHC complexes are described elsewhere in Sec. 4.2.3.2, while the current section focuses on the basic pipeline for modeling binary TCR Vα:Vβ structures.

The class of flow control modules includes program components, that apply other modules of a pipeline step on a problem. A problem of a pipeline step and its dimensionality is defined by the sequence of applied SOs. With DynaDom, two different modules for flow control are implemented, namely a multi-dimensional sampling module and a multi-dimensional minimizer module. The sampling module applies all permutations of values from a discrete range and determines the combination of values for which the multi-dimensional objective function (*i. e.* the energy) is the lowest. The multi-dimensional minimization module allows for an application of all

optimization algorithms implemented in the GSL (see Sec. 4.1.2.1) on the problem of a pipeline [100].

SO modules contribute to the objective (energy) function and provide a gradient for each of the parameters on which they operate. The gradient and energy computations are based on a molecular FF (see Chapter 3) and for each SO a distinct area is taken into account for the interactions. Some SOs are capable of directly manipulating atom coordinates, whereas other SOs are auxiliary modules, which are coupled with other modules and manipulate their energy and gradient contributions.

For the modeling of TCR Vα/Vβ domain orientations in absence of a pMHC ligand, three different SOs are employed. The first SO is the *RMZ* operator module, which is the core of the method (for details see p. 223). This operator is based on the work by Mirzaei *et al.* (see Sec. 4.1.2.2) and allows for the six-dimensional RB optimization of single domains [200]. The energy and gradient contribution of this SO requires only non-bonded interactions (see Sec. 3.1.3.1) between RBs. The second SO module is the restraint operator *RPS* (details on p. 224), which is coupled to the *RMZ*. The *RPS* operator monitors a defined point of a RB to stay in a distinct range by manipulating the energy contribution and the translational gradient of the coupled *RMZ* operator. In the case of the TCR Vβ domain RB optimization the distance between the conserved Q and the CoR_β is restrained. The third SO, the *AQR* operator, is responsible for the correction and adaption of the Q carboxamide group orientation at the CoR_β (details on p. 224). This SO applies a rotation of the nitrogen and oxygen atoms around an axis. The energy routines compute the interaction between the rotated atoms and their environment using bonded and non-bonded energy terms (see Sec. 3.1.3.1). For the one-dimensional gradient of the rotation angle the forces of the affected atoms are combined to a tangential force (for details see p. 223).

Furthermore an initialization module and an evaluation module are part of the modeling pipeline. The initialization module transforms the complex from the zero position to an arbitrary orientation and is used to create the starting orientation according to the cluster representatives. The evaluation module is used to compute the FF energies of separate components of the complex as well as of the whole protein in order to determine the binding energy, required for the ranking of the models.

4.2.2 TCR Inter-Domain Prediction Pipeline Evaluation

4.2.2.1 Structural Data Set

75 *biological unit*s (BUs) were selected originating from 48 different crystal structures from the set of MHC bound TCRs ($\mathcal{S}_{bound}$) as already used for the analysis presented in Chapter 2. Structures with unresolved atoms in the variable do-

Table 4.1: **Different Orientations in the Structural Data Sets for the Modeling of TCRs and TCRpMHC Complexes** (Sections 4.2.2.1 and 4.2.3.3)

Data Set	Vα Ori.[a]	Rel. Vβ Ori.[b]	Rel. pMHC Ori.[c]	Size[d]
$\mathbf{DS_{TC}}$	2bnu	crystal	n/a	75
$\mathbf{DS_{T0}}$	2bnu	zero/2bnu	n/a	75
$\mathbf{DS_{CC}}$	2bnu	crystal	crystal	53
$\mathbf{DS_{M0}}$	2bnu	crystal	zero/3e3q	53
$\mathbf{DS_{C0}}$	2bnu	zero/2bnu	zero/3e3q	53

[a] All Vα domains were aligned to the reference structure 2bnu.
[b] Crystal: Vβ orientation is only affected by the Vα superpositioning.
Zero: The Vβ domain were artificially placed according to the orientation found in the reference structure 2bnu.
[c] n/a: Not available due to removed pMHC.
Crystal: pMHC orientation were only affected by the Vα superpositioning.
Zero: The pMHC domains were artificially placed according to the orientation found in the reference structure 3e3q/RSTP after superpositioning to the reference structure 2bnu Vα domain.
[d] Number of different biological units contained in the data sets.

mains were not taken into account in order to avoid bias in the modeling process, even though standard side chain placement or modeling tools such as *MODELLER*, *PSFgen*, *SCWRL*, or *IRECS* could easily be used for construction of the missing heavy atoms (see Sec. 3.2.2.1) [110, 157, 215, 241]. These 48 different TCRpMHC crystal structures represent 17 murine and 31 human TCRs of 22 different TCR clonotypes (mutations not accounted). Details are listed in Tab. C.S1, pp. 227.

All the BUs were superposed onto conserved residues of the Vα domain as described in Sec. 2.2.1, p. 22. The 75 structures contained only the TCR variable domains, whereas the TCR constant domains, all MHC domains as well as bound peptides, and all non-protein atoms were removed from the structures. Hydrogen atoms were added and topologies were created for the OPLS-AA FF (see Sec. 3.1.3.2, p. 41), using the *pdb2gmx* tool (Version 4.5.6) [119, 139, 141].

Two different variants of this data set were used. In the first variant, DS_{TC}, the orientations of the Vβ domains relative to the Vα domains were kept as found in the corresponding original crystal structures. In the second variant, DS_{T0}, the Vα and the Vβ domains were transformed to the orientation of the reference structure 2bnu [46]. The orientation used in the latter set is further referred to as *zero orientation* of the TCR. In Tab. 4.1, both data sets for the TCR modeling without the pMHC are listed together with the sets for the TCRpMHC complex modeling (following in Sec. 4.2.3.3).

4.2.2.2 Evaluations

The pipeline for the prediction of the TCR Vα/Vβ orientation was applied in two different categories of *evaluation setting* (ES). The first category is further referred

to as *main* and the second one as *control*. For both categories the influence of the concurrent *glutamine carboxamide group optimization* (QO) and of the application of *positions restraints* (PR) on the prediction accuracy was evaluated. For this purpose four different variants of each ES category were used: (i) QO and PR were both disabled, (ii) either QO or (iii) PR was switched off, and (iv) QO as well as PR were enabled. The different variants are listed in Tab. D.1, p. 240. Two different success criteria were used for determine the prediction rates, as explained in detail below.

The main evaluations settings, M_T, were used to assess the performance of the pipeline according to the methods prediction capability by remodeling the TCRs. For each of the 75 BUs of the data set DS_{T0}, 11 models were built and evaluated by two success criteria $\mathcal{C}_\mathcal{R}$ und $\mathcal{C}_\mathcal{E}$ (see below). In these main ES, the Vβ domains were moved to the eleven different angular starting conformations of Vβ. Afterwards, the positions of the Vβ domains were optimized. The goal of these experiments was to discriminate between the eleven different predicted conformations after RB energy minimization. A prediction was considered as successful according to criterion $\mathcal{C}_\mathcal{E}$, if the conformation with the lowest inter-domain interaction energy belongs to a predicted structure with an alpha carbon *root mean square deviation* (RMSD) less than 2.0 Å with respect to the crystal structure. The criterion $\mathcal{C}_\mathcal{R}$ accounts success, if at least one model with an RMSD below 2.0 Å was found for a BU. The prediction rates according to the criteria $\mathcal{C}_\mathcal{R}$ und $\mathcal{C}_\mathcal{E}$ were determined for the full set DS_{T0} and for a reduced set DS^*_{T0}. In the reduced set DS^*_{T0} only one BU per experimental structure was taken into account.

The control ES C_C, were used to test the stability and robustness of the approach by applying the optimization steps of the modeling pipeline to the structures in their original experimental conformations (data set DS_{TC}). The goal of the control ES was to confirm that the domains in the crystal conformation do not escape from their original structure after application of the approach. In accordance with the success criterion $\mathcal{C}_\mathcal{R}$ used in the main ES, a structure was considered to have escaped from its starting crystal conformation, if the alpha carbon RMSD with respect to the crystal structure exceeded 2.0 Å. The control ES differs from the main ES in the Pipeline Steps 1 and 3 (Fig. 4.1). In the first step the crystal orientation was used instead of the zero orientation. The third step was skipped and only one model was built. In order to examine the influence of the QO and the PR on the general achieved model quality, the evaluation setting M_T for modeling the TCRs was carried out in different combinations in which the features were switched on or off (*QO*, *PR*, *QO+PR*, *none*). In the analysis described in the following the RMSD values with respect to the experimental structures were computed for all eleven models build for the complete set of BUs in the test sets. Three different subsets of the data set DS_{T0} were used: (i) the subset *mispaired* contained all the cases, in which a

mispaired interaction of the glutamine residues at the CoR_{β} was identified in the experimental structures (see column *Pairing state – ES* in Tab. C.S2, p. 229), (ii) the subset *paired* contained the opposite cases in which the glutamine residues were identified to be paired, and (iii) combines both cases. For all models of the sets (i) to (iii) the quality was assessed by accounting RMSD values below a threshold (1, 2, and 3 Å) in each of the four feature categories *QO*, *PR*, *QO+PR*, and *none*. Alternatively the RMSD values of all created models were partitioned in ranges of 0.1 Å to be plotted in a histogram.

4.2.3 Extension of the Pipeline for the Prediction of Domain Orientations in TCRpMHC Complexes

The prediction pipeline for TCR Vα/Vβ inter-domain association angles based on the DynaDom method was introduced and tested in Secs. 4.2.1 and 4.2.2. In what follows this pipeline is now extended in order to concurrently predict the orientations of the TCR variable domains as well as of the pMHC ligand in the whole TCRpMHC complex. First, this extension is generally described in Sec. 4.2.3.1, followed by naming the required additional modules in Sec. 4.2.3.2. Finally, the performance of the extended approach and particular parts of the pipeline are exhaustively assessed in Sec. 4.2.3.3.

4.2.3.1 Pipeline Overview

The pipeline for the modeling of TCR Vα/Vβ association angles in absence of the pMHC ligand was described before (Sec. 4.2.1.1) with six stages: (1) complex assembly, (2) single glutamine correction, (3) start orientation adaption, (4) glutamine bifurcation adaption, (5) restrained RB optimization with concurrent Q-Q bifurcation adaption, and (6) model energy evaluation. The modular design of the DynaDom method allowed to easily extend the pipeline by additional steps or to extend such pipeline steps by additional SOs. To additionally take account of the pMHC orientation in the TCRpMHC complex, the assembling step and the RB optimization step necessitated modifications. Furthermore, after the step adapting the TCR Vβ orientation to one of the cluster orientations, a preplacement step for the pMHC ligand was required (compare Figs. 4.1 and 4.2). This preplacement of the pMHC performed two concurrent optimizations starting from the FFs called *zero orientation* (see Secs. 2.3.2.1 and 4.2.3.3.1 below). The one optimization transformed the ligand along an perpendicular axis running through the *Center of Rotation* of the pMHC ligand (CoR_{μ}) (Sec. 2.3.2.1) while the other optimization rotated the ligand around this axis. This procedure prevented clashes between the TCR and the pMHC in the starting orientation for subsequent RB optimization. Furthermore, the procedure reduced initial repulsions of the receptor and its ligand due to opposed charges.

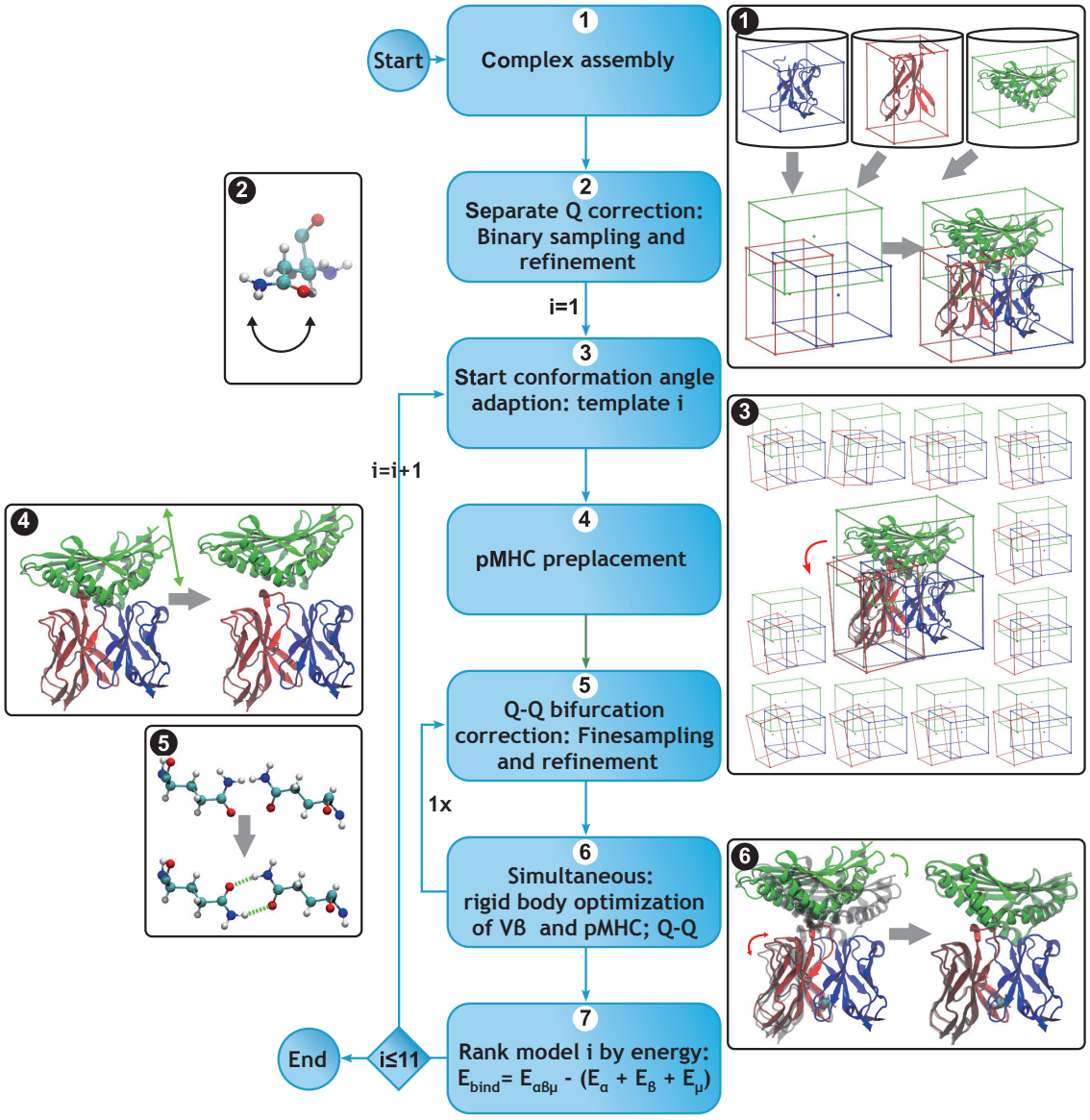


Figure 4.2: **Extended Prediction Pipeline for the Orientations in TCRpMHC Complexes.** (1) The complex was pre-assembled from the separated components Vα, Vβ, and pMHC (blue, red, and green) in the *zero orientation* and then (3) the TCR domains were transformed to one of eleven start orientations (*cluster representatives*). (4) The pMHC ligand was pre-placed by translation along and rotation around a perpendicular axis, while the distance was restrained. (6) The variable domains and the pMHC ligand were treated by a *rigid body* (RB) optimization while the conserved glutamine (Q) residues were (2,5, and 6) optimized in their interaction and restrained in their distance. (7) For each cluster representative a model was computed and the best model was selected according to the energy rank.

This restrained RB optimization in the main part of the pipeline (Fig. 4.2, Step 6) was carried out concurrently for both RBs, TCR Vβ domain and the pMHC ligand, while simultaneously the Q-Q interactions at the CoR_{β} were optimized. For the RB optimization of the pMHC, the CoR_{μ} (Sec. 2.3.2) was chosen as a pivot point. Besides Fig. 4.2 the modeling pipeline is also illustrated by an animation (Movie C.S1), which is embedded in the electronic version of this document to be played with an appropriate PDF-viewer.

4.2.3.2 Additional Modules

To extend the pipeline for TCRpMHC modeling in practice, only three additional SOs were required, while the existing algorithmic modules (Sec. 4.2.1.2) could be used again. The additionally implemented SOs involved (i) the axis based RB rotation module *RAR*, (ii) the axis based RB translation module *RAT*, and (iii) the restraint operator *RPC*. All three SOs were used for the preplacement procedure (Fig. 4.2, Step 4) and the *RPC* operator was additionally coupled with the *RMZ* operator in the main step (5) for the purpose of restraining the translation of the pMHC during the RB optimization.

The *RAT* operator is capable of translating a RB along a defined axis according to a one-dimensional parameter. The corresponding energy contribution and the gradient were only based on non-bonded interactions (Sec. 3.1.3.1) between the RBs. Furthermore, the one-dimensional gradient was computed by projecting the involved forces vectors onto the axis (details on p. 223).

The *RAR* operator featured the rotation of a RB around an axis according to an one-dimensional angular parameter. As described for *RAT* only non-bonded interactions between RBs were considered. The one-dimensional gradient was computed by means of tangential forces, which were already employed in the context of the *AQR* operator (see p. 223).

The *RPS* operator, which was used for restraining the TCR Vβ position, requires a defined atom of the domain for the purpose of monitoring the atom distance to a reference point. In the case of the TCR Vβ such an atom can easily be located in the vicinity of the CoR_{β}, since conserved Q residues constitute this center. In contrast, such a mapping of a defined residue or even atom is difficult in the case of the MHC molecule, because the CoR_{μ} is located in the peptide binding groove apart from the MHC residues. Since, the bound peptide is anything but conserved, as a workaround an anchor point was constructed by the geometric mean between the backbone *termini* of the peptide and was used by the *RPC* operator (details p. 224).

4.2.3.3 Computational Settings and Test Datasets

4.2.3.3.1 Data Sets for the TCRpMHC Modeling The data sets DS_{CC}, DS_{C0}, and DS_{M0}, which were used for the modeling of the complexes containing TCR as well as the pMHC ligand, were assembled similarly to TCR test sets DS_{T0} and DS_{TC} (Sec. 4.2.2.1). In contrast to the latter two sets, the structures include the TCR variable domains and additionally a peptide as well as the presenting MHC class I α_1- and α_2-domains. Tab. 4.1 lists the differences of all five data sets for the modeling of TCRs (Sec. 4.2.2.1) and TCRpMHC complexes.

The DS_{CC}, DS_{C0}, and DS_{M0} were derived from a set of 34 TCRpMHC crystal structures, which were also contained in DS_{TC}. The sets contain 18 BUs representing 3 different murine TCR types and 35 BUs of 11 different human TCR types. The sets represent 21 different peptide/MHC combinations, 8 different MHC class I alleles, 21 different peptide sequences. Structures of MHC class II molecules and structures with unresolved atoms in the two TCR variable domains, in the peptide, in the α_1-, or in the α_2-domain, were discarded. Details are listed in Tab. C.S1, p. 227.

The relative orientations of the Vα, Vβ, and the pMHC ligand in the structures of data set DS_{CC} were kept unchanged with respect to the original crystal structures. In the structures of the data set DS_{M0}, the orientations of the TCR domains were not changed, but the pMHC ligand was transformed to its zero orientation. This zero orientation of the pMHC ligand was found in the analysis presented in Sec. 2.3.2.1 (3e3q, chains R, S, T, and P) [138].

4.2.3.3.2 Evaluation of the TCRpMHC Modeling Method As described for the TCRs in absence of the pMHC ligands (Sec. 4.2.2.2), the pipeline for the modeling of TCRpMHC complex was evaluated with *main* and *control* ES. Different variants of the ES were applied (Tab. D.1), in which the concurrent optimization of the CoR_β carboxamide groups or the position restraints were enabled or disabled (if applicable or of interest). For the determination of the prediction rates, two different success criteria $\mathcal{C}_\mathcal{R}$ and $\mathcal{C}_\mathcal{E}$ were used as defined in Sec. 4.2.2.2.

In the main ES, M_C, the Vβ domains in the 53 BUs of data set DS_{C0} were moved to the eleven different angular starting orientations. For the pMHC ligand a single predefined start orientation, namely the zero orientation, was used. Then the pMHC ligand was pre-optimized by rotating around and moving along an axis (Fig. 4.2). In Step 6 of the pipeline, the Vβ and the pMHC ligand were treated in a concurrent RB optimization. Finally, the eleven models were evaluated according to the success criteria $\mathcal{C}_\mathcal{R}$ and $\mathcal{C}_\mathcal{E}$ as described in Sec. 4.2.2.2.

Several different control ES were carried out in order to assess the influence of the different parts of the TCRpMHC prediction pipeline. These ES differ in the initial orientations of the Vβ domain or the pMHC ligand (*e. g.* zero or crystal

orientation), and in an immobilization of complex components. The control ES are listed in Tab. D.1 and their differences and purposes are briefly described in the following. In accordance with the success criterion $\mathcal{C}_{\mathcal{R}}$ used in the main ES, a case was considered to be correct, if the alpha carbon RMSD with respect to the crystal structure did not exceeded 2.0 Å.

In the ES C_{cMc} and C_{mMc} the dataset DS_{CC} was used, *i. e.* the complex was assembled in the orientation as found in the experimental structure. Steps 3 and 4 of the pipeline were skipped, and therefore only one model was computed instead of eleven and the axis based preplacement of the pMHC was ignored. The two control ES differ in the Pipeline Step 5: In the ES C_{cMc} both the pMHC and the Vβ domain were treated as movable RBs. In contrast, in the ES C_{mMc}, only the pMHC was transformed, but the Vβ was fixed. In other words, it was tested, whether the pMHC escapes from the crystal position and whether the TCR flexibility influences their behavior.

The two ES C_{mpMc} and C_{cpMc} differed from the latter two by additionally using the Pipeline Step 4, which is the axis based preplacement of the pMHC. The data set DS_{CC} was used and thus the preplacement of the pMHC starts from the crystal orientation. Step 3 of the pipeline, which was the start orientation adaption of the Vβ, was skipped. Therefore the RB optimization starts with the Vβ in the crystal orientation. The ES C_{mpMc} and C_{cpMc} differ in Step 5, the RB optimization, which is applied only on the pMHC in the first ES and additionally on the Vβ as well in the latter case. The ES were performed to investigate the influence of the preplacement procedure.

In the C_{cpM0} the procedure was comparable to the ES C_{cpMc} above, with the difference that the dataset DS_{M0} was used. In other words, the complex was assembled with the TCR in the crystal structure orientation and the pMHC in the artificial zero orientation. Before the multi RB optimization of the Vβ and the pMHC was carried out, the axis based preplacement procedure was applied on the pMHC . RB restraints and the glutamine optimization were switched on. Comparing the two ES C_{cpM0} and C_{cpMc} brought insight in the influence of the initial position of the pMHC in the modeling process.

In order to investigate the influence of the pMHC starting position, for a particular case, the pipeline was applied for several different starting orientations of the pMHC and the TCR in the crystal orientation. In this ES, C_{1oga}, the complex 1oga was taken from the dataset DS_{M0}, *i. e.* the TCR adopted a orientation as found in the experimental crystal structure, whereas the pMHC initially was placed in the zero orientation. After the assembly of the complex with the pMHC in the zero orientation, 27 different starting orientations of the pMHC were generated by the rotations of $+5.0°$, $0.0°$, or $-5.0°$ for all three *Euler* angle components (all permutations). For each of these 27 starting orientations a model was built starting with

the axis preplacement of the pMHC. Both the pMHC and the Vβ were treated as movable RBs. RB restraints and the glutamine optimization were switched on.

4.3 Results and Discussion

4.3.1 Publication 3: DynaDom: A rigid body optimization-based approach for the prediction of T-cell receptor inter-domain and T-cell receptor-peptide-MHC association angles

The results of the so far described study were sent to the peer reviewed open access journal BMC Structural Biology for publication:
Thomas Hoffmann*, Antoine Marion, Iris Antes, Heinrich Leonhardt, (2016) *DynaDom: A rigid body optimization-based approach for the prediction of T-cell receptor inter-domain and T-cell receptor-peptide-MHC association angles.* BMC Structural Biology (submitted)
Submitted: April 19, 2016

In the following two sections the article is summarized and my contribution to this work is declared. The full article and the supplementary material are included in Appendix C, p. 171 of this thesis.

4.3.1.1 Summary

Due to the high diversity of *T-cell receptor* (TCR) molecules the adaptive immune system is enabled to recognize a tremendous number of pathogenic peptides presented by *major histocompatibility complex* (MHC) molecules, as introduced in Sec. 1.1. Structural models of the TCRpMHC complex assist in the fields of vaccine design and the adoptive T-cell therapy, but reliable homology modeling approaches are requiered since experimental methods for structure elucidation are to time consumptive to be applied on such a enormous amount of different complexes. In Chapter 2 it was shown that relative orientation of the TCR variable domains and the docking orientation of the MHC molecule increase the structural variety of TCRpMHC complexes. This flexibility on domain level makes the structure prediction more difficult and is so far not explicitly taken into account in current homology modeling approaches for these complexes (Sec. 4.1.1).

Thus, the focus of Publication 3 is the development of the method *DynaDom* for the prediction of inter-domain orientations and other properties of protein complexes. The approach is based on two substantial aspects derived from the broad analysis of TCRpMHC complexes provided in Chapter 2. On the one hand, for a modeling case several models are created from different starting orientations based

on the angular clusters found in the analysis and represented as combinations of cuboids (Fig. C.3 A, p. 220). On the other hand, a *rigid body* (RB) optimization method is deployed, which requires the definition of a pivot point that was chosen to be the *Center of Rotation* of the variable β domain (CoR_β) discovered in the analysis (Fig. C.1, p. 218). This CoR_β is taken into account for the modeling of the TCR variable α (Vα)/variable β (Vβ) orientations by two more mechanisms: One mechanism uses the CoR_β as a reference point for the application of RB position restraints. A second mechanism ensured the formation of bifurcated interactions of the conserved glutamine (Q) residues located at the CoR_β area of both TCR chains (Fig. C.3 B+C, p. 220). In this context a limitation found in many structures resolved by X-ray crystallography effects the modeling: Similarities in the electron densities of oxygen and nitrogen cause an ambiguity in the conformation of carboxamide groups of glutamine residues. Thus in many TCR structures mispairing of the strong attractive bifurcated Q-Q interactions are observed to induce an opposite repulsive effect (Tab. C.S2, Col. ES, p. 229). As shown in Tab. C.S3, p. 232, standard methods for the correction of the so *NQ-flipping problem* (Sec. 4.1.3 and p. 230) have unfortunately only a limited applicability in the present modeling cases. This limitation is not caused by the current correction methods themself, but by the fact that the environmental context of the Q residues may change when the separated variable domains associate with each other. Thus, it was necessary to develop a new NQ-flip correction method with prediction performance comparable to the standard methods (comparison: Tab. C.S2, p. 229) and to be applicable during the modeling process in order to adapt the Q conformations in dependency of the changing position of the RBs.

A design goal for the DynaDom method was a modular implementation, which allows to tailor and extend system specific modeling pipelines. Hence, it was possible to extend the modeling pipeline of the TCR Vα/Vβ association angle prediction in order to concurrently predict the docking geometry of the pMHC ligand. Both pipelines, including the TCR modeling and the TCRpMHC modeling, are illustrated in Fig. C.2, p. 219 and an optimization run is animated in Movie C.S1 (embedded in the electronic version of the present document at p. 226).

The performance of both pipelines was tested on broad sets of 75 TCR and 53 TCRpMHC structures, respectively. In these tests two aspects were examined with two different criteria. On the one hand, the methods capability of constructing models with a sufficient quality was of interest. On the other hand, the methods power to discriminate between models within one modeling case ranked by a *force field* (Chapter 3) based binding energy was evaluated (Fig. C.S1, p. 235). With different evaluation settings the influence of the pipeline components were investigated, such as the Q-Q optimization, the position restrains (Fig. 2, p. 235), and the preplacement procedure of the pMHC ligand. As shown in Tab. C.2, p. 215, in 96% of

the modeling cases of the TCR Vα/Vβ at least one model was constructed with an RMSD less than 2.0 Å and in 89% of the cases such sufficient models were selected according to the energy ranking. Based on the success criterion a small improvement of the prediction rates can be observed when the Q-Q optimization and the RB position restraints were used. Taking all built models into account, the diagram in Fig. C.4, p. 221) demonstrates that generally the model quality increases. In particular, these pipeline features effect cases for which the Q residues were observed in the experimental structures to be in a mispaired state. For the prediction of orientations in the TCRpMHC complexes it was possible to build at least one sufficient model with an RMSD below 2.0 Å and the energy based selection was successful for 72% of the cases. Further explorations with modifications in the pre-placement procedure of the pMHC ligand suggest potential improvement focused on this particular pipeline step (Tabs. C.4 and C.S5, pp. 216 and 236).

4.3.1.2 Contribution

I am first author of this publication having designed and performed all programming and analyzed the resulting data with the following exceptions: Antoine Marion created Figs. C.4 and C.1, analyzed the data for Fig. C.4, analyzed the ES M_T*11 contributing that part to Tab. C.2, and edited the manuscript for publication. Hanieh Mirzaei provided C code of her RB minimizer, which was adopted to the DynaCell C++ code.

Chapter 5

Conclusions and Outlook

This work focused on the structural analysis and modeling of *T-cell receptors* (TCRs) and their interaction with pMHC complexes. TCRs play a key role in the adaptive immune response by forming complexes with pMHC ligands consisting of peptides presented by *major histocompatibility complex* (MHC) molecules (see Figs. C.1, 1.1, and 1.2). Structural models of these complexes are of special importance to investigate T-cell signaling, to develop new vaccines, and to design TCRs in the context of adoptive T-cell therapy. The models need to be highly accurate, since the smallest alteration of sequences of TCRpMHC complex components may have a dramatic impact on the resulting signal [71]. Experimental methods such as NMR or X-ray crystallography are indispensable for structure elucidation. However, these methods are not sufficient for this purpose, as they are very time consuming and an essential feature of TCRs is their high diversity [12, 44, 284]. Thus computational methods are necessary to obtain structural and thus homology modeling methods became of particular interest.

This work elaborates a new methodology focusing on an important characteristic of TCRpMHC complexes that was not taken into account in previous homology modeling approaches. In this context, first an exhaustive analysis of the available TCR structures was performed. The analysis is based on the representation of the structurally diverse different TCR variable α (Vα) and variable β (Vβ) domains as unified cuboids. This cuboid-method was used to analyze the Vα/Vβ association angles, which were clustered according to an *Euler angle distance* (EAD) metric. Afterwards, a grid-based approach was applied for the determination of a fulcrum, which serves as *Center of Rotation* (CoR) for the two domains. The CoR was further analyzed using sequence data from an *multiple sequence alignment* (MSA). The results of this novel cuboid-based measurement performed on a up-to-data and very broad data set confirm the results of previous studies on individual TCRs or small structural subsets thereof that TCRs adopt different Vα/Vβ orientations [73, 125, 188]. The EAD-based clustering showed that TCRs of the same clonotype

adapt to the same clusters of orientations upon binding the pMHC ligand and may differ in unbound states. Thus it can be concluded from this analysis study that the TCR Vα/Vβ orientations are clonotype specific and an adaption of the orientation occurs in a two-step binding process. In addition, the results of the grid analysis allowed to identify a previously unknown conserved CoR as a new feature of TCRs. This CoR is formed and stabilized by the interaction of conserved glutamine (Q) residues. This observation supports the assumption that the rotatability of the TCR Vα/Vβ domains is an essential feature of these molecules.

Based on these analysis results a new modeling method, DynaDom, was developed, which takes the two main results of the analysis into account in order to predict the association geometry of the TCR variable domains: different models were built starting from the orientations obtained by the clustering and the carboxamide groups of the conserved Q residues were kept rotatable. The method, which has been incorporated into the *DynaCell* program, is based on energy minimizing techniques and uses a *rigid body* (RB) optimization approach by Mirzaei *et al.* [6, 200]. In order to test the approach, a remodeling study with 75 cases was carried out and resulted in the remarkable prediction rate of 96%. The newly discovered CoR was treated semi-flexible by automatically correcting local inaccuracies in the Q side chain orientations in the experimental crystal structures. The application of this new *NQ-flipping* correction additionally improved the quality of the models and further supports the importance of the Q-Q based CoR. The results show that the method is robust and carries potential for future blind modeling and for the extension to predict the orientations in TCRpMHC complexes.

Based on the successful remodeling of TCR Vα and Vβ domain orientations this approach was further extended for the prediction of the whole complex containing the TCR variable domains as well as the pMHC ligand. To find an appropriate starting orientation for the pMHC proved a challenging task. This is in accordance with another study, which showed that MHC molecules adapt their orientation in an unbound yet already structurally close conformation to the TCR based on long-range interactions [82]. This aspect was taken into account in the new extended prediction pipeline, which achieved a prediction rate of 74% in a remodeling study. The question occurs why the new method performs so much better for the prediction of Vα/Vβ domain orientation than for pMHC placement. Based on an extensive series of tests, the reason could be identified in the pMHC preplacement. For an individual case for which a computationally more demanding pre-placement of the pMHC was applied by using different initial angular orientations, the results could be improved considerably, which is in accordance with other studies [219].

The DynaDom method was implemented in a modular manner, which turned out to have several advantages. Arbitrary modules could be assembled to various problem-specific prediction pipelines and new functionalities could be added easily.

This flexibility enabled the easy extension of the TCR prediction pipeline to the modeling of the orientation of the pMHC ligands.

The application to other molecule classes is also feasible due to the general design of the approach. A possibly straightforward application would be the modeling of *antibody* (AB) Fab fragments for two reasons: First, it was shown that ABs can also adopt different domain orientations in a similar manner to TCRs [73]. An application of the new cuboid-based analysis method would allow identifying clusters of possible starting orientations. Second, the existence of a CoR of ABs is likely, since sequence analyses provide proof of conserved Q-Q interactions for this molecule class as well [43, 49]. Another extension would be to take the TCR constant domains into account. This, however, would require more implementation and computational efforts. Nevertheless it would allow several advantages: First, the degrees of freedom of the variable domains would be reduced, since the search space is limited by avoiding artificial unphysical moves of the variable domains, which is currently achieved by RB restraints. Second, the influence of scissoring effects, *i. e.* the motions of the constant domains with respect to that of the variable domains, could be studied in the context of T-cell signaling. In particular, the *A-B loop* located in the constant domains (Fig. 1.3, p. 7) was discussed to be possibly involved in the signal transduction [19]. Further additional requirements for that purpose include the modeling of the elbow loop region connecting the variable and the constant domain of a chain, as well as the treatment of coupled RBs. For the latter aspect a further development of the Mirzaei-approach could be applied [201]. For the purpose of loop modeling, fine-grained modules are needed in order to move residues, side chains, or atoms relatively to the RB motions. The feasibility of such RB dependent fine-grained modules has been demonstrated in the present thesis by the successful Q-optimization at the CoR. Such fine-grained modules could also be useful for the modeling of the adaption of hyper-variable *complementary determining region* (CDR) loops of the TCR to the bound ligand.

The incorporation of the DynaDom method into the *DynaCell* program has the further advantage that the OPMD approach can be applied [6]. This approach enables to perform energy minimizations starting with initial atom-atom overlaps, this was shown to lead to more efficient optimizations and thus a reduced number of starting orientations and therefore a reduction of the computational effort. Furthermore this technique could be used for the optimization of the CDR loops with respect to the pMHC ligand. In order to take the domain orientations of TCRpMHC complexes into account during homology modeling, two different approaches using DynaDom might be possible: First, the individual components of the complex could be separately modeled with standard homology modeling tools such as *MODELLER* and then subsequently assembled and refined with DynaDom/DynaCell [241]. Second, a pipeline combining the DynaDom method including fine-grained modules,

the *IRECS* side chain placement tool, and an OPMD refinement could possibly be used as a full homology modeling technique [6, 110].

Such a modeling tool could be used for the prediction of TCR Vα:Vβ pairings or miss-parings, which is particularly important for adoptive T-cell therapy. In this context, the prediction of backbone or side chain conformations located at the Vα:Vβ interface requires the implementation of further modules. Prediction of such mis-pairings is currently not possible, since the required experimental binding data of mispaired or unpaired TCR chains are only sparsely available.

To summarize, TCRpMHC complexes were analyzed by a cuboid-based method revealing two general properties of these molecules, namely a clonotype specific orientation of the TCRs Vα/Vβ domains and a conserved CoR. Since current homology modeling methods do not take these properties of TCRs into account, the development of a new prediction method for domain orientations has become a necessity. In this thesis an approach named DynaDom was developed using molecular mechanics force fields and a rigid body optimization method. With this new approach it was able to predict TCR domain orientations as well as the orientation of the pMHC ligand in TCRpMHC complexes with a high accuracy in a remodeling assessment. Due to its modular design, the method can easily be further improved, which is necessary in particular for the pMHC placement. Furthermore the approach is ready to be extended to a blind modeling tool for TCRpMHC and to be applied on other classes of molecules. ■

Bibliography

[1] M. Achour, X. Jacq, P. Ronde, M. Alhosin, C. Charlot, T. Chataigneau, M. Jeanblanc, M. Macaluso, A. Giordano, A. D. Hughes, V. B. Schini-Kerth, and C. Bronner. The interaction of the SRA domain of ICBP90 with a novel domain of DNMT1 is involved in the regulation of VEGF gene expression. *Oncogene*, 27(15):2187–2197, Apr 2008

[2] B. J. Alder and T. E. Wainwright. Phase Transition for a Hard Sphere System. *The Journal of Chemical Physics*, 27(5):1208–1209, Nov. 1957

[3] J. C. Almagro, E. Vargas-Madrazo, F. Lara-Ochoa, and E. Horjales. Molecular modeling of a T-cell receptor bound to a major histocompatibility complex molecule: implications for T-cell recognition. *Protein Sci*, 4(9):1708–17, 1995

[4] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *Journal of molecular biology*, 215(3):403–410, Oct 1990

[5] H. C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72(4):2384–2393, 1980

[6] I. Antes. DynaDock: A new molecular dynamics-based algorithm for protein-peptide docking including receptor flexibility. *Proteins*, 78(5):1084–1104, 2010

[7] I. Antes, S. W. I. Siu, and T. Lengauer. DynaPred: a structure and sequence based method for the prediction of MHC class I binding peptide sequences and conformations. *Bioinformatics (Oxford, England)*, 22(14):e16–e24, Jul 2006

[8] J. Aqvist, C. Medina, and J. E. Samuelsson. A new method for predicting binding affinity in computer-aided drug design. *Protein engineering*, 7(3):385–391, Mar 1994

[9] J. Aqvist, V. B. Luzhkov, and B. O. Brandsdal. Ligand binding affinities from MD simulations. *Accounts of chemical research*, 35(6):358–365, Jun 2002

[10] J. K. Archbold, W. A. Macdonald, S. Gras, L. K. Ely, J. J. Miles, M. J. Bell, R. M. Brennan, T. Beddoe, M. C. J. Wilce, C. S. Clements, A. W. Purcell, J. McCluskey, S. R. Burrows, and J. Rossjohn. Natural micropolymorphism in human leukocyte antigens provides a basis for genetic control of antigen recognition. *J Exp Med*, 206 (1):209–219, Jan 2009

[11] K. Arita, M. Ariyoshi, H. Tochio, Y. Nakamura, and M. Shirakawa. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature*, 455(7214):818–821, Oct 2008

[12] T. P. Arstila, A. Casrouge, V. Baron, J. Even, J. Kanellopoulos, and P. Kourilsky. A direct estimate of the human alphabeta T cell receptor diversity. *Science (New York, N.Y.)*, 286(5441):958–961, Oct 1999

[13] G. V. Avvakumov, J. R. Walker, S. Xue, Y. Li, S. Duan, C. Bronner, C. H. Arrowsmith, and S. Dhe-Paganon. Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature*, 455(7214):822–825, Oct 2008

[14] D. Baker and A. Sali. Protein Structure Prediction and Structural Genomics. *Science*, 294(5540):93–96, 2001

[15] N. Ban, C. Escobar, K. W. Hasel, J. Day, A. Greenwood, and A. McPherson. Structure of an anti-idiotypic Fab against feline peritonitis virus-neutralizing antibody and a comparison with the complexed Fab. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 9(1):107–114, Jan 1995

[16] A. S. Bayden, M. Fornabaio, J. N. Scarsdale, and G. E. Kellogg. Web application for studying the free energy of binding and protonation states of protein-ligand complexes based on HINT. *J Comput Aided Mol Des*, 23(9):621–32, 2009

[17] P. B. Becker, S. Ruppert, and G. Schütz. Genomic footprinting reveals cell type-specific DNA binding of ubiquitous factors. *Cell*, 51(3):435–443, Nov 1987

[18] E. M. Beckman, S. A. Porcelli, C. T. Morita, S. M. Behar, S. T. Furlong, and M. B. Brenner. Recognition of a lipid antigen by CD1-restricted alpha beta+ T cells. *Nature*, 372(6507):691–694, Dec 1994

[19] T. Beddoe, Z. Chen, C. S. Clements, L. K. Ely, S. R. Bushell, J. P. Vivian, L. Kjer-Nielsen, S. S. Pang, M. A. Dunstone, Y. C. Liu, W. A. Macdonald, M. A. Perugini, M. C. J. Wilce, S. R. Burrows, A. W. Purcell, T. Tiganis, S. P. Bottomley, J. McCluskey, and J. Rossjohn. Antigen ligation triggers a conformational change within the constant domain of the alphabeta T cell receptor. *Immunity*, 30(6):777–788, Jun 2009

[20] D. Beeman. Some multistep methods for use in molecular dynamics calculations. *Journal of Computational Physics*, 20(2):130–139, Feb. 1976

[21] G. M. Bendle, C. Linnemann, A. I. Hooijkaas, L. Bies, M. A. de Witte, A. Jorritsma, A. D. M. Kaiser, N. Pouw, R. Debets, E. Kieback, W. Uckert, J.-Y. Song, J. B. A. G. Haanen, and T. N. M. Schumacher. Lethal graft-versus-host disease in mouse models of T cell receptor gene therapy. *Nat Med*, 16(5):565–570, May 2010

[22] H. Berendsen. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1-3):43–56, Sept. 1995

[23] H. Berendsen, J. Postma, W. van Gunsteren, and J. Hermans. *Interaction Models for Water in Relation to Protein Hydration*, volume 14 of *The Jerusalem Symposia on Quantum Chemistry and Biochemistry*, pages 331–342. Springer Netherlands, 1981. ISBN 978-90-481-8368-5

[24] H. Berendsen, J. Postma, W. van Gunsteren, A. DiNola, and J. Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81 (8):3684–3690, 1984

[25] D. X. Beringer, F. S. Kleijwegt, F. Wiede, A. R. van der Slik, K. L. Loh, J. Petersen, N. L. Dudek, G. Duinkerken, S. Laban, A. Joosten, J. P. Vivian, Z. Chen, A. P. Uldrich, D. I. Godfrey, J. McCluskey, D. A. Price, K. J. Radford, A. W. Purcell, T. Nikolic, H. H. Reid, T. Tiganis, B. O. Roep, and J. Rossjohn. T cell receptor reversed polarity recognition of a self-antigen major histocompatibility complex. *Nature immunology*, 16(11):1153–1161, Nov 2015

[26] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The Protein Data Bank. *Nucleic acids research*, 28 (1):235–242, Jan 2000

[27] F. C. Bernstein, T. F. Koetzle, G. J. Williams, J. Meyer, E. F., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol*, 112(3): 535–42, 1977

[28] T. N. Bhat, G. A. Bentley, T. O. Fischmann, G. Boulot, and R. J. Poljak. Small rearrangements in structures of Fv and Fab fragments of antibody D1.3 on antigen binding. *Nature*, 347(6292):483–485, Oct 1990

[29] T. N. Bhat, G. A. Bentley, G. Boulot, M. I. Greene, D. Tello, W. Dall'Acqua, H. Souchon, F. P. Schwarz, R. A. Mariuzza, and R. J. Poljak. Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proceedings of the National Academy of Sciences of the United States of America*, 91(3): 1089–1093, Feb 1994

[30] S. Bietz, S. Urbaczek, B. Schulz, and M. Rarey. Protoss: a holistic approach to predict tautomers and protonation states in protein-ligand complexes. *J Cheminform*, 6:12, 2014

[31] A. Bird. DNA methylation patterns and epigenetic memory. *Genes Dev*, 16(1):6–21, Jan 2002

[32] M. E. Birnbaum, J. L. Mendoza, D. K. Sethi, S. Dong, J. Glanville, J. Dobbins, E. Ozkan, M. M. Davis, K. W. Wucherpfennig, and K. C. Garcia. Deconstructing the peptide-MHC specificity of T cell recognition. *Cell*, 157(5):1073–1087, May 2014

[33] O. Y. Borbulevych, K. H. Piepenbrink, B. E. Gloor, D. R. Scott, R. F. Sommese, D. K. Cole, A. K. Sewell, and B. M. Baker. T cell receptor cross-reactivity directed by antigen-dependent tuning of peptide-MHC molecular flexibility. *Immunity*, 31(6): 885–896, Dec 2009

[34] O. Y. Borbulevych, K. H. Piepenbrink, and B. M. Baker. Conformational melding permits a conserved binding geometry in TCR recognition of foreign and self molecular mimics. *J Immunol*, 186(5):2950–2958, Mar 2011

[35] M. Bostick, J. K. Kim, P.-O. Esteve, A. Clark, S. Pradhan, and S. E. Jacobsen. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science*, 317(5845):1760–1764, Sep 2007

[36] A. Brero, H. P. Easwaran, D. Nowak, I. Grunewald, T. Cremer, H. Leonhardt, and M. C. Cardoso. Methyl CpG-binding proteins induce large-scale chromatin reorganization during terminal differentiation. *J Cell Biol*, 169(5):733–743, Jun 2005

[37] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4(2):187–217, 1983

[38] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and

M. Karplus. CHARMM: the biomolecular simulation program. *Journal of computational chemistry*, 30(10):1545–1614, Jul 2009

[39] W. J. Browne, A. C. North, D. C. Phillips, K. Brew, T. C. Vanaman, and R. L. Hill. A possible three-dimensional structure of bovine alpha-lactalbumin based on that of hen's egg-white lysozyme. *Journal of molecular biology*, 42(1):65–86, May 1969

[40] C. G. Broyden. The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations. *IMA Journal of Applied Mathematics*, 6(1):76–90, 1970

[41] S. R. Burrows, Z. Chen, J. K. Archbold, F. E. Tynan, T. Beddoe, L. Kjer-Nielsen, J. J. Miles, R. Khanna, D. J. Moss, Y. C. Liu, S. Gras, L. Kostenko, R. M. Brennan, C. S. Clements, A. G. Brooks, A. W. Purcell, J. McCluskey, and J. Rossjohn. Hard wiring of T cell receptor specificity for the major histocompatibility complex is underpinned by TCR adaptability. *Proc Natl Acad Sci U S A*, 107(23):10608–10613, Jun 2010

[42] J. Buslepp, H. Wang, W. E. Biddison, E. Appella, and E. J. Collins. A correlation between TCR Valpha docking on MHC and CD8 dependence: implications for T cell selection. *Immunity*, 19(4):595–606, Oct 2003

[43] D. A. Calarese, C. N. Scanlan, M. B. Zwick, S. Deechongkit, Y. Mimura, R. Kunert, P. Zhu, M. R. Wormald, R. L. Stanfield, K. H. Roux, J. W. Kelly, P. M. Rudd, R. A. Dwek, H. Katinger, D. R. Burton, and I. A. Wilson. Antibody domain exchange is an immunological solution to carbohydrate cluster recognition. *Science*, 300(5628): 2065–2071, Jun 2003

[44] A. Casrouge, E. Beaudoing, S. Dalle, C. Pannetier, J. Kanellopoulos, and P. Kourilsky. Size estimate of the alpha beta TCR repertoire of naive mouse splenocytes. *Journal of immunology (Baltimore, Md. : 1950)*, 164(11):5782–5787, Jun 2000

[45] A. K. Chakraborty and A. Weiss. Insights into the initiation of TCR signaling. *Nat Immunol*, 15(9):798–807, Sep 2014

[46] J.-L. Chen, G. Stewart-Jones, G. Bossi, N. M. Lissin, L. Wooldridge, E. M. L. Choi, G. Held, P. R. Dunbar, R. M. Esnouf, M. Sami, J. M. Boulter, P. Rizkallah, C. Renner, A. Sewell, P. A. van der Merwe, B. K. Jakobsen, G. Griffiths, E. Y. Jones, and V. Cerundolo. Structural and kinetic basis for heightened immunogenicity of T cell vaccines. *J Exp Med*, 201(8):1243–1255, Apr 2005

[47] V. B. Chen, r. Arendall, W. B., J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 1):12–21, 2010

[48] C. Chothia and A. M. Lesk. The relation between the divergence of sequence and structure in proteins. *The EMBO journal*, 5(4):823–826, Apr 1986

[49] C. Chothia, D. R. Boswell, and A. M. Lesk. The outline structure of the T-cell alpha beta receptor. *The EMBO journal*, 7(12):3745–3755, Dec 1988

[50] K. Choudhuri, A. Kearney, T. R. Bakker, and P. A. van der Merwe. Immunology: how do T cells recognize antigen? *Curr Biol*, 15(10):R382–5, May 2005

[51] L. S. Chuang, H. I. Ian, T. W. Koh, H. H. Ng, G. Xu, and B. F. Li. Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science (New York, N.Y.)*, 277(5334):1996–2000, Sep 1997

[52] D. K. Cole, K. M. Miles, F. Madura, C. J. Holland, A. J. A. Schauenburg, A. J. Godkin, A. M. Bulek, A. Fuller, H. J. E. Akpovwa, P. G. Pymm, N. Liddy, M. Sami, Y. Li, P. J. Rizkallah, B. K. Jakobsen, and A. K. Sewell. T-cell receptor (TCR)-peptide specificity overrides affinity-enhancing TCR-major histocompatibility complex interactions. *J Biol Chem*, 289(2):628–638, Jan 2014

[53] L. A. Colf, A. J. Bankovich, N. A. Hanick, N. A. Bowerman, L. L. Jones, D. M. Kranz, and K. C. Garcia. How a single T cell receptor recognizes both self and foreign MHC. *Cell*, 129(1):135–146, Apr 2007

[54] P. M. Colman, W. G. Laver, J. N. Varghese, A. T. Baker, P. A. Tulloch, G. M. Air, and R. G. Webster. Three-dimensional structure of a complex of antibody with influenza virus neuraminidase. *Nature*, 326(6111):358–363, Mar 1987

[55] W. Cornell, P. Cieplak, C. Bayly, I. Gould, K. Merz, D. Ferguson, D. Spellmeyer, T. Fox, J. Caldwell, and P. Kollman. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, 1995

[56] E. A. Coutsias, C. Seok, and K. A. Dill. Using quaternions to calculate RMSD. *Journal of computational chemistry*, 25(15):1849–1857, Nov 2004

[57] M. Crowley, M. Williamson, and R. Walker. CHAMBER: Comprehensive support for CHARMM force fields within the AMBER software. *International Journal of Quantum Chemistry*, 109(15):3767–3772, 2009

[58] M. A. Cuendet and O. Michielin. Protein-protein interaction investigated by steered molecular dynamics: the TCR-pMHC complex. *Biophys J*, 95(8):3575–3590, Oct 2008

[59] M. A. Cuendet, V. Zoete, and O. Michielin. How T cell receptors interact with peptide-MHCs: a multiple steered molecular dynamics study. *Proteins*, 79(11):3007–3024, Nov 2011

[60] S. Dai, E. S. Huseby, K. Rubtsova, J. Scott-Browne, F. Crawford, W. A. Macdonald, P. Marrack, and J. W. Kappler. Crossreactive T Cells spotlight the germline rules for alphabeta T cell-receptor interactions with MHC molecules. *Immunity*, 28(3):324–334, Mar 2008

[61] T. Darden, D. York, and L. Pedersen. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, June 1993

[62] I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, r. Arendall, W. B., J. Snoeyink, J. S. Richardson, and D. C. Richardson. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res*, 35(Web Server issue):W375–83, 2007

[63] S. J. Davis and P. A. van der Merwe. The structure and ligand interactions of CD2: implications for T-cell function. *Immunology today*, 17(4):177–187, Apr 1996

[64] S. J. Davis and P. A. van der Merwe. The kinetic-segregation model: TCR triggering and beyond. *Nature immunology*, 7(8):803–809, Aug 2006

[65] M. C. De Rosa, B. Giardina, C. Bianchi, C. Carelli Alinovi, D. Pirolli, G. Ferraccioli, M. De Santis, G. Di Sante, and F. Ria. Modeling the ternary complex TCR-Vbeta/CollagenII(261-273)/HLA-DR4 associated with rheumatoid arthritis. *PLoS One*, 5(7):e11550, 2010

[66] D. Defays. An efficient algorithm for a complete link method. *The Computer Journal*, 20(4):364–366, 1977

[67] M. Degano, K. C. Garcia, V. Apostolopoulos, M. G. Rudolph, L. Teyton, and I. A. Wilson. A functional hot spot for antigen recognition in a superagonist TCR/MHC complex. *Immunity*, 12(3):251–261, Mar 2000

[68] L. Deng, R. J. Langley, P. H. Brown, G. Xu, L. Teng, Q. Wang, M. I. Gonzales, G. G. Callender, M. I. Nishimura, S. L. Topalian, and R. A. Mariuzza. Structural basis for the recognition of mutant self by a tumor-specific, MHC class II-restricted T cell receptor. *Nat Immunol*, 8(4):398–408, Apr 2007

[69] M. M. Deza and E. Deza. *Encyclopedia of Distances*, page 313p. Springer Science + Business Media, 2009

[70] Y. H. Ding, K. J. Smith, D. N. Garboczi, U. Utz, W. E. Biddison, and D. C. Wiley. Two human T cell receptors bind in a similar diagonal mode to the HLA-A2/Tax peptide complex using different TCR amino acids. *Immunity*, 8(4):403–411, Apr 1998

[71] Y. H. Ding, B. M. Baker, D. N. Garboczi, W. E. Biddison, and D. C. Wiley. Four A6-TCR/peptide/HLA-A2 structures that generate very different T cell signals are nearly identical. *Immunity*, 11(1):45–56, Jul 1999

[72] R. B. DuBridge, P. Tang, H. C. Hsia, P. M. Leong, J. H. Miller, and M. P. Calos. Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. *Molecular and cellular biology*, 7(1):379–387, Jan 1987

[73] J. Dunbar, A. Fuchs, J. Shi, and C. M. Deane. ABangle: characterising the VH-VL orientation in antibodies. *Protein Eng Des Sel*, 26(10):611–620, Oct 2013

[74] J. Dunbar, B. Knapp, A. Fuchs, J. Shi, and C. M. Deane. Examining variable domain orientations in antigen receptors gives insight into TCR-like antibody design. *PLoS Comput Biol*, 10(9):e1003852, Sep 2014

[75] S. M. Dunn, P. J. Rizkallah, E. Baston, T. Mahon, B. Cameron, R. Moysey, F. Gao, M. Sami, J. Boulter, Y. Li, and B. K. Jakobsen. Directed evolution of human T cell receptor CDR2 residues by phage display dramatically enhances affinity for cognate peptide-MHC without increasing apparent cross-reactivity. *Protein Sci*, 15(4):710–721, Apr 2006

[76] B. Efron. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1):1–26, Jan. 1979

[77] F. Ehrenmann, Q. Kaas, and M.-P. Lefranc. IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF. *Nucleic acids research*, 38(Database issue):D301–D307, Jan 2010

[78] B. S. Everitt, S. Landau, M. Leese, and D. Stahl. *Cluster Analysis*. John Wiley & Sons, Ltd, jan 2011

[79] A. P. Feinberg and B. Tycko. The history of cancer epigenetics. *Nature reviews. Cancer*, 4(2):143–153, Feb 2004

[80] S. Feller, Y. Zhang, R. Pastor, and B. Brooks. Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics*, 103(11): 4613–4621, 1995

[81] D. Feng, C. J. Bond, L. K. Ely, J. Maynard, and K. C. Garcia. Structural evidence for a germline-encoded T cell receptor-major histocompatibility complex interaction 'codon'. *Nat Immunol*, 8(9):975–983, Sep 2007

[82] M. Ferber, V. Zoete, and O. Michielin. T-cell receptors binding orientation over peptide/MHC class I is driven by long-range interactions. *PLoS One*, 7(12):e51943, 2012

[83] R. Fletcher. A new approach to variable metric algorithms. *The Computer Journal*, 13(3):317–322, 1970

[84] R. Fletcher. *Practical Methods of Optimization*. John Wiley & Sons, Chichester, Sussex, 2004. ISBN 978-0-471-49463-8

[85] D. R. Flower. *Vaccines: Structural Approaches*, pages 217–255. John Wiley and Sons, Ltd, 2009. ISBN 9780470699836

[86] C. Frauer and H. Leonhardt. A versatile non-radioactive assay for DNA methyltransferase activity and DNA binding. *Nucleic Acids Res*, 37(3):e22, Feb 2009

[87] S. D. Gadola, M. Koch, J. Marles-Wright, N. M. Lissin, D. Shepherd, G. Matulis, K. Harlos, P. M. Villiger, D. I. Stuart, B. K. Jakobsen, V. Cerundolo, and E. Y. Jones. Structure and binding kinetics of three different human CD1d-alpha-galactosylceramide-specific T cell receptors. *J Exp Med*, 203(3):699–710, Mar 2006

[88] S. J. Gagnon, O. Y. Borbulevych, R. L. Davis-Harrison, R. V. Turner, M. Damirjian, A. Wojnarowicz, W. E. Biddison, and B. M. Baker. T cell receptor recognition via cooperative conformational plasticity. *J Mol Biol*, 363(1):228–243, Oct 2006

[89] D. N. Garboczi, P. Ghosh, U. Utz, Q. R. Fan, W. E. Biddison, and D. C. Wiley. Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature*, 384(6605):134–141, Nov 1996

[90] K. C. Garcia, M. Degano, R. L. Stanfield, A. Brunmark, M. R. Jackson, P. A. Peterson, L. Teyton, and I. A. Wilson. An alphabeta T cell receptor structure at 2.5 A and its orientation in the TCR-MHC complex. *Science (New York, N.Y.)*, 274 (5285):209–219, Oct 1996

[91] K. C. Garcia, M. Degano, L. R. Pease, M. Huang, P. A. Peterson, L. Teyton, and I. A. Wilson. Structural basis of plasticity in T cell receptor recognition of a self peptide-MHC antigen. *Science (New York, N.Y.)*, 279(5354):1166–1172, Feb 1998

[92] N. R. J. Gascoigne. T cell receptor structures: three for the price of one. *Immunity*, 35(1):1–3, Jul 2011

[93] M. Gehring, W. Reik, and S. Henikoff. DNA demethylation by DNA repair. *Trends in genetics : TIG*, 25(2):82–90, Feb 2009

[94] D. Gil, A. G. Schrum, B. Alarcon, and E. Palmer. T cell receptor engagement by peptide-MHC ligands induces a conformational change in the CD3 complex of thymocytes. *J Exp Med*, 201(4):517–522, Feb 2005

[95] V. Giudicelli, D. Chaume, and M.-P. Lefranc. IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res*, 33(Database issue):D256–61, Jan 2005

[96] D. Globisch, M. Munzel, M. Muller, S. Michalakis, M. Wagner, S. Koch, T. Bruckl, M. Biel, and T. Carell. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One*, 5(12):e15367, 2010

[97] D. Goldfarb. A Family of Variable-Metric Methods Derived by Variational Means. *Mathematics of Computation*, 24(109):23–26, 1970

[98] H. Goldstein, C. P. Poole, and J. L. Safko. *Classical Mechanics (3rd Edition)*. Addison-Wesley, 3 edition, June 2001. ISBN 0201657023

[99] M. G. Goll and T. H. Bestor. Eukaryotic cytosine methyltransferases. *Annu Rev Biochem*, 74:481–514, 2005

[100] B. Gough. *GNU Scientific Library Reference Manual - Third Edition*. Network Theory Ltd., 3rd edition, 2009. ISBN 0954612078, 9780954612078

[101] S. Gras, S. R. Burrows, L. Kjer-Nielsen, C. S. Clements, Y. C. Liu, L. C. Sullivan, M. J. Bell, A. G. Brooks, A. W. Purcell, J. McCluskey, and J. Rossjohn. The shaping of T cell receptor recognition by self-tolerance. *Immunity*, 30(2):193–203, Feb 2009

[102] S. Gras, X. Saulquin, J.-B. Reiser, E. Debeaupuis, K. Echasserieau, A. Kissenpfennig, F. Legoux, A. Chouquet, M. Le Gorrec, P. Machillot, B. Neveu, N. Thielens, B. Malissen, M. Bonneville, and D. Housset. Structural bases for the affinity-driven selection of a public TCR against a dominant human cytomegalovirus epitope. *J Immunol*, 183(1):430–437, Jul 2009

[103] S. Gras, Z. Chen, J. J. Miles, Y. C. Liu, M. J. Bell, L. C. Sullivan, L. Kjer-Nielsen, R. M. Brennan, J. M. Burrows, M. A. Neller, R. Khanna, A. W. Purcell, A. G. Brooks, J. McCluskey, J. Rossjohn, and S. R. Burrows. Allelic polymorphism in the T cell receptor and its impact on immune responses. *J Exp Med*, 207(7):1555–1567, Jul 2010

[104] J. J. Gray, S. Moughon, C. Wang, O. Schueler-Furman, B. Kuhlman, C. A. Rohl, and D. Baker. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J Mol Biol*, 331(1):281–99, 2003

[105] J. Greer. Comparative model-building of the mammalian serine proteases. *Journal of molecular biology*, 153(4):1027–1042, Dec 1981

[106] J. U. Guo, Y. Su, C. Zhong, G.-l. Ming, and H. Song. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell*, 145(3):423–434, Apr 2011

[107] M. Hahn, M. J. Nicholson, J. Pyrdol, and K. W. Wucherpfennig. Unconventional topology of self peptide-major histocompatibility complex binding by a human autoimmune T cell receptor. *Nat Immunol*, 6(5):490–496, May 2005

[108] J. N. Haidar, B. Pierce, Y. Yu, W. Tong, M. Li, and Z. Weng. Structure-based design of a T-cell receptor leads to nearly 100-fold improvement in binding affinity for pepMHC. *Proteins*, 74(4):948–60, 2009

[109] M. Harkiolaki, S. L. Holmes, P. Svendsen, J. W. Gregersen, L. T. Jensen, R. McMahon, M. A. Friese, G. van Boxel, R. Etzensperger, J. S. Tzartos, K. Kranc, S. Sainsbury, K. Harlos, E. D. Mellins, J. Palace, M. M. Esiri, P. A. van der Merwe, E. Y. Jones, and L. Fugger. T cell-mediated autoimmune disease due to low-affinity crossreactivity to common microbial peptides. *Immunity*, 30(3):348–357, Mar 2009

[110] C. Hartmann, I. Antes, and T. Lengauer. IRECS: a new algorithm for the selection of most probable ensembles of side-chain conformations in protein models. *Protein Sci*, 16(7):1294–1307, 2007

[111] H. Hasegawa and L. Holm. Advances and pitfalls of protein structural alignment. *Curr Opin Struct Biol*, 19(3):341–348, Jun 2009

[112] H. Hashimoto, J. R. Horton, X. Zhang, M. Bostick, S. E. Jacobsen, and X. Cheng. The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature*, 455(7214):826–829, Oct 2008

[113] Y.-F. He, B.-Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C.-X. Song, K. Zhang, C. He, and G.-L. Xu. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science (New York, N.Y.)*, 333(6047):1303–1307, Sep 2011

[114] S. Henikoff and J. G. Henikoff. Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America*, 89 (22):10915–10919, Nov 1992

[115] J. Hennecke and D. C. Wiley. Structure of a complex of the human alpha/beta T cell receptor (TCR) HA1.7, influenza hemagglutinin peptide, and major histocompatibility complex class II molecule, HLA-DR4 (DRA*0101 and DRB1*0401): insight into TCR cross-restriction and alloreactivity. *The Journal of experimental medicine*, 195(5):571–581, Mar 2002

[116] J. Hennecke, A. Carfi, and D. C. Wiley. Structure of a covalently stabilized complex of a human alphabeta T-cell receptor, influenza HA peptide and MHC class II molecule, HLA-DR1. *EMBO J*, 19(21):5611–5624, Nov 2000

[117] J. N. Herron, X. M. He, D. W. Ballard, P. R. Blier, P. E. Pace, A. L. Bothwell, E. W. J. Voss, and A. B. Edmundson. An autoantibody to single-stranded DNA: comparison of the three-dimensional structures of the unliganded Fab and a deoxynucleotide-Fab complex. *Proteins*, 11(3):159–175, 1991

[118] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18 (12):1463–1472, 1997

[119] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. GROMACS: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, 2008

[120] H. D. Hickman, A. D. Luis, R. Buchli, S. R. Few, M. Sathiamurthy, R. S. VanGundy, C. F. Giberson, and W. H. Hildebrand. Toward a definition of self: proteomic evaluation of the class I peptide repertoire. *Journal of immunology (Baltimore, Md. : 1950)*, 172(5):2944–2952, Mar 2004

[121] A. Hildebrandt, A. K. Dehof, A. Rurainski, A. Bertsch, M. Schumann, N. C. Toussaint, A. Moll, D. Stöckel, S. Nickels, S. C. Mueller, H.-P. Lenhof, and O. Kohlbacher. BALL - biochemical algorithms library 1.3. *BMC Bioinformatics*, 11(1):1–5, 2010

[122] K. L. Ho, I. W. McNae, L. Schmiedeberg, R. J. Klose, A. P. Bird, and M. D. Walkinshaw. MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol Cell*, 29(4):525–531, Feb 2008

[123] H. L. Hoare, L. C. Sullivan, G. Pietra, C. S. Clements, E. J. Lee, L. K. Ely, T. Beddoe, M. Falco, L. Kjer-Nielsen, H. H. Reid, J. McCluskey, L. Moretta, J. Rossjohn, and A. G. Brooks. Structural basis for a major histocompatibility complex class Ib-restricted T cell response. *Nat Immunol*, 7(3):256–264, Mar 2006

[124] R. W. Hockney. The Potential Calculation and Some Applications. *Methods Comput. Phys.*, 9:136–210, 1970

[125] T. Hoffmann, A. M. Krackhardt, and I. Antes. Quantitative Analysis of the Association Angle between T-cell Receptor Valpha/Vbeta Domains Reveals Important Features for Epitope Recognition. *PLoS Comput Biol*, 11(7):e1004244, 2015

[126] L. Holm and C. Sander. Protein structure comparison by alignment of distance matrices. *J Mol Biol*, 233(1):123–138, Sep 1993

[127] R. W. Hooft, C. Sander, and G. Vriend. Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. *Proteins*, 26(4):363–76, 1996

[128] W. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31(3):1695–1697, 1985

[129] W. Humphrey, A. Dalke, and K. Schulten. VMD: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–8, 27–8, Feb 1996

[130] D. Huynh. Metrics for 3D rotations: Comparison and analysis. *Journal of Mathematical Imaging and Vision*, 35(2):155–164, 2009

[131] International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011):931–945, Oct 2004

[132] K. Iqbal, S.-G. Jin, G. P. Pfeifer, and P. E. Szabo. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc Natl Acad Sci U S A*, 108(9):3642–3647, Mar 2011

[133] D. J. Irvine, M. A. Purbhoo, M. Krogsgaard, and M. M. Davis. Direct observation of ligand recognition by T cells. *Nature*, 419(6909):845–849, Oct 2002

[134] J. Ishizuka, G. B. E. Stewart-Jones, A. van der Merwe, J. I. Bell, A. J. McMichael, and E. Y. Jones. The structural dynamics and energetics of an immunodominant T cell receptor are programmed by its Vbeta domain. *Immunity*, 28(2):171–182, Feb 2008

[135] S. Ito, A. C. D'Alessio, O. V. Taranova, K. Hong, L. C. Sowers, and Y. Zhang. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, 466(7310):1129–1133, Aug 2010

[136] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, and Y. Zhang. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science (New York, N.Y.)*, 333(6047):1300–1303, Sep 2011

[137] C. A. Jeneway. *Immunobiology : the immune system in health and disease*. Garland, New York, 2001. ISBN 081533642X 9780815336426

[138] L. L. Jones, L. A. Colf, J. D. Stone, K. C. Garcia, and D. M. Kranz. Distinct CDR3 conformations in TCRs determine the level of cross-reactivity for diverse antigens, but not the docking orientation. *Journal of immunology (Baltimore, Md. : 1950)*, 181(9):6255–6264, Nov 2008

[139] W. Jorgensen and J. Tirado-Rives. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, 110(6):1657–1666, 1988

[140] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983

[141] W. L. Jorgensen, D. S. Maxwell, and T. Rives. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.*, 118(45):11225–11236, 1996

[142] W. Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, 32(5):922–923, Sept. 1976

[143] K. Katoh, K. Misawa, K.-i. Kuma, and T. Miyata. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research*, 30(14):3059–3066, Jul 2002

[144] L. Kaufman, P. J. Rousseeuw, and W. I. O. service). Finding groups in data an introduction to cluster analysis, 1990

[145] B. Kessler, O. Michielin, C. L. Blanchard, I. Apostolou, C. Delarbre, G. Gachelin, C. Gregoire, B. Malissen, J. C. Cerottini, F. Wurm, M. Karplus, and I. F. Luescher. T cell recognition of hapten. Anatomy of T cell receptor binding of a H-2kd-associated photoreactive peptide derivative. *J Biol Chem*, 274(6):3622–31, 1999

[146] J. K. Kim, P.-O. Esteve, S. E. Jacobsen, and S. Pradhan. UHRF1 binds G9a and participates in p21 transcriptional regulation in mammalian cells. *Nucleic Acids Res*, 37(2):493–505, Feb 2009

[147] L. Kjer-Nielsen, C. S. Clements, A. G. Brooks, A. W. Purcell, J. McCluskey, and J. Rossjohn. The 1.5 A crystal structure of a highly selected antiviral T cell receptor provides evidence for a structural basis of immunodominance. *Structure (London, England : 1993)*, 10(11):1521–1532, Nov 2002

[148] L. Kjer-Nielsen, C. S. Clements, A. W. Purcell, A. G. Brooks, J. C. Whisstock, S. R. Burrows, J. McCluskey, and J. Rossjohn. A structural basis for the selection of dominant alphabeta T cell receptors in antiviral immunity. *Immunity*, 18(1):53–64, Jan 2003

[149] L. Kjer-Nielsen, N. A. Borg, D. G. Pellicci, T. Beddoe, L. Kostenko, C. S. Clements, N. A. Williamson, M. J. Smyth, G. S. Besra, H. H. Reid, M. Bharadwaj, D. I. Godfrey, J. Rossjohn, and J. McCluskey. A structural basis for selection and cross-species reactivity of the semi-invariant NKT cell receptor in CD1d/glycolipid recognition. *J Exp Med*, 203(3):661–673, Mar 2006

[150] J. Kleinjung and F. Fraternali. Design and application of implicit solvent models in biomolecular simulations. *Current opinion in structural biology*, 25:126–134, Apr 2014

[151] B. Knapp, G. Dorffner, and W. Schreiner. Early relaxation dynamics in the LC 13 T cell receptor in reaction to 172 altered peptide ligands: a molecular dynamics simulation study. *PLoS One*, 8(6):e64464, 2013

[152] B. Knapp, J. Dunbar, and C. M. Deane. Large scale characterization of the LC13 TCR and HLA-B8 structural landscape in reaction to 172 altered peptide ligands: a molecular dynamics simulation study. *PLoS Comput Biol*, 10(8):e1003748, Aug 2014

[153] P. Koehl. Protein Structure Classification. In K. B. Lipkowitz, T. R. Cundari, and D. B. Gillet, V. J. B. Boyd, editors, *Reviews in Computational Chemistry*, volume 22, chapter 1, page 30. John Wiley & Sons, Inc., 2006. ISBN 9780471780366

[154] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, and T. E. Cheatham. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accounts of chemical research*, 33(12): 889–897, Dec 2000

[155] R.-Y. Konig, Huang and R. N. Germain. MHC class II interaction with CD4 mediated by a region analogous to the MHC class I binding site for CD8. *Nature*, 356(6372): 796–798, 04 1992

[156] S. Kriaucionis and N. Heintz. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, 324(5929):929–930, May 2009

[157] G. G. Krivov, M. V. Shapovalov, and J. Dunbrack, R. L. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins*, 77(4):778–95, 2009

[158] G. J. A. Kroon, H. Mo, M. A. Martinez-Yamout, H. J. Dyson, and P. E. Wright. Changes in structure and dynamics of the Fv fragment of a catalytic antibody upon binding of inhibitor. *Protein Sci*, 12(7):1386–1394, Jul 2003

[159] P. Labute. Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins*, 75(1):187–205, 2009

[160] M. B. Lascombe, P. M. Alzari, G. Boulot, P. Saludjian, P. Tougard, C. Berek, S. Haba, E. M. Rosen, A. Nisonoff, and R. J. Poljak. Three-dimensional structure of Fab R19.9, a monoclonal murine antibody specific for the p-azobenzenearsonate group. *Proceedings of the National Academy of Sciences of the United States of America*, 86(2):607–611, Jan 1989

[161] A. Leach. *Molecular Modelling: Principles and Applications (2nd Edition)*. Prentice Hall, 2 edition, Apr. 2001. ISBN 0582382106

[162] F. S. Lee, Z. T. Chu, M. B. Bolger, and A. Warshel. Calculations of antibody-antigen interactions: microscopic and semi-microscopic evaluation of the free energies of binding of phosphorylcholine analogs to McPC603. *Protein engineering*, 5(3):215–228, Apr 1992

[163] M.-P. Lefranc, C. Pommié, M. Ruiz, V. Giudicelli, E. Foulquier, L. Truong, V. Thouvenin-Contet, and G. Lefranc. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Developmental and comparative immunology*, 27(1):55–77, Jan 2003

[164] M.-P. Lefranc, C. Pommié, Q. Kaas, E. Duprat, N. Bosc, D. Guiraudou, C. Jean, M. Ruiz, I. Da Piédade, M. Rouard, E. Foulquier, V. Thouvenin, and G. Lefranc. IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. *Developmental and comparative immunology*, 29 (3):185–203, 2005

[165] H. Lei, S. P. Oh, M. Okano, R. Jüttermann, K. A. Goss, R. Jaenisch, and E. Li. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development (Cambridge, England)*, 122(10):3195–3205, Oct 1996

[166] A. Leimgruber, M. Ferber, M. Irving, H. Hussain-Kahn, S. Wieckowski, L. Derre, N. Rufer, V. Zoete, and O. Michielin. TCRep 3D: an automated in silico approach to study the structural properties of TCR repertoires. *PLoS One*, 6(10):e26301, 2011

[167] H. Leonhardt, A. W. Page, H. U. Weier, and T. H. Bestor. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell*, 71(5):865–873, Nov 1992

[168] E. Li, T. H. Bestor, and R. Jaenisch. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, 69(6):915–926, Jun 1992

[169] H. Li, M. I. Lebedeva, E. S. Ward, and R. A. Mariuzza. Dual conformations of a T cell receptor V alpha homodimer: implications for variability in V alpha V beta domain association. *J Mol Biol*, 269(3):385–394, Jun 1997

[170] X. Li, M. P. Jacobson, K. Zhu, S. Zhao, and R. A. Friesner. Assignment of polar states for protein amino acid residues using an interaction cluster decomposition algorithm and its application to high resolution protein structure modeling. *Proteins*, 66(4): 824–37, 2007

[171] Y. Li, Y. Huang, J. Lue, J. A. Quandt, R. Martin, and R. A. Mariuzza. Structure of a human autoimmune TCR bound to a myelin basic protein self-peptide and a multiple sclerosis-associated MHC class II molecule. *EMBO J*, 24(17):2968–2979, Sep 2005

[172] Y. Li, G. Parry, L. Chen, J. A. Callahan, D. E. Shaw, E. J. Meehan, A. P. Mazar, and M. Huang. An anti-urokinase plasminogen activator receptor (uPAR) antibody: crystal structure and binding epitope. *J Mol Biol*, 365(4):1117–1129, Jan 2007

[173] C. Ling and L. Groop. Epigenetics: a molecular link between environmental factors and type 2 diabetes. *Diabetes*, 58(12):2718–2725, Dec 2009

[174] T. Lippert and M. Rarey. Fast automated placement of polar hydrogen atoms in protein-ligand complexes. *J Cheminform*, 1(1):13, 2009

[175] D. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(1-3):503–528, 1989

[176] J. G. Luz, M. Huang, K. C. Garcia, M. G. Rudolph, V. Apostolopoulos, L. Teyton, and I. A. Wilson. Structural comparison of allogeneic and syngeneic T cell receptor-peptide-major histocompatibility complex complexes: a buried alloreactive mutation subtly alters peptide presentation substantially increasing V(beta) Interactions. *The Journal of experimental medicine*, 195(9):1175–1186, May 2002

[177] J. Ma and S. Wang. Algorithms, Applications, and Challenges of Protein Structure Alignment. volume 94 of *Advances in Protein Chemistry and Structural Biology*, chapter 5, pages 121 – 175. Academic Press, 2014

[178] W. A. Macdonald, Z. Chen, S. Gras, J. K. Archbold, F. E. Tynan, C. S. Clements, M. Bharadwaj, L. Kjer-Nielsen, P. M. Saunders, M. C. J. Wilce, F. Crawford, B. Stadinsky, D. Jackson, A. G. Brooks, A. W. Purcell, J. W. Kappler, S. R. Burrows, J. Rossjohn, and J. McCluskey. T cell allorecognition via molecular mimicry. *Immunity*, 31(6):897–908, Dec 2009

[179] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry. B*, 102(18):3586–3616, Apr 1998

[180] A. D. MacKerell, N. Banavali, and N. Foloppe. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers*, 56(4):257–265, 2001 2000

[181] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 281–297, Berkeley, Calif., 1967. University of California Press

[182] M. W. Mahoney and W. L. Jorgensen. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *The Journal of Chemical Physics*, 112(20):8910–8922, May 2000

[183] A. Marchler-Bauer, M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz, L. Y. Geer, R. C. Geer, J. He, M. Gwadz, D. I. Hurwitz, C. J. Lanczycki, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, Z. Wang, R. A. Yamashita, D. Zhang, C. Zheng, and S. H. Bryant. CDD: NCBI's conserved domain database. *Nucleic acids research*, 43(Database issue):D222–D226, Jan 2015

[184] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries. The MARTINI force field: coarse grained model for biomolecular simulations. *J Phys Chem B*, 111(27):7812–24, 2007

[185] G. J. Martyna, D. J. Tobias, and M. L. Klein. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics*, 101(5):4177–4189, 1994

[186] J. Maynard, K. Petersson, D. H. Wilson, E. J. Adams, S. E. Blondelle, M. J. Boulanger, D. B. Wilson, and K. C. Garcia. Structure of an autoimmune T cell receptor complexed with class II peptide-MHC: insights into MHC bias and antigen specificity. *Immunity*, 22(1):81–92, Jan 2005

[187] C. Mazza, N. Auphan-Anezin, C. Gregoire, A. Guimezanes, C. Kellenberger, A. Roussel, A. Kearney, P. A. van der Merwe, A.-M. Schmitt-Verhulst, and B. Malissen. How much can a T-cell antigen receptor adapt to structurally distinct antigenic peptides? *EMBO J*, 26(7):1972–1983, Apr 2007

[188] C. McBeth, A. Seamons, J. C. Pizarro, S. J. Fleishman, D. Baker, T. Kortemme, J. M. Goverman, and R. K. Strong. A new twist in TCR diversity revealed by a forbidden alphabeta TCR. *J Mol Biol*, 375(5):1306–1319, 2008

[189] J. A. McCammon, B. R. Gelin, and M. Karplus. Dynamics of folded proteins. *Nature*, 267(5612):585–590, Jun 1977

[190] I. K. McDonald and J. M. Thornton. Satisfying hydrogen bonding potential in proteins. *J Mol Biol*, 238(5):777–93, 1994

[191] L. L. McQuitty. Elementary Linkage Analysis for Isolating Orthogonal and Oblique Types and Typal Relevancies. *Educational and Psychological Measurement*, 17(2): 207–229, jul 1957

[192] D. Meilinger, K. Fellinger, S. Bultmann, U. Rothbauer, I. M. Bonapace, W. E. F. Klinkert, F. Spada, and H. Leonhardt. Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. *EMBO Rep*, 10(11):1259–1264, Nov 2009

[193] M. A. Messih, R. Lepore, and A. Tramontano. LoopIng: a template-based tool for predicting the structure of protein loops. *Bioinformatics (Oxford, England)*, 31(23): 3767–3772, Dec 2015

[194] R. Meyers. *Epigenetic Regulation and Epigenomics: Advances in Molecular Biology and Medicine*. Wiley-Blackwell, 2012. ISBN 9783527326822

[195] D. L. Michels and M. Desbrun. A Semi-analytical Approach to Molecular Dynamics. *J. Comput. Phys.*, 303(C):336–354, Dec. 2015

[196] O. Michielin and M. Karplus. Binding free energy differences in a TCR-peptide-MHC complex induced by a peptide mutation: a simulation analysis. *Journal of molecular biology*, 324(3):547–569, Nov 2002

[197] O. Michielin, I. Luescher, and M. Karplus. Modeling of the TCR-MHC-peptide complex. *J Mol Biol*, 300(5):1205–1235, Jul 2000

[198] J. J. Miles, A. M. Bulek, D. K. Cole, E. Gostick, A. J. A. Schauenburg, G. Dolton, V. Venturi, M. P. Davenport, M. P. Tan, S. R. Burrows, L. Wooldridge, D. A. Price, P. J. Rizkallah, and A. K. Sewell. Genetic and structural basis for selection of a ubiquitous T cell receptor deployed in Epstein-Barr virus infection. *PLoS Pathog*, 6 (11):e1001198, 2010

[199] P. J. Miller, Y. Pazy, B. Conti, D. Riddle, E. Appella, and E. J. Collins. Single MHC mutation eliminates enthalpy associated with T cell receptor binding. *J Mol Biol*, 373(2):315–327, Oct 2007

[200] H. Mirzaei, D. Beglov, I. C. Paschalidis, S. Vajda, P. Vakili, and D. Kozakov. Rigid Body Energy Minimization on Manifolds for Molecular Docking. *J Chem Theory Comput*, 8(11):4374–4380, 2012

[201] H. Mirzaei, S. Zarbafian, E. Villar, S. Mottarella, D. Beglov, S. Vajda, I. C. Paschalidis, P. Vakili, and D. Kozakov. Energy Minimization on Manifolds for Docking Flexible Molecules. *Journal of chemical theory and computation*, 11(3):1063–1076, Mar 2015

[202] S. Miyamoto and P. A. Kollman. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.*, 13(8):952–962, Oct. 1992

[203] S. Monaco-Malbet, C. Berthet-Colominas, A. Novelli, N. Battaï, N. Piga, V. Cheynet, F. Mallet, and S. Cusack. Mutual conformational adaptations in antigen and antibody upon complex formation between an Fab and HIV-1 capsid protein p24. *Structure (London, England : 1993)*, 8(10):1069–1077, Oct 2000

[204] R. A. Morgan, M. E. Dudley, J. R. Wunderlich, M. S. Hughes, J. C. Yang, R. M. Sherry, R. E. Royal, S. L. Topalian, U. S. Kammula, N. P. Restifo, Z. Zheng, A. Nahvi, C. R. de Vries, L. J. Rogers-Freezer, S. A. Mavroukakis, and S. A. Rosenberg. Cancer regression in patients after transfer of genetically engineered lymphocytes. *Science*, 314(5796):126–129, Oct 2006

[205] S. Mosimann, R. Meleshko, and M. N. G. James. A critical assessment of comparative molecular modeling of tertiary structures of proteins*. *Proteins: Structure, Function, and Bioinformatics*, 23(3):301–317, 1995

[206] J. Moult, K. Fidelis, A. Kryshtafovych, T. Schwede, and A. Tramontano. Critical assessment of methods of protein structure prediction (CASP) —round x. *Proteins*, 82(0 2):1–6, 02 2014

[207] M. T. Nelson, W. Humphrey, A. Gursoy, A. Dalke, L. V. Kalé, R. D. Skeel, and K. Schulten. NAMD: a Parallel, Object-Oriented Molecular Dynamics Program. *International Journal of High Performance Computing Applications*, 10(4):251–268, 1996

[208] E. W. Newell, L. K. Ely, A. C. Kruse, P. A. Reay, S. N. Rodriguez, A. E. Lin, M. S. Kuhns, K. C. Garcia, and M. M. Davis. Structural basis of specificity and cross-reactivity in T cell receptors specific for cytochrome c-I-E(k). *J Immunol*, 186(10): 5823–5832, May 2011

[209] S. Nosé. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics*, 81(1):511–519, July 1984

[210] M. Okano, D. W. Bell, D. A. Haber, and E. Li. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, 99(3):247–257, Oct 1999

[211] K. Okonechnikov, O. Golosova, M. Fursov, and UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics (Oxford, England)*, 28(8):1166–1167, Apr 2012

[212] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem*, 25(13):1656–76, 2004

[213] D. Pearlman. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications*, 91(1-3):1–41, Sept. 1995

[214] D. G. Pellicci, A. P. Uldrich, J. Le Nours, F. Ross, E. Chabrol, S. B. G. Eckle, R. de Boer, R. T. Lim, K. McPherson, G. Besra, A. R. Howell, L. Moretta, J. McCluskey, M. H. M. Heemskerk, S. Gras, J. Rossjohn, and D. I. Godfrey. The molecular bases of delta/alphabeta T cell-mediated antigen recognition. *J Exp Med*, 211(13): 2599–2615, Dec 2014

[215] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. Scalable molecular dynamics with NAMD. *J Comput Chem*, 26(16):1781–1802, Dec 2005

[216] K. H. Piepenbrink, O. Y. Borbulevych, R. F. Sommese, J. Clemens, K. M. Armstrong, C. Desmond, P. Do, and B. M. Baker. Fluorine substitutions in an antigenic peptide selectively modulate T-cell receptor binding in a minimally perturbing manner. *Biochem J*, 423(3):353–361, Nov 2009

[217] B. G. Pierce and Z. Weng. A flexible docking approach for prediction of T cell receptor-peptide-MHC complexes. *Protein Sci*, 22(1):35–46, 2013

[218] B. G. Pierce, L. M. Hellman, M. Hossain, N. K. Singh, C. W. Vander Kooi, Z. Weng, and B. M. Baker. Computational design of the affinity and specificity of a therapeutic T cell receptor. *PLoS Comput Biol*, 10(2):e1003478, Feb 2014

[219] B. G. Pierce, T. Vreven, and Z. Weng. Modeling T cell receptor recognition of CD1-lipid and MR1-metabolite complexes. *BMC bioinformatics*, 15:319, 2014

[220] K. S. Pitzer. Potential energies for rotation about single bonds. *Discuss. Faraday Soc.*, 10:66–73, 1951

[221] E. Polak and G. Ribiere. Note sur la convergence de méthodes de directions conjuguées. *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique*, 3(R1):35–43, 1969

[222] S. Porcelli, M. B. Brenner, J. L. Greenstein, S. P. Balk, C. Terhorst, and P. A. Bleicher. Recognition of cluster of differentiation 1 antigens by human CD4-CD8-cytolytic T lymphocytes. *Nature*, 341(6241):447–450, Oct 1989

[223] A. Portela and M. Esteller. Epigenetic modifications and human disease. *Nature biotechnology*, 28(10):1057–1068, Oct 2010

[224] W. Qiu-Dong. The global solution of the N-body problem. *Celestial Mechanics and Dynamical Astronomy*, 50(1):73–88, 1990

[225] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0

[226] M. Rasool, A. Malik, M. I. Naseer, A. Manan, S. Ansari, I. Begum, M. H. Qazi, P. Pushparaj, A. M. Abuzenadah, M. H. Al-Qahtani, M. A. Kamal, and S. Gan. The role of epigenetics in personalized medicine: challenges and opportunities. *BMC medical genomics*, 8 Suppl 1:S5, Jan 2015

[227] E. L. Reinherz, K. Tan, L. Tang, P. Kern, J. Liu, Y. Xiong, R. E. Hussey, A. Smolyar, B. Hare, R. Zhang, A. Joachimiak, H. C. Chang, G. Wagner, and J. Wang. The crystal structure of a T cell receptor in complex with peptide and MHC class II. *Science (New York, N.Y.)*, 286(5446):1913–1921, Dec 1999

[228] J. B. Reiser, C. Darnault, A. Guimezanes, C. Gregoire, T. Mosser, A. M. Schmitt-Verhulst, J. C. Fontecilla-Camps, B. Malissen, D. Housset, and G. Mazza. Crystal structure of a T cell receptor bound to an allogeneic MHC molecule. *Nat Immunol*, 1(4):291–297, Oct 2000

[229] J. B. Reiser, C. Grégoire, C. Darnault, T. Mosser, A. Guimezanes, A. M. Schmitt-Verhulst, J. C. Fontecilla-Camps, G. Mazza, B. Malissen, and D. Housset. A T cell receptor CDR3beta loop undergoes conformational changes of unprecedented magnitude upon binding to a peptide/MHC class I complex. *Immunity*, 16(3):345–354, Mar 2002

[230] J.-B. Reiser, C. Darnault, C. Gregoire, T. Mosser, G. Mazza, A. Kearney, P. A. van der Merwe, J. C. Fontecilla-Camps, D. Housset, and B. Malissen. CDR3 loop flexibility contributes to the degeneracy of TCR recognition. *Nat Immunol*, 4(3): 241–247, Mar 2003

[231] J. M. Rini, U. Schulze-Gahmen, and I. A. Wilson. Structural evidence for induced fit as a mechanism for antibody-antigen recognition. *Science (New York, N.Y.)*, 255 (5047):959–965, Feb 1992

[232] K. D. Robertson. DNA methylation and human disease. *Nature reviews. Genetics*, 6(8):597–610, Aug 2005

[233] J. Robinson, J. A. Halliwell, J. D. Hayhurst, P. Flicek, P. Parham, and S. G. E. Marsh. The IPD and IMGT/HLA database: allele variant databases. *Nucleic acids research*, 43(Database issue):D423–D431, Jan 2015

[234] D. Rognan, A. Stryhn, L. Fugger, S. Lyngbaek, J. Engberg, P. S. Andersen, and S. Buus. Modeling the interactions of a peptide-major histocompatibility class I ligand with its receptors. I. Recognition by two alpha beta T cell receptors. *Journal of computer-aided molecular design*, 14(1):53–69, Jan 2000

[235] A. Rottach, H. Leonhardt, and F. Spada. DNA methylation-mediated epigenetic control. *J Cell Biochem*, 108(1):43–51, Sep 2009

[236] A. Rottach, C. Frauer, G. Pichler, I. M. Bonapace, F. Spada, and H. Leonhardt. The multi-domain protein Np95 connects DNA methylation and histone modification. *Nucleic Acids Res*, 38(6):1796–1804, Apr 2010

[237] M. G. Rudolph, R. L. Stanfield, and I. A. Wilson. How TCRs bind MHCs, peptides, and coreceptors. *Annu Rev Immunol*, 24:419–66, 2006

[238] C. Rueckert and C. A. Guzman. Vaccines: from empirical development to rational design. *PLoS Pathog*, 8(11):e1003001, 2012

[239] V. E. A. Russo, R. A. Martienssen, and A. D. Riggs. *Epigenetic mechanisms of gene regulation*, pages 1–4. Plainview, N.Y.: Cold Spring Harbor Laboratory Press, 1996. ISBN 0879694904; 9780879694906

[240] J.-P. Ryckaert, G. Ciccotti, and H. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3):327–341, 1977

[241] A. Sali and T. L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*, 234(3):779–815, Dec 1993

[242] M. Saline, K. E. J. Rodstrom, G. Fischer, V. Y. Orekhov, B. G. Karlsson, and K. Lindkvist-Petersson. The structure of superantigen complexed with TCR and MHC reveals novel insights into superantigenic T cell activation. *Nat Commun*, 1: 119, 2010

[243] R. D. Salter, R. J. Benjamin, P. K. Wesley, S. E. Buxton, T. P. J. Garrett, C. Clayberger, A. M. Krensky, A. M. Norment, D. R. Littman, and P. Parham. A binding site for the T-cell co-receptor CD8 on the [alpha]3 domain of HLA-A2. *Nature*, 345 (6270):41–46, may 1990

[244] M. Sami, P. J. Rizkallah, S. Dunn, P. Molloy, R. Moysey, A. Vuidepot, E. Baston, P. Todorov, Y. Li, F. Gao, J. M. Boulter, and B. K. Jakobsen. Crystal structures of high affinity human T-cell receptors bound to peptide major histocompatibility complex reveal native diagonal binding geometry. *Protein Eng Des Sel*, 20(8):397–403, Aug 2007

[245] N. Sasai and P.-A. Defossez. Many paths to one goal? The proteins that recognize methylated DNA in eukaryotes. *Int J Dev Biol*, 53(2-3):323–334, 2009

[246] T. Schlick. *Molecular Modeling and Simulation: An Interdisciplinary Guide*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2002. ISBN 038795404X

[247] T. Schwede, A. Sali, N. Eswar, and M. C. Peitsch. *Protein Structure Modeling*, pages 3–35. WORLD SCIENTIFIC, 2016/04/11 2008. ISBN 978-981-277-877-2

[248] D. K. Sethi, D. A. Schubert, A.-K. Anders, A. Heroux, D. A. Bonsor, C. P. Thomas, E. J. Sundberg, J. Pyrdol, and K. W. Wucherpfennig. A highly tilted binding mode by a self-reactive T cell receptor results in altered engagement of peptide and MHC. *J Exp Med*, 208(1):91–102, Jan 2011

[249] D. F. Shanno. Conditioning of Quasi-Newton Methods for Function Minimization. *Mathematics of Computation*, 24(111):647–656, July 1970

[250] J. Sharif, M. Muto, S.-i. Takebayashi, I. Suetake, A. Iwamatsu, T. A. Endo, J. Shinga, Y. Mizutani-Koseki, T. Toyoda, K. Okamura, S. Tajima, K. Mitsuya, M. Okano, and H. Koseki. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature*, 450(7171):908–912, Dec 2007

[251] K. T. Simons, C. Kooperberg, E. Huang, and D. Baker. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J Mol Biol*, 268(1):209–25, 1997

[252] P. H. A. Sneath. The application of computers to taxonomy. *Journal of general microbiology*, 17(1):201–226, Aug 1957

[253] R. R. Sokal and C. D. Michener. A statistical method for evaluating systematic relationships. *University of Kansas Scientific Bulletin*, 28:1409–1438, 1958

[254] J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman, and D. A. Case. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices. *Journal of the American Chemical Society*, 120(37):9401–9409, Aug. 1998

[255] R. L. Stanfield, M. Takimoto-Kamimura, J. M. Rini, A. T. Profy, and I. A. Wilson. Major antigen-induced domain rearrangements in an antibody. *Structure (London, England : 1993)*, 1(2):83–93, Oct 1993

[256] A. Stavrakoudis. Insights into the structure of the LC13 TCR/HLA-B8-EBV peptide complex with molecular dynamics simulations. *Cell Biochem Biophys*, 60(3):283–295, Jul 2011

[257] F. J. Stevens, C. H. Chang, and M. Schiffer. Dual conformations of an immunoglobulin light-chain dimer: heterogeneity of antigen specificity and idiotope profile may result from multiple variable-domain interaction mechanisms. *Proceedings of the National Academy of Sciences of the United States of America*, 85(18):6895–6899, Sep 1988

[258] G. B. E. Stewart-Jones, A. J. McMichael, J. I. Bell, D. I. Stuart, and E. Y. Jones. A structural basis for immunodominant human T cell receptor recognition. *Nat Immunol*, 4(7):657–663, Jul 2003

[259] F. H. Stillinger and A. Rahman. Improved simulation of liquid water by molecular dynamics. *The Journal of Chemical Physics*, 60(4):1545–1557, Feb. 1974

[260] R. Sun, S. E. Shepherd, S. S. Geier, C. T. Thomson, J. M. Sheil, and S. G. Nathenson. Evidence that the antigen receptors of cytotoxic T lymphocytes interact with a common recognition pattern on the H-2Kb molecule. *Immunity*, 3(5):573–82, 1995

[261] R. Suzuki and H. Shimodaira. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, 22(12):1540–1542, Jun 2006

[262] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.*, 76(1): 637–649, Jan. 1982

[263] A. Szwagierczak, S. Bultmann, C. S. Schmidt, F. Spada, and H. Leonhardt. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res*, 38(19):e181, Oct 2010

[264] M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, and A. Rao. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, 324(5929):930–935, May 2009

[265] O. Takeuchi, K. Hoshino, T. Kawai, H. Sanjo, H. Takada, T. Ogawa, K. Takeda, and S. Akira. Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components. *Immunity*, 11(4):443–451, Oct 1999

[266] A. Teplyakov, G. Obmolova, T. Malia, and G. Gilliland. Antigen recognition by antibody C836 through adjustment of V(L)/V(H) packing. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, 67(Pt 10):1165–1167, Oct 2011

[267] F. E. Tynan, S. R. Burrows, A. M. Buckle, C. S. Clements, N. A. Borg, J. J. Miles, T. Beddoe, J. C. Whisstock, M. C. Wilce, S. L. Silins, J. M. Burrows, L. Kjer-Nielsen, L. Kostenko, A. W. Purcell, J. McCluskey, and J. Rossjohn. T cell receptor recognition of a 'super-bulged' major histocompatibility complex class I-bound peptide. *Nat Immunol*, 6(11):1114–1122, Nov 2005

[268] F. E. Tynan, H. H. Reid, L. Kjer-Nielsen, J. J. Miles, M. C. J. Wilce, L. Kostenko, N. A. Borg, N. A. Williamson, T. Beddoe, A. W. Purcell, S. R. Burrows, J. McCluskey, and J. Rossjohn. A T cell receptor flattens a bulged antigenic peptide presented by a major histocompatibility complex class I molecule. *Nat Immunol*, 8 (3):268–276, Mar 2007

[269] R. G. Urdinguio, J. V. Sanchez-Mut, and M. Esteller. Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies. *The Lancet. Neurology*, 8 (11):1056–1072, Nov 2009

[270] V. Valinluck and L. C. Sowers. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res*, 67 (3):946–950, Feb 2007

[271] V. Valinluck, H.-H. Tsai, D. K. Rogstad, A. Burdzy, A. Bird, and L. C. Sowers. Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic Acids Res*, 32(14):4100–4108, 2004

[272] G. I. van Boxel, S. Holmes, L. Fugger, and E. Y. Jones. An alternative conformation of the T-cell receptor alpha constant region. *J Mol Biol*, 400(4):828–837, Jul 2010

[273] P. A. van der Merwe and O. Dushek. Mechanisms for T cell receptor triggering. *Nat Rev Immunol*, 11(1):47–55, Jan 2011

[274] P. A. van der Merwe, S. J. Davis, A. S. Shaw, and M. L. Dustin. Cytoskeletal polarization and redistribution of cell-surface molecules during T cell antigen recognition. *Seminars in immunology*, 12(1):5–21, Feb 2000

[275] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: fast, flexible, and free. *Journal of computational chemistry*, 26(16):1701–1718, Dec 2005

[276] L. Verlet. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.*, 159:98–103, Jul 1967

[277] B. Wallner and A. Elofsson. All are not equal: a benchmark of different homology modeling programs. *Protein science : a publication of the Protein Society*, 14(5): 1315–1327, May 2005

[278] S. Wan, P. V. Coveney, and D. R. Flower. Molecular basis of peptide recognition by the TCR: affinity differences calculated using large scale computing. *Journal of immunology (Baltimore, Md. : 1950)*, 175(3):1715–1723, Aug 2005

[279] S. Wan, D. R. Flower, and P. V. Coveney. Toward an atomistic understanding of the immune synapse: large-scale molecular dynamics simulation of a membrane-embedded TCR-pMHC-CD4 complex. *Mol Immunol*, 45(5):1221–1230, Mar 2008

[280] J. Wang, K. Lim, A. Smolyar, M. Teng, J. Liu, A. G. Tse, J. Liu, R. E. Hussey, Y. Chishti, C. T. Thomson, R. M. Sweet, S. G. Nathenson, H. C. Chang, J. C. Sacchettini, and E. L. Reinherz. Atomic structure of an alphabeta T cell receptor (TCR) heterodimer in complex with an anti-TCR fab fragment derived from a mitogenic antibody. *EMBO J*, 17(1):10–26, Jan 1998

[281] L. Wang, Y. Zhao, Z. Li, Y. Guo, L. L. Jones, D. M. Kranz, W. Mourad, and H. Li. Crystal structure of a complete ternary complex of TCR, superantigen and peptide-MHC. *Nat Struct Mol Biol*, 14(2):169–171, Feb 2007

[282] W. Wang, J. Wang, and P. A. Kollman. What determines the van der Waals coefficient beta in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations? *Proteins*, 34(3):395–402, Feb 1999

[283] J. H. Ward. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301):236–244, Mar. 1963

[284] R. L. Warren, J. D. Freeman, T. Zeng, G. Choe, S. Munro, R. Moore, J. R. Webb, and R. A. Holt. Exhaustive T-cell repertoire sequencing of human peripheral blood samples reveals signatures of antigen selection and a directly measured repertoire size of at least 1 million clonotypes. *Genome Res*, 21(5):790–7, 2011

[285] C. X. Weichenberger and M. J. Sippl. Self-consistent assignment of asparagine and glutamine amide rotamers in protein crystal structures. *Structure*, 14(6):967–72, 2006

[286] C. X. Weichenberger and M. J. Sippl. NQ-Flipper: recognition and correction of erroneous asparagine and glutamine side-chain rotamers in protein structures. *Nucleic Acids Res*, 35(Web Server issue):W403–6, 2007

[287] C. X. Weichenberger, P. Byzia, and M. J. Sippl. Visualization of unfavorable interactions in protein folds. *Bioinformatics*, 24(9):1206–7, 2008

[288] P. K. Weiner and P. A. Kollman. AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. *J. Comput. Chem.*, 2(3):287–303, 1981

[289] G. Werlen and E. Palmer. The T-cell receptor signalosome: a dynamic structure with expanding complexity. *Current opinion in immunology*, 14(3):299–305, Jun 2002

[290] L. V. Woodcock. Isothermal molecular dynamics calculations for liquid salts. *Chemical Physics Letters*, 10(3):257–261, Aug. 1971

[291] J. M. Word, S. C. Lovell, J. S. Richardson, and D. C. Richardson. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol*, 285(4):1735–47, 1999

[292] M. Wossidlo, T. Nakamura, K. Lepikhov, C. J. Marques, V. Zakhartchenko, M. Boiani, J. Arand, T. Nakano, W. Reik, and J. Walter. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat Commun*, 2: 241, 2011

[293] Y. Yin, Y. Li, M. C. Kerzic, R. Martin, and R. A. Mariuzza. Structure of a TCR with high affinity for self-antigen reveals basis for escape from negative selection. *EMBO J*, 30(6):1137–1148, Mar 2011

[294] M. Yoneya, H. J. C. Berendsen, and K. Hirasawa. A Non-Iterative Matrix Method for Constraint Molecular Dynamics Simulations. *Molecular Simulation*, 13(6):395–405, 1994

[295] K. Yoshida, A. L. Corper, R. Herro, B. Jabri, I. A. Wilson, and L. Teyton. The diabetogenic mouse MHC class II molecule I-Ag7 is endowed with a switch that modulates TCR affinity. *J Clin Invest*, 120(5):1578–1590, May 2010

[296] X.-Z. Yu, P. J. Martin, and C. Anasetti. CD28 signal enhances apoptosis of CD8 T cells after strong TCR ligation. *Journal of immunology (Baltimore, Md. : 1950)*, 170(6):3002–3006, Mar 2003

[297] L. Zhang, X. Lu, J. Lu, H. Liang, Q. Dai, G.-L. Xu, C. Luo, H. Jiang, and C. He. Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nature chemical biology*, 8(4):328–330, Apr 2012

[298] Y. Zhang. I-TASSER server for protein 3D structure prediction. *BMC bioinformatics*, 9:40, Jan 2008

[299] V. Zoete and O. Michielin. Comparison between computational alanine scanning and per-residue binding free energy decomposition for protein-protein association using MM-GBSA: application to the TCR-p-MHC complex. *Proteins*, 67(4):1026–1047, Jun 2007

[300] V. Zoete, M. B. Irving, and O. Michielin. MM-GBSA binding free energy decomposition and T cell receptor engineering. *J Mol Recognit*, 23(2):142–152, Mar-Apr 2010

[301] V. Zoete, M. Irving, M. Ferber, M. A. Cuendet, and O. Michielin. Structure-Based, Rational Design of T Cell Receptors. *Front Immunol*, 4:268, 2013

Appendix A

Embedded Publication 1

A.1 Copy Permissions

A.2 Full Article

OPEN ACCESS Freely available online

PLoS one

Recognition of 5-Hydroxymethylcytosine by the Uhrf1 SRA Domain

Carina Frauer[1,4☯], Thomas Hoffmann[2,4☯], Sebastian Bultmann[1,4], Valentina Casa[3], M. Cristina Cardoso[3], Iris Antes[2,4]*, Heinrich Leonhardt[1,4]*

1 Department of Biology II, Ludwig Maximilians University Munich, Planegg-Martinsried, Germany, **2** Department of Life Sciences, Technical University Munich, Freising-Weihenstephan, Germany, **3** Department of Biology, Technical University Darmstadt, Darmstadt, Germany, **4** Center for Integrated Protein Science Munich (CIPSM), Munich, Germany

Abstract

Recent discovery of 5-hydroxymethylcytosine (5hmC) in genomic DNA raises the question how this sixth base is recognized by cellular proteins. In contrast to the methyl-CpG binding domain (MBD) of MeCP2, we found that the SRA domain of Uhrf1, an essential factor in DNA maintenance methylation, binds 5hmC and 5-methylcytosine containing substrates with similar affinity. Based on the co-crystal structure, we performed molecular dynamics simulations of the SRA:DNA complex with the flipped cytosine base carrying either of these epigenetic modifications. Our data indicate that the SRA binding pocket can accommodate 5hmC and stabilizes the flipped base by hydrogen bond formation with the hydroxyl group.

Citation: Frauer C, Hoffmann T, Bultmann S, Casa V, Cardoso MC, et al. (2011) Recognition of 5-Hydroxymethylcytosine by the Uhrf1 SRA Domain. PLoS ONE 6(6): e21306. doi:10.1371/journal.pone.0021306

Editor: Shuang-yong Xu, New England Biolabs, Inc., United States of America

Received March 18, 2011; **Accepted** May 25, 2011; **Published** June 22, 2011

Funding: This work was supported by grants from the Nanosystems Initiative Munich (NIM, http://www.nano-initiative-munich.de), the Center for NanoSciences (CeNS, http://www.cens.de) Munich, and by grants from the Deutsche Forschungsgemeinschaft (DFG, SFB TR5, http://www.dfg.de) to HL. CF gratefully acknowledges support by the Elite Network of Bavaria (International Doctorate Program NanoBioTechnology, IDK-NBT, http://www.cens.de/doctorate-program) and the International Max Planck Research School for Molecular and Cellular Life Sciences (IMPRS-LS, http://www.imprs-ls.de). SB was supported by the Graduate School for Life Sciences Munich (LSM, http://www.lsm.bio.lmu.de). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: antes@wzw.tum.de (IA); h.leonhardt@lmu.de (HL)

☯ These authors contributed equally to this work.

Introduction

DNA methylation is an epigenetic modification that is well known to control eukaryotic gene expression [1,2]. In fact, methylation of regulatory sequences often correlates with a transcriptionally silent state. DNA methylation in mammals occurs as 5-methylcytosine (5mC) within CpG dinucleotides and is catalyzed by a family of DNA methyltransferases (Dnmts) [3]. Dnmt members are distinguished by their function; while the *de novo* methyltransferases Dnmt3a and Dnmt3b establish methylation patterns during development and cellular differentiation [4,5], the *maintenance* methyltransferase Dnmt1 copies these patterns during DNA replication [6,7,8]. Although DNA methylation per se can prevent binding of transcriptional regulators [9], the main mechanism by which transcriptional repression is achieved appears to involve 5mC binding proteins (MBPs). MBPs specifically recognize methylation marks and consequently stabilize silent chromatin states by recruitment of histone modifying enzymes and chromatin remodeling factors [10].

There are three families of MBPs known to date: the methyl-CpG binding domain (MBD) family, the Uhrf family and the Kaiso protein family. In contrast to the members of the MBD and Kaiso families that specifically recognize fully methylated CpG sites, Uhrf1, the best characterized member of the Uhrf family, preferentially binds hemimethylated DNA, the substrate of maintenance methylation [11,12,13,14]. Notably, crystal structures of the DNA binding domains of MeCP2 and Uhrf1 in complex with DNA revealed striking differences: whereas the MeCP2 MBD recognizes methylated CpG sites based on hydration of the DNA major groove, the Uhrf1 (Set and Ring associated) SRA domain uses a base-flipping mechanism to bind DNA containing hemimethylated CpG sites [11,12,14,15]. Interestingly, Uhrf1 recently emerged as essential cofactor for maintenance methylation potentially by recruiting Dnmt1 to its target sites [13,16,17].

In addition to 5mC, genomic DNA has been recently shown to contain 5-hydroxy-methylcytosine (5hmC), which results from oxidation of 5mC catalyzed by Tet proteins [18,19,20]. This new modification has been implicated in DNA demethylation, either passively as 5hmC containing DNA is not a substrate for Dnmt1 [21], or actively by so far unknown mechanisms. The central questions remain which proteins recognize 5hmC modified DNA and whether 5hmC has a direct role in gene regulation similar to its analog 5mC.

In this study, we characterized the 5mC/5hmC DNA binding properties of two representative 5mC binding protein domains, the MBD of MeCP2 and the SRA domain of Uhrf1. We found that in contrast to the MBD, the SRA domain binds hydroxymethylated DNA substrates with similar affinity as methylated substrates. We investigated the binding mode and energies of Uhrf1 to DNA substrates containing 5mC and 5hmC using molecular dynamics simulations of the respective SRA:DNA complexes.

Results

Uhrf1 binds DNA substrates containing hydroxymethylated CpG sites

Using a newly established DNA binding assay [22,23] as well as electrophoretic mobility shift assays, we investigated the DNA binding activity of Uhrf1, its SRA domain (SRA^{Uhrf1}) and the MBD of MeCP2 (MBD^{MeCP2}) to methylated and hydroxymethylated DNA in direct competition (Figure 1, Supplementary Figure S1; note that all supplementary information can also be found in the Combined Supporting Information File S1). We found that the Uhrf1 constructs bind 5mC and 5hmC containing substrates with similar affinities independent of whether one or both cytosine residues of the palindromic CpG site were modified. Control experiments performed with hemimethylated DNA in competition with either unmethylated substrates or substrates containing no CpG site showed that the observed binding activity to methylated and hydroxymethylated DNA is indeed specific (Supplementary Figure S2). In stark contrast to Uhrf1, we found that MBD^{MeCP2} clearly discriminates between methylation and hydroxymethylation, which is in accordance with previous reports [21,24].

Molecular dynamics simulations of SRA:DNA complexes with 5mC and 5hmC

To investigate the binding mode of the SRA domain to DNA containing 5mC or 5hmC, we performed molecular dynamics simulations for both SRA:DNA complexes. Consistent with the *in vitro* DNA binding data, modeling of an additional hydroxyl group into the complex structure of the Uhrf1 SRA domain with DNA containing hemimethylated CpG sites revealed no spatial constraints for accommodation of the flipped 5hmC nucleotide within the binding pocket (Figure 2). Based on these initial models of the bound conformation, we performed molecular dynamics simulations for a time interval of 57 ns and monitored the RMSD and RMSF values (Supplementary Figures S3 and S4). In both systems equilibrium was reached after 20 to 30 ns. To assure evaluation of equilibrated systems, we continued the equilibrium simulations for another 27 ns and used only the last 10 ns for subsequent interaction energy analysis [25]. To evaluate the stability of the flipped nucleotides within the binding site, we monitored the occurrence and stability of all hydrogen bonds in the vicinity of the binding site with respect to the progress of the simulations (Figure 3).

Before starting the simulations, all water molecules from the X-ray structure were removed and new water molecules were placed by the setup solvation algorithm of NAMD [26]. Therefore, no water molecules were present in the vicinity of the flipped nucleotides at the beginning of the simulations. Interestingly, in both simulations, water molecules from the water-filled simulation box moved into the nucleotide binding site within the first couple of nanoseconds (Figures 3C and 3D, hydrogen bonds 14 to 18). During the remainder of the simulation time, one water molecule was stabilized within the binding site by formation of distinct hydrogen bonds with protein and DNA. Notably, the position of this water molecule in the 5mC complex corresponds to that of a conserved water molecule in the experimental structure (Supplementary Figure S5), confirming the stability and accuracy of our simulations.

Despite the presence of a conserved water molecule in the binding pockets of both complexes, the corresponding hydrogen bond networks showed interesting differences. In the 5mC complex, this water molecule forms hydrogen bonds with the phosphodiester group of the methylated nucleotide as well as with the SRA residues I454 and G453, thereby bridging the DNA

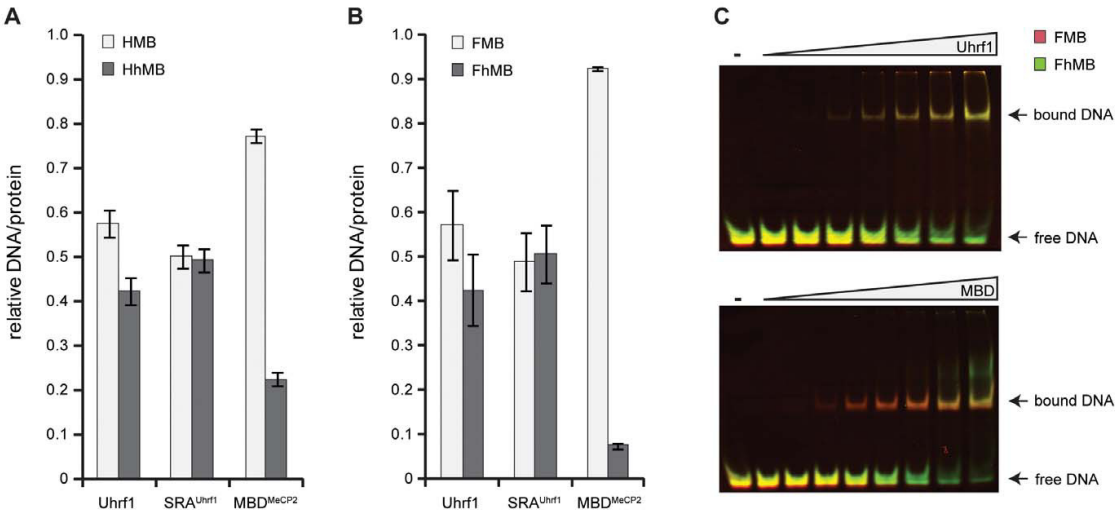


Figure 1. DNA binding specificity of 5-methylcytosine binding proteins. (**A+B**) Relative DNA/protein ratios of Uhrf1, its SRA domain (SRA^{Uhrf1}) and the MBD of MeCP2 (MBD^{MeCP2}) with two differentially labeled DNA substrates in direct competition. (A) Binding to DNA substrates containing a hemimethylated or hemihydroxymethylated CpG site (HMB versus HhMB, respectively). (B) Binding to DNA substrates containing a fully methylated or fully hydroxymethylated CpG site (FMB versus FhMB, respectively). Results are shown as means of three independent experiments with standard deviation error bars. Note that MBD^{MeCP2} preferentially binds to FMB, whereas the Uhrf1 constructs do not discriminate between FMB and FhMB. (**C**) Electrophoretic mobility shift assays were performed with Uhrf1 or MBD^{MeCP2} and equimolar amounts of FMB (red) and FhMB (green) in competition. The overlay of the two substrate channels reveals simultaneous shifting of both DNA substrates with Uhrf1, whereas with MBD^{MeCP2} the FMB substrate shifts at a lower protein concentration than the FhMB substrate, confirming differential binding.
doi:10.1371/journal.pone.0021306.g001

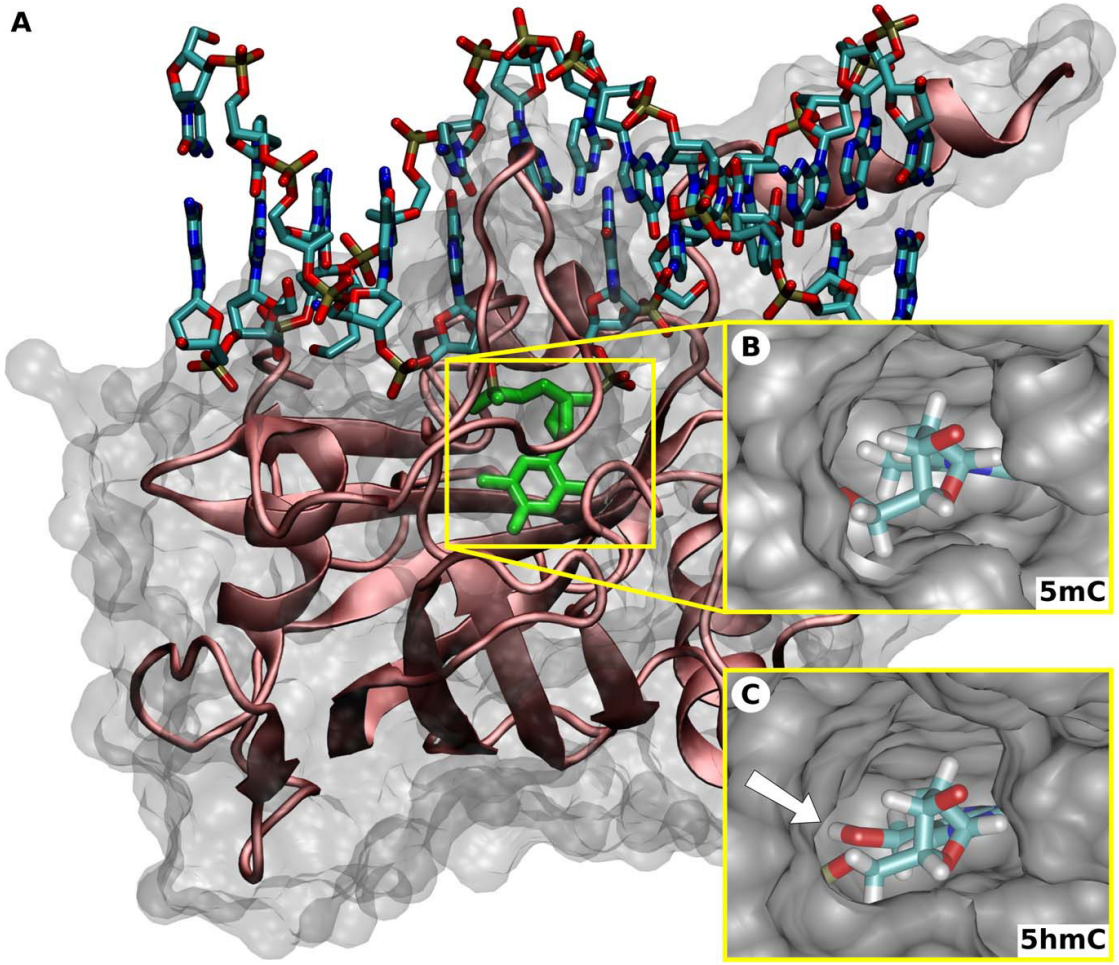


Figure 2. Structure of the Uhrf1 SRA domain in complex with hemimethylated and hemihydroxymethylated DNA. (**A**) Experimental structure of the Uhrf1 SRA domain in complex with hemimethylated DNA (PDB-ID:3fde, [14]). The protein is shown in cartoon and the DNA in licorice representation. The 5mC nucleotide is highlighted in green. Note that the 5mC residue is flipped out of the DNA double helix. (**B+C**) Models of the SRA binding pocket with bound 5mC (B) and 5hmC (C) serving as starting points for the molecular dynamics simulations. The location of the hydroxyl group in the 5hmC complex is highlighted by the white arrow. The view is from the top of the binding site (DNA backbone) and rotated by 90 degrees compared to (A).
doi:10.1371/journal.pone.0021306.g002

backbone:protein interaction (Figure 3A–C, hydrogen bonds 14–16, Figure 4A). Furthermore, direct hydrogen bonds between the 5mC DNA backbone and the protein are formed involving residues G453, S486, and R489 (hydrogen bonds 1–4).

The hydrogen bond network of the 5hmC complex is more stable compared to the 5mC complex (Figure 3D, compare with 3C). Most prominently, one additional and very stable hydrogen bond is formed between the conserved water molecule and the hydroxyl group of the 5hmC nucleotide (hydrogen bond 17). This interaction seems to specifically stabilize the hydrogen bonding network between the DNA backbone and the binding pocket residues G453, S486, and R489 (hydrogen bonds 1–4). Interestingly, these hydrogen bonds have been previously identified to be important for DNA binding [14] and possibly stabilize the flipped conformation of the nucleotide within the binding site. In addition, the hydrogen bond network within the protein involving residues V466 and G453 as well as residues T484 and D474 is stabilized in the 5hmC complex (hydrogen bonds 11–13).

Since water dynamics and to some extent also DNA dynamics can depend on the ion concentration parameters used in the molecular dynamics simulation, we performed a second simulation of the 5hmC complex with a higher ion concentration (Supplementary Figure S6). Consistent to the first simulation with 5hmC, we observed the same overall water dynamics and hydrogen bonding patterns including hydrogen bond formation between the hydroxyl group of the 5hmC nucleotide and the conserved water molecule within the SRA structure. Notably, the stable hydrogen bonding between protein residue S486 and the DNA backbone in the first simulation (hydrogen bonds 2a and 2b) seems to be replaced by a stable hydrogen bond of S486 with the water

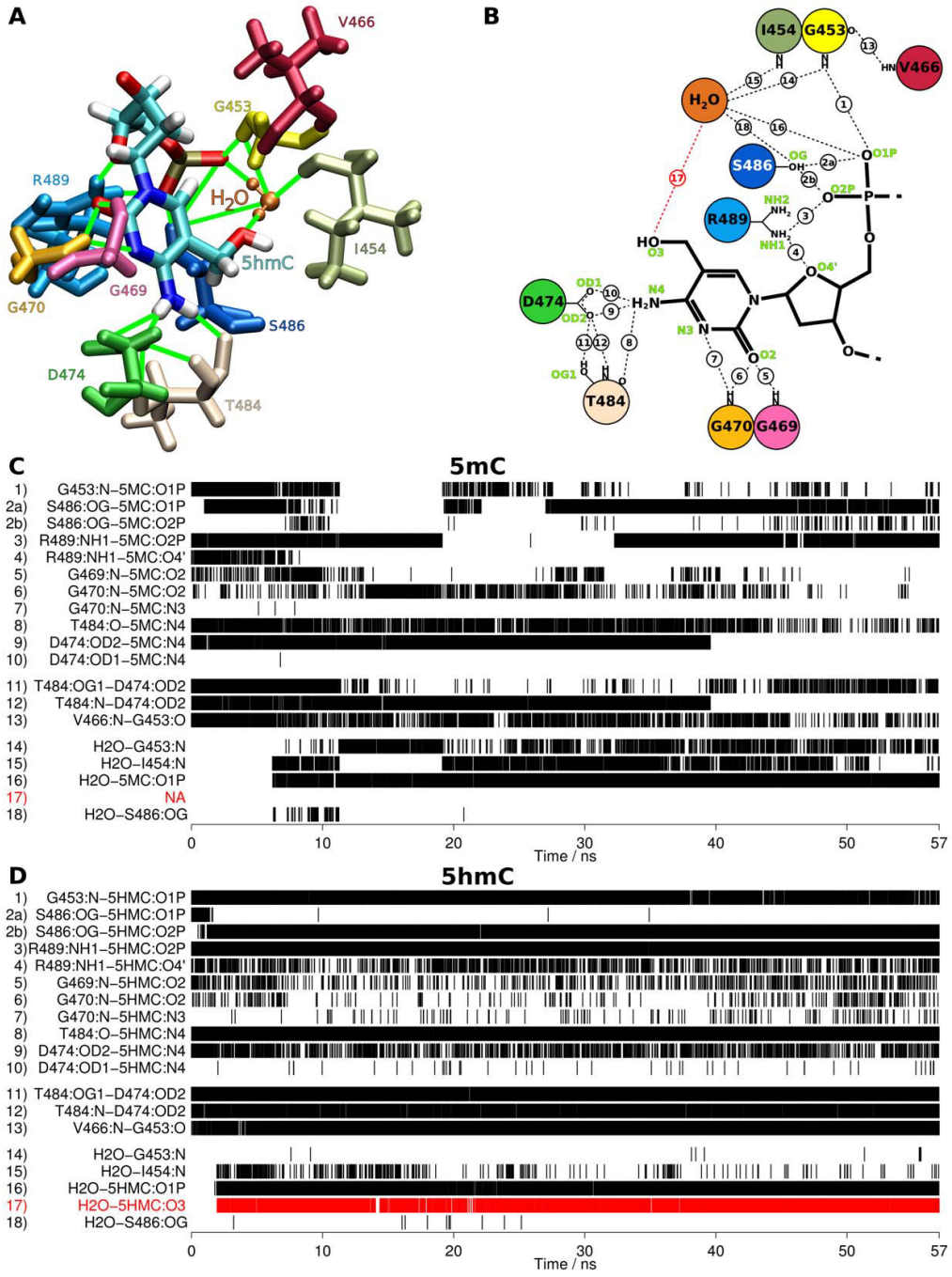


Figure 3. Molecular dynamics simulations of the SRA domain in complex with 5mC and 5hmC containing DNA. (**A+B**) Three and two-dimensional schematic drawings summarizing the hydrogen bond networks between the nucleotides, the SRA binding pocket, and a conserved water molecule during the simulations. The numbers in (B) correspond to the numbering in (C+D). (**C+D**) Hydrogen bond occurrences during the molecular dynamics simulations of the SRA domain in complex with either 5mC (C) or 5hmC containing DNA (D). Each vertical line represents a single observed hydrogen bond. The hydrogen bond between 5hmC and the conserved water is highlighted in red.
doi:10.1371/journal.pone.0021306.g003

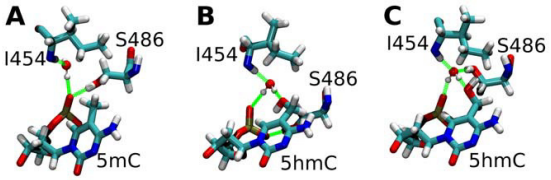


Figure 4. Hydrogen bond networks stabilizing 5mC and 5hmC within the SRA binding pocket. (**A**) SRA complex with DNA containing 5mC. (**B+C**) SRA complex with DNA containing 5hmC. In the 5hmC complex, the water molecule stably interacts with the hydroxyl group of the nucleotide, but two alternative conformations of the SRA binding pocket exist depending on the ion concentration. In the absence of salt, binding involves an interaction of the S486 residue with the phosphate group of the flipped nucleotide (B), whereas in the presence of 0.5 M NaCl, residue S486 interacts with the conserved water molecule (C).
doi:10.1371/journal.pone.0021306.g004

molecule in the second simulation (hydrogen bond 18), indicating two alternative interaction patterns for the S486 residue in the 5hmC complex (Figures 4B and 4C, compare Figure 3D and Supplementary Figure S6B). In conclusion, these data suggest that stable, water bridged hydrogen bond formation of the hydroxyl group of the flipped 5hmC nucleotide with its surrounding occurs in and stabilizes this DNA:SRA complex.

Similar interaction energies for SRA complexes with 5mC and 5hmC containing DNA

To estimate the binding affinity between the Uhrf1 SRA domain and DNA containing either 5mC or 5hmC, we calculated the respective interaction energies using the linear interaction energy (LIE) approach [25]. To exclude energy contributions due to base-flipping when comparing the interaction of the DNA with the protein (bound state) or with the solvent (unbound state), we simulated the DNA in a flipped state in both cases. We determined the difference between the binding energies of the two complexes ($\Delta\Delta G = \Delta G_{5mC} - \Delta G_{5hmC}$). We included either i) the whole DNA and SRA structure ($\Delta\Delta G = -7.94$ kcal/mol) or ii) the flipped nucleotide with its five neighboring nucleotides and the binding pocket of the protein, defined as all residues within a distance of 15 Å from the nucleotide in the starting conformation ($\Delta\Delta G = -6.65$ kcal/mol). These values suggest that the slight difference in binding affinity is predominantly due to interaction of the flipped nucleotide with the proximal protein residues that form the binding site. Considering the estimated uncertainty of about 3–4 kcal/mol in our calculations, these values indicate that both 5mC and 5hmC containing DNA substrates bind with very similar affinity to the SRA domain of Uhrf1.

Discussion

In summary, we observed fundamentally different binding specificities for the DNA binding domains of representative 5mC binding proteins. Hydroxylation of 5mC clearly interferes with DNA binding by the MBD of MeCP2 and might prevent subsequent establishment of repressive chromatin structures in a cellular context, thereby changing the cellular interpretation of an epigenetic modification. Notably, MeCP2 expression is highest in brain tissues where also 5hmC levels are highest [18,27,28]. In stark contrast, Uhrf1, a key factor in maintenance methylation, recognizes 5hmC as well as 5mC. The results of our molecular dynamics simulations provide a structural explanation for recognition of 5hmC. Interestingly, the flipped 5hmC base not only fits into the binding pocket of the Uhrf1 SRA domain, but is specifically stabilized by hydrogen bond formation involving the 5hmC hydroxyl group. This interaction is bridged by a conserved water molecule present within the SRA binding pocket and seems to stabilize the overall hydrogen bond network of the 5hmC complex. Also in the 5mC complex a conserved water molecule is found in the vicinity of the flipped cytosine, which in this case, however, only interacts with the SRA domain and the backbone of the DNA and not with the flipped nucleotide itself.

The specific binding of Uhrf1 to 5hmC containing DNA was clearly unexpected and puts the existing hypothesis on Uhrf1 function into a new perspective. Knock-out studies in mouse embryonic stem cells and embryos revealed that Uhrf1 is essential for maintenance DNA methylation by Dnmt1 [17]. Based on the specific binding of Uhrf1 to hemimethylated CpG sites and its interaction with Dnmt1, Uhrf1 was suggested to operate by recruiting Dnmt1 to its target sites [11,12,13,14,17]. Recent studies suggested a role of hydroxymethylation in passive [21] and/or active [29,30,31] DNA demethylation. The binding of Uhrf1 to hydroxymethylated DNA reported in this study now raises the question how Uhrf1 contributes to change or maintenance of methylation *in vivo*. In this context it should also be noted that the preferential binding of Uhrf1 to hemimethylated DNA is relatively weak, especially if compared to the intrinsic preference of Dnmt1 for methylation of these substrates [22,23]. Moreover, multiple interactions of Uhrf1 with repressive histone tail modifications [23] as well as other heterochromatin associated proteins [32,33] seem to be required for the specific localization and targeting of Uhrf1 *in vivo*. Together, these data strongly argue for a more complex mechanism of Uhrf1 function in living cells and emphasize the need for further studies to understand the pivotal role of Uhrf1 in the establishment, maintenance and change of genome-wide methylation patterns.

Using a combination of *in vitro* and *in silico* studies, we clearly demonstrate that Uhrf1 can bind 5hmC containing DNA. It still remains elusive whether or in which specific context Uhrf1 binds 5hmC modified DNA substrates in living cells. Uhrf1 binding to 5hmC and possible functional consequences *in vivo* are likely to depend on additional interacting factors. Comparison of genome-wide Uhrf1 ChIP profiles with 5mC and 5hmC distribution should help to clarify the interactions and functions of Uhrf1 *in vivo*. Finally, it is interesting to note that Uhrf1 is the only base-flipping protein with so far unknown catalytic function on DNA. The direct interaction of a water molecule with the hydroxyl group of 5hmC within the SRA binding pocket might possibly point towards a role of Uhrf1 in the further modification of this sixth DNA base. In conclusion, our study provides new perspectives on the cellular interpretation and possible further metabolism of this new epigenetic DNA modification.

Materials and Methods

Expression constructs, cell culture and transfection

Mammalian expression constructs for enhanced green fluorescent protein (GFP), Uhrf1 (GFP-Uhrf1), the SRA domain of Uhrf1 (GFP-SRA^{Uhrf1}) and the MBD of MeCP2 (MBD^{MeCP2}-YFP) were described previously [22,23,34]. Note that all constructs encode fusion proteins of either GFP or yellow fluorescent protein (YFP). HEK293T cells [35] were cultured in DMEM supplemented with 50 µg/ml gentamicin and 10% fetal calf serum. For expression of GFP/YFP fusion proteins, HEK293T cells were transfected with the corresponding expression constructs using polyethylenimine (Sigma).

DNA substrate preparation

Fluorescently labeled DNA substrates were prepared by mixing two HPLC-purified DNA oligonucleotides (IBA GmbH, Supplementary Tables S1 and S2) in equimolar amounts, denaturation

for 30 sec at 92°C and slow cool-down to 25°C allowing hybridization. After purification by 15% non-denaturing PAGE, DNA substrates were resuspended in binding buffer (20 mM TrisHCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM DTT).

Pull-down DNA binding assay

In vitro DNA binding assays were performed as described previously [22,23]. In brief, GFP/YFP fusions were purified from HEK293T extracts using the GFP-Trap® (ChromoTek GmbH) and incubated with two differentially labeled DNA substrates at a final concentration of 200 nM DNA/50–100 nM immobilized protein for 45 min at room temperature in binding buffer. After removal of unbound substrate, the amounts of protein and DNA were determined by fluorescence intensity measurements with a Tecan Infinite M1000 plate reader. Binding ratios were calculated dividing the concentration of bound DNA substrate by the concentration of GFP/YFP fusion on the beads, corrected by values from a control experiment using DNA substrates of the same sequence but with different fluorescent labels, and normalized by the total amount of bound DNA.

Electrophoretic mobility shift assay

For competitive electrophoretic mobility shift assays, equimolar amounts of two differentially labeled DNA substrates (250 nM each) were incubated with increasing amounts of GFP/YFP fusion protein (Supplementary Figure S1), subjected to 6% non-denaturing PAGE and analyzed with a Typhoon scanner (GE Healthcare), which allowed separate detection of DNA substrates and protein by ATTO labels and GFP tag, respectively, using the following laser/filter settings: 532 nm/580 nm (ATTO550), 633 nm/none (ATTO700), 488 nm/520 nm (GFP/YFP).

Molecular dynamics simulations

Molecular dynamics simulations were performed based on the X-ray structure of the Uhrf1 SRA domain with the PDB identifier 3FDE [14], using the program NAMD 2.7b1 [26] and the CHARMM22/27 force field [36,37]. Binding free energies were estimated using the Linear Interaction Energy (LIE) model [25].

After energy minimization of 50,000 steps, one hydrogen atom of the methyl group of the protein-bound 5-methylcytosine (5mC) residue was substituted by a hydroxyl group using the tool psfgen. CHARMM22 force field parameters were available for 5mC (patch: PRES 5MC2), but not for 5-hydroxymethylcytosine (5hmC). Therefore, a new 5hmC residue was created based on the 5mC parameters and topology. For this purpose, one hydrogen atom of the 5mC methyl group was exchanged by a hydroxyl group. The charges of the hydroxyl group were subsequently set to charges of the hydroxyl group of a serine residue according to the CHARMM27; the charges of the CH_2 group were adjusted accordingly (Supplementary Table S3). After solvation, the 5mC and 5hmC structures were further energy minimized for 50,000 steps. For each structure, two simulations were performed, in which the charges were either neutralized or a salt concentration of 0.5 M was used.

Each simulation was performed using periodic boundary conditions and particle-mesh-ewald summation [38] for long range non-bonded interactions. The non-bonded cutoff was set to 14 Å with a switching/shifting distance of 12 Å. A stepsize of 1 fs was chosen. The systems were heated from 0 to 200 K for 160 ps under constant volume. Harmonic restraints (1000 kcal mol^{-1} nm^{-2}) were applied to all atoms of the complex. The heat up was continued without harmonic restraints from 200 to 300 K for 80 ps under constant pressure conditions, using a Nose-Hoover barostat [39,40] with a target pressure of 1.01325 bar, an oscillation time scale of 100 fs, and a damping time scale of 50 fs. The temperature was maintained by Langevin dynamics using a damping coefficient of 5/ps. The temperature bath was not coupled to hydrogen atoms. After the heat up procedure, the simulations were continued for 57 ns. During the simulations, all bond lengths were constrained to ideal values using the Shake algorithm [41,42].

For analysis of the simulation results, all hydrogen bonds formed by the flipped nucleotides and the binding site were identified and monitored throughout the simulations and the occurrence of water molecules in and around the binding site was monitored every 5 ps. In order to estimate the difference in the binding free energy of the two nucleotides, we performed three further simulations in which the protein and the two DNA molecules were simulated separately using the conditions described above. To keep the DNA in the flipped state, we additionally applied harmonic restraints to the whole DNA backbone (atom names: C4′, P, O1P, O2P, O5′, C5′, C3′, O3′). The solvated single protein was simulated for 34 ns and the separated DNA molecules were simulated for 20 ns.

To estimate the binding affinity of the two DNA molecules to the protein, we estimated the binding free energy according to the Linear Interaction Energy (LIE) model [25]:

$$\Delta G_{bind} = \alpha \Delta \langle V^{vdw}_{DNA-s} \rangle + \beta \Delta \langle V^{el}_{DNA-s} \rangle + \gamma \quad (1)$$

$$\Delta \langle V^{el/vdw}_{DNA-s} \rangle = \langle V^{el/vdw}_{bound} \rangle - \langle V^{el/vdw}_{unbound} \rangle \quad (2)$$

In this approach the binding free energy is approximated by the difference between the interaction energies ΔV^{el} and ΔV^{vdw} of the ligand in the protein-ligand complex (bound state) and in solution (unbound state). The <> denotes the average values obtained from the simulation trajectories. According to the linear response approximation the weights α and β were set to 1 and 0.5, respectively. We calculated the DNA-(protein+solvent) (bound state) and the DNA-solvent (free state) interaction energies from the trajectories of the DNA/SRA and the DNA/solvent simulations, using the average energy over the last 10 ns.

Supporting Information

Figure S1 Electrophoretic mobility shift assays with methylated and hydroxymethylated DNA substrates. Increasing amounts of Uhrf1, its SRA domain (SRA^{Uhrf1}) or the MBD domain of MeCP2 (MBD^{MeCP2}) were incubated with two differentially ATTO-labeled DNA substrates, which contain either one central fully methylated or fully hydroxymethylated CpG site (FMB-ATTO700 or FhMB-ATTO550, respectively), in direct competition. Samples were subjected to 6% non-denaturing PAGE and analyzed with a Typhoon scanner (GE Healthcare). The first, second and third columns show the scans for GFP/YFP, ATTO700 and ATTO550 fluorescence, respectively. The overlay of the two ATTO channels is shown in the fourth column (FMB: red, FhMB:green).
(PDF)

Figure S2 DNA binding specificity of Uhrf1. Relative DNA/Uhrf1 ratios are shown for two differentially labeled fluorescent DNA substrates in direct competition. (A) Binding of Uhrf1 to DNA substrates containing no CpG site or one central hemimethylated CpG site (noCGB versus HMB, respectively). (B) Binding of Uhrf1 to DNA substrates containing one central un- or hemimethylated CpG site (UMB versus HMB, respectively). Results are shown as means of three independent experiments with standard deviation error bars. DNA substrates were prepared

by hybridization as described in the main text, except for noCGB, which was prepared by primer extension as described previously [22]. See Supplementary Tables S1 and S2 for DNA oligonucleotide sequences and purification grade of the used substrates.
(PDF)

Figure S3 Atom-positional root-mean-square deviation of the protein and DNA backbone atoms during the simulations. The terminal DNA and protein residues were excluded from the calculations in the "subset" sets (red and black lines).
(PDF)

Figure S4 Atom-positional root-mean-square fluctuations of the protein (A, C) and both DNA strands (B, D) during two simulation periods. Note that both structures show the same flexibility pattern during both simulation periods and are overall stable during both periods. This is in agreement with the RMSD data in Figure S3, which shows that equilibration is reached after 30 ns of simulation time.
(PDF)

Figure S5 Superposition of the equilibrated 5mC structure after simulation (atom-name specific coloring) and the crystal structure (PDB-ID:3fde [14], green). The 5mC nucleotide, the residue I454 of the SRA binding pocket and the conserved water molecule are shown. Note that the distance between the oxygen atoms of the conserved water molecules in the two structures is only 1.1 Å.
(PDF)

Figure S6 Molecular dynamics simulations of the Uhrf1 SRA domain in complex with 5mC (A) and 5hmC (B) containing DNA in 0.5 M NaCl. Hydrogen bond occurrences during the simulation of the SRA:DNA complex using a concentration of 0.5 M NaCl.
(PDF)

Table S1 Sequences of DNA oligonucleotides used for preparation of double stranded fluorescent DNA substrates. M: 5-methylcytosine. X: 5-hydroxymethylcytosine.
(PDF)

Table S2 DNA substrates used for the DNA binding assays.
(PDF)

Table S3 Residue Topology File and parameters used for the 5hmC residue during the simulations.
(PDF)

File S1 Combined supporting figures and tables.
(PDF)

Author Contributions

Conceived and designed the experiments: CF HL IA MCC TH. Performed the experiments: CF TH. Analyzed the data: CF HL IA TH. Wrote the paper: CF HL IA. Performed initial experiments: SB VC.

References

1. Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev 16: 6–21.
2. Rottach A, Leonhardt H, Spada F (2009) DNA methylation-mediated epigenetic control. J Cell Biochem 108: 43–51.
3. Goll MG, Bestor TH (2005) Eukaryotic cytosine methyltransferases. Annu Rev Biochem 74: 481–514.
4. Lei H, Oh S, Okano M, Juttermann R, Goss K, et al. (1996) De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development 122: 3195–3205.
5. Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99: 247–257.
6. Leonhardt H, Page AW, Weier HU, Bestor TH (1992) A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell 71: 865–873.
7. Chuang LS-H, Ian H-I, Koh T-W, Ng H-H, Xu G, et al. (1997) Human DNA-(Cytosine-5) Methyltransferase-PCNA Complex as a Target for p21WAF1. Science 277: 1996–2000.
8. Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69: 915–926.
9. Becker PB, Ruppert S, Schutz G (1987) Genomic footprinting reveals cell type-specific DNA binding of ubiquitous factors. Cell 51: 435–443.
10. Sasai N, Defossez PA (2009) Many paths to one goal? The proteins that recognize methylated DNA in eukaryotes. Int J Dev Biol 53: 323–334.
11. Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M (2008) Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature 455: 818–821.
12. Avvakumov GV, Walker JR, Xue S, Li Y, Duan S, et al. (2008) Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature 455: 822–825.
13. Bostick M, Kim JK, Esteve P-O, Clark A, Pradhan S, et al. (2007) UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. Science 317: 1760–1764.
14. Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, et al. (2008) The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature 455: 826–829.
15. Ho KL, McNae IW, Schmiedeberg L, Klose RJ, Bird AP, et al. (2008) MeCP2 binding to DNA depends upon hydration at methyl-CpG. Mol Cell 29: 525–531.
16. Achour M, Jacq X, Ronde P, Alhosin M, Charlot C, et al. (2008) The interaction of the SRA domain of ICBP90 with a novel domain of DNMT1 is involved in the regulation of VEGF gene expression. Oncogene 27: 2187–2197.
17. Sharif J, Muto M, Takebayashi SI, Suetake I, Iwamatsu A, et al. (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature.
18. Kriaucionis S, Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 324: 929–930.
19. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, et al. (2009) Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. Science 324: 930–935.
20. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, et al. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature 466: 1129–1133.
21. Valinluck V, Sowers LC (2007) Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. Cancer Res 67: 946–950.
22. Frauer C, Leonhardt H (2009) A versatile non-radioactive assay for DNA methyltransferase activity and DNA binding. Nucleic Acids Res 37: e22.
23. Rottach A, Frauer C, Pichler G, Bonapace IM, Spada F, et al. (2010) The multi-domain protein Np95 connects DNA methylation and histone modification. Nucleic Acids Res 38: 1796–1804.
24. Valinluck V, Tsai HH, Rogstad DK, Burdzy A, Bird A, et al. (2004) Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). Nucleic Acids Res 32: 4100–4108.
25. Aqvist J, Luzhkov VB, Brandsdal BO (2002) Ligand binding affinities from MD simulations. Acc Chem Res 35: 358–365.
26. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26: 1781–1802.
27. Globisch D, Munzel M, Muller M, Michalakis S, Wagner M, et al. (2010) Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS One 5: e15367.
28. Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H (2010) Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res 38: e181.
29. Guo JU, Su Y, Zhong C, Ming GL, Song H (2011) Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain. Cell 145: 423–434.
30. Iqbal K, Jin SG, Pfeifer GP, Szabo PE (2011) Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proc Natl Acad Sci U S A 108: 3642–3647.
31. Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, Zakhartchenko V, et al. (2011) 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat Commun 2: 241.
32. Kim JK, Esteve PO, Jacobsen SE, Pradhan S (2009) UHRF1 binds G9a and participates in p21 transcriptional regulation in mammalian cells. Nucleic Acids Res 37: 493–505.
33. Meilinger D, Fellinger K, Bultmann S, Rothbauer U, Bonapace IM, et al. (2009) Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b,

and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. EMBO Rep 10: 1259–1264.
34. Brero A, Easwaran HP, Nowak D, Grunewald I, Cremer T, et al. (2005) Methyl CpG-binding proteins induce large-scale chromatin reorganization during terminal differentiation. J Cell Biol 169: 733–743.
35. DuBridge RB, Tang P, Hsia HC, Leong PM, Miller JH, et al. (1987) Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. Mol Cell Biol 7: 379–387.
36. MacKerell AD, Jr., Banavali N, Foloppe N (2000) Development and current status of the CHARMM force field for nucleic acids. Biopolymers 56: 257–265.
37. MacKerell AD, Bashford D, Bellott, Dunbrack RL, Evanseck JD, et al. (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins†. The Journal of Physical Chemistry B 102: 3586–3616.
38. Darden T, York D, Pedersen L (1993) PARTICLE MESH EWALD - AN N.LOG(N) METHOD FOR EWALD SUMS IN LARGE SYSTEMS. Journal of Chemical Physics 98: 10089–10092.
39. Martyna GJ, Tobias DJ, Klein ML (1994) CONSTANT-PRESSURE MOLECULAR-DYNAMICS ALGORITHMS. Journal of Chemical Physics 101: 4177–4189.
40. Feller SE, Zhang YH, Pastor RW, Brooks BR (1995) CONSTANT-PRESSURE MOLECULAR-DYNAMICS SIMULATION - THE LANGEVIN PISTON METHOD. Journal of Chemical Physics 103: 4613–4621.
41. Miyamoto S, Kollman PA (1992) SETTLE - AN ANALYTICAL VERSION OF THE SHAKE AND RATTLE ALGORITHM FOR RIGID WATER MODELS. Journal of Computational Chemistry 13: 952–962.
42. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) NUMERICAL-INTEGRATION OF CARTESIAN EQUATIONS OF MOTION OF A SYSTEM WITH CONSTRAINTS - MOLECULAR-DYNAMICS OF N-ALKANES. Journal of Computational Physics 23: 327–341.

A.3 Supporting Information

Supplementary Material

Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain

Carina Frauer, Thomas Hoffmann, Sebastian Bultmann, Valentina Casa, M. Cristina Cardoso, Iris Antes and Heinrich Leonhardt

Supplementary Tables

Supplementary Table S1. Sequences of DNA oligonucleotides used for preparation of double stranded fluorescent DNA substrates.
M: 5-methylcytosine. X: 5-hydroxymethylcytosine.

Name	Sequence
CGup	5′- CTCAACAACTAACTACCATCCGGACCAGAAGAGTCATCATGG -3′
MGup	5′- CTCAACAACTAACTACCATCMGGACCAGAAGAGTCATCATGG -3′
hmCGup	5'- CTCAACAACTAACTACCATCXGGACCAGAAGAGTCATCATGG -3'
noCGup	5′- CTCAACAACTAACTACCATCTGGACCAGAAGAGTCATCATGG -3′
um550	5′- ATTO550-CCATGATGACTCTTCTGGTCCGGATGGTAGTTAGTTGTTGAG -3′
um590	5′- ATTO590-CCATGATGACTCTTCTGGTCCGGATGGTAGTTAGTTGTTGAG -3′
um647N	5′- ATTO647N-CCATGATGACTCTTCTGGTCCGGATGGTAGTTAGTTGTTGAG -3′
um700	5′- ATTO700-CCATGATGACTCTTCTGGTCCGGATGGTAGTTAGTTGTTGAG -3′
mC700	5'- ATTO700-CCATGATGACTCTTCTGGTCMGGATGGTAGTTAGTTGTTGAG -3'
hmC550	5'- ATTO550-CCATGATGACTCTTCTGGTCXGGATGGTAGTTAGTTGTTGAG -3'
550-Fill-In	5′- ATTO550-CCATGATGACTCTTCTGGTC -3′

Supplementary Table S2. DNA substrates used for the DNA binding assays.

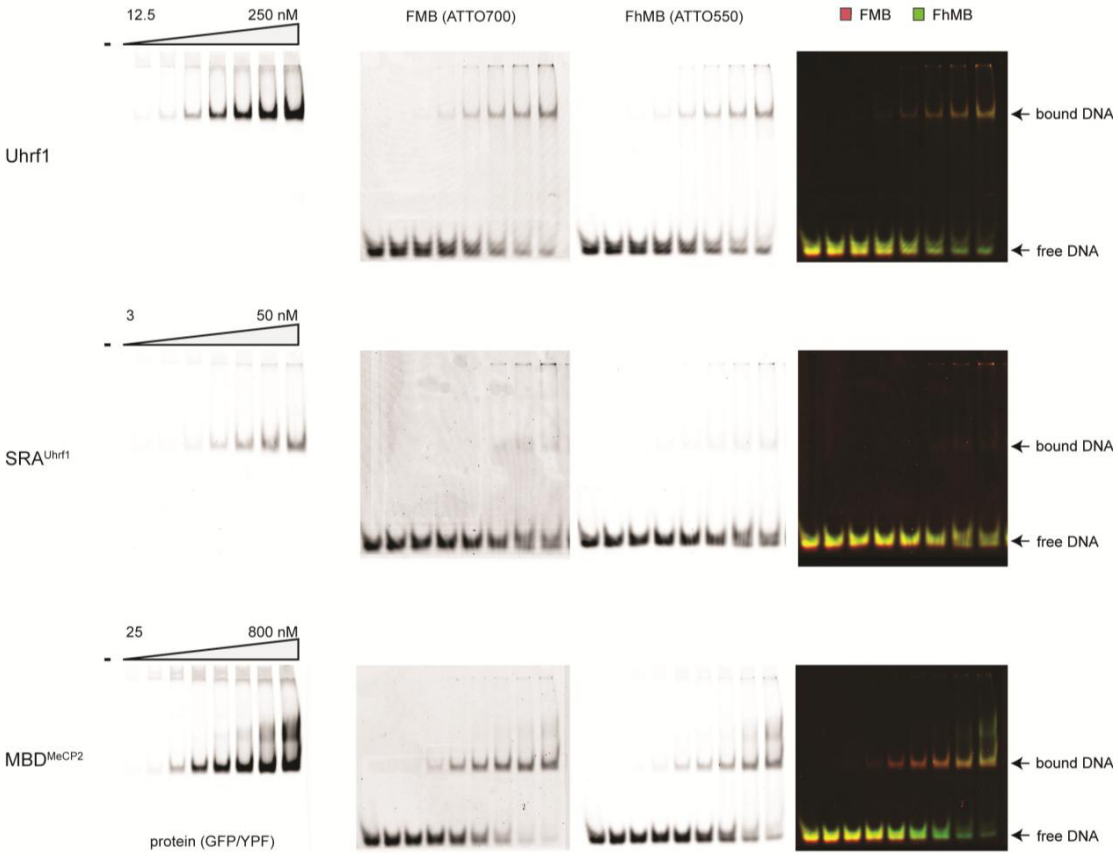
Name	CpG site	Label	Oligo I	Oligo II	Purification grade and use
HMB550	hemimethylated	ATTO550	MGup	um550	· hybridization of HPLC-purified oligos · gel-purification · used for data in figure 2 and supplementary figure 1
HMB700	hemimethylated	ATTO700	MGup	um700	
HhMB700	hemihydroxymethylated	ATTO700	hmCGup	um700	
FMB700	fully methylated	ATTO700	MGup	mC700	
FhMB550	fully hydroxymethylated	ATTO550	hmCGup	hmC550	
noCG550	no CpG site	ATTO550	noCGup	550-Fill-In	· primer extension for noCG550 · hybridization of HPLC-purified oligos for HMB substrates · gel-purification · used for data in supplementary figure 2A
HMB550	hemimethylated	ATTO550	MGup	um550	
HMB647N	hemimethylated	ATTO647N	MGup	um647N	
UMB550	unmethylated	ATTO550	CGup	um550	· hybridization of HPLC-purified oligos · used for data in supplementary figure 2B, n=2
UMB590	unmethylated	ATTO590	CGup	um590	
HMB590	hemimethylated	ATTO590	MGup	um590	
UMB647N	unmethylated	ATTO647N	CGup	um647N	· hybridization of PAGE-purified oligos · used for data in supplementary figure 2B, n=1
UMB700	unmethylated	ATTO700	CGup	um700	
HMB700	hemimethylated	ATTO700	MGup	um700	

Supplementary Table S3. Residue Topology File and parameters used for the 5hmC residue during the simulations.

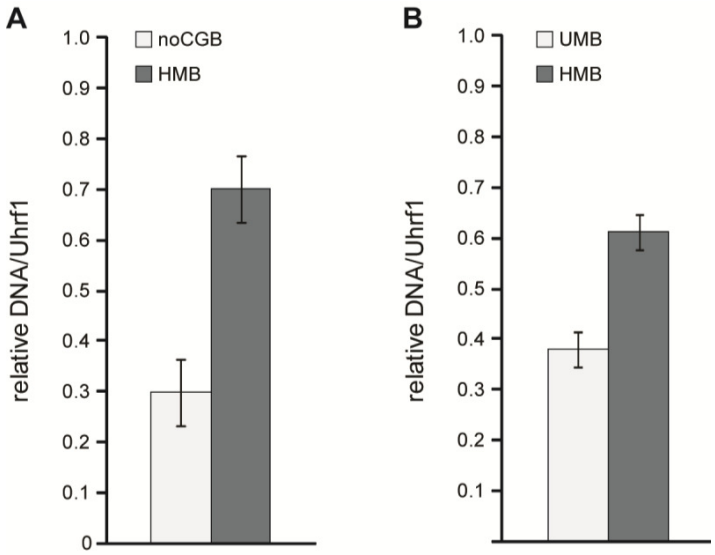
```
================================================================================
TOPOLOGY (based on 5mC topology from patches: PRES 5MC2 and PRES DEO1)
================================================================================
! 5-hydroxy-methyl cytosine
RESI 5HMC        -1.00  !
ATOM P    P       1.50  !
ATOM O1P  ON3    -0.78  !                      H42  H41
ATOM O2P  ON3    -0.78  !                        \  /
ATOM O5'  ON2    -0.57  !                         N4
ATOM C5'  CN8B   -0.08  !                         |
ATOM H5'  HN8     0.09  !              H3-O3      C4
ATOM H5'' HN8     0.09  !                 \     /  \\
GROUP                   !              H5M2 -C5M-C5   N3
ATOM C4'  CN7     0.16  !                   /    ||   |
ATOM H4'  HN7     0.09  !               H5M1  H6-C6   C2
ATOM O4'  ON6    -0.50  !                        \  / \\
ATOM C1'  CN7B    0.16  !                         N1   O2
ATOM H1'  HN7     0.09  !                          \
GROUP                   !                          \
ATOM N1   NN2    -0.13  !                           \
ATOM C6   CN3     0.05  !        O1P    H5' H4'  O4'  \
ATOM H6   HN3     0.17  !         |      |    \ /   \  \
ATOM C5   CN3D   -0.11  !        -P-O5'-C5'---C4'    C1'
ATOM C5M  CN9     0.10  !         |      |     \     / \
ATOM H5M1 HN9     0.09  !        O2P    H5''   C3'--C2' H1'
ATOM H5M2 HN9     0.09  !                     / \   / \
ATOM O3   OH1    -0.66  !                 O3' H3' H2' H2''
ATOM H3   H       0.43  !                  |
ATOM C2   CN1     0.52  !                  |
ATOM O2   ON1C   -0.49  !
ATOM N3   NN3    -0.66  !
ATOM C4   CN2     0.65  !
ATOM N4   NN1    -0.75  !
ATOM H41  HN1     0.37
ATOM H42  HN1     0.33
GROUP
ATOM C2'  CN8    -0.18  !
ATOM H2'' HN8     0.09  !
ATOM H2'  HN8     0.09  !
GROUP
ATOM C3'  CN7     0.01
ATOM H3'  HN7     0.09
ATOM O3'  ON2    -0.57
BOND P    O1P       P    O2P       P     O5'
BOND O5'  C5'       C5'  C4'       C4'  O4'      C4'  C3'    O4'  C1'
BOND C1'  N1        C1'  C2'       N1   C2       N1   C6
BOND C2   N3        C4   N4        N4   H41      N4   H42
BOND C4   C5        C2'  C3'       C3'  O3'      O3'  +P
BOND C1'  H1'       C2'  H2''      C2'  H2'      C3'  H3'    C4'  H4'    C5'  H5'
BOND C5'  H5''      C6   H6
BOND C5   C5M  C5M  H5M1  C5M H5M2  C5M  O3  O3 H3
ANGL C4   C5   C5M   C6   C5   C5M
ANGL C5   C5M  H5M1  C5   C5M  H5M2  C5   C5M  O3  C5M O3 H3
ANGL H5M1 C5M  H5M2  H5M1 C5M  O3    H5M2 C5M  O3
DIHE C5M  C5   C4   N3   C5M  C5   C4   N4
DIHE C5M  C5   C6   H6   C5M  C5   C6   N1
DIHE H5M1 C5M  C5   C4   H5M1 C5M  C5   C6
DIHE H5M2 C5M  C5   C4   H5M2 C5M  C5   C6
DIHE O3   C5M  C5   C4   O3   C5M  C5   C6
DIHE H3   O3   C5M  C5 H3   O3   C5M  H5M2
DIHE H3   O3   C5M  H5M1
DOUBLE   C2   O2    C5   C6        N3   C4
IMPR C2   N1   N3   O2        C4   N3   C5   N4
IMPR N4   C4   H41  H42
DONO H42  N4
DONO H41  N4
DONO H3   O3
ACCE O2   C2
ACCE N3
ACCE O1P  P
ACCE O2P  P
```

```
ACCE O3'
ACCE O4'
ACCE O5'
ACCE O3

BILD -O3' P    O5'  C5'    1.6001  101.45  -46.90  119.00   1.4401 !alpha
BILD -O3' O5'  *P   O1P    1.6001  101.45 -115.82  109.74   1.4802
BILD -O3' O5'  *P   O2P    1.6001  101.45  115.90  109.80   1.4801
BILD  P   O5'  C5'  C4'    1.5996  119.00 -146.00  110.04   1.5160 !beta
BILD O5'  C5'  C4'  C3'    1.4401  108.83   60.00  116.10   1.5284 !gamma
BILD C5'  C4'  C3'  O3'    1.5160  116.10  140.00  115.12   1.4212 !delta
BILD C4'  C3'  O3'  +P     1.5284  111.92  155.00  119.05   1.6001 !epsilon
BILD C3'  O3'  +P   +O5'   1.4212  119.05  -95.20  101.45   1.5996 !zeta
BILD O4'  C3'  *C4' C5'    1.4572  104.06 -120.04  116.10   1.5160
BILD C2'  C4'  *C3' O3'    1.5284  100.16 -124.08  115.12   1.4212
BILD C4'  C3'  C2'  C1'    1.5284  100.16  -30.00  102.04   1.5251
BILD C3'  C2'  C1'  N1     1.5284  101.97  147.89  113.71   1.4896
BILD O4'  C1'  N1   C2     1.5251  113.71  -97.2   125.59   1.3783 !chi
BILD C1'  C2   *N1  C6     1.4896  117.79 -180.00  120.6    1.364
BILD C2   N1   C6   C5     1.399   120.6     0.0   121.0    1.337
BILD C6   N1   C2   N3     1.364   120.6     0.0   118.9    1.356
BILD N1   N3   *C2  O2     1.399   118.9   180.0   121.9    1.237
BILD N1   C2   N3   C4     1.399   118.9     0.0   120.0    1.334
BILD C5   N3   *C4  N4     1.426   121.8   180.00  118.9    1.337
BILD N3   C4   N4   H41    1.337   117.9     0.00  118.9    1.01
BILD H41  C4   *N4  H42    1.01    118.9   180.00  120.7    1.01
BILD N1   C5   *C6  H6     0.0       0.0   180.0     0.0    0.0
BILD C1'  C3'  *C2' H2'    1.5284  102.04 -114.67  110.81   1.01
BILD O4'  C2'  *C1' H1'    0.0       0.0  -115.0     0.0    0.0
BILD C1'  C3'  *C2' H2''   0.0       0.0   115.0     0.0    0.0
BILD C1'  C3'  *C2' H2'    0.0       0.0  -115.0     0.0    0.0
BILD C2'  C4'  *C3' H3'    0.0       0.0   115.0     0.0    0.0
BILD C3'  O4'  *C4' H4'    0.0       0.0  -115.0     0.0    0.0
BILD C4'  O5'  *C5' H5'    0.0       0.0  -115.0     0.0    0.0
BILD C4'  O5'  *C5' H5''   0.0       0.0   115.0     0.0    0.0
BILD C6   C4   *C5  C5M    0.0       0.0   180.0     0.0    0.0
BILD C4   C5   C5M  H5M1   0.0       0.0   180.0     0.0    0.0
BILD C5   H5M1 *C5M H5M2   0.0       0.0  -115.0     0.0    0.0
BILD H5M1 H5M2 *C5M O3     0.0       0.0   115.0     0.0    0.0
BILD C4   C5   C5M  O3     0.0       0.0    60.0     0.0    0.0
BILD C5   C5M  O3   H3     0.0       0.0   180.0     0.0    0.0
================================================================================

================================================================================
FORCEFIELD PARAMETERS:
================================================================================
...
BONDS
...
!added for 5HMC TU_TCB TH und ISA
OH1  CN9   428.000     1.4200 ! ACC. TO OH1-CT3
...
ANGLES
...
!added for 5HMC TU_TCB TH und ISA
OH1  CN9  CN3D   75.700   110.1000 !ACC. TO  OH1-CT2-CT2
OH1  CN9  HN9    45.900   108.8900 !ACC. TO  OH1-CT3-HA
H    OH1  CN9    57.500   106.0000 !ACC. TO  H-OH1-CT2
...
DIHEDRALS
...
!added for 5HMC TU_TCB TH und ISA
H    OH1  CN9  CN3D     1.3000  1     0.00 !ACC. TO  H-OH1-CT2-CT2
H    OH1  CN9  CN3D     0.3000  2     0.00 !ACC. TO  H-OH1-CT2-CT2
H    OH1  CN9  CN3D     0.4200  3     0.00 !ACC. TO  H-OH1-CT2-CT2
CN3  CN3D CN9  OH1      0.0     3     0.0  !ACC. TO  CN3-CN3D-CN9-HN9
CN2  CN3D CN9  OH1      0.35    3     0.0  !ACC. TO  CN3-CN3D-CN9-HN9
HN9  CN9  OH1  H        0.1400  3     0.00 !ACC. TO  X-CT2-OH1-X
...
IMPROPER
...
```

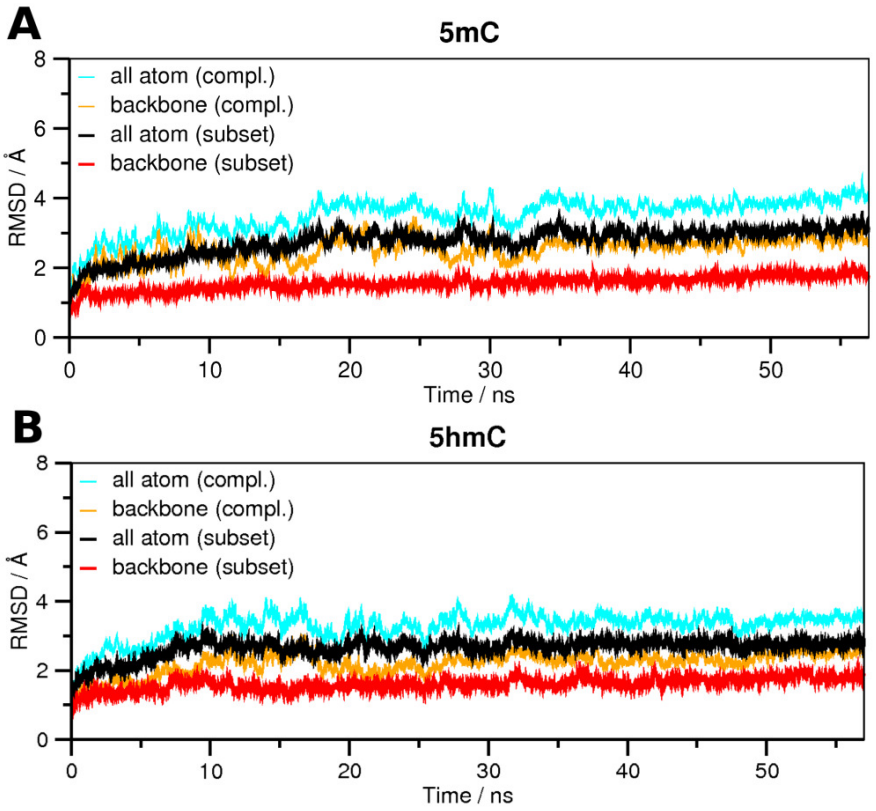
Supplementary Figures



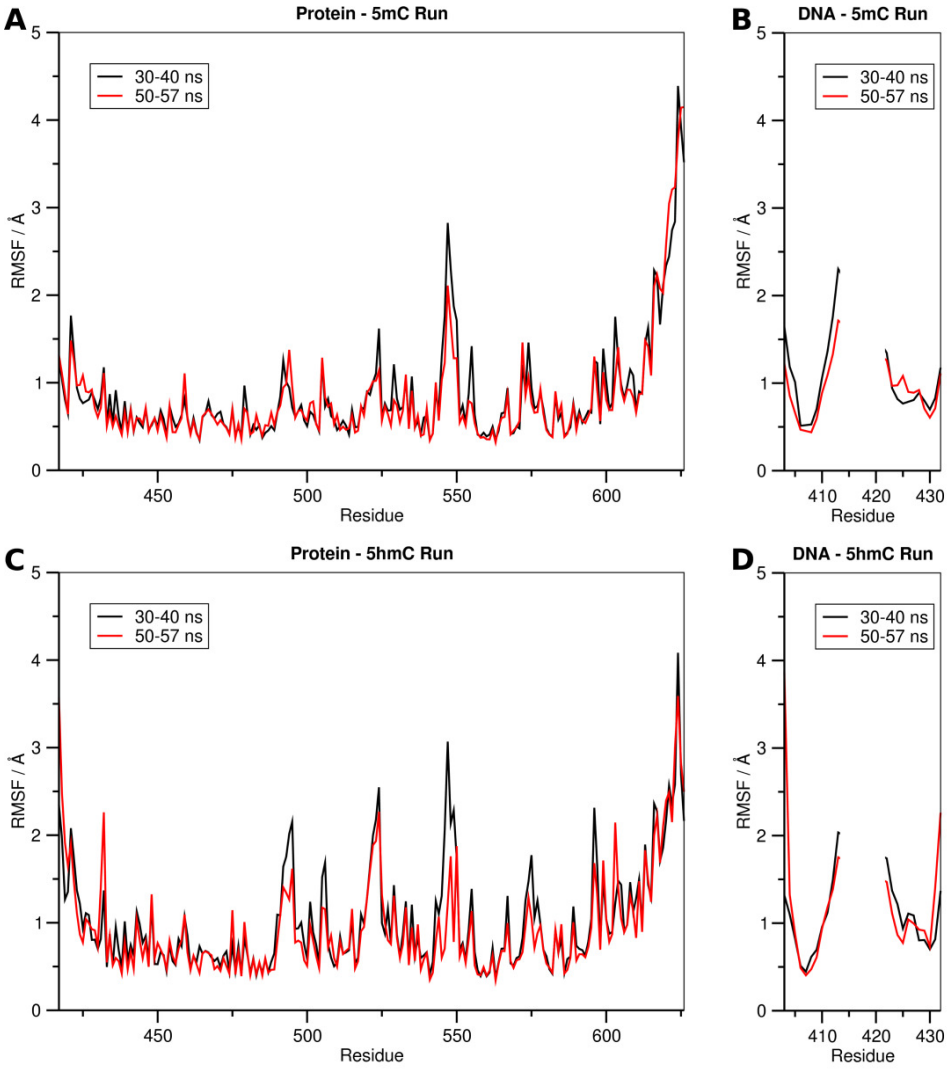
Supplementary Figure S1. Electrophoretic mobility shift assays with methylated and hydroxymethylated DNA substrates. Increasing amounts of Uhrf1, its SRA domain (SRA^{Uhrf1}) or the MBD domain of MeCP2 (MBD^{MeCP2}) were incubated with two differentially ATTO-labeled DNA substrates, which contain either one central fully methylated or fully hydroxymethylated CpG site (FMB-ATTO700 or FhMB-ATTO550, respectively), in direct competition. Samples were subjected to 6 % non-denaturing PAGE and analyzed with a Typhoon scanner (GE Healthcare). The first, second and third columns show the scans for GFP/YFP, ATTO700 and ATTO550 fluorescence, respectively. The overlay of the two ATTO channels is shown in the fourth column (FMB: red, FhMB:green).



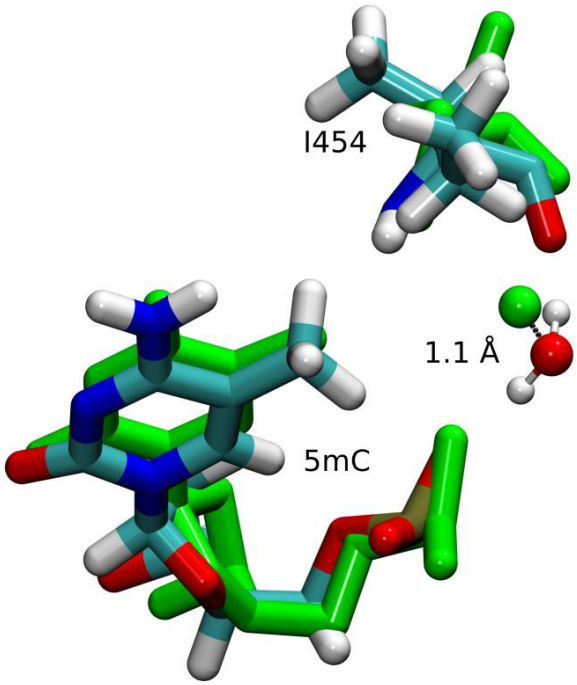
Supplementary Figure S2. DNA binding specificity of Uhrf1. Relative DNA/Uhrf1 ratios are shown for two differentially labeled fluorescent DNA substrates in direct competition. (A) Binding of Uhrf1 to DNA substrates containing no CpG site or one central hemimethylated CpG site (noCGB versus HMB, respectively). (B) Binding of Uhrf1 to DNA substrates containing one central un- or hemimethylated CpG site (UMB versus HMB, respectively). Results are shown as means of three independent experiments with standard deviation error bars. DNA substrates were prepared by hybridization as described in the main text, except for noCGB, which was prepared by primer extension as described previously [1]. See Supplementary Tables 1 and 2 for DNA oligonucleotide sequences and purification grade of the used substrates.



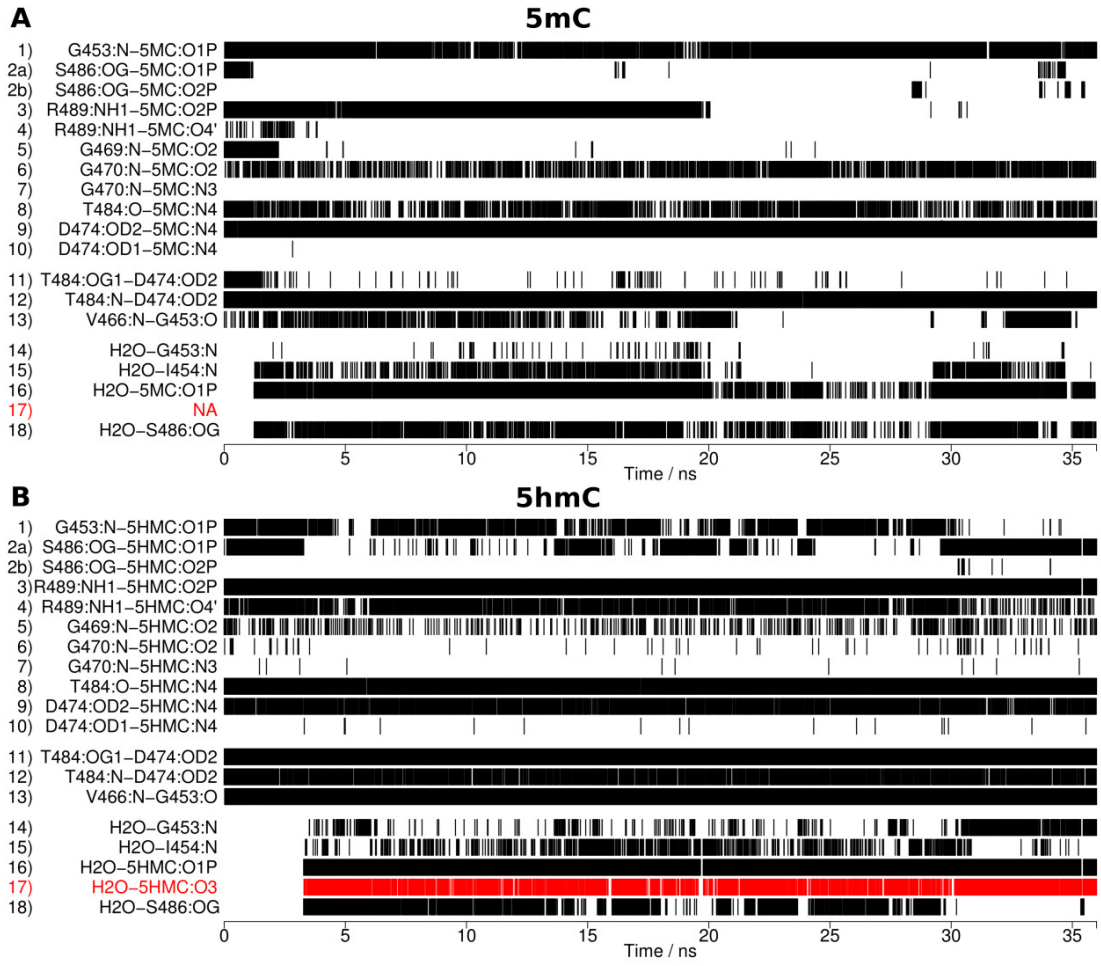
Supplementary Figure S3. Atom-positional root-mean-square deviation of the protein and DNA backbone atoms during the simulations. The terminal DNA and protein residues were excluded from the calculations in the "subset" sets (red and black lines).



Supplementary Figure S4. Atom-positional root-mean-square fluctuations of the protein (A, C) and both DNA strands (B, D) during two simulation periods. Note that both structures show the same flexibility pattern during both simulation periods and are overall stable during both periods. This is in agreement with the RMSD data in Figure S3, which shows that equilibration is reached after 30 ns of simulation time.



Supplementary Figure S5. Superposition of the equilibrated 5mC structure after simulation (atom-name specific coloring) and the crystal structure (PDB-ID:3fde [2], green). The 5mC nucleotide, the residue I454 of the SRA binding pocket and the conserved water molecule are shown. Note that the distance between the oxygen atoms of the conserved water molecules in the two structures is only 1.1 Å.



Supplementary Figure S6. Molecular dynamics simulations of the Uhrf1 SRA domain in complex with 5mC (A) and 5hmC (B) containing DNA in 0.5 M NaCl. Hydrogen bond occurrences during the simulation of the SRA:DNA complex using a concentration of 0.5 M NaCl.

Supplementary References

1. Frauer C, Leonhardt H (2009) A versatile non-radioactive assay for DNA methyltransferase activity and DNA binding. Nucleic Acids Res 37: e22.
2. Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, et al. (2008) The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature 455: 826-829.

Appendix B

Embedded Publication 2

B.1 Copy Permissions

B.2 Full Article

PLOS | COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Quantitative Analysis of the Association Angle between T-cell Receptor Vα/Vβ Domains Reveals Important Features for Epitope Recognition

Thomas Hoffmann[1], Angela M. Krackhardt[2,3,4], Iris Antes[1]*

1 Department of Biosciences and Center for Integrated Protein Science Munich,Technische Universität München, Freising-Weihenstephan, Germany, **2** Medizinische Klinik III, Innere Medizin mit Schwerpunkt Hämatologie und Onkologie, Technische Universität München, Munich, Germany, **3** Clinical Cooperation Group, Antigen specific T cell therapy, Helmholtz Zentrum München (GmbH), German Center for Environmental Health, Munich, Germany, **4** German Cancer Consortium (DKTK), Munich, Germany

* antes@tum.de

OPEN ACCESS

Citation: Hoffmann T, Krackhardt AM, Antes I (2015) Quantitative Analysis of the Association Angle between T-cell Receptor Vα/Vβ Domains Reveals Important Features for Epitope Recognition. PLoS Comput Biol 11(7): e1004244. doi:10.1371/journal.pcbi.1004244

Editor: Bjoern Peters, La Jolla Institute for Allergy and Immunology, UNITED STATES

Received: November 6, 2014

Accepted: March 17, 2015

Published: July 17, 2015

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: TH and IA were funded by: CIPSM cluster of excellence; http://www.dfg.de/ SFB1035; http://www.dfg.de/. AMK was funded by: DJCLS R11/23; https://www.carreras-stiftung.de/ F2-F5121.7.3-10c/23932; http://www.bayimmunet.de/ 110281; http://www.krebshilfe.de. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

T-cell receptors (TCR) play an important role in the adaptive immune system as they recognize pathogen- or cancer-based epitopes and thus initiate the cell-mediated immune response. Therefore there exists a growing interest in the optimization of TCRs for medical purposes like adoptive T-cell therapy. However, the molecular mechanisms behind T-cell signaling are still predominantly unknown. For small sets of TCRs it was observed that the angle between their Vα- and Vβ-domains, which bind the epitope, can vary and might be important for epitope recognition. Here we present a comprehensive, quantitative study of the variation in the Vα/Vβ interdomain-angle and its influence on epitope recognition, performing a systematic bioinformatics analysis based on a representative set of experimental TCR structures. For this purpose we developed a new, cuboid-based superpositioning method, which allows a unique, quantitative analysis of the Vα/Vβ-angles. Angle-based clustering led to six significantly different clusters. Analysis of these clusters revealed the unexpected result that the angle is predominantly influenced by the TCR-clonotype, whereas the bound epitope has only a minor influence. Furthermore we could identify a previously unknown center of rotation (CoR), which is shared by all TCRs. All TCR geometries can be obtained by rotation around this center, rendering it a new, common TCR feature with the potential of improving the accuracy of TCR structure prediction considerably. The importance of Vα/Vβ rotation for signaling was confirmed as we observed larger variances in the Vα/Vβ-angles in unbound TCRs compared to epitope-bound TCRs. Our results strongly support a two-step mechanism for TCR-epitope: First, preformation of a flexible TCR geometry in the unbound state and second, locking of the Vα/Vβ-angle in a TCR-type specific geometry upon epitope-MHC association, the latter being driven by rotation around the unique center of rotation.

Competing Interests: The authors have declared that no competing interests exist.

Author Summary

The recognition of antigenic peptides by cytotoxic T-cells is one of the crucial steps during the adaptive immune response. Thus a detailed understanding of this process is not only important for elucidating the mechanism behind T-cell signaling, but also for various emerging new medical applications like T-cell based immunotherapies and designed biotherapeutics. However, despite the fast growing interest in this field, the mechanistic basis of the immune response is still largely unknown. Previous qualitative studies suggested that the T-cell receptor (TCR) Vα/Vβ-interdomain angle plays a crucial role in epitope recognition as it predetermines the relative position of its antigen-recognizing CDR1-3 loops and thus TCR specificity. In the manuscript we present a systematic bioinformatic analysis of the structural characteristics of bound and unbound TCR molecules focusing on the Vα/Vβ-angle. Our results demonstrate the importance of this angle for signaling, as several distinct Vα/Vβ-angle based structural clusters could be observed and larger angle flexibilities exist for unbound TCRs than for bound TCRs, providing quantitative proof for a two-step locking mechanism upon epitope recognition. In this context, we could identify a unique rotational point, which allows a quantitative, yet intuitive description of all observed angle variations and the structural changes upon epitope binding.

Introduction

T-cells play a major role in cell-mediated adaptive immune responses necessary for the defense against foreign invaders and transformed malignant cells. Heterodimeric T-cell receptors (TCR) recognize antigenic peptides presented on the surface of cells by major histocompatibility complex (MHC) molecules. Recognition of MHC molecules presenting foreign peptides induces TCR signaling leading to T cell expansion and specific T cell functions such as elimination of virus-infected or transformed target cells. Therefore the immune system needs to balance the subtle distinction between self-restriction and self-tolerance and responses may reach extremes from multifunctional T-cell activation to tolerance induction. Due to the complexity of the signaling process, its mechanistic details are still not well understood. Several mechanisms of signal transduction have been proposed, which can be classified into (i) aggregation-, (ii) conformational change-, and (iii) segregation-models [1–3]. These three classes are not mutually exclusive. A conformational change in the TCR associated CD3 molecule was observed to be a basic early event in the signaling cascade [4]. In this context, mechanical forces applied by the TCR domains to the associated coreceptors are a suggested explanation [5]. Recent studies showed an antigen-specific conformational change of the A-B loop of the TCR constant α (Cα) domain for at least two TCR types. However, neither the structural details of this inter-subunit communication nor its initiation mechanism are yet known [4]. In order to provide the TCRs complex functions required for the signaling process, a variety of regulatory elements are involved in the process. Among those are the conformational changes within the TCR that are triggered during the early stage binding to the peptide-MHC (pMHC) complex. TCRs structurally consist of two membrane-anchored chains (α and β chain), which form two domains with an immunoglobulin-like (IG-) fold, one constant and one variable domain (Cα, Cβ, Vα, and Vβ). The variable domains of the two chains associate to the Vα:Vβ-complex, which binds to the pMHC-complexes and thus is responsible for antigen recognition. The overall structure is Fab-fragment like and each Vα and Vβ domain consists of a framework region and three antigen-MHC specific recognition loops, the CDR1 to CDR3 loops (Fig 1*A* and Fig 2*B*).

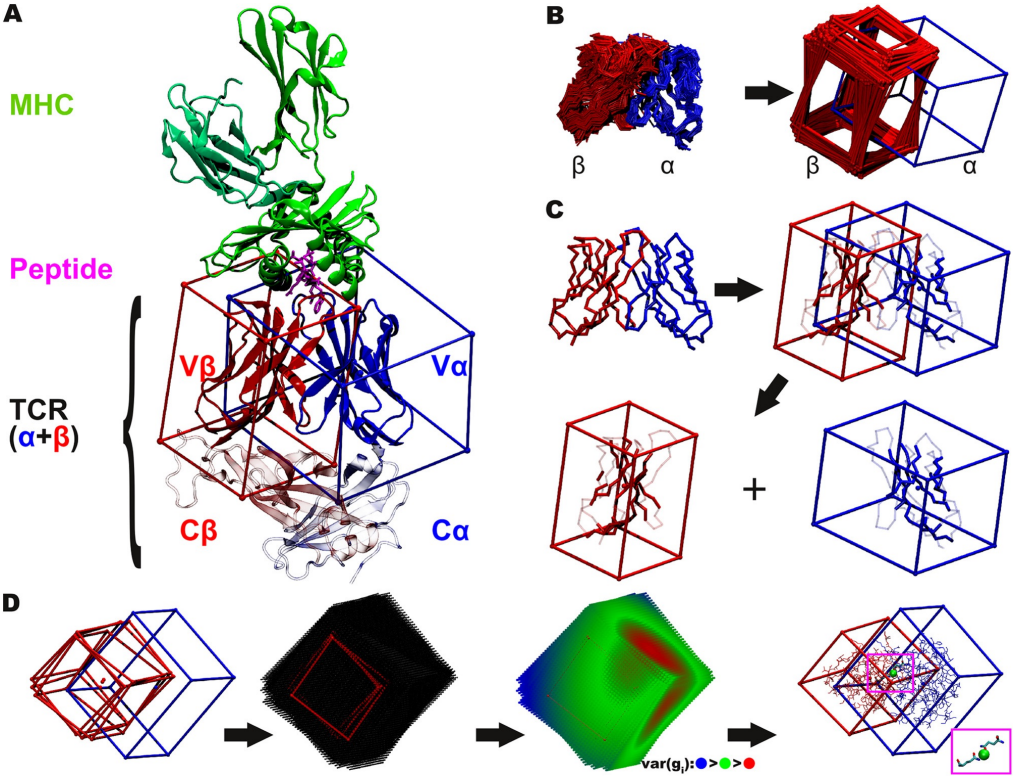


Fig 1. Cuboid and grid representations of the T–cell receptor geometries. (A) Localisation of the considered Vα and Vβ variable domains within the ternary TCR:pMHC complex. A TCR consists of two chains, the α and the β chain (blue and red). Each chain is partitioned into two domains, the constant domain (Cα and Cβ shown transparently) and a variable domain (Vα and Vβ, here surrounded by cuboids). The Vα and Vβ domains form the binding interface to the major histocompatibility complex (MHC) molecule (green) presenting an antigenic peptide (magenta) to the TCR. This work focuses on the variable domains. (B) Superimposition of the TCR variable domains. (i) The TCR structures were superimposed on the Vα domains leading to displaced Vβ domains. (ii) Cuboids were placed around the superimposed Vα and Vβ domains. This unified description of the different domains allows a quantitative analysis of the displacement. (C) Preparation of the cuboid placement templates. Vα (blue) and Vβ (red) domains of the structure 2bnu are used as reference structure. Both chains are surrounded with cuboids of the size of their spatial extent. Residues considered for superimposition are determined in an iterative process (unused residues are depicted transparently). These residues are used to compute the angular displacement of the Vβ domain relative to the Vα domain. (D) Center of Rotation (CoR). (i) Different geometries of (only three for clearness) β-cuboid geometries (red), superimposed on the α-cuboids (blue). (ii) Grids were fit into the β-cuboids. (iii) For each grid point *i*, the sum of pairwise distances and the variance was computed according to Formula 2. (iv) The residues at the center of rotation (CoR, green sphere) were investigated. For most of the structures, a conserved pair hydrogen bond interaction between the α and the β chain is located directly at the CoR. These hydrogen bonds are established by conserved Q residues.

doi:10.1371/journal.pcbi.1004244.g001

The capability of the immune system to recognize many different pMHC complexes is achieved by a vast variety of different TCRs. The T-cell repertoire was estimated to $2.5x10^7$ for one human individual [6], and to $2x10^6$ for mice [7]. This genetic variety together with the associated conformational differences within the TCRs seem to contribute to the structural and functional plasticity of TCRs [8]. The highly variable CDR3 loops encoded by VDJ recombinations are responsible for specific peptide recognition. Conformational changes within the CDR3 region after assembly with the pMHC complex have been demonstrated to provide an adaption to distinct peptide-MHC pairs which may additionally be influenced by CD8 co-

PLOS | COMPUTATIONAL BIOLOGY

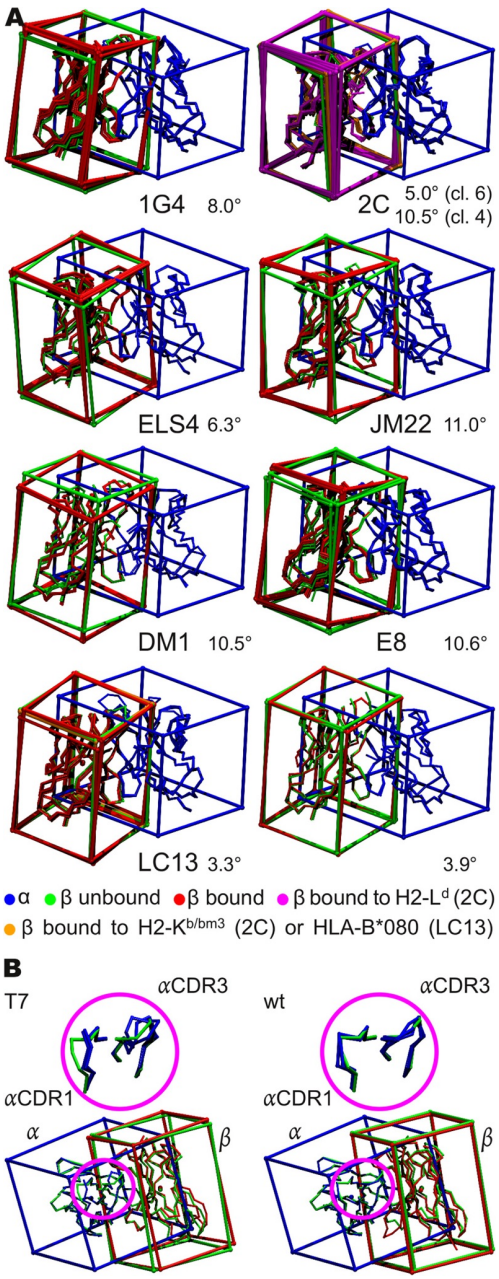


Fig 2. Differences in the TCR chain association geometries. (A) Differences between the bound and unbound geometries. Shown are seven different receptor types in their unbound state as well as their bound

state. Notably, for the 1G4 receptor the two unbound states are derived from different crystal structures, but are very similar. For some receptors, such as the 2C receptor, crystal structures including different ligands are available. In case of the 2C TCR, wild type (wt), 2C T7 and 2C T7 mutants are shown. In case of the E8 TCR, the two unbound states are derived from the same crystal structure. (B) Different conformations of the bound 2C TCR structures and their variants. The magnifications show the different CDR1/3 conformations observed in the m67 variant structure 2e2h (green), with respect to the 2C T7 variants (right, blue, 2oi9, 3e3q, and 2e7l) and the 2C wt structures (left, blue, 1g6r, 1mwa, and 2ckb). In the lower figures both variable domains are shown together with the placed cuboids for the structures 2e2h (m67, green, left+right) in comparison to 2e7l (T7 m6, blue/red, right) and 1mwa (2C wt, blue/red, left). The αCDR3 loops of the T7 variants differ in sequence and thus in their backbone conformations, whereas the CDR1 loop conformation is the same for the T7-wt, m6, and m13, but differs for m67 (upper left magnification). In the case of the m67 variant the CDR3 and CDR1 loop conformations are consistent with the 2C wt conformations (upper right magnification).

doi:10.1371/journal.pcbi.1004244.g002

receptor binding [9]. Moreover, the presence or absence of co-receptors as well as co-stimulatory molecules can have opposite effects on distinct TCR and T-cells suggesting an additional module for regulation [10]. Recently, it has been described that TCR Vα and Vβ domains can switch among alternate conformations when binding to MHC class I or II peptide complexes. A flex point in the FGXG motif of the J element has been proposed as swivel point for adjusting the interaction of Vα and Vβ [11]. In 1997 Li *et al.* proposed the capability of TCRs to increase their plasticity by rearranging the relative orientation of the Vα/Vβ domains, analogous to several known rearrangements of the V_L/V_H domains of antibodies [12]. Later works of Gagnon *et al.* reported a shift in the Vα/Vβ orientation of the A6 TCR bound to different ligands and influence of these shifts on the constant domains of the TCRs [13]. When the first structure of an A6:Tax:HLA-A2 complex was resolved, small variations in the Vα/Vβ interdomain angles could be determined [14,15]. This system was further studied with different agonistic or antagonistic peptides [16] and it was found, that different peptides induced these minor changes in the relative Vα/Vβ association geometries. Studies of the fluorination of the Tax-Peptide to increase the affinity confirmed this effect and also showed an alteration of the relative angle of the constant domains [13]. These scissoring effects were also observed for other receptors with different ligands or comparing the bound and unbound state: 2C [17], HA1.7 [18], LC13 [19], JM22 [20], DM1 [21], sc1.D9.B2 [22] and also for an invariant natural killer T cell receptor (NKT) [23].

The conformational changes were rather seen as further degree of freedom of the TCRs to adapt to the shape of their ligands [20,22]. A direct relationship between different conformational Vα/Vβ adjustments was not found [16].

In 2008, McBeth *et al.* systematically determined the Vα/Vβ interdomain angles for 35 TCR structures and concluded, that this angle is a general property of TCRs, which expands the repertoire of specificity [22]. Similarly, two recently published studies of Dunbar *et al.* investigate the interdomain geometries of antibodies [24] and compare them to the geometries of a non-redundant set of 39 structures [25]. The structure of TCRs is similar to Fab-fragments of antibodies [26], whereby the antibody V_L/V_H correspond to the TCR Vα/Vβ domains. In early and recent studies of antibody structures a rearrangement of the V_L/V_H upon ligand binding was considered and later confirmed [27–39]. Knowledge about TCR chain interactions might not only be important for the understanding of different TCR functions but may additionally provide information for reliable prediction of chain pairing. This is particularly interesting in T-cell based immunotherapy in which TCRs are considered as therapeutic tools for viral diseases and various cancers. For this purpose T cells redirected after genetic transfer of TCR chains with defined specificity are applied [40]. Understanding of TCR chain assembly is highly important in this regard as incorrect binding of introduced TCR chains with an endogenous TCR α and β chain may result in severe morbidity [41].

In the field of rational TCR engineering and optimization, homology modeling of these receptors gained in importance. Michielin *et al.* early created a homology model of the T1 TCR [42] using the MODELLER tool [43]. Later, other distinct TCR:pMHC models were investigated using more elaborate techniques including molecular dynamics (MD), computational alanine scan, or free energy calculations to study the influence of single mutations in the TCR or in the ligand, or to study differences of similar systems [44–54], and since recently, the automated modeling approach TCRep 3D is available to predict arbitrary TCR:pMHCI complex structures [55]. Recently Knapp *et al.* applied the ABangle methodology to a broad range of MD simulations of the LC13 TCR bound to 172 different ligands [24,56]. However, none of the previous modeling approaches explicitly includes any features concerning potential alterations in the Vα/Vβ interdomain angles, thus the presented structural analysis can help to improve the performance of the existing TCR modeling approaches.

In this work we perform a systematic, quantitative analysis of the Vα/Vβ interdomain angles in experimental TCR structures. For this purpose we developed a new structure-based method, which allows a systematic and very accurate quantitative comparison of the differences in the Vα/Vβ interdomain angles and introduces a new distance measure for clustering leading to a more accurate structural alignment of the TCRs than the approaches used in previous studies. The determination of TCR interdomain geometries is complicated by the fact that structural data is only available for a small subset of the vast variety of TCRs and that the TCRs for which structural data is available differ considerably in their loop structure and chain length, rendering the location of common conserved structural elements difficult. To solve these complications our method transfers all TCR variable domains into a unified geometric scaffold and performs a systematic analysis of the TCR structure geometries for 85 representative structures with respect to their Vα/Vβ interdomain geometries and interactions.

Results

To analyze relative positions of the Vα and Vβ domains of all bound and unbound TCR structures in the dataset (Table 1, S1 Table and S2 Table) we introduced a new methodology which assigns uniquely defined cuboid-based frames to the individual Vα and Vβ domains of the TCRs and thus allows an unambiguous analysis of their relative geometries (for details see Methods). Based on this method we first examined the relative positions of the two domains with respect to each other and then performed a throughout analysis of the structural basis of the obtained observations.

Cluster analysis of the TCR Vα/Vβ association angles

For the analysis of the relative Vα and Vβ domain geometries we superposed the Vα domain of these structures and investigated the differences in the position of the corresponding Vβ domains using their assigned cuboid frames. For this purpose a conserved framework region was identified in both TCR chains and cuboids were placed around each variable domain centered on the framework region (Fig 1*B–1D*). Afterwards the relative Euler angles of the Vβ cuboids were measured with respect to the superposed Vα domains.

The analysis showed that the relative positions of the Vα and Vβ domains of the TCRs differ considerably with respect to each other (Fig 3), which is consistent with former qualitative studies on small subsets or individual TCRs [22]. In Fig 1*B* it can be observed that if the central β-sheets of the Vα domain are superposed very well, the backbone positions of the corresponding Vβ domains differ significantly featuring interdomain Euler-angle distances d_E up to 30° (see Methods). Therefore, the two TCR binding domains can adopt different orientations (Fig 1*B*) with respect to each other. To analyze these differences in more detail we clustered all

Table 1. All TCR structures used for the analysis.

Name	S[a]	BS[b]	PDB	Name	S[a]	BS[b]	PDB
1G4	h	u	2bnu	LC13	h	1	3kps
1G4	h	1	2bnq	MEL5	h	1	3hg1
1G4	h	1	2bnr	OB.1A12	h	2	2wbj
1G4 AV-wt	h	1	2f54	OB.1A12	h	2	1ymm
1G4 c5c1	h	u	2pyf	RA14	h	1	3gsn
1G4 c5c1	h	1	2pye	SB27	h	1	2ak4
1G4 c49c50	h	1	2f53	SB27[K16Dα]	h	1	3kxf
1G4 c58c62	h	1	2p5w	TCR MS2-3C8	h	2	3o6f
1G4 c58c61	h	1	2p5e	TK3 wt	h	1	3mv7
3A6	h	2	1zgl	TK3 Q55H	h	1	3mv8
A6	h	1	2gj6	TK3 Q55A	h	1	3mv9
A6	h	1	1qsf	1934,4	m	2	2pxy
A6	h	1	1qse	1F1E8	m	u	3mff
A6	h	1	3d3v	226 TCR	m	2	3qiu
A6	h	1	3d39	226 TCR	m	2	3qiw
A6	h	1	1qrn	2B4	m	1	3qib
A6	h	1	1ao7	2B4	m	u	3qjf
A6	h	1	3h9s	2C	m	u	1tcr
A6	h	1	3pwp	2C	m	1	1g6r
AS01	h	1	3o4l	2C	m	1	1mwa
B7	h	1	1bd2	2C	m	1	2ckb
cf34	h	1	3ffc	2C T7	m	s,(2) [c]	2icw
DM1	h	1	3dxa	2C [T7-wt-s] [d]	m	1	2oi9
DM1	h	u	3dx9	2C m13 [T7-s] [d]	m	1	3e3q
E8	h	2	2ian	2C m6 [T7-s][d]	m	1	2e7l
E8	h	2	2iam	2C m67 [T7-s] [d]	m	1	3e2h
E8	h	u	2ial	2W20	m	2	3c6l
ELS4	h	1	2nx5	5c.c7	m	u	3qjh
ELS4	h	u	2nw2	AHIII12.2	m	1	2uwe
HA1.7	h	2	1fyt	AHIII12.2	m	1	2jcc
HA1.7	h	2	1j8h	AHIII12.2	m	1	1lp9
Hy.1B1	h	1	3pl6	B3K506	m	2	3c5z
JM22	h	1	2vlj	BM3.3	m	1	1nam
JM22	h	1	2vlk	BM3.3	m	1	1fo0
JM22	h	1	1oga	BM3.3	m	1	2ol3
JM22	h	u	2vlm	cl19	m	2	2z31
JM22	h	s,(2)[c]	2xn9	D10	m	2	1d9k
JM22	h	s	2xna	KB5-C20	m	1	1kj2
JM22 [S99βA]	h	1	2vlr	N15	m	u,a	1nfd
KK50.4	h	1	2esv	TCR 21.30	m	2	3mbe
LC13	h	1	1mi5	TCR172.10	m	2	1u3h
LC13	h	u	1kgc	YAe62	m	2	3c60
LC13	h	1	3kpr				

[a] Species: h = *homo sapiens*, m = *mus musculus*.

[b] Bound state: u = unbound, 1 = MHCI, 2 = MHCII, s = superantigen. More detailed information is available in S1 Table.

[c] Structure 2xn9 and 2icw are not considered as MHC II bound TCRs, since the TCRs only contact the super-antigens.

[d] WT with solubility mutations acc. to ref. [63]. More detailed information is available in S2 Table.

doi:10.1371/journal.pcbi.1004244.t001

PLOS | COMPUTATIONAL BIOLOGY

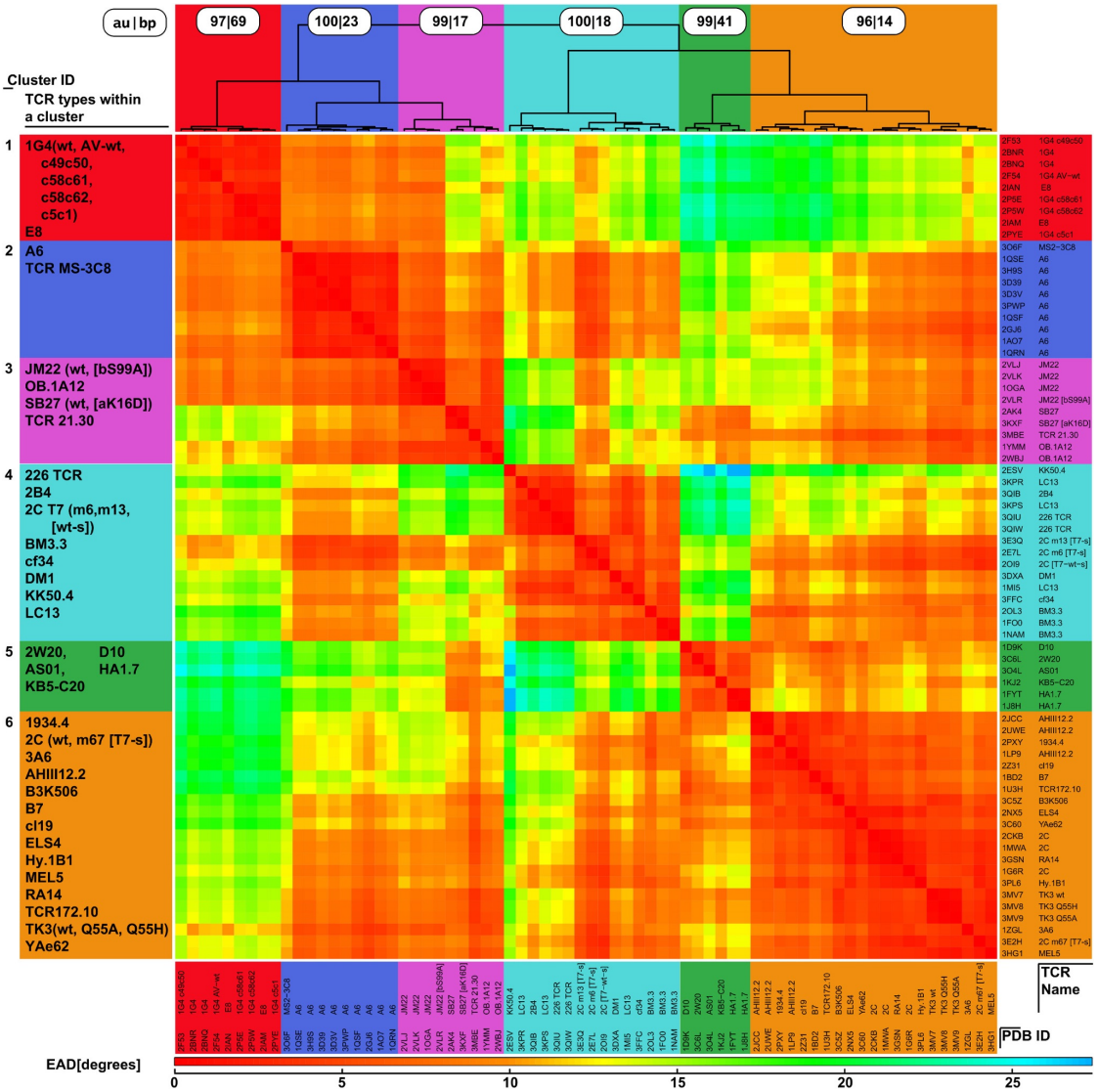


Fig 3. Geometry clusters of pMHC bound TCRs. Pairwise Euler-angle distances (EAD) were determined for all pMHC-bound TCR structures according to Formula 1. The distance matrix was hierarchically clustered using the Ward update formula. We identified six significant clusters, using a bootstrapping approach [58]. Notably, in most of the cases, TCRs of the same type occur in the same cluster. Upper panel: Clustering dendrogram with bootstrapping results (au = approximately unbiased, bp = bootstrapping probability). Left panel: TCR types occurring within a cluster. Right/lower panel: PDB identifiers and corresponding TCR names. Central panel: Pairwise Euler-angle distances (EAD). The color key is provided in the bottom of the figure.

doi:10.1371/journal.pcbi.1004244.g003

superposed (i) MHC-bound structures (Fig 3 and S1 Fig) as well as (ii) all MHC bound and unbound structures (S2 Fig and S3 Fig) according to their angular deviations in the Vβ

domains with respect to the corresponding Vα domain using the Ward clustering algorithm [57]. Afterwards we performed a bootstrapping analysis (Fig 3) and identified six clusters with a significance greater than 95% [58]. Nearly all structures of TCRs of the same type from which different X-ray structures exist were placed in the same cluster (93% of the TCR types with more than one MHC bound crystal structure; except 2C TCR). This shows that the observed phenomenon is not caused by the variation of the crystallographic conditions and that the clustering is robust, describing a phenomenon which is caused by biological differences within different types of TCRs.

Structural analysis of the TCR cluster geometries

After the angle-based cluster analysis of the superposed structures we analyzed the structural features leading to the different interdomain geometries observed. For this purpose we used a grid-based analysis of the superimposed cuboid structures (Fig 1*D*, for details see Methods). This analysis showed that all TCR structures share an area, which is invariant towards rotation and translation of the TCR variable domains. At the center of this region a rotation point (Center of Rotation, CoR) can be identified, which exists in all TCR structures. The core region around this CoR is situated at the center between the two Vα and Vβ domains (Fig 1*D*). Notably, the average CoR position (x = 27.768Å, y = 36.783Å, z = 55.723Å) with respect to the reference coordinate system (2bnu) is located directly between or close to a twofold hydrogen bond between two conserved residues (Q for most of the structures), one from each chain (Fig 1*D*, magenta box). These hydrogen bonds connecting the two chains are known to be conserved through all TCRs [59]. As similar structural constraints were observed for antibodies [59,60], these features (CoR stabilized by conserved H-bonds) seem to be characteristic for Fab-fragment like domains in general.

To investigate the conservation of these two residues, we performed a sequence-based analysis with the sequences of all currently known functional variable αβ TCR gene segments as found in the database IMGT/GENE-DB [61]. In total 342 α chain and 164 β chain sequences were analyzed (six of 348 α sequences were incomplete). This analysis shows, that in contrast to antibodies, different residues can be found at the CoR position (Table 2 and S3 Table). Table 2 provides the absolute number of the observed amino acids at the CoR position separately for the known α- and β-alleles. The investigated CoR position corresponds to sequence position 44 in the IMGT unique numbering [62] scheme for both, the Vα and the Vβ domains.

Table 2. Conservation at the CoR position.

AA[a]	Freq. α [%]		Freq. β [%]	
Q	89.2	(305)	98.2	(161)
R	26.0	(9)	1.2	(2)
E	0.3	(1)		
H	5.0	(17)		
W	0.6	(2)		
K	1.8	(6)	0.6	(1)
L	0.6	(2)		

Relative (and absolute) Frequency of the AA at the α or β CoR position, based on an multiple sequence alignments of all functional variable TCR gene segments alleles of the α (342 sequences) or β (164 sequences) locus obtained from the IMGT/Gene-DB [61].

[a)] amino acid type

doi:10.1371/journal.pcbi.1004244.t002

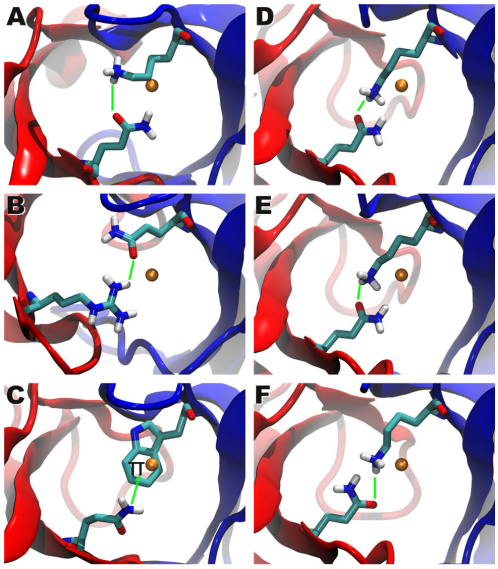


Fig 4. Exceptional structural examples of the center of rotation. Region around the Center of Rotation (CoR), the Vα domain is shown in blue and the Vβ domain in red. Hydrogen atoms were added for the end-groups of the interacting amino acids. The average center of rotation is drawn as an orange sphere and the interacting residues are shown in licorice representation. CoR stabilizing interactions are drawn as a green line. For these six structures the highly conserved Q-Q interaction between the α and the β chains is replaced by the following residues (shown in licorice style): (A, D-F) αK (PDB-IDs: 1bd2, 3qiu, 3qiw, 3qjh), (B) βR (PDB-ID: 2esv), (C) αW (PDB-ID: 3gsn).

doi:10.1371/journal.pcbi.1004244.g004

In case of the α chain the amino acids Q, H, R, K, L, W, and E can be observed, whereas in case of the β chain Q is overrepresented, but is occasionally replaced by R and K. The lower amount of different residues found in the β chain alleles might be a statistical artifact, since for the α chain about twice as many sequences are known than for the β chain.

Structural investigation of the interaction pattern of these alternative residues observed at position 44 showed that all of them can form strong interactions with their interacting partner residue in the complementary chain and thus compensate for the lost hydrogen bonds of the Q-Q interaction (see Fig 4): In some cases the Q residue of the α chain is replaced by an apolar W or L residue. In the case of W the Π-system of its indolyl group forms strong interactions with the Q residue from the opposite chain. In most cases Q is replaced by R or K and therefore the formation of the interchain hydrogen bonds can still be observed, as shown in Fig 4. As no structures are available for the replacement of Q by E or H no structural analysis is possible for these mutations. The same holds for the K mutant in the β chain, as in all available structures the conserved position in the β chain is occupied by Q except for the TCR KK50.4 (structure 2esv), where Q is replaced by R (Fig 4*B*). In this structure the side chain oxygen atom of the Q residue of the α chain forms a hydrogen bond with the guanidine group of the R residue. Compared to structures with Q-Q interactions, the Q residue is slightly displaced towards a neighboring loop, due to the size of the interacting R residue. This displacement allows a further interaction of the Q residue with a backbone carbonyl-oxygen of the neighboring loop. The α chain offers more diversity: K residues are found at the α-CoR position in the TCRs B7 (PDB

ID 1bd2), 226 TCR (PDB IDs 3qiu, 3qiw), and 5c.c7 (PDB IDs 3qjh) (Fig 4*A*, 4*D*–4*F*). The rare W residue at the α-CoR position can be observed in the RA14 TCR (PDB ID: 3gsn, Fig 4*C*). In the β chain of the B7 TCR (1bd2) the CoR position is occupied by a conserved Q residue, the side chain oxygen is directed towards the side chain nitrogen atom of the K residue at the CoR position of the α chain. The distance between the two atoms is 3.59 Å. The K residue is drawn towards a neighboring loop, such that the amino group can also interact with an oxygen atom of the backbone of the loop (distance N-O: 2.89 Å). This additional interaction stabilizes the conformation of the K residue. For both "226 TCR" structures (3qui and 3qiw) very similar conformations of K- and the Q-residue can be observed: the side chain nitrogen atom of the K of the α chain is directed towards the side chain oxygen of the Q (distance in 3qui: 2.39 Å; 2.60 Å for 3qiw). In the 5c.c7 structure (3qjh) the atomic coordinates of the two observed Q and K residues are very similar compared to the two "226 TCR" structures. However, the oxygen atom and the nitrogen atom of the Q residue are swapped in one of the BUs, so that the K nitrogen is directed towards the nitrogen atom of the Q. In the RA14 TCR (PDB ID 3gsn), where W occurs at the CoR position, the conformation of the W residue is stabilized by a hydrogen bond (distance 2.15 Å) between the nitrogen NE1 and a backbone carbonyl-oxygen of the neighboring loop of the α chain. The W residue flanks the hydrophobic core of the TCR. The Q residue of the β chain pushes towards the solvent, due to the size of the W residue. For E and H we also expect the formation of hydrogen bonds with the β chain, however, no structure exist of this case yet.

Overall, the existence of the conserved CoR in such close proximity to the conserved αQ-βQ interactions confirms the hypothesis of a rotation-driven mechanism of α:β-association leading to the differences in the association angles of the Vα and Vβ domains. However, due to the low amount of mutated sequences available it was not possible to investigate the influence of the different amino acids occupying the conserved position 44 on the TCR interdomain geometry and the TCR specificity in a comprehensive manner. In general the above observations suggest an association mechanism of the Vα and Vβ domains in which the hydrogen bond interaction between the conserved residues are formed first and afterwards the domains arrange each other around this pivot point, adopting different relative association angles.

Detailed analysis of specific TCR:pMHC structures

Next we performed a more detailed functional and structural analysis of the clustering results. For several different TCR types (2C, A6, 1G4, JM22, BM3.3, AHIII12.2, TK3) there exist more than one structure within the analyzed dataset in which the TCR is either bound to different MHC alleles and/or different peptides or different variants of the same TCR were crystallized.

These structures differ in several features: i) mutations in the TCR framework (S2 Table) or CDR-loop regions, ii) different presenting MHC molecules (including different alleles, single point mutations, and different MHC classes), and iii) different peptides presented to the TCR (including single point mutations). Furthermore, the data set includes the two similar TCRs "2B4" and "226", which both share the gene loci for their variable segments of their α chains as well as their β chains, but differ in the loci for the joining segments.

Based on our cluster analysis we can distinguish between *major* and *minor* angular differences. According to our definition minor differences between two TCR structures occur if both structures can be found within the same cluster. Major angular differences between two TCR structures can be found for structures assigned to two distinct clusters.

Analyzing the clustering behavior of the different structures available for the same TCR types and their variants we observed the interesting results that for all except one TCR type (2C) all structures belonging to the same TCR type are located within one cluster (Fig 3). Thus,

we investigated this phenomenon in more detail. In this section we briefly summarize the main results and refer the interested reader to a more detailed description in the supporting material (S1 Text).

Detailed analysis of the different structures available for the TCR types A6, 1G4, JM22, BM3.3, AHIII12.2, and TK3 shows that neither mutations within the TCRs nor the binding to different peptidic ligands of varying immunogenicity (e.g. A6 [13–16], S1 Table) or MHC alleles lead to major angular differences. However, minor angular differences are frequently observed within the individual sets (see Supporting Material).

In contrary, the 2C TCR can be found in two different angular clusters (Fig 3, Table 3). This is in agreement with the original publications, which show that depending on the pMHC bound, the 2C TCR can adopt two distinct docking orientations [63], but that on the other hand mutations in the CDR3 loop of the TCR do not lead to significant changes in its orientation with respect to the pMHC ligand, if the same pMHC is bound [64]. Regarding the bound pMHC alleles in our two clusters, we find that all TCRs bound to the MHC molecule H2-K1^{b} associate in cluster 6, whereas all TCRs bound to the MHC molecule H2-L^{d} are located in cluster 4, except for the m67 variant, which is bound to H2-L^{d}, but located in cluster 6. As the reason for this unusual behavior of the m67 variant it was found that its mutation of the αCDR3 loop sterically enforces a conformation of the neighboring αCDR1 loop (binding the MHC molecule), which leads to a shift between the Vα domain and thus to different interdomain angles, closer to cluster 6 (Fig 2). The same conformation shift is also observed in the experimental publication, but as the docking orientation of the Vβ-domain on the MHC is retained and only the relative Vα loops shift, no significant changes are observed with respect to the overall docking orientation [64]. Due to this surprising result we had a closer look at the structures and discovered that actually two subtypes of the 2C TCR were crystalized: the wild type (wt) and the 2C T7 TCR, which differ in the framework region (S2 Table). The wt 2C TCRs are all bound to the MHC molecule H2-K1^{b} and associate in cluster 6, whereas the T7 TCRs are all bound to the MHC molecule H2-L^{d} and belong to cluster 4, except the m67 variant. Therefore the two TCR structures compared in [63] actually belong to two different variants and thus the

Table 3. Pairwise Euler Angle Distances [°] of the bound and free 2C TCR variants.

PDB	C[a]	S[b]	L[c]	1tcr	2ckb	1mwa	1g6r	3e2h	2e7l	2oi9	3e3q
1tcr	-	wt	U	0,0	3,4	4,1	5,9	6,8	9,9	9,5	12,1
2ckb	6	wt	*K*E	3,4	**0,0**	**0,8**	**2,9**	**3,5**	6,6	6,2	8,8
1mwa	6	wt	*K*E	4,1	**0,8**	**0,0**	**2,1**	**2,8**	5,9	5,4	8,0
1g6r	6	wt	*K*S	5,9	**2,9**	**2,1**	**0,0**	**2,0**	4,3	3,8	6,3
3e2h	6	m67	*L*Q	6,8	**3,5**	**2,8**	**2,0**	**0,0**	3,5	2,8	5,5
2e7l	4	m6	*L*Q	9,9	6,6	5,9	4,3	3,5	**0,0**	**2,0**	**2,3**
2oi9	4	[T7-wt-s]	*L*Q[d]	9,5	6,2	5,4	3,8	2,8	**2,0**	**0,0**	**3,0**
2e3q	4	m13	*L*Q	12,1	8,8	8,0	6,3	5,5	**2,3**	**3,0**	**0,0**

The structures were superimposed to the α variable domains. All Euler angle distances are given in degrees in respect to averaged geometries of all biological units. Unlike other 2C T7 variants (m6, T7-wt-s, m13) the m67 variant (underlined) affiliates to cluster (C) 6 occupied by the 2C wt TCRs bound to a different ligand. The two clusters are emphasized by bold typesetting.

[a)] Cluster affiliation.

[b)] Subtypes.

[c)] Ligands: U = unbound, *K*E = H2-K1^{b}+EQYKFYSV, *K*S = H2-K1^{b}+SIYRYYG, *L*Q = H2-L^{d}+QLSPFPFDL. The ligand main type (MHC) is indicated by the first letter in italics.

[d)] MHC mutation: (F9Y)(V12T)(I23T).

doi:10.1371/journal.pcbi.1004244.t003

results and conclusions of that publication, namely that the different docking orientations are solely caused by the different pMHC ligand bound, need to be regarded with caution. Unfortunately, as for both variants only bound structures to the same pMHC are available and no "cross" TCR type ⇔ pMHC allele structures, it can not clearly be distinguished if the differences in the docking orientation and Vα/Vβ angles are caused by the framework mutations or the different pMHCs bound. However, regarding the above discussed 2C T7 m67 TCR, which is bound to H2-L^d and has the T7 framework mutations, but still adopts an angular conformation closer to the 2C-wt:H2-$K1^b$ (Table 3, underlined) structures belonging to cluster 6 (Fig 2*B*) [64], it seems that neither of the above features (framework mutation or pMHC allele bound) seems to induce unsurmountable restrictions on the final TCR conformation. As the m67 variant is the only T7 variant, which adopts the 2C wt Vα/Vβ angle, the induced changes in its αCDR1 conformation, which are not present in the other T7 variants (Fig 2*B*), seem to play a crucial role for the Vα/Vβ association angle, whereas the αCDR3 conformation influences the angle only indirectly.

These results indicate that the 2C TCR can in principle adopt two distinct conformations, which can be modulated by framework as well as the CDR mutations and presumably also its binding partner.

This is in agreement with previous qualitative observations about the overall TCR:pMHC association angle, stating that this angle is mainly dependent on the nature of the MHC allele and the TCR type rather than the antigenic peptide molecule [22,63–65]. As the CDR1 and CDR2 loops are interacting with the MHC molecule and the CDR3 loop predominantly with the bound peptide, the observed CDR1 dependent structural changes are in agreement with these former studies and might be a complementary feature to the CDR3-peptide binding in the process of TCR signaling. However, as these observations are based on one TCR only, these conclusions should be taken with caution.

Comparison of bound and unbound TCRs

To further investigate the influence of the pMHC complex on the overall TCR structure we compared the structural features of unbound and bound TCR structures of the same type (Fig 2*A*, S2 Fig and S3 Fig). We observed that in most of the cases the orientations of the unbound TCRs slightly differ from the bound TCRs. The seven TCR types 1G4, 2C, DM1, ELS4, JM22, 2B4, and LC13 can be found in the unbound state as well as in the MHC bound state in our data set–TCRs bound to superantigens are not considered. Only in the case of the 2B4 TCRs and the LC13 TCRs both states associate in the same clusters. In the other cases, the angular deviation of the unbound TCRs is between 5° to 11°, leading to an association to a different cluster than the bound variants. Comparing all examined structures of bound and unbound TCRs it can be observed that the differences in the Vβ domain orientations are considerably larger for the unbound TCRs (S3 Fig).

In Fig 2*A* the differences between the bound and the unbound structures are illustrated for several TCR types. The repertoire of analyzed 1G4 TCRs contains nine structures of wt TCRs and mutants. Two different structures are available in the unbound state: (i) The structure 2bnu is the wt and (ii) 2pyf is the variant c5c1, which differs from the wt in the αCDR3-, βCDR2-, βCDR3-loops, and in three positions of the framework region [66, 67]. The subset of bound 1G4 TCRs contains wt TCRs (2bnq and 2bnr), the variant wt-AV (2f54, contains solubility mutations in the framework region [68]), and variants, which contain mutations in the framework region and the αCDR2-, αCDR3-, βCDR2-, βCDR3 loops: c5c1 (2pye), c49c50 (2f53), c58c62 (2p5w), and c58c61 (2p5e)–S2 Table lists the mutations in detail. All ligands of

Table 4. Pairwise Euler Angle Distances [°] of the bound and free 1G4 TCR variants.

PDB	S[a]	L[b]	2bnq	2bnr	2f53	2f54	2p5e	2pye	2p5w	2pyf	2bnu
2bnq	W	v	0,0	1,3	1,5	1,2	2,2	2,4	2,5	7,2	7,9
2bnr	W	c	**1,3**	0,0	2,7	2,2	2,5	2,2	2,9	6,1	6,7
2f53	B	c	**1,5**	**2,7**	0,0	2,1	1,7	2,5	1,8	8,7	9,3
2f54	V	c	**1,2**	**2,2**	**2,1**	0,0	3,2	3,6	3,5	7,3	8,4
2p5e	D	c	**2,2**	**2,5**	**1,7**	**3,2**	0,0	1,0	0,4	8,5	8,7
2pye	A	c	**2,4**	**2,2**	**2,5**	**3,6**	**1,0**	0,0	1,2	7,8	7,8
2p5w	C	c	**2,5**	**2,9**	**1,8**	**3,5**	**0,4**	**1,2**	0,0	8,8	9,0
2pyf	A	u	7,2	6,1	8,7	7,3	8,5	7,8	8,8	0,0	2,5
2bnu	W	u	7,9	6,7	9,3	8,4	8,7	7,8	9,0	2,5	0,0

The structures were superimposed to the α variable domains. All Euler angle distances given in degrees. Averaged angle distances: Inter unbound: 2.5°, inter bound (bold): 2.1°, bound vs. unbound (underlined): 8.0°.

[a] Subtypes: W = wild type, V = AV-wt, A = c5c1, B = c49c50, C = c58c62, D = C58c61

[b] Ligands: u = unbound, v = SLLMWITQV+HLA-A*0201, c = SLLMWITQC+HLA-A*0201.

doi:10.1371/journal.pcbi.1004244.t004

the bound 1G4 TCR structures are the MHC molecule HLA-A*0201 presenting the peptide SLLMWITQ**C** (except 2bnq: SLLMWITQ**V**).

Notably, the two unbound orientations differ only by 2.5°, but have an average distance of 8.0° to the bound structures (Table 4). On the other hand, all bound 1G4 TCR structures are very similar (2.1°, var = 0.7°). This indicates a shift in the relative orientation of the two domains upon binding of the TCR to the peptide-MHC complex. Both unbound structures associate in cluster 2, differently to the bound 1G4 TCRs, which are found in cluster 1.

For both, the DM1 and the JM22 types, an angular deviation between the bound and the unbound state of 10.5/11° can be observed. The JM22 TCR was reported to reveal a considerably greater scissoring motion than other TCRs [20], which is consistent with our findings. The unbound JM22 TCR (2vlm) can be found in cluster 6, whereas the bound JM22 TCR structures are located in cluster 2. The bound (3dxa) or unbound (3dx9) DM1 TCRs can be found in cluster 4, respectively cluster 2. The unbound E8 (wt) structure (2ial) associates to cluster 1 and differs by 10.6° from the bound (wt) variants (2ian, 2iam), which associate to cluster 3. The bound variant of ELS4 (2nx5) is located in cluster 6 and differs to the unbound variant (2nw2, cluster 4) by 6.3°.

In the case of the 2C TCR, the unbound wt (1tcr) associates with cluster 6, which contains the bound 2C wt structures and the exceptional bound 2C T7 m67 variant (see above). The average angular distance of the unbound wt to these bound structures is 5.0° whereas it's distance to the bound 2C T7 variants in cluster 4 is 10.5°. In contrast, for the LC13 and the 2B4 TCRs the angular difference between the bound and the unbound structure is low (3.3° and 3.9°, cluster 4).

Thus by including the unbound TCRs into the clustering process (S1 Fig), a tendency towards smaller significant clusters can be observed. This means, that the pMHC-ligand stabilizes the TCR variable domain geometries in a favored position. The TCRs' ability to adopt multiple geometries might play an important role in the signal transduction and the loss of flexibility upon pMHC binding might induce an initial event in the signaling cascade.

Another interesting point is that structures from human and mouse are found in the same clusters, no differences were observed in their clustering behavior.

Discussion

We performed a comprehensive quantitative analysis of the structural features of T-cell receptors in their bound and unbound states. For this purpose, we introduced a new cuboid-based method, which allowed us to obtain a unique quantitative measure for the Vα/Vβ association angles and thus the previously observed rotation between the two TCR domains. As our method is based on highly conserved framework residues and ignores the loop regions it can be applied to all possible chain combinations and we performed a detailed analysis based on a representative set of all currently available TCR structures in the PDB Database.

Differences in the TCR Vα/Vβ association angles were first recognized for the A6:Tax:HLA-A2 complex by Ding *et al.* [16]. Since then the same phenomenon was also observed for other TCR clonotypes by several groups [17–21,23]. The first more comprehensive analysis of the angular space of the TCR Vα/Vβ association was performed by McBeth *et al.* [22], who analyzed 38 TCR structures (biological units), including unbound TCRs, MHC I- or MHC II-bound TCRs, and three NKT TCRs. The analysis was based on three angles: two angles were defined as the pitch of a pseudodyad axis and a third angle described the rotation around this axis when superimposing the two variable domains.

The results of the study of McBeth showed that different TCRs adopt a broad range of orientations and that the orientation of TCRs of the same type in the bound and unbound states can differ. Furthermore, the authors observed angular differences between TCRs differing only in a few amino acids, concluding that the variation of the interdomain angle potentially has an effect on the TCRs specificity or polyspecificity [22].

The pseudodyad-based method used by McBeth *et al.* is a classical approach of crystallographers to determine the relative orientation of antibody V domains or to determine the antibody elbow angle in Fab fragments. The computation of the pseudodyad-axis is achieved by superimposing of the Vα onto the Vβ domain. The drawback of this approach is that the precision of this process depends on the similarity of the two domains and it can be expected that the cross-chain similarity of the variable domains is lower than the similarity between two variable domains of the same chain type (either Vα or Vβ). Thus, two variable domains of the same chain type can be structurally aligned more precisely than superimposing similar cross-chain domains. Due to these limitations we developed a new method for superpositioning, which allows a unique definition of the interdomain rotational angle by superimposing domains of the same type using structurally highly conserved regions for the superimpositioning process. Our method describes the orientation of the Vβ domain relative to the Vα domain by a unified cuboid instead of a pseudodyad-axis as used by McBeth. The cuboid-based description provides several benefits. First, only one angle is necessary to describe the interdomain rotation, which is not only intuitively accessible, but also forms the simplest description of the phenomenon and allows a straightforward bioinformatics structural analysis. Second, the Euler angle distance can be computed between cuboids, which can be used as a measure for clustering. Third, cuboid geometry combinations can be used as a template for an arbitrary cross type chain assembly in a modeling process.

Since 2008 the number of TCR structures available in the PDB increased considerably. Therefore, we were also able to base our analysis on a much broader data set. The data set of McBeth included 18 non-NKT and 3 NKT structures (38 BUs), whereas our set contains 37 different non-NKT TCR types (mutants not counted, 136 BUs). In both studies free, MHC I bound, and MHC II bound TCRs originating from human or mouse were studied. However, the recent data allowed us to compare additional TCRs in bound and unbound state (*e.g.* DM1, JM22, LC13, E8, 2C, 1G4). For other TCRs the new dataset contains structures with additional different pMHC ligands (*e.g.* A6, SB27, 1G4, 2C, LC13, Ob.1A1).

Recently another study was published by Dunbar *et al.* [25], which also analyzes the TCR Vα/Vβ interdomain angle. However, the focus of that study is on the comparison of TCR and antibody geometries, because of its importance in the field of rational design of TCR-like antibodies. This is quite different from our goal of a systematic comparison of the interdomain angle variations within different TCR structures depending on their surrounding and thus the publication of Dunbar *et al.* sheds light on an important, yet complementary aspect of TCR architecture. Due to its different focus, the study also differs in the methodology applied as well as the data set used and the results discussed. Regarding the data set Dunbar *et al.* examine a smaller set of 39 structures, which does not contain TCR type-based redundant structures to avoid statistical bias in the comparison with the antibodies. In contrast, the inclusion of different structures of the same TCR is a desired feature of our data set as our analysis focuses on the differences between the available TCR structures in dependence of their binding state and partners. However, in contrast to Dunbar's study, we excluded NKT receptors (binding CD1d ligands) as well as structures containing superantigens as they show a different binding behavior and function and thus are not representative for TCR:pMHC complex structures.

Further, as Dunbar's study focuses on the comparison of TCRs with antibodies, it is based on an adaption of the ABangle methodology to TCRs, which was originally developed for antibodies [24]. Although our method and ABangle have in common that they use the conserved positions in the IG-like domain for structural alignment, the ABangle method describes the rotation and translation by five angles, a distance, and a precomputed axis. A benefit of this method is the ability to inspect each component of the transformation separately and therefore it allowed identifying the main difference between the antibodies and the TCRs angular space, which lies in the HC2 (twist) angle. In contrast, the major goal of our analysis is to analyze possible orientations the two TCR variable domains can adopt depending on their type and state, functional mutations and the bound ligand. For this purpose we introduced a specialized robust method for the applied cluster analysis, which differs considerably from the method of Dunbar *et al.*: It reduces the variable domains to cuboids, to allow easy visualization of the transformational differences between two TCRs. Further, we describe the rotation of these cuboids by Euler angles, from which an Euclidean distance can be calculated, which is needed to obtain a robust clustering, as the commonly used RMSD-based measure was found to be too insensitive to capture the partially rather small angular differences between two TCRs. The Euler angle based measure showed a more robust performance and is, in addition, independent of protein translation. In contrary, the study of Dunbar *et al.* is based on the RMSD of the relative domain orientations, which is accurate enough to clearly distinguish between the two molecule classes (TCR and AB).

Finally, instead of several independent components we use one center of rotation (CoR) to describe the angular differences, which is a necessary prerequisite for the cuboid-based clustering and provides an intuitive measure for presenting and discussing our results. In addition, we use bootstrapping to confirm the significance of our clustering results [58].

As the focus of the study of Dunbar is on the comparison of the Vα/Vβ interdomain angles of TCRs and with the V_H/V_L angles of antibodies, the study leads to the important result that TCRs and antibodies differ significantly in their interdomain angles. However, it also demonstrates that TCR-like antibodies, which were specially designed for pMHC binding, can adopt TCR-like geometries. Thus the study provides an important contribution to a better, detailed understanding of the structural features and characteristics of immunoglobulin-like folds and should therefore be very helpful for the rational of protein-based pharmaceuticals.

In agreement with the majority of the previous, predominantly experimental studies on small sets of TCRs, our comprehensive cluster analysis of the bound structures of the TCRs showed that TCRs of the same kind normally occupy the same structural cluster. Only one

exception was observed, in this case two different clusters were found for the wt and mutant form (T7) of the same TCR (2C) and a specific combination of mutations in the framework and the CDR3 loop led to a shift of one mutant structure (T7-m67) into the cluster of the corresponding wt structures. For all other TCR types only one cluster was observed for both wt and mutated MHC-bound structures. These observations indicate that the differences in the Vα and Vβ interdomain angles of the bound TCR structures are predominantly determined by the preformed chain combination and subtype dependent interdomain angles of the unbound TCR structures and that neither the type of bound peptide nor the presenting MHC molecule lead to a significant angle shift. This geometry can be altered within the range of the subtype structures of the same TCR by mutations in the CDR loops as observed for the 2C m67 TCR. However, the analysis of the 1G4 structures showed that most changes in the CDR sequences do not have a significant effect on the interdomain angles. The same holds true for the binding of different MHC-peptide complexes to the same TCR, *e.g.* as shown for the A6 TCRs. This is in agreement with the previous studies of these TCRs, which observed only small differences in the Vα/Vβ interdomain angles for the different variants and bound pMHC complexes studies of these TCRs [14–16].

Comparing bound and unbound structures of the same TCR, a strong shift in the interdomain angles was observed in most cases upon binding of the TCR to a pMHC complex, as the bound and unbound structures of the same TCR were observed in different structural clusters. Further analysis showed that the differences in the interdomain Euler angles between the bound and unbound structures of the same TCR were often significantly higher than the variation of these angles within the bound or unbound structure set.

As basis for the observed differences in the association angles a so-called Center of Rotation (CoR) could be identified. This CoR is situated in the vicinity of two to four conserved residues (mainly Q), which interact via hydrogen bonding or charged interactions thus stabilizing the rotation center. Sequence analysis of these conserved residues showed that in contrary to antibodies in TCRs different amino acids can occupy these positions. However, all observed side chain types share the capability to form directed interactions such as hydrogen bonds. Due to the limited amount of TCR structures featuring other residues than Q at these positions, analysis of a correlation between the occurrence of specific residues at these position and the observed interdomain angles was not possible. The observation of a CoR is in agreement with previous studies of individual TCR types, as *e.g.* Ishizuka *et al.* [20] observed that for the JM22 TCR a binding hotspot of Vα/Vβ could be a center of motion or rocking. In this study, all JM22 structures were superimposed to the Vβ domains and the hinge was located at the salt bridge Q38α-Q39β [20]. In addition, already in the first described Vα/Vβ complex structure (2C) the (i) conserved Q-Q interaction between Vα and Vβ was observed at the Vα/Vβ binding interface, as well as water mediated hydrogen bonds between conserved residues of both domains and a symmetric hydrophobic core consisting of further conserved residues [26].

These individual results are considerably substantiated by our broad analysis. Throughout our dataset only minor variations were observed in the position of the CoR, which is highly conserved. In addition, nearly no shifting motions were observed, which seem to play only a minor role in the adjustment of the variable domains compared to the angular displacement.

Next to the conserved hydrogen bonds around the CoR, the contact area between the Vα and the Vβ domain is dominated by hydrophobic residues and is shaped similar to a saddle joint. This shape allows a certain rotation and translation of the Vβ domain sliding on the Vα-interface. As found by our grid analysis, the center of rotation is located at this area, but is slightly flexible. In contrast to the constant domains, the variable domains are not bound by a rigid disulfide bridge, but are kept together more loosely at the center of rotation by conserved Q-Q H-bond interactions. Our sequence analysis showed, that the Q-Q interaction is highly

conserved, but in minor cases Q can be exchanged by other H-bond donors or acceptors or charged residues. Amino acids, which neither form H-bonds nor salt-bridges occur very seldom. In the latter case, the CoR possibly is shifted to other less conserved residues in this area, such as Y. Thus, the conserved residues at the CoR area keep the CoR at a defined position, but nevertheless the nature of the interactions permits flexibility that leads to the different orientations of the variable domains.

These observations are consistent with most other studies, discussing this topic, as a twofold hydrogen bond interaction between the Q-residues of the Vα and the Vβ domain was already reported for the first resolved TCR structure (1tcr, 2C) [26] and the involved residues are highly conserved for TCRs as well as for V_L/V_H domains of antibodies [59]. Similarly, for antibodies it was proposed that in contrast to the constant domains the absence of a disulfide bond between the two variable domains is evolutionary preferred to allow for their rearrangement [38]. However, there exists one publication in which it was claimed that for A6 TCRs neither hydrogen bonds nor salt bridges can be observed between Vα/Vβ [13] and the authors propose that the diversity in Vα/Vβ rearrangement might be a result of the slippery hydrophobic interactions between the two variable domains. This is not only in contrast to the above discussed results from literature, but also our data, since we also observed the above described Q-Q interactions for all A6 TCRs (*e.g.* distances between the opposing atoms in structure 2gj6: D:Q37:NE2-E:Q37:OE1 = 3.06 Å, E:Q37:NE2-D:Q37:OE1 = 3.14 Å). Therefore, these electrostatic interactions, which are a magnitude weaker than a covalent disulfide bond, are highly conserved and most likely function as a flexible constraint, which keeps the two variable domains of the TCR in a preferred position, but at the same time allows for the necessary flexibility for their rearrangement upon binding to a specific pMHC complex.

Our analysis shows that TCRs of the same type bound to different ligands are normally found in the same clusters, whereas a significant change in the association angle can be observed upon binding of the TCR to the pMHC complex. Thus the question arises about the consequences of this behavior for the signal transduction cascade. According to our results two statements can be made. First, there seems to be a locking step upon pMHC assembly, during which the TCR is locked into a TCR clonotype specific geometry. Second, as the differences in the Vα/Vβ interdomain angles between the same TCRs bound to different binding partners, are rather small, the locking motion can be expected to be important during the signal transduction, whereas the differences in the absolute association angles are either not that significant or, assuming a signal is induced by the domain adjustment, only minor changes might be necessary. This agrees with most previous observations [22, 64], which show *e.g.* for the A6 TCR that peptide ligands with different affinities induce only minor changes in the relative positions of the variable domains [14–16]. In our analysis all A6 TCRs feature a very similar orientation of the Vβ domains and the structures all associate in the same cluster.

Due to its comprehensiveness our analysis puts these individual results on a common basis and provides thus a general picture of how pMHC binding influences TCR structure and function. Since many peptides with varying immunogenicity presented to the same TCR type only induce *minor* angular differences (see results), we conclude that signaling does not directly depend on a *major* change in the Vα/Vβ interdomain angle. In contrary, according to our analysis already minor changes of the Vα/Vβ interdomain geometry might have a significant influence on the triggering of the signaling cascade. These observations agree with the computational results of Knapp *et al.* who performed large scale MD simulations of 172 peptides of known immunogenicity presented to the LC13 TCR [56]. In that study, several features between a set of more immunogenic and less immunogenic peptides were compared, such as the Vα/Vβ geometries and the orientation of the TCR towards the pMHC, the solvent accessible surface area, the binding affinities, hydrogen bond footprints, and structural root mean

square fluctuations (RMSF). The study confirms our results as the examined LC 13 TCR it was observed to adopt only a slightly more "open" binding site when recognizing more immunogenic peptides, which is consistent with the minor changes we see upon the binding of different ligands.

However, when discussing the topic of signal transduction, it needs to be stated that in our study we did not investigate positional changes of the constant domains. Such changes were observed in single studies and were postulated to have an influence on the minor changes in CD3 binding or activation [17]. On the other hand investigations of the relationships of constants domains of bound and unbound A6 TCRs showed no alteration [15]. However, possibly the conformational adjustment of the A6 constant domains might be very small. Different conformations of the Cα A-B loop dependent on the antigenic ligand were described and it was speculated that this alteration might induce the conformational changes of the CD3 molecule [69]. Possibly, the antigenic ligand induces first an adjustment of the variable domains, leading to a change of the relative positions of the constant domains and finally to the observed conformational change of the Cα A-B loop. This effect could either be achieved mechanically or by an alteration of the surrounding forces. For the JM22 TCR it was observed experimentally that the temperature factors of the constant TCR domains increase upon ligand binding [20]. This observation supports the idea of the Cα A-B loop becoming more flexible after other parts of the TCRs loose flexibility, as observed in this study through the locking motion upon pMHC binding. Regarding the structural analysis of the 2C TCR, this locking motion seems to be caused by interplay between the Vα/Vβ association angles and the bound-conformation of the MHC-binding CDR loops. MD simulations similar to the study of Knapp *et al.* [56] could be used to study these effects.

The structures of TCRs are generally similar to Fab-fragments of antibodies [14,26], whereby the AB V_L/V_H correspond to the TCR Vα/Vβ domains. In early and recent studies of AB structures a rearrangement of the V_L/V_H upon ligand binding was considered [25,27–39]. Computer-aided methods including MD simulations were carried out to investigate the changes of the elbow angle between the variable and the constant domains of antibodies [70–72]. TCRs feature a lower diversity in the variable loops 1 and 2 but a higher diversity in the CDR3 loop compared to ABs, resulting in a smaller diversity in the overall shape of the TCRs [14]. However, it was shown that ABs V_L/V_H association angles are generally incompatible to the angular space of TCRs binding to MHC molecules [25]. It remains interesting to apply our method on ABs investigating whether this molecule class also shares a CoR and if ABs of the same type adapt to similar association angles.

Conclusion

Since flexibility of the TCR Vα/Vβ interdomain association was considered for the first time in the end of the 90s of the last century [12], it took one decade to examine this phenomenon comparing the angular space of different TCR types due to the initial difficulties of obtaining experimental structures [22]. Now, immunologist can benefit from two independent new studies of the TCR interdomain association geometry by Dunbar *et al.* [25] and our present one. Both papers complement by focusing on different topics and methodologies. Whereas Dunbar *et al.* focus on the comparison between TCRs and antibodies, in this study we performed a systematic, exhaustive analysis of the Vα/Vβ interdomain angle for a representative set of experimental TCR structures. Our results are in agreement with the majority of previous experimental studies on small sets of TCRs. However, due to the comprehensiveness of our analysis we were able to put these individual observations on a broader, more general basis.

This allowed us to deduce general features describing the relationship between TCR interdomain angle variations and pMHC binding and signaling.

First, our data clearly shows that the Vα/Vβ interdomain angle of pMHC-bound TCR structures can vary considerably, but is in most cases well conserved within the same TCR clonotype and its variants, independent of the ligand (pMHC) bound and individual mutations within the TCR. Nevertheless, there are individual exceptions like the 2C TCR, which show larger variations in their angle repertoire. Analysis of the 2C TCR structures revealed correlations between the Vα/Vβ interdomain angle, specific framework mutations, and conformational changes in the MHC-binding Vα-CDR loops due to Vα-CDR3 mutations. This is in accordance with previous experimental studies on individual TCRs, indicating that the Vα/Vβ interdomain angle is mainly influenced by the bound MHC allele and not the peptide. Unfortunately, due to the currently still sparse structural data available, no generalizable conclusions can be drawn about the dynamic mechanisms behind such Vα/Vβ angle switches.

Second, through a systematic analysis of the structural basis for the observed angular deviations we could identify a central point of rotation (CoR) common to all TCR structures independent of their state (bound or unbound) and type, which is stabilized by electrostatic and hydrogen bonding interactions. As in all previous studies the Vα/Vβ interdomain angle was described by at least three geometric quantities, the identification of one CoR, which allows a simple yet intuitive description of this functionally important, variable angle, sheds new light on the structural features and also the functional dynamics of TCRs and will also be important for the improvement of existing and for future TCR modeling approaches.

Third, analyzing bound versus unbound TCR structures, we observed that the angle variations between bound and unbound structures are more significant than between TCR structures bound to different MHC-peptide complexes or even mutated TCR structures with different specificities. This suggests that binding of the TCR to the pMHC complex is accompanied by a dynamic lock mechanism during which the two TCR variable domains are driven into a TCR-specific binding geometry leading to a stabilization of the TCR variable domain upon pMHC binding. In a previous study it was found that with a rigidification of the variable region the constant region becomes more flexible [20]. Furthermore, the influence of constant domain shifts was considered to be involved in CD3 activation [17] as well as a conformational change in the A-B loop of the Cα domain [69]. Supported by these observations we propose that locking of the variable domains upon ligand binding might enhance the motions of the constant domains in this oscillating system. The change of motion of the constant domains could then induce the conformational changes, which lead to CD3 activation and thus initiate T-cell signaling.

Based on these results the TCR/pMHC binding mechanism can be envisioned as a two-state process: First, preformation of the general α/β domain geometry in the free state and second, locking of this angle in a specific geometry upon association with the MHC-peptide complex. The last step might be an important feature during signal transduction upon binding.

Methods

Data set

A set of 85 X-ray crystal structures was acquired from the Protein Data Bank (PDB) [73]. The used structures contain bound and unbound TCRs from *H. sapiens* and *Mus musculus*. Receptors of invariant natural killer cells (iNKT) were not considered in our analysis. Although the iNKT receptors share sequential and structural similarity with other αβ TCRs, these special TCRs do not recognize pMHC ligands, but detect lipids presented by the MHC like CD1d molecule [74,75]. Thus iNKT:CD1d complexes must be treated separately. For the analysis each

biological unit (BU) was treated as a separate structure, leading to a total amount of 136 different TCR complexes. For some analysis steps we computed averages over all BUs of the same crystal. This set is further referred to as *S*. The used structures are listed together with the names used in the literature and their bound state in Table 1. For some receptor types (*e.g.* A6, 1G4 etc.) several entries in the PDB are available. Furthermore, in some cases the names used in literature are ambiguous, since some of these structures bear mutations. Therefore, we compared the sequences of all structures against the wild type (wt), and we extended the TCR type names to indicate deviations. The sequence differences are presented in S1 Table and S2 Table in the supplementary material. This table also contains information about the appearing gene segments, differences in the CDR3 loops, and the bound ligands including the peptide sequence.

Cuboid-based superpositioning approach

In a first preparatory step, all TCR structures were reduced to their variable binding domains (V). The constant domains of the TCRs were not considered for two reasons. First, in some of the TCR structures data are only available for the variable domains but not for the constant domains. Second, the constant and the variable domains are connected by a flexible loop. Superimpositioning the complete TCR chains onto their variable domains showed a visible displacement of the constant domains. Thus, including the constant domain into our superpositioning template would influence the structural alignment of the Vα domains.

Cuboid construction and placement

Then we defined unified cuboids for each V domain of the different TCR chains. The cuboid templates (CTα and CTβ; Fig 1) comprise Vα or Vβ domain and a reduced set of the 2bnu Vα or Vβ domain framework residues. The reduced set was defined to allow for a robust superpositioning of the experimental structures during our future modeling procedure and during the geometrical measurements. For the superpositioning the tool DaliLite [76,77] was used together with the defined subset of V-framework residues as templates (Fig 1C). The Dali algorithm uses Cα-Cα distance matrices and does not depend on sequence information. Due to the high homology of the TCR framework regions, a solely structure based superpositioning method is indispensable for our needs.

To define the subset of residues contained in the superpositioning template, in a preparatory step the structures of each variable domain were superimposed separately. For this purpose all loops and turns were removed and the template residues were determined iteratively from the remaining residues, such that the set of mapped residues used as superpositioning anchors in DaliLite converged and the variance of the backbone root mean square deviation (RMSD) over all superposed structures was low. After the subsets were identified, the following procedure was used to superpose the combined variable domains and to place the cuboids. First, the Vα:Vβ-complexes were superimposed based on their Vα domains using the above defined α-subset and the tool DaliLite. All structures were superimposed to the high resolution (1.4 Å) structure with the PDB ID 2bnu. This step leads to a set of TCR-structures, which is further referred to as Sα and a corresponding set of cuboid templates (CTα), containing cuboids, which were placed around the Vα domains based on the positions of the α-subset residues. Second, the same procedure was used to place cuboid templates (CT, Fig 1C) additionally around each Vβ domain contained in the set Sα according to the relative position of the β chains towards their paired, superposed Vα domains resulting in a set of cuboid templates around the Vβ domains (CTβ). This step results in the set C consisting of the Vβ domain cuboids (CTβ) and the corresponding Vβ domain structures.

Comparison of the relative Vα/Vβ domain geometries

We computed the Euler angles for each cuboid-based geometry with respect to a reference coordinate system. The calculation was implemented using the GNU generic math template library; all angles were computed in xyz-order. The reference coordinate system was chosen to be the coordinate system of the 2bnu structure (Fig 1C). Since all structures of the set S were superimposed to conserved framework residues of the reference structure 2bnu, the rotation of the Vβ domain is computed relative to the orientation of the Vβ of 2bnu. The similarity between two geometries we defined as the Euclidean distance of the Euler angles [78]:

$$d_E(i,j) = \sqrt{(\Phi_i - \Phi_j)^2 + (\Psi_i - \Psi_j)^2 + (\Theta_i - \Theta_j)^2} \quad (1)$$

The distance matrix D was clustered hierarchically, using Ward's minimum variance method [57,79]. In the case of structures containing multiple BUs each Euler angle component Φ, Ψ, and θ was averaged, leading to an artificial unified geometry. These unified averaged geometries cluster together with other TCRs of the same type, describing a general geometrical state of a TCR. In contrast, the differing geometries found within one structure indicate, that TCRs may adopt an ensemble of different geometries. The significance of the found clusters was confirmed using a bootstrap analysis [58]. The number of bootstrap replica was set to 10^6.

Center of rotation and translation

The common center of rotation (CoR) was computed for all structures (BUs were treated independently), and subsequently the residues situated in the CoR-areas were analyzed. The conservation of these CoR-residues was confirmed by multiple sequence alignment of the TCR.

The CoR was determined for each Vβ domain of the set *C*. For this purpose, grids were fitted into the cuboids (as illustrated in Fig 1D) and the grid points were indexed in the same manner for each cuboid. For all grid points of the same index *i*, pairwise distances were computed. The variance

$$\mathrm{var}(g_i) = \frac{1}{n^2-1}\sum_{k=1}^{n}\sum_{l=1}^{n}\left(\delta_S(g_{i,k},g_{i,l}) - \frac{\sum_{r=1}^{n}\sum_{s=1}^{n}\delta_S(g_{i,r},g_{i,s})}{n^2}\right) \quad (2)$$

was determined; $\delta_S(g_{i,x}g_{i,y})$ is the Euclidean spatial distance between a grid point with the index *i* of the structure *x* and the equivalent grid point of structure *y*; *n* denotes the number of observed structures. We define the CoR grid index point as I_{min} = argmin(min{var(g_i)|$1 \le i \le n$}). The corresponding coordinate for $I_{\min}$ was computed according to the reference coordinate system of the reference structure 2bnu. The computations were performed on the whole set *C* and on subsets of this set. The first subset C^b contained only MHC bound TCRs, whereas the second subset C^u contained unbound TCRs. Furthermore, we investigated the structural environment of the location of the corresponding $I_{\min}$ coordinates. The grids were implemented in a cubic shape with an amount of 33,076,161 grid points and a minimum distance between each point of 0.1 Å, resulting in a grid size of 32 Å in each dimension. In contrast to the intersect method for the CoR calculation, the grid method allows highlighting invariant areas and is more robust against deviations caused by geometrical translation. Sequence similarity and conservation of amino acids located at the CoR was explored by creating multiple sequence alignments (MSAs) of all known human and murine (functional) TCR variable segment sequences. For this purpose, the tool MAFFT [80] (linsi: localpair, maxiter 1000, Blosum62 [81]) was applied on sequences obtained from the IMGT GENE-DB [61].

Supporting Information

S1 Text. Analysis of specific TCR:pMHC structures: Detailed discussion for the individual TCR types.
(PDF)

S1 Table. Properties of available TCR structures. For each X-ray structure TCR names used in the literature are listed and the relevant TCR chains within one biological unit are indicated. Loci and alleles of the TRAV, TRBV, TRAJ, and TRBJ were assigned using the IMGT Gene/DB [61] and the sequences of the CDR loops are listed for comparison of the different subtypes (Mutations within the framework region are summarized separately in S2 Table). For the bound TCRs the loci/allele of the MHC (-like) molecule and it's ligand is provided; mutations are indicated.
(PDF)

S2 Table. Mutations within the framework region of the TCR structures. The sequences of the TCR structures were compared to the corresponding wild type (WT). Differences to the alleles provided in the IMGT Gene-DB [61] are provided as mutation pairs in brackets. Naming of the sheets and loops, and the residue indices follow the IMGT unique numbering [62].
(PDF)

S3 Table. Multiple sequence alignments at the COR.
(PDF)

S4 Table. Epitopes of the bound TCR structures.
(PDF)

S5 Table. References to the TCR structures used for the analysis.
(PDF)

S1 Fig. Bootstrapping dendrogram of the clustering of the MHC bound TCRs. Pairwise Euler-angle distances (EAD) were determined for all MHC-bound TCR structures and the free TCR structures according to Formula 1. Structures containing more than one biological unit were merged to one unique geometry. The distance matrix was hierarchically clustered using the Ward update formula. For each subtree of the dendrogram, the au (approximately biased) and the bp (bootstrapping probability) according to the bootstrapping method [58] are provided. We identified six significant clusters of an au-value greater than 95%. The clusters are marked by colored boxes.
(PDF)

S2 Fig. Bootstrapping dendrogram of the clustering of the free TCRs together with the MHC bound TCRs. For details see S1 Fig. The bootstrapping dendrogram was computed for the bound and free TCRs. The clusters of the unbound case are marked by colored boxes for comparison. Significant clusters are only found for smaller subtrees.
(PDF)

S3 Fig. Geometry clusters of bound and unbound TCRs. Pairwise Euler-angle distances (EAD) were determined for all MHC-bound TCR structures and the free TCR structures according to Formula 2 (see Methods). Structures containing more than one biological unit were merged to one unique geometry. The distance matrix was hierarchically clustered using the Ward update formula. We identified six significant clusters, using a bootstrapping approach [58]. Notably, in most of the cases, TCRs of the same type occur in the same cluster. Upper panel: Clustering dendrogram with bootstrapping results (au = approximately unbiased,

bp = bootstrapping probability) Left panel: TCR types occurring within a cluster. Right/lower panel: Structure PDB identifiers and corresponding TCR names. MHC unbound TCR structures are indicated by bold-italics fonts. Central panel: Pairwise Euler-angle distances (EAD). The color key is provided in the bottom of the figure.
(PDF)

Acknowledgments

We like to thank Sabine Schweizer for very interesting discussions and Atanas Patronov for many helpful suggestions and proofreading the manuscript.

Author Contributions

Conceived and designed the experiments: TH AMK IA. Performed the experiments: TH. Analyzed the data: TH IA. Contributed reagents/materials/analysis tools: TH IA. Wrote the paper: TH AMK IA.

References

1. Choudhuri K, Kearney A, Bakker TR, van der Merwe PA. Immunology: how do T cells recognize antigen? Current biology: CB. 2005 May 24; 15(10):R382–5. PMID: 15916940.
2. van der Merwe PA, Dushek O. Mechanisms for T cell receptor triggering. Nature reviews Immunology. 2011 Jan; 11(1):47–55. PMID: 21127503. doi: 10.1038/nri2887
3. Chakraborty AK, Weiss A. Insights into the initiation of TCR signaling. Nat Immunol. 2014 Aug 19; 15 (9):798–807. PMID: 25137454. doi: 10.1038/ni.2940
4. Gil D, Schrum AG, Alarcon B, Palmer E. T cell receptor engagement by peptide-MHC ligands induces a conformational change in the CD3 complex of thymocytes. J Exp Med. 2005 Feb 21; 201(4):517–22. PMID: 15728235. Pubmed Central PMCID: 1868566.
5. Choudhuri K, van der Merwe PA. Molecular mechanisms involved in T cell receptor triggering. Seminars in immunology. 2007 Aug; 19(4):255–61. PMID: 17560121.
6. Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J, Kourilsky P. A direct estimate of the human alphabeta T cell receptor diversity. Science. 1999 Oct; 286(5441):958–61. PMID: 10542151
7. Casrouge A, Beaudoing E, Dalle S, Pannetier C, Kanellopoulos J, Kourilsky P. Size estimate of the alpha beta TCR repertoire of naive mouse splenocytes. J Immunol. 2000 Jun 1; 164(11):5782–7. PMID: 10820256.
8. Garcia KC, Degano M, Pease LR, Huang M, Peterson PA, Teyton L, et al. Structural basis of plasticity in T cell receptor recognition of a self peptide-MHC antigen. Science. 1998 Feb; 279(5354):1166–72. PMID: 9469799
9. Reiser JB, Grégoire C, Darnault C, Mosser T, Guimezanes A, Schmitt-Verhulst AM, et al. A T cell receptor CDR3beta loop undergoes conformational changes of unprecedented magnitude upon binding to a peptide/MHC class I complex. Immunity. 2002 Mar; 16(3):345–54. PMID: 11911820
10. Yu X- Z, Martin PJ, Anasetti C. CD28 signal enhances apoptosis of CD8 T cells after strong TCR ligation. J Immunol. 2003 Mar; 170(6):3002–6. PMID: 12626553
11. Gascoigne NRJ. T cell receptor structures: three for the price of one. Immunity. 2011 Jul; 35(1):1–3. doi: 10.1016/j.immuni.2011.07.001 PMID: 21777791
12. Li H, Lebedeva MI, Ward ES, Mariuzza RA. Dual conformations of a T cell receptor V alpha homodimer: implications for variability in V alpha V beta domain association. J Mol Biol. 1997 Jun 13; 269(3):385–94. PMID: 9199407.
13. Gagnon SJ, Borbulevych OY, Davis-Harrison RL, Turner RV, Damirjian M, Wojnarowicz A, et al. T cell receptor recognition via cooperative conformational plasticity. J Mol Biol. 2006 Oct; 363(1):228–43. PMID: 16962135
14. Garboczi DN, Ghosh P, Utz U, Fan QR, Biddison WE, Wiley DC. Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. Nature. 1996 Nov; 384(6605):134–41. PMID: 8906788
15. Ding YH, Smith KJ, Garboczi DN, Utz U, Biddison WE, Wiley DC. Two human T cell receptors bind in a similar diagonal mode to the HLA-A2/Tax peptide complex using different TCR amino acids. Immunity. 1998 Apr; 8(4):403–11. PMID: 9586631

16. Ding YH, Baker BM, Garboczi DN, Biddison WE, Wiley DC. Four A6-TCR/peptide/HLA-A2 structures that generate very different T cell signals are nearly identical. Immunity. 1999 Jul; 11(1):45–56. PMID: 10435578

17. Degano M, Garcia KC, Apostolopoulos V, Rudolph MG, Teyton L, Wilson IA. A functional hot spot for antigen recognition in a superagonist TCR/MHC complex. Immunity. 2000 Mar; 12(3):251–61. PMID: 10755612

18. Hennecke J, Carfi A, Wiley DC. Structure of a covalently stabilized complex of a human alphabeta T-cell receptor, influenza HA peptide and MHC class II molecule, HLA-DR1. EMBO J. 2000 Nov; 19 (21):5611–24. PMID: 11060013

19. Kjer-Nielsen L, Clements CS, Brooks AG, Purcell AW, McCluskey J, Rossjohn J. The 1.5 A crystal structure of a highly selected antiviral T cell receptor provides evidence for a structural basis of immunodominance. Structure. 2002 Nov; 10(11):1521–32. PMID: 12429093

20. Ishizuka J, Stewart-Jones GBE, van der Merwe A, Bell JI, McMichael AJ, Jones EY. The structural dynamics and energetics of an immunodominant T cell receptor are programmed by its Vbeta domain. Immunity. 2008 Feb; 28(2):171–82. doi: 10.1016/j.immuni.2007.12.018 PMID: 18275829

21. Archbold JK, Macdonald WA, Gras S, Ely LK, Miles JJ, Bell MJ, et al. Natural micropolymorphism in human leukocyte antigens provides a basis for genetic control of antigen recognition. J Exp Med. 2009 Jan; 206(1):209–19. doi: 10.1084/jem.20082136 PMID: 19139173

22. McBeth C, Seamons A, Pizarro JC, Fleishman SJ, Baker D, Kortemme T, et al. A new twist in TCR diversity revealed by a forbidden alphabeta TCR. J Mol Biol. 2008 Feb; 375(5):1306–19. PMID: 18155234

23. Gadola SD, Koch M, Marles-Wright J, Lissin NM, Shepherd D, Matulis G, et al. Structure and binding kinetics of three different human CD1d-alpha-galactosylceramide-specific T cell receptors. J Exp Med. 2006 Mar 20; 203(3):699–710. PMID: 16520393. Pubmed Central PMCID: 2118257.

24. Dunbar J, Fuchs A, Shi J, Deane CM. ABangle: characterising the VH-VL orientation in antibodies. Protein Eng Des Sel. 2013 Oct; 26(10):611–20. PMID: 23708320. doi: 10.1093/protein/gzt020

25. Dunbar J, Knapp B, Fuchs A, Shi J, Deane CM. Examining variable domain orientations in antigen receptors gives insight into TCR-like antibody design. PLoS computational biology. 2014 Sep; 10(9): e1003852. PMID: 25233457. Pubmed Central PMCID: 4168974. doi: 10.1371/journal.pcbi.1003852

26. Garcia KC, Degano M, Stanfield RL, Brunmark A, Jackson MR, Peterson PA, et al. An alphabeta T cell receptor structure at 2.5 A and its orientation in the TCR-MHC complex. Science. 1996 Oct; 274 (5285):209–19. PMID: 8824178

27. Ban N, Escobar C, Hasel KW, Day J, Greenwood A, McPherson A. Structure of an anti-idiotypic Fab against feline peritonitis virus-neutralizing antibody and a comparison with the complexed Fab. FASEB journal: official publication of the Federation of American Societies for Experimental Biology. 1995 Jan; 9(1):107–14. PMID: 7821749.

28. Bhat TN, Bentley GA, Boulot G, Greene MI, Tello D, Dall'Acqua W, et al. Bound water molecules and conformational stabilization help mediate an antigen-antibody association. Proc Natl Acad Sci U S A. 1994 Feb 1; 91(3):1089–93. PMID: 8302837. Pubmed Central PMCID: 521459.

29. Bhat TN, Bentley GA, Fischmann TO, Boulot G, Poljak RJ. Small rearrangements in structures of Fv and Fab fragments of antibody D1.3 on antigen binding. Nature. 1990 Oct 4; 347(6292):483–5. PMID: 2215663.

30. Colman PM, Laver WG, Varghese JN, Baker AT, Tulloch PA, Air GM, et al. Three-dimensional structure of a complex of antibody with influenza virus neuraminidase. Nature. 1987 Mar 26-Apr 1; 326 (6111):358–63. PMID: 2436051.

31. Herron JN, He XM, Ballard DW, Blier PR, Pace PE, Bothwell AL, et al. An autoantibody to single-stranded DNA: comparison of the three-dimensional structures of the unliganded Fab and a deoxynucleotide-Fab complex. Proteins. 1991; 11(3):159–75. PMID: 1749770.

32. Kroon GJ, Mo H, Martinez-Yamout MA, Dyson HJ, Wright PE. Changes in structure and dynamics of the Fv fragment of a catalytic antibody upon binding of inhibitor. Protein Sci. 2003 Jul; 12(7):1386–94. PMID: 12824485. Pubmed Central PMCID: 2323930.

33. Lascombe MB, Alzari PM, Boulot G, Saludjian P, Tougard P, Berek C, et al. Three-dimensional structure of Fab R19.9, a monoclonal murine antibody specific for the p-azobenzenearsonate group. Proc Natl Acad Sci U S A. 1989 Jan; 86(2):607–11. PMID: 2911596. Pubmed Central PMCID: 286521.

34. Li Y, Parry G, Chen L, Callahan JA, Shaw DE, Meehan EJ, et al. An anti-urokinase plasminogen activator receptor (uPAR) antibody: crystal structure and binding epitope. J Mol Biol. 2007 Jan 26; 365 (4):1117–29. PMID: 17101149.

35. Monaco-Malbet S, Berthet-Colominas C, Novelli A, Battai N, Piga N, Cheynet V, et al. Mutual conformational adaptations in antigen and antibody upon complex formation between an Fab and HIV-1 capsid protein p24. Structure. 2000 Oct 15; 8(10):1069–77. PMID: 11080628.

36. Rini JM, Schulze-Gahmen U, Wilson IA. Structural evidence for induced fit as a mechanism for antibody-antigen recognition. Science. 1992 Feb 21; 255(5047):959–65. PMID: 1546293.

37. Stanfield RL, Takimoto-Kamimura M, Rini JM, Profy AT, Wilson IA. Major antigen-induced domain rearrangements in an antibody. Structure. 1993 Oct 15; 1(2):83–93. PMID: 8069628.

38. Stevens FJ, Chang CH, Schiffer M. Dual conformations of an immunoglobulin light-chain dimer: heterogeneity of antigen specificity and idiotope profile may result from multiple variable-domain interaction mechanisms. Proc Natl Acad Sci U S A. 1988 Sep; 85(18):6895–9. PMID: 3137576. Pubmed Central PMCID: 282085.

39. Teplyakov A, Obmolova G, Malia T, Gilliland G. Antigen recognition by antibody C836 through adjustment of V(L)/V(H) packing. Acta crystallographica Section F, Structural biology and crystallization communications. 2011 Oct 1; 67(Pt 10):1165–7. PMID: 22102019. Pubmed Central PMCID: 3212354.

40. Morgan RA, Dudley ME, Wunderlich JR, Hughes MS, Yang JC, Sherry RM, et al. Cancer regression in patients after transfer of genetically engineered lymphocytes. Science. 2006 Oct 6; 314(5796):126–9. PMID: 16946036. Epub 2006/09/02. eng.

41. Bendle GM, Linnemann C, Hooijkaas AI, Bies L, de Witte MA, Jorritsma A, et al. Lethal graft-versus-host disease in mouse models of T cell receptor gene therapy. Nat Med. 2010 May; 16(5):565–70, 1p following 70. doi: 10.1038/nm.2128 PMID: 20400962

42. Michielin O, Luescher I, Karplus M. Modeling of the TCR-MHC-peptide complex. J Mol Biol. 2000 Jul 28; 300(5):1205–35. PMID: 10903865.

43. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol. 1993 Dec 5; 234(3):779–815. PMID: 8254673.

44. Cuendet MA, Michielin O. Protein-protein interaction investigated by steered molecular dynamics: the TCR-pMHC complex. Biophysical journal. 2008 Oct; 95(8):3575–90. PMID: 18621828. Pubmed Central PMCID: 2553100. doi: 10.1529/biophysj.108.131383

45. Cuendet MA, Zoete V, Michielin O. How T cell receptors interact with peptide-MHCs: a multiple steered molecular dynamics study. Proteins. 2011 Nov; 79(11):3007–24. PMID: 21989928. doi: 10.1002/prot.23104

46. De Rosa MC, Giardina B, Bianchi C, Carelli Alinovi C, Pirolli D, Ferraccioli G, et al. Modeling the ternary complex TCR-Vbeta/CollagenII(261–273)/HLA-DR4 associated with rheumatoid arthritis. PloS one. 2010; 5(7):e11550. PMID: 20644721. Pubmed Central PMCID: 2904365. doi: 10.1371/journal.pone.0011550

47. Michielin O, Karplus M. Binding free energy differences in a TCR-peptide-MHC complex induced by a peptide mutation: a simulation analysis. J Mol Biol. 2002 Nov 29; 324(3):547–69. PMID: 12445788.

48. Rognan D, Stryhn A, Fugger L, Lyngbaek S, Engberg J, Andersen PS, et al. Modeling the interactions of a peptide-major histocompatibility class I ligand with its receptors. I. Recognition by two alpha beta T cell receptors. Journal of computer-aided molecular design. 2000 Jan; 14(1):53–69. PMID: 10702925.

49. Stavrakoudis A. Insights into the structure of the LC13 TCR/HLA-B8-EBV peptide complex with molecular dynamics simulations. Cell biochemistry and biophysics. 2011 Jul; 60(3):283–95. PMID: 21253892. doi: 10.1007/s12013-011-9151-2

50. Wan S, Flower DR, Coveney PV. Toward an atomistic understanding of the immune synapse: large-scale molecular dynamics simulation of a membrane-embedded TCR-pMHC-CD4 complex. Molecular immunology. 2008 Mar; 45(5):1221–30. PMID: 17980430.

51. Zoete V, Irving MB, Michielin O. MM-GBSA binding free energy decomposition and T cell receptor engineering. Journal of molecular recognition: JMR. 2010 Mar-Apr; 23(2):142–52. PMID: 20151417. doi: 10.1002/jmr.1005

52. Zoete V, Michielin O. Comparison between computational alanine scanning and per-residue binding free energy decomposition for protein-protein association using MM-GBSA: application to the TCR-p-MHC complex. Proteins. 2007 Jun 1; 67(4):1026–47. PMID: 17377991.

53. Wan S, Coveney PV, Flower DR. Molecular basis of peptide recognition by the TCR: affinity differences calculated using large scale computing. J Immunol. 2005 Aug 1; 175(3):1715–23. PMID: 16034112.

54. Knapp B, Dorffner G, Schreiner W. Early relaxation dynamics in the LC 13 T cell receptor in reaction to 172 altered peptide ligands: a molecular dynamics simulation study. PloS one. 2013; 8(6):e64464. PMID: 23762240. Pubmed Central PMCID: 3675092.

55. Leimgruber A, Ferber M, Irving M, Hussain-Kahn H, Wieckowski S, Derre L, et al. TCRep 3D: an automated in silico approach to study the structural properties of TCR repertoires. PLoS One. 2011; 6(10):e26301. PMID: 22053188. Pubmed Central PMCID: 3203878. doi: 10.1371/journal.pone.0026301

56. Knapp B, Dunbar J, Deane CM. Large Scale Characterization of the LC13 TCR and HLA-B8 Structural Landscape in Reaction to 172 Altered Peptide Ligands: A Molecular Dynamics Simulation Study. PLoS computational biology. 2014 Aug; 10(8):e1003748. PMID: 25101830. Pubmed Central PMCID: 4125040. doi: 10.1371/journal.pcbi.1003748

57. Ward JH Jr. Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association. 1963; 58(301):pp. 236–44.

58. Suzuki R, Shimodaira H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics. 2006 Jun; 22(12):1540–2. PMID: 16595560

59. Chothia C, Boswell DR, Lesk AM. The outline structure of the T-cell alpha beta receptor. EMBO J. 1988 Dec; 7(12):3745–55. PMID: 3208747

60. Calarese DA, Scanlan CN, Zwick MB, Deechongkit S, Mimura Y, Kunert R, et al. Antibody domain exchange is an immunological solution to carbohydrate cluster recognition. Science. 2003 Jun 27; 300 (5628):2065–71. PMID: 12829775. Epub 2003/06/28. eng.

61. Giudicelli V, Chaume D, Lefranc M-P. IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. Nucleic Acids Res. 2005 Jan; 33(Database issue): D256–D61. PMID: 15608191

62. Lefranc MP, Pommie C, Ruiz M, Giudicelli V, Foulquier E, Truong L, et al. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. Developmental and comparative immunology. 2003 Jan; 27(1):55–77. PMID: 12477501. Epub 2002/12/13. eng.

63. Colf LA, Bankovich AJ, Hanick NA, Bowerman NA, Jones LL, Kranz DM, et al. How a single T cell receptor recognizes both self and foreign MHC. Cell. 2007 Apr; 129(1):135–46. PMID: 17418792

64. Jones LL, Colf LA, Stone JD, Garcia KC, Kranz DM. Distinct CDR3 conformations in TCRs determine the level of cross-reactivity for diverse antigens, but not the docking orientation. J Immunol. 2008 Nov; 181(9):6255–64. PMID: 18941216

65. Rudolph MG, Stanfield RL, Wilson IA. How TCRs bind MHCs, peptides, and coreceptors. Annual review of immunology. 2006; 24:419–66. PMID: 16551255.

66. Chen J-L, Stewart-Jones G, Bossi G, Lissin NM, Wooldridge L, Choi EML, et al. Structural and kinetic basis for heightened immunogenicity of T cell vaccines. J Exp Med. 2005 Apr; 201(8):1243–55. PMID: 15837811

67. Sami M, Rizkallah PJ, Dunn S, Molloy P, Moysey R, Vuidepot A, et al. Crystal structures of high affinity human T-cell receptors bound to peptide major histocompatibility complex reveal native diagonal binding geometry. Protein Eng Des Sel. 2007 Aug; 20(8):397–403. PMID: 17644531

68. Dunn SM, Rizkallah PJ, Baston E, Mahon T, Cameron B, Moysey R, et al. Directed evolution of human T cell receptor CDR2 residues by phage display dramatically enhances affinity for cognate peptide-MHC without increasing apparent cross-reactivity. Protein Sci. 2006 Apr; 15(4):710–21. PMID: 16600963

69. Beddoe T, Chen Z, Clements CS, Ely LK, Bushell SR, Vivian JP, et al. Antigen ligation triggers a conformational change within the constant domain of the alphabeta T cell receptor. Immunity. 2009 Jun 19; 30(6):777–88. PMID: 19464197. doi: 10.1016/j.immuni.2009.03.018

70. Sotriffer CA, Liedl KR, Linthicum DS, Rode BM, Varga JM. Ligand-induced domain movement in an antibody Fab: molecular dynamics studies confirm the unique domain movement observed experimentally for Fab NC6.8 upon complexation and reveal its segmental flexibility. J Mol Biol. 1998 May 1; 278 (2):301–6. PMID: 9571052.

71. Sotriffer CA, Rode BM, Varga JM, Liedl KR. Elbow flexibility and ligand-induced domain rearrangements in antibody Fab NC6.8: large effects of a small hapten. Biophysical journal. 2000 Aug; 79 (2):614–28. PMID: 10919996. Pubmed Central PMCID: 1300962.

72. Stanfield RL, Zemla A, Wilson IA, Rupp B. Antibody elbow angles are influenced by their light chain class. J Mol Biol. 2006 Apr 14; 357(5):1566–74. PMID: 16497332.

73. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000 Jan; 28(1):235–42. PMID: 10592235

74. Porcelli S, Brenner MB, Greenstein JL, Balk SP, Terhorst C, Bleicher PA. Recognition of cluster of differentiation 1 antigens by human CD4-CD8-cytolytic T lymphocytes. Nature. 1989 Oct; 341(6241):447–50. PMID: 2477705

75. Beckman EM, Porcelli SA, Morita CT, Behar SM, Furlong ST, Brenner MB. Recognition of a lipid antigen by CD1-restricted alpha beta+ T cells. Nature. 1994 Dec; 372(6507):691–4. PMID: 7527500

76. Hasegawa H, Holm L. Advances and pitfalls of protein structural alignment. Curr Opin Struct Biol. 2009 Jun; 19(3):341–8. doi: 10.1016/j.sbi.2009.04.003 PMID: 19481444

77. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. J Mol Biol. 1993 Sep 5; 233(1):123–38. PMID: 8377180. Epub 1993/09/05. eng.

78. Huynh DQ. Metrics for 3D Rotations: Comparison and Analysis. J Math Imaging Vis. 2009 October; 35 (2):155–64.

79. Sokal RR, Michener CD. A statistical method for evaluating systematic relationships. University of Kansas Scientific Bulletin. 1958; 28:1409–38.

80. Katoh K, Misawa K, ichi Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002 Jul; 30(14):3059–66. PMID: 12136088

81. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A. 1992 Nov; 89(22):10915–9. PMID: 1438297

B.3 Supporting Information

S1 Text. Analysis of specific TCR:pMHC structures: Detailed discussion for the individual TCR types

In addition to the comparative cluster analysis we performed a more detailed functional and structural analysis of the clustering results. For several of the analysed TCR types there exist different structures in which the TCR is either bound to different MHC alleles and/or different peptides or different variants of the same TCR were crystallized. Analyzing the clustering behavior of these different structures of the same TCR types and their variants we observed that all except one (2C) are located in the same clusters.

Thus, we investigated the 2C TCR geometries in more detail. A closer look at the alloreactive murine 2C receptor shows that it can be observed in two different clusters (Figure 2: orange cluster (4) and cyan cluster (6)), whereas for all other TCR types all structures are members of the same cluster. The repertoire of the resolved 2C TCR structures is quite large in comparison to the other TCRs and contains different mutants of this receptor as well as different ligands (1tcr, 1g6r, 1mwa, 2ckb 2icw, 2oi9, 3e3q, 2e7l, 3e2h) [1-7] In Table 3 all pairwise Euler angle distances are provided together with the receptor's subtype and its ligand. We observe clustering in two significantly different geometries. The 2C structure repertoire contains wild type (2C wt) TCRs (1tcr, 1g6r, 1mwa, 2ckb), as well as four different variants of the 2C T7 TCR (2oi9, 3e3q, 2e7l, 3e2h). Next to several MHC-bound structures the 2C wt TCR is also available in its free state (1tcr).

In this context we would like to mention that the 2C T7 TCR crystal structure 2icw [7], in which the receptor is bound to a superantigen in association with a human MHC class II molecule, was not considered in this analysis, because in this structure direct TCR:MHC interaction is prevented by the superantigen which due to its size bridges all interactions. Therefore, this structure has a special binding mode, which is different from all other structures in our analysis and was thus discarded. The studied 2C T7 TCR variants (m13, m16, and m67) share the same mutations in the αC', αE, βB, and βC' strands (Table S2) with respect to the 2C wt TCR sequence as the original T7 variant, but additionally differ in their sequences of the Vα-CDR3 loop (Table S1), whereas no such differences exist in the original T7 variant. The MHC-bound 2C wt structures are crystallized together with different murine MHC class I molecules (H2-K^b (2ckb) or it's mutant form H2-K^{bm3} (1mwa)) and different bound peptides (EQYKFYSV (1mwa, 2ckb) or SIYRYYGL (1g6r)). The MHC-bound T7 (2oi9, 3e3q, 2e7l, 3e2h) structures are all associated with the same murine MHC class I molecule H2-L^d and the peptide QLSPFPFDL. Regarding Figure 2 it can be observed that all 2C wt TCR (2ckb, 1mwa, 1g6r) structures are members of cluster 6, whereas the T7 structures of the variants m6, m13, and T7-wt (3e3q, 2e7l, 2oi9) associate in cluster 4. However, the T7 structure of the m67 (3e2h) variant is member of cluster 6 along with the 2C wt TCRs, *i.e.* the m67 variant structure has the same Vα/Vβ association angle as the 2C wt TCR and not as the T7 TCRs. To elucidate the structural reason behind the structural differences within the 2C TCR structures a thorough structural and sequence analysis of all 2C structures was performed:

First, to investigate the potential influence of the different pMHC complexes on the structural differences in the TCRs, we analyzed the pMHC complexes to which the 2C wt and T7 structures are bound (Table 3). It can be observed that all 2C wt TCRs are bound to different MHC alleles as well as different peptide molecules, however, these do not influence the TCR geometry. In contrast, all T7 structures are bound to

the same pMHC complex and feature the same Vα/Vβ angle, except the m67 variant, which adopts the 2C wt geometry although the 2c wt TCRs are bound to a different pMHC complex. Thus the data set does not allow any judgment if the observed differences are caused by different pMHC alleles or rather by the differences within the TCR variants.

Second, analyzing the structures in more detail it can first be observed that the CDR loops of the Vβ domains adopt similar conformations for all 2C and T7 structures and show the same interaction geometry with the peptide and the MHC molecule. This is in contrast to the CDR1 and CDR3 loop conformations of the Vα domains, which differ between the 2C and T7 structures (Figure 4B), leading to different pMHC binding geometries and thus differences in the overall Vα/Vβ angle. Considering that the 2C wt and T7 structures only differ sequence-wise in the framework region, mutations in this region seem to have a direct influence on the TCRs CDR conformations.

Third, in the cases of the T7 variants, all variants are only mutated in their CDR3 loops and for the m6 and m13 variants, the mutations in their Vα-CDR3 loop only alter the conformation of this loop, but leave the Vα-CDR1 loop in its original T7 conformation. In contrary in the m67 variant, the conformational changes in its Vα-CDR3 loop lead to structural changes in its Vα-CDR1 loop, which adopts the conformation of the 2C wt structure (Figure 4B). As the m67 variant is the only T7 variant, which adopts the 2C wt Vα/Vβ angle, the CDR1 conformation seems to play a crucial role for the Vα/Vβ association angle, whereas the CDR3 conformation influences the angle only indirectly. Therefore, two effects could be observed within the available 2C TCR structural ensemble: Differences in the TCR framework sequences can lead to different Vα/Vβ association angles which can cause different CDR loop conformations and altered pMHC binding. On the other hand, differences in the sequences of the CDR loops, which lead to changes in the MHC binding CDRs (1 and 2) can back-lead to differences in the Vα/Vβ association angles.

A second TCR type, 1G4, for which various structures are available, contains several subtypes (Table S1 and S2), which differ in the αCDR1, αCDR3, βCDR1, and the βCDR3 loops. In contrast to the 2C variants above, for the 1G4 TCR variants cluster affiliations do not differ. Thus, in this case, all TCR structures (2bnu, 2bnq, 2bnr, 2f54, 2pyf, 2pye, 2f53, 2p5w, 2p5e), which are all bound to HLA-A*0201, are accumulated in cluster 1. In all structures the receptors are bound to the peptide SLMWITQC, except in the wild type (wt) structure 2bnq, which contains the peptide SLMWITQV.

In the case of the A6 TCR all nine structures (2gj6, 1qsf, 1qse, 3d3v, 3d39, 1qrn, 1ao7, 3h9s, 3pwp) can be found in the same cluster (cluster 2). The averaged angular distance between the nine structures is 2.1° (var=0.8°). All these structures share exactly the same TCR bound to HLA-A*0201, but differ in the bound peptide. The two structures 3d39 and 3d3v both contain (double)fluorinated derivatives of the peptide mutant LLFGFPVYV to achieve an increased affinity (y5f^{4F} or y5f^{3,4FF}) [8]. Furthermore, in the structure 2gj6 [9] the peptide LLFGKFVYV, where the K-residue is linked to 4-(3-indolyl)-butric acid, is presented.

The three BM3.3 bound structures differ in average by 3.1° and can be found in cluster 5. The structures mainly differ by the presented peptides (SQYYYNSL, INFDFNTI, and RGYVYQGL) presented by the MHC molecules H2-K1^{b} or H2-K1^{bm8}. Similar to the case of the A6 TCRs; the different peptides do not significantly alter the TCR geometry.

For the JM22 TCR four different structures (2vlj, 2vlk, 2vlr, and 1oga) with pMHC ligands are available. All of them share the same ligand, which is the peptide GILGFVFTL presented by the MHC molecule HLA-A*0201. The sequences of the TCRs of the four structures do not differ, except for 2vlr, which contains the mutation S99A in the αCDR3-loop. The average angular distance between all 4 structures is low (1.5º) and as expected they associate all in the same cluster (cluster 3).
A set of AHIII12.2 (wt) TCRs (2jcc, 2uwe, and 1lp9), can be found in cluster 6. To the three TCRs the peptide ALWGFFPVL is presented by the MHC molecule HLA-A*0201. In two cases a single amino acid is mutated within the binding pockets of the MHC molecules (2jcc: W176V, 2uwe: T163A) [10]. However, the differences in the MHC binding pockets obviously do not significantly alter the TCR geometry (average angular distance: 1.3º)
Three different structures of the TK3 TCR are available, the wt structure (3mv7), the mutant Q55H (3mv8), and the mutant Q55A (3mv9), all bound to a HLA-B*3501:HPVGEADYFEY complex. The mutations of the two variants are located in the αC'-sheet of the framework region. This allelic polymorphism (TRBV9*01 vs. TRBV9*02) influences the charge complementarity at the binding interface to the pMHC ligand [11]. Although these framework mutations alter the activity of the TCR, the exchange of Q55 to H or A does not influence the TCR interdomain geometry: all TK3 structures associate in cluster 6 (average angular distance: 0.7º).
The two different TCRs 2B4 (3qib) and 226 (3qiu, 3qiw) can both be found in cluster 4. Both TCRs bind to the murine MHC class II molecule H2-Ea^{k}:H2-Eb^{k} either presenting the peptide ADLIAYLKQATKG (3qib, 3qiu) or ADLIAYLEQATKG (3qiw). The 2B4 TCR facilitates the alleles TRAV4D-4*02, TRAJ56*01, TRBV26*01, and TRBJ2-5*01, whereas the 226 TCR uses the alleles TRAV4D-4*01, TRAJ16*01, TRBV26*01, and TRBJ1-2*01 (Table S1). Thus, the different TCRs have a similar framework region in common and share the CDR1/2 loops of both chains. Even though, both receptors differ in their CDR3 loops and the J-segments and an averaged angular distance of only 1.5º can be measured.
More TCRs of the same type in the same cluster are listed in Table S4, but are not further discussed, since for each of them only two representatives are available (E8 (cl. 1), SB27 (cl. 3), OB1A12 (cl. 3), and HA1.7 (cl. 5)).

SI References:

1. Colf LA, Bankovich AJ, Hanick NA, Bowerman NA, Jones LL, et al. (2007) How a single T cell receptor recognizes both self and foreign MHC. Cell 129: 135-146.
2. Degano M, Garcia KC, Apostolopoulos V, Rudolph MG, Teyton L, et al. (2000) A functional hot spot for antigen recognition in a superagonist TCR/MHC complex. Immunity 12: 251-261.
3. Garcia KC, Degano M, Pease LR, Huang M, Peterson PA, et al. (1998) Structural basis of plasticity in T cell receptor recognition of a self peptide-MHC antigen. Science 279: 1166-1172.
4. Garcia KC, Degano M, Stanfield RL, Brunmark A, Jackson MR, et al. (1996) An alphabeta T cell receptor structure at 2.5 A and its orientation in the TCR-MHC complex. Science 274: 209-219.
5. Jones LL, Colf LA, Stone JD, Garcia KC, Kranz DM (2008) Distinct CDR3 conformations in TCRs determine the level of cross-reactivity for diverse antigens, but not the docking orientation. J Immunol 181: 6255-6264.

6. Luz JG, Huang M, Garcia KC, Rudolph MG, Apostolopoulos V, et al. (2002) Structural comparison of allogeneic and syngeneic T cell receptor-peptide-major histocompatibility complex complexes: a buried alloreactive mutation subtly alters peptide presentation substantially increasing V(beta) Interactions. J Exp Med 195: 1175-1186.
7. Wang L, Zhao Y, Li Z, Guo Y, Jones LL, et al. (2007) Crystal structure of a complete ternary complex of TCR, superantigen and peptide-MHC. Nat Struct Mol Biol 14: 169-171.
8. Piepenbrink KH, Borbulevych OY, Sommese RF, Clemens J, Armstrong KM, et al. (2009) Fluorine substitutions in an antigenic peptide selectively modulate T-cell receptor binding in a minimally perturbing manner. Biochem J 423: 353-361.
9. Gagnon SJ, Borbulevych OY, Davis-Harrison RL, Turner RV, Damirjian M, et al. (2006) T cell receptor recognition via cooperative conformational plasticity. J Mol Biol 363: 228-243.
10. Miller PJ, Pazy Y, Conti B, Riddle D, Appella E, et al. (2007) Single MHC mutation eliminates enthalpy associated with T cell receptor binding. J Mol Biol 373: 315-327.
11. Gras S, Chen Z, Miles JJ, Liu YC, Bell MJ, et al. (2010) Allelic polymorphism in the T cell receptor and its impact on immune responses. J Exp Med 207: 1555-1567.

S1 Table. Properties of available TCR structures. For each X-ray structure TCR names used in the literature are listed and the relevant TCR chains within one biological unit are indicated. Loci and alleles of the TRAV, TRBV, TRAJ, and TRBJ were assigned using the IMGT Gene/DB [1] and the sequences of the CDR loops are listed for comparison of the different subtypes (Mutations within the framework region are summarized separately in Table S2). For the bound TCRs the loci/allele of the MHC (-like) molecule and it's ligand is provided; mutations are indicated.

Name	Bound State[a]	PDB-ID	Chain Pairs	Spec.	Literature PMID	IMGT-Name		Alpha			Beta			CDR-Loops[b]			Pep/Lip[c]	Ligand	
						V-Segment	J-Segment	1	2	3	V-Segment	J-Segment	1	2	3		MHC I/II alpha	MHC II beta
1G4	u	2BNU	AB	H	15837811	TRAV21*01	TRAJ6*01	DSA.....IYN	IQSS...QRE	AVRPTSGGSYIPT	TRBV6-5*01	TRBJ2-2*01	NNH.....EY	SVG....AGI	ASSYVG.NTGELF			
1G4	1	2BNQ	DE	H	15837811	TRAV21*01	TRAJ6*01	DSA.....IYN	IQSS...QRE	AVRPTSGGSYIPT	TRBV6-5*01	TRBJ2-2*01	NNH.....EY	SVG....AGI	ASSYVG.NTGELF	SLLMWITQV	HLA-A*0201	
1G4	1	2BNR	DE	H	15837811	TRAV21*01	TRAJ6*01	DSA.....IYN	IQSS...QRE	AVRPTSGGSYIPT	TRBV6-5*01	TRBJ2-2*01	NNH.....EY	SVG....AGI	ASSYVG.NTGELF	SLLMWITQC	HLA-A*0201	
1G4 AV-wt	1	2F54	DE:KL	H	16600963	TRAV21*01	TRAJ6*01	DSA.....IYN	IQSS...QRE	AVRPTSGGSYIPT	TRBV6-5*01	TRBJ2-2*01	NNH.....EY	SVG....AGI	ASSYVG.NTGELF	SLLMWITQC	HLA-A*0201	
1G4 c5c1	u	2PYF	AB	H	17644531	TRAV21*01	TRAJ6*01	DSA.....IYN	IQSS...QRE	AVRPTLDGTYIPT	TRBV6-5*01	TRBJ2-2*01	NNH.....EY	SVG....AGT	ASSYLG.NTGELF			
1G4 c5c1	1	2PYE	DE	H	17644531	TRAV21*01	TRAJ6*01	DSA.....IYN	IQSS...QRE	AVRPTLDGTYIPT	TRBV6-5*01	TRBJ2-2*01	NNH.....EY	SVG....AGT	ASSYLG.NTGELF	SLLMWITQC	HLA-A*0201	
1G4 c49c50	1	2F53	DE	H	16600963	TRAV21*01	TRAJ6*01	DSA.....IYN	IPEN...QRE	AVRPTSGGSYIPT	TRBV6-5*01	TRBJ2-2*01	NNH.....EY	SVG....YGH	ASSYVG.NTGELF	SLLMWITQC	HLA-A*0201	
1G4 c58c61	1	2P5W	DE	H	17644531	TRAV21*01	TRAJ6*01	DSA.....IYN	IPEN...QRE	AVRPTLDGTYIPT	TRBV6-5*01	TRBJ2-2*01	NNH.....EY	SVG....YGH	ASSYLG.NTGELF	SLLMWITQC	HLA-A*0201	
1G4 c58c61	1	2P5E	DE	H	17644531	TRAV21*01	TRAJ6*01	DSA.....IYN	IPEN...QRE	AVRPTLDGTYIPT	TRBV6-5*01	TRBJ2-2*01	NNH.....EY	SVA....IGT	ASSYLG.NTGELF	SLLMWITQC	HLA-A*0201	
3A6	2	1ZGL	ST.MP.QR.UV	H	16079912	TRAV9-2*02	TRAJ24*01	ATG.....YPS	ATKA...DDK	ALSGGD.SSYKLI	TRBV5-1*01	TRBJ1-1*01	SGH.....RS	YFN....ETQ	ASSLADRVNTEAF	VHFFKNIVTPRTP	HLA-DRA*0101	HLA-DRB5*0101
A6	1	2GJ6	DE	H	16962135	TRAV12-2*02	TRAJ24*02	DRG.....SQS	IYS....NGD	AVTTDS..WGKLQ	TRBV6-5*01	TRBJ2-7*01	NNH.....EY	SVG....AGI	ASRPGLAGGRPEQY	LLFGKPVYV	HLA-A*0201	
A6	1	1QSF	DE	H	10435578	TRAV12-2*02	TRAJ24*02	DRG.....SQS	IYS....NGD	AVTTDS..WGKLQ	TRBV6-5*01	TRBJ2-7*01	NNH.....EY	SVG....AGI	ASRPGLAGGRPEQY	LLFGYPVAV	HLA-A*0201*	
A6	1	1QSE	DE	H	10435578	TRAV12-2*02	TRAJ24*02	DRG.....SQS	IYS....NGD	AVTTDS..WGKLQ	TRBV6-5*01	TRBJ2-7*01	NNH.....EY	SVG....AGI	ASRPGLAGGRPEQY	LLFGYPRYV	HLA-A*0201	
A6	1	3D3V	DE	H	19698083	TRAV12-2*02	TRAJ24*02	DRG.....SQS	IYS....NGD	AVTTDS..WGKLQ	TRBV6-5*01	TRBJ2-7*01	NNH.....EY	SVG....AGI	ASRPGLAGGRPEQY	LLFGFPVYV	HLA-A*0201	
A6	1	3D39	DE	H	19698083	TRAV12-2*02	TRAJ24*02	DRG.....SQS	IYS....NGD	AVTTDS..WGKLQ	TRBV6-5*01	TRBJ2-7*01	NNH.....EY	SVG....AGI	ASRPGLAGGRPEQY	LLFGFPVYV	HLA-A*0201	
A6	1	1QRN	DE	H	10435578	TRAV12-2*02	TRAJ24*02	DRG.....SQS	IYS....NGD	AVTTDS..WGKLQ	TRBV6-5*01	TRBJ2-7*01	NNH.....EY	SVG....AGI	ASRPGLAGGRPEQY	LLFGYAVYV	HLA-A*0201	
A6	1	1AO7	DE	H	8906788	TRAV12-2*02	TRAJ24*02	DRG.....SQS	IYS....NGD	AVTTDS..WGKLQ	TRBV6-5*01	TRBJ2-7*01	NNH.....EY	SVG....AGI	ASRPGLAGGRPEQY	LLFGYPVYV	HLA-A*0201	
A6	1	3H9S	DE	H	20064447	TRAV12-2*02	TRAJ24*02	DRG.....SQS	IYS....NGD	AVTTDS..WGKLQ	TRBV6-5*01	TRBJ2-7*01	NNH.....EY	SVG....AGI	ASRPGLAGGRPEQY	MLMGYLQYV	HLA-A*0201	
A6	1	3PWP	DE	H	21282516	TRAV12-2*02	TRAJ24*02	DRG.....SQS	IYS....NGD	AVTTDS..WGKLQ	TRBV6-5*01	TRBJ2-7*01	NNH.....EY	SVG....AGI	ASRPGLAGGRPEQY	LGYGFVNYI	HLA-A*0201	
AS01	1	3O4L	DE	H	21124993	TRAV5*01	TRAJ31*01	DSS.....STY	IFSN...MDM	AEDNN....ARIM	TRBV20-1*01	TRBJ1-2*01	DFQ.....ATT	SNEG...SKA	SARDGT..GNGYT	GLCTLVAML	HLA-A*0201	
B7	1	1BD2	DE	H	9586631	TRAV29*01	TRAJ54*01	NSM.....FDY	ISSI...KDK	AAMEG...AQKLV	TRBV6-5*01	TRBJ2-7*01	NNH.....EY	SVG....AGI	ASSYPGGFYEQY	LLFGYPVYV	HLA-A*0201	
c134	1	3FFC	DE:KJ	H	19167249	TRAV14*01	TRAJ49*01	MSDP.....SYG	QGSY..DQGN	AMREDT..GNQFY	TRBV11-2*01	TRBJ2-3*01	SGH.....AT	FQN....NKV	ASSFTWTSGGATD	FLRGRAYGL	HLA-B*0801	
DM1	1	3DXA	DE.IJ.NO	H	19139173	TRAV26-1*02	TRAJ13*02	TISG.....NN	IWGG...YQKVT	AVQASG.GSYIPT	TRBV7-9*01	TRBJ1-1*01	SEH.....NR	FQN....EAQ	ASRYRDDSYNEQF	EENLLDFVRF	HLA-B*4405	
DM1	u	3DX9	AB.CD	H	19139173	TRAV26-1*02	TRAJ13*02	TISG.....NEY	GLK.....NN	IVWGG...YQKVT	TRBV7-9*01	TRBJ2-1*01	SEH.....NR	FQN....EAQ	ASRYRDDSYNEQF			
E8	2	2IAN	DE.IJ.NO.ST	H	17334368	TRAV22*01	TRAJ54*01	DSV.....NN	IPS.....GT	AALIQG..AQKLV	TRBV6-6*01	TRBJ1-3*01	NNH.....NY	SVG....AGI	ASTYH....GTGY	GELIGTLNAAKVPAD	HLA-DRA*0101	HLA-DRB1*0101
E8	2	2IAM	CD	H	17334368	TRAV22*01	TRAJ54*01	DSV.....NN	IPS.....GT	AALIQG..AQKLV	TRBV6-6*01	TRBJ1-3*01	NNH.....NY	SVG....AGI	ASTYH....GTGY	GELIGILNAAKVPAD	HLA-DRA*0101	HLA-DRB1*0101
E8	u	2IAL	AB.CD	H	17334368	TRAV22*01	TRAJ54*01	DSV.....NN	IPS.....GT	AALIQG..AQKLV	TRBV6-6*01	TRBJ1-3*01	NNH.....NY	SVG....AGI	ASTYH....GTGY			
ELS4	1	2NX5	DE.IJ.NP.TU	H	17259989	TRAV1-2*01	TRAJ6*01	TSG.....FNG	NVL.....DGL	AVQASG.GSYIPT	TRBV10-3*01	TRBJ1-5*01	EHN.....RY	SVG....VKD	ATGTGD.SNQPQH	EPLPQGQLTAY	HLA-B*3501	
ELS4___	u	2NW2	AB	H	17259989	TRAV1-2*01	TRAJ6*01	TSG.....FNG	NVL.....DGL	AVQASG.GSYIPT	TRBV10-3*01	TRBJ1-5*01	EHN.....RY	SVG....VKD	ATGTGD.SNQPQH			
HA1.7	2	1FYT	DE	H	11060013	TRAV8-4*05	TRAJ48*01	SSV.....PFY	YTSA...ATIV	AVSESPFGNEKLT *	TRBV28*01	TRBJ1-2*01	NDH.....EN	SYD....VKM	ASSSTG.LPYGYT	PKYVKQNTLKLAT	HLA-DRA*0101	HLA-DRB1*0101
HA1.7	2	1J8H	DE	H	11877480	TRAV8-4*05	TRAJ48*01	SSV.....PFY	YTSA...ATIV	AVSESPFGNEKLT *	TRBV28*01	TRBJ1-2*01	NDH.....EN	SYD....VKM	ASSSTG.LPYGYT	PKYVKQNTLKLAT	HLA-DRA*0101	HLA-DRB1*0401
Hy.1B1	2	3PL6	CD	H	21199956	TRAV13-1*02	TRAJ48*01	DSA.....SNY	IRSN...VGE	AASSFG...NEKLT *	TRBV7-3*01	TRBJ2-3*01	SGH.....TA	FQG....TGA	ATSAL...GDTQY	NPVVHFFKNIVTPR (L)	HLA-DQA1*0102	HLA-DQB1*0501
JM22	1	2VLJ	DE	H	18275829	TRAV27*01	TRAJ42*01	SVF.....SS	VVTG...GEV	AGAGS...QGNLI	TRBV19*01	TRBJ2-7*01	LNH.....DA	SQI....VND	ASSSRS..SYEQY	GILGFVFTL	HLA-A*0201	
JM22	1	2VLK	DE	H	18275829	TRAV27*01	TRAJ42*01	SVF.....SS	VVTG...GEV	AGAGS...QGNLI	TRBV19*01	TRBJ2-7*01	LNH.....DA	SQI....VND	ASSSRS..SYEQY	GILGFVFTL	HLA-A*0201	
JM22	1	1OGA	DE	H	12796775	TRAV27*01	TRAJ42*01	SVF.....SS	VVTG...GEV	AGAGS...QGNLI	TRBV19*01	TRBJ2-7*01	LNH.....DA	SQI....VND	ASSSRS..SYEQY	GILGFVFTL	HLA-A*0201	
JM22	u	2VLM	DE	H	18275829	TRAV27*01	TRAJ42*01	SVF.....SS	VVTG...GEV	AGAGS...QGNLI	TRBV19*01	TRBJ2-7*01	LNH.....DA	SQI....VND	ASSSRS..SYEQY			
JM22	2,s	2XN9	AB	H	21081917	TRAV27*01	TRAJ42*01	SVF.....SS	VVTG...GEV	AGAGS...QGNLI	TRBV19*01	TRBJ2-7*01	LNH.....DA	SQI....VND	ASSSRS..SYEQY	PKYVKQNTLKLAT	HLA-DRA*0101	HLA-DRB1*0101
JM22	s	2XNA	AB	H	21081917	TRAV27*01	TRAJ42*01	SVF.....SS	VVTG...GEV	AGAGS...QGNLI	TRBV19*01	TRBJ2-7*01	LNH.....DA	SQI....VND	ASSSRS..SYEQY			
JM22 [S99βA]	1	2VLR	DE.IJ	H	18275829	TRAV27*01	TRAJ42*01	SVF.....SS	VVTG...GEV	AGAGS...QGNLI	TRBV19*01	TRBJ2-7*01	LNH.....DA	SQI....VND	ASSSRA..SYEQY	GILGFVFTL	HLA-A*0201	
KK50.4	1	2ESV	DE	H	16474394	TRAV26-1*01	TRAJ37*01	TISG.....NEY	GLK.....NN	IVYRSS..NTGKLI	TRBV14*01	TRBJ2-3*01	SGH.....DN	FYK....ESK	ASSQD..RDTQY	VMAPRTLIL	HLA-E*0101	
LC13	1	1MI5	DE	H	12530975	TRAV26-2*01	TRAJ52*01	TISG.....TDY	GLT.....SN	ILPLAGGTSYGKLT *	TRBV7-8*01	TRBJ2-7*01	SGH.....VS	FQN....EAQ	ASSLGQ..AYEQY	FLRGRAYGL	HLA-B*0801	
LC13	u	1KGC	DE	H	12429093	TRAV26-2*01	TRAJ52*01	TISG.....TDY	GLT.....SN	ILPLAGGTSYGKLT *	TRBV7-8*01	TRBJ2-7*01	SGH.....VS	FQN....EAQ	ASSLGQ..AYEQY			
LC13	1	3KPR	DE.IJ	H	20064448	TRAV26-2*01	TRAJ52*01	TISG.....TDY	GLT.....SN	ILPLAGGTSYGKLT *	TRBV7-8*01	TRBJ2-7*01	SGH.....VS	FQN....EAQ	ASSLGQ..AYEQY	EEYLKWTF	HLA-B*4405	
LC13	1	3KPS	DE	H	20064448	TRAV26-2*01	TRAJ52*01	TISG.....TDY	GLT.....SN	ILPLAGGTSYGKLT *	TRBV7-8*01	TRBJ2-7*01	SGH.....VS	FQN....EAQ	ASSLGQ..AYEQY	EEYLQAFTY	HLA-B*4405	
MEL5	1	3HG1	DE	H	19605354	TRAV12-2*01	TRAJ27*01	DRG.....SQS	IYS....NGD	AVNVA....GKST	TRBV30*01	TRBJ2-2*01	GTS.....NPN	SVG....TG	AWSETGLGTGELF	ELAGIGILTV	HLA-A*0201	
NKT12 [H]	u	2EYR	AB	H	16505140	TRAV10*01	TRAJ18*01	VSP.....FSN	MTFS...ENT	VVSDRGSTLGRIY	TRBV25-1*01	TRBJ2-7*01	MGH.....DK	SYG....VNS	ASPSRR.GSYEQY			
iNKT [H,A]	u	2CDE	AB.CD.EF	H	16520393	TRAV10*01	TRAJ18*01	VSP.....FSN	MTFS...ENT	VVSDRGSTLGRIY	TRBV25-1*01	TRBJ2-7*01	MGH.....DK	SYG....VNS	ASSEHIGTNYEQY			
nkt15 [H]	1	3HUJ	EF.GH	H	19592275	TRAV10*01	TRAJ18*01	VSP.....FSN	MTFS...ENT	VVSDRGSTLGRIY	TRBV25-1*01	TRBJ2-7*01	MGH.....DK	SYG....VNS	ASSGLRDRGLYEQY	CID 2826713	CD1d	
nkt15 [H]	1	2PO6	CD.GH	H	17581592	TRAV10*01	TRAJ18*01	VSP.....FSN	MTFS...ENT	VVSDRGSTLGRIY	TRBV25-1*01	TRBJ2-7*01	MGH.....DK	SYG....VNS	ASSGLRDRGLYEQY	CID 2826713	CD1d	
nkt15 [H]	u	2EYS	AB	H	16505140	TRAV10*01	TRAJ18*01	VSP.....FSN	MTFS...ENT	VVSDRGSTLGRIY	TRBV25-1*01	TRBJ2-7*01	MGH.....DK	SYG....VNS	ASSGLRDRGLYEQY			
nkt15 [H]	u	2EYT	AB.CD	H	16505140	TRAV10*01	TRAJ18*01	VSP.....FSN	MTFS...ENT	VVSDRGSTLGRIY	TRBV25-1*01	TRBJ2-7*01	MGH.....DK	SYG....VNS	ASSGLRDRGLYEQY			
nkt 5B [H]	u	2CDG	AB	H	16520393	TRAV17*01	TRAJ18*01	TSI.....NN	IRSN...ERE	AEFDRGSTLGRIY	TRBV25-1*01	TRBJ1-2*01	MGH.....DK	SYG....VNS	ASSESRTGINYGYT			
nkt 5E [H]	u	2CDF	AB	H	16520393	TRAV27*01	TRAJ18*01	SVF.....SS	VVTG...GEV	AGADRGSTLGRIY	TRBV25-1*01	TRBJ1-4*01	MGH.....DK	SYG....VNS	ASSERPDGNEKLF			

a) Bound states: a=antibody; u =unbound, s=superantigen;1=MHC class I; 2=MHC class II

b) Differences to the IMGT AA sequence in CDR1/2 loops are underlined; differences of subtypes in the CDR3 loops are underlinded. (*) IMGTDomainGapAlign results were manually adjusted.

c) Peptides linked to MHC or TCR are marked with (L); For non-peptidic ligands the pubchem ID is given.

d) Y5 is mutated to K and linked to 4-(3-Indolyl)-butyric acid.

e) Mutation I219A in MHC α_1 domain

f) Mutation A245V in the MHC α3 domain.

g) Mutations Q65A, T69A, Q155A.

h) H2-K$^{b(m3)}$ and H2-K$^{b(mm8}$ are natural mutants of H2-K^{b}

i) WT with solubility mutations.

j) Solubility Mutations (F9Y, V12T, I23T)

k) Mutation R189K

l) Mutation T163A in the MHC binding pocket

m) Mutation W167V in th MHC binding pocket

n) F5: 3,4-difluoro-phenylalanine.

o) F5: 4-fluoro-phenylalanine

S1 Table. Properties of available TCR structures (continued).

Name	Bound State[a]	PDB-ID	Chain Pairs	Spec.	Literature PMID	IMGT-Name V-Segment	IMGT-Name J-Segment	Alpha CDR Loops[b] 1	Alpha CDR Loops[b] 2	Alpha CDR Loops[b] 3	Beta IMGT-Name V-Segment	Beta IMGT-Name J-Segment	Beta CDR-Loops[b] 1	Beta CDR-Loops[b] 2	Beta CDR-Loops[b] 3	Pep/Lig[c]	Ligand MHC I/II alpha	Ligand MHC II beta
OB.1A12	2	2WBJ	CD,GH	H	19303388	TRAV17*01	TRAJ40*01	TSI.....NN	IRSN...ERE	AYDYTS.GYKYI	TRBV20-1*01	TRBJ2-1*01	DFQ.....ATT	SNEG...SKA	SARDIJTSGANNEQ	FARVHFISALHGS (L)	HLA-DRA*0101	HLA-DRB1*1501
OB.1A12	2	1YMM	DE	H	15821740	TRAV17*01	TRAJ40*01	TSI.....NN	IRSN...ERE	AYDYTS.GYKYI	TRBV20-1*01	TRBJ2-1*01	DFQ.....ATT	SNEG...SKA	SARDIJTSGANNEQ	ENFVVHFFKNIVTPR (L)	HLA-DRA*0101	HLA-DRB1*1501
RA14	1	3GSN	AB	H	19542454	TRAV24*01	TRAJ49*01	SSN.....FYA	WTLN...GDE	ARNTG....NQFY	TRBV6-5*01	TRBJ1-2*01	NNH.....EY	SVG....AGI	ASSPVTGGIYGYT	NLVPMVATV	HLA-A*0201[f]	
SB27	1	2AK4	DE,IJ,NP,TU	H	16188824	TRAV19*01	TRAJ34*01	TRDT.....TYY	RNSF..DEQN	ALSGFY.NYDKLI	TRBV6-1*01	TRBJ2-7*01	NNH.....NS	SAS....EGT	ASPGIA.GEYFGY	LPEPLPQGQLTAY	HLA-B*3508	
SB27 [K16Dq]	1	3KXF	DE,GH,NP,MO	H	20483993	TRAV19*01	TRAJ34*01	TRDT.....TYY	RNSF..DEQN	ALSGFY.NYDKLI	TRBV6-1*01	TRBJ2-7*01	NNH.....NS	SAS....EGT	ASPGIA.GEYFGY	LPEPLPQGQLTAY	HLA-B*3508[g]	
TCR MS2-3C8	2	3O6F	CD,GH	H	21297580	TRAV26-2*01	TRAJ32*01	TISG.....TDY	GLT.....SN	TVYGGA..TNKLI *	TRBV20-1*01	TRBJ1-6*01	DFQ.....ATT	SNEG...SKA	SARGGS.YNSPLH	FSWGABSQPPGFG (L)	HLA-DRA*0101	HLA-DRB1*0401
TK3 WT	1	3MV7	DE	H	PMC2901058	TRAV20*01	TRAJ52*01	VSG.....LRG	IYSA...GEE	AVQDLGTSGSRLT *	TRBV9*01	TRBJ2-2*01	SGD.....LS	YYN....GEE	ASSAR...SGELF	HPVGEADYFEY	HLA-B*3501	
TK3 Q55H	1	3MV8	DE	H	PMC2901058	TRAV20*01	TRAJ52*01	VSG.....LRG	IYSA...GEE	AVQDLGTSGSRLT *	TRBV9*02	TRBJ2-2*01	SGD.....LS	YYN....GEE	ASSAR...SGELF	HPVGEADYFEY	HLA-B*3501	
TK3 Q55A	1	3MV9	DE	H	PMC2901058	TRAV20*01	TRAJ52*01	VSG.....LRG	IYSA...GEE	AVQDLGTSGSRLT *	TRBV9*01	TRBJ2-2*01	SGD.....LS	YYN....GEE	ASSAR...SGELF	HPVGEADYFEY	HLA-B*3501	
1934.4	2	2PXY	AB	M	17694060	TRAV6D-7*01	TRAJ48*01	ASG.....YPA	ASRD...KEK	ALSENY.GNEKLT	TRBV13-2*01	TRBJ2-3*01	NNH.....NN	SYG....AGS	ASGDAS.GAETLY	HSRGGASQYRPSQ	H2-Aa(u)	H2-Ab(u)
1F1E8	u	3MFF	AB,CD	M	20630474	TRAV14D-3*2	TRAJ52*01	NSA.....FDY	IRSV...SDK	AATGAN..TGKLT	TRBV29*01	TRBJ2-7*01	NSH.....ET	SYD....VDS	ASSLAGTGNYEQY			
226 TCR	2	3QIU	CD	M	21490152	TRAV4D-4*01	TRAJ16*01	TTM.....RA	LAS.....GT	AAEPSS..GQKLV	TRBV26*01	TRBJ1-2*01	KGH.....PV	FQN....QEV	ASSLNN.ANSDYT	ADLIAYLKQATKG	H2-Ea(k)	H2-Eb(k)
226 TCR	2	3QIW	CD	M	21490152	TRAV4D-4*01	TRAJ16*01	TTM.....RA	LAS.....GT	AAEPSS..GQKLV	TRBV26*01	TRBJ1-2*01	KGH.....PV	FQN....QEV	ASSLNN.ANSDYT	ADLIAYLEQATKG	H2-Ea(k)	H2-Eb(k)
2B4	1	3QIB	CD	M	21490152	TRAV4D-4*02	TRAJ56*01	TTM.....RA	LAS.....GT	AALRATGGNNKLT	TRBV26*01	TRBJ2-5*01	KGH.....PV	FQN....QEV	ASSLNW.SQDTQY	ADLIAYLKQATKG	H2-Ea(k)	H2-Eb(k)
2B4	u	3QJF	AB,CD	M	21490152	TRAV4D-4*02	TRAJ56*01	TTM.....RA	LAS.....GT	AALRATGGNNKLT	TRBV26*01	TRBJ2-5*01	KGH.....PV	FQN....QEV	ASSLNW.SQDTQY			
2C	u	1TCR	AB	M	8824178	TRAV9-4*01	TRAJ35*02	YSA.....TPY	YYSG..DPVV	AVSGF...ASALT	TRBV13-2*01	TRBJ2-4*01	NNH.....NN	SYG....AGS	ASGGG....GTLY	STYRYYGL	H2-K1(b)	
2C	1	1GRA	AB,CD	M	10755612	TRAV4D-4*01	TRAJ35*02	YSA.....TPY	YYSG..DPVV	AVSGF...ASALT	TRBV13-2*01	TRBJ2-4*01	NNH.....NN	SYG....AGS	ASGGG....GTLY			
2C	1	1MWA	AB, CD	M	11994422	TRAV9-4*01	TRAJ35*02	YSA.....TPY	YYSG..DPVV	AVSGF...ASALT	TRBV13-2*01	TRBJ2-4*01	NNH.....NN	SYG....AGS	ASGGG....GTLY	EQYKFYSV	H2-K1(bm3)[h]	
2C	1	2CKB	AB,CD	M	9469799	TRAV9-4*01	TRAJ35*02	YSA.....TPY	YYSG..DPVV	AVSGF...ASALT	TRBV13-2*01	TRBJ2-4*01	NNH.....NN	SYG....AGS	ASGGG....GTLY	EQYKFYSV	H2-K1(b)	
2C T7	S.2	2ICW	IJ,KL	M	17220897	TRAV9-4*01	TRAJ35*02	YSA.....TPY	YYSG..DPVV	AVSGF...ASALT	TRBV13-2*01	TRBJ2-4*01	NNH.....NN	SYG....AGS	ASGGG....GTLY	PRYVKQNTLKLAT	H2A-DRA*0101	HLA-DRB1*0101
2C [T7-wt-s][i]	1	2OI9	BC	M	17418792	TRAV9-4*01	TRAJ35*02	YSA.....TPY	YYSG..DPVV	AVSGF...ASALT	TRBV13-2*01	TRBJ1-2*01	KGH.....PV	FQN....QEV	ASSLNN.ANSDYT	QLSPFPFDL	H2-L(d)[i]	
2C m13 [T7-s]	↑	3E3Q	DE,CF,IJ,MN,R S,VW,Zade	M	18941216	TRAV9-4*01	TRAJ35*02	YSA.....TPY	YYSG..DPVV	AVSDE...PELLT	TRBV13-2*01	TRBJ2-4*01	NNH.....NN	SYG....AGS	ASGGG....GTLY	QLSPFPFDL	H2-L(d)[i]	
2C m6 [T7-s]	1	2E7L	AD,BC	M	17418792	TRAV9-4*01	TRAJ35*02	YSA.....TPY	YYSG..DPVV	AVSHQ...GEYLT	TRBV13-2*01	TRBJ2-4*01	NNH.....NN	SYG....AGS	ASGGG....GTLY	QLSPFPFDL	H2-L(d)	
2C m67 [T7-s]	1	3E2H	BC	M	18941216	TRAV9-4*01	TRAJ35*02	YSA.....TPY	YYSG..DPVV	AVSHE...KEFLT	TRBV13-2*01	TRBJ2-4*01	NNH.....NN	SYG....AGS	ASGGG....GTLY	QLSPFPFDL	H2-L(d)[k]	
2W20	2	3C6L	AB,EF	M	18308592	TRAV14-2*02	TRAJ31*01	DST.....FDY	IRPV...SNK	AASD.....NRIF	TRBV13-2*01	TRBJ2-7*01	NNH.....NN	SYG....AGS	ASGDAW..GYEQY	FEAQKAKANKAVD (L)	H2-Aa(b)	H2-Ab(b)[k]
5c.c7	u	3QJH	AB,CD	M	21490152	TRAV4D-4*03	TRAJ34*02	TTM.....RA	LAS.....GT	AAEAAN..TNKVV	TRBV26*01	TRBJ1-2*01	KGH.....PV	FQN....QEV	ASSLNN.ANSDYT			
AHIII12.2	1	2UVE	EF,LM	M	17825839	TRAV12D-2*01	TRAJ50*01	STY.....SPF	SFTD...NKR	ALFLASGSFSKLV	TRBV13-3*01	TRBJ2-7*01	NNH.....DY	SYV....ADS	ASSDWV..SYEQY	ALWGFFPVL	HLA-A*0201[m]	
AHIII12.2	1	2JCC	EF,LM	M	17825839	TRAV12D-2*01	TRAJ50*01	STY.....SPF	SFTD...NKR	ALFLASGSFSKLV	TRBV13-3*01	TRBJ2-7*01	NNH.....DY	SYV....ADS	ASSDWV..SYEQY	ALWGFFPVL	HLA-A*0201[m]	
AHIII12.2	1	1LP9	EF,LM	M	14563323	TRAV12D-2*01	TRAJ50*01	STY.....SPF	SFTD...NKR	ALFLASGSFSKLV	TRBV13-3*01	TRBJ2-7*01	NNH.....DY	SYV....ADS	ASSDWV..SYEQY	ALWGFFPVL	HLA-A*0201	
B3K506	2	3C6Z	AB,EF	M	18308592	TRAV6-7*01	TRAJ34*02	ASG.....YPA	ASRD...KEK	ALVISN..TNKVV	TRBV13-3*01	TRBJ1-3*01	NNH.....DY	SYV....ADS	ASIDSS..GNTLY	FEAQKAKANKAVD (L)	H2-Aa(b)	H2-Ab(b)[k]
BM3.3	1	1NAM	AB	M	12563259	TRAV16*01	TRAJ32*01	TQDS.....STF	QDSY..KKEN	AMRGDYGGSGNKLI *	TRBV1*01	TRBJ1-3*01	NSQ.....YFW	LRS....PGD	TCSADR.VGNTLY	RGYVYQGL	H2-K1(b)	
BM3.3	1	1FO0	AB	M	11017099	TRAV16*01	TRAJ32*01	TQDS.....STF	QDSY..KKEN	AMRGDYGGSGNKLI *	TRBV1*01	TRBJ1-3*01	NSQ.....YFW	LRS....PGD	TCSADR.VGNTLY	INFDFNTI	H2-K1(b)	
BM3.3	1	2OL3	AB	M	17363906	TRAV16*01	TRAJ32*01	TQDS.....STF	QDSY..KKEN	AMRGDYGGSGNKLI *	TRBV1*01	TRBJ1-3*01	NSQ.....YFW	LRS....PGD	TCSADR.VGNTLY	SQYYYNSL	H2-K1(bm8)[h]	
cl19	2	2Z31	AB	M	17694060	TRAV6D-7*01	TRAJ48*01	ASG.....YPA	ASRD...KEK	ALSENY.GNEKTT	TRBV13-2*01	TRBJ2-4*01	NNH.....NN	SYG....AGS	ASGDAS.GGNTLY	RGGASQYRPSQ	H2-Aa(u)	H2-Ab(u)
D10	2	1D9K	AB,EF	M	10583947	TRAV14D-2*01	TRAJ11*01	DST.....FDY	ISLV...SNK	AATGS...FNKLT	TRBV13-2*01	TRBJ2-1*01	NNH.....NN	SYG....AGS	ASGGQG..RAEQF	GNSHRGATEVPSCIESG	H2-Aa(k)	
KB5-C20	1	1KJ2	AB,DE	M	11911820	TRAV14-1*01	TRAJ15*01	EDSTI.....FN	SIRSV...SD	CAARYQ.GGRALI	TRBV1*01	TRBJ2-3*01	NSQ.....YFW	LRS....PGD	TCSAAPDWGASAE	KVTTFIDL	H2-K1(b)	
N15	u,a	1NFD	AB,CD	M	9427737	TRAV12D-1*01	TRAJ40*01	TTYI.....TIA	SSTD...NKR	ALSEGG..NYKVV	TRBV12-1*01	TRBJ2-7*01	SGH.....SN	HYE....KVE	ASSLRW..GDEQY			
nkt [M.A]	1	3O8X	CD	M	20921281	TRAV11*02	TRAJ18*01	VTP.....DNH	LVDQ...KDK	VVGDRGSALGRLH	TRBV13-2*01	TRBJ2-5*01	NNH.....NN	SYG....AGS	ASGQE...GYTQY	CID 6420173	CD1d1	
nkt [M.A]	1	3O9W	CD	M	20921281	TRAV11*02	TRAJ18*01	VTP.....DNH	LVDQ...KDK	VVGDRGSALGRLH	TRBV13-2*01	TRBJ2-5*01	NNH.....NN	SYG....AGS	ASGQE...GYTQY	CID 44593851	CD1d1	
nkt [M.A]	1	3QUY	CD	M	21562205	TRAV11*02	TRAJ18*01	VTP.....DNH	LVDQ...KDK	VVGDRGSALGRLH	TRBV13-2*01	TRBJ2-5*01	NNH.....NN	SYG....AGS	ASGQE...GYTQY	?	CD1d1	
nkt [M.A]	1	3QUX	CD	M	21552205	TRAV11*02	TRAJ18*01	VTP.....DNH	LVDQ...KDK	VVGDRGSALGRLH	TRBV13-2*01	TRBJ2-5*01	NNH.....NN	SYG....AGS	ASGQE...GYTQY	CID 9811322	CD1d1	
Nkt [M.A]	1	3QUZ	CD	M	21552205	TRAV11*02	TRAJ18*01	VTP.....DNH	LVDQ...KDK	VVGDRGSALGRLH	TRBV13-2*01	TRBJ2-5*01	NNH.....NN	SYG....AGS	ASGQE...GYTQY	CID 25228911	CD1d1	
nkt [M.B]	u	2Q86	AB,CD	M	18295796	TRAV11*02	TRAJ18*01	VTP.....DNH	LVDQ...KDK	VVGDR...GSALI	TRBV13-2*01	TRBJ2-5*01	NNH.....NN	SYG....AGS	ASGQE...GYTQY			
nkt [M.C]	1	3ARB	CD	M	21376639	TRAV11*02	TRAJ18*01	VTP.....DNH	LVDQ...KDK	VVGDRGSALGRLH	TRBV13-2*01	TRBJ2-1*01	NNH.....NN	SYG....AGS	ASGDAGGNYAEQF	CID 2826714	CD1d1	
nkt [M.C]	1	3ARD	CD	M	21376639	TRAV11*02	TRAJ18*01	VTP.....DNH	LVDQ...KDK	VVGDRGSALGRLH	TRBV13-2*01	TRBJ2-1*01	NNH.....NN	SYG....AGS	ASGDAGGNYAEQF	CID 45273283)	CD1d1	
nkt [M.C]	1	3ARE	CD	M	21376639	TRAV11*02	TRAJ18*01	VTP.....DNH	LVDQ...KDK	VVGDRGSALGRLH	TRBV13-2*01	TRBJ2-1*01	NNH.....NN	SYG....AGS	ASGDAGGNYAEQF	CID 45269857	CD1d1	
nkt [M.C]	1	3ARF	CD	M	21376639	TRAV11*02	TRAJ18*01	VTP.....DNH	LVDQ...KDK	VVGDRGSALGRLH	TRBV13-2*01	TRBJ2-1*01	NNH.....NN	SYG....AGS	ASGDAGGNYAEQF	CID 50990900	CD1d1	
nkt [M.C]	1	3ARG	CD	M	21376639	TRAV11*02	TRAJ18*01	VTP.....DNH	LVDQ...KDK	VVGDRGSALGRLH	TRBV13-2*01	TRBJ2-1*01	NNH.....NN	SYG....AGS	ASGDAGGNYAEQF	CID 50990901	CD1d1	
nkt [M.C]	1	3HE6	CD	M	19592275	TRAV11*02	TRAJ18*01	VTP.....DNH	LVDQ...KDK	VVGDRGSALGRLH	TRBV13-2*01	TRBJ2-1*01	NNH.....NN	SYG....AGS	ASGDAGGNYAEQF	CID 2826713	CD1d1	
nkt [M.C]	1	3HE7	CD	M	19592275	TRAV11*02	TRAJ18*01	VTP.....DNH	LVDQ...KDK	VVGDRGSALGRLH	TRBV29*01	TRBJ2-1*01	NSH.....ET	SYD....VDS	ASSSTG..TAFTQY	CID 2826713	CD1d1	
nkt [M.D]	2	3MBE	CD,GH	M	20407212	TRAV11*02	TRAJ18*01	TTM.....KS	IAQ.....GT	AAEDSG.SGNKLI	TRBV13-3*01	TRBJ1-3*01	NNH.....DY	SYV....ADS	ASWSDR.AGNTLY	GAMKRHGLDNYRGTSLGN	H2-Aa(d)	H2-Ab(NOD)
TCR1T2.10	2	1U3H	AB,EF	M	15664161	TRAV14D-3*08	TRAJ13*01	ENSA.....FD	SILSV...SD	CAASAN.SGTYQR	TRBV13-2*01	TRBJ2-7*01	NNH.....NN	SYG....AGS	ASGDAG.GGYEQY	SRGGASQGTRFSQ	H2-Aa(u)	H2-Ab(u)
YAe62	2	3C60	AB,EF	M	18308592	TRAV6-3*01	TRAJ13*01	TPG.....YPT	VTPA...NNK	AANSG.....TYQR	TRBV12-1*01	TRBJ2-4*01	NNH.....NN	SYG....AGS	ASGDPW..GDTLY	PEAQKAKANKAVD (L)	H2-Aa(b)	H2-Ab(b)[k]

a) Bound states: a=antibody; u =unbound, s=superantigen;1=MHC class I; 2=MHC class II

b) Differences to the IMGT AA sequence in CDR1/2 loops are underlined; differences of subtypes in the CDR3 loops are underlined. (*) IMGTDomainGapAlign results were manually adjusted.

c) Peptides linked to MHC or TCR are marked with (L); For non-peptidic ligands the pubchem ID is given.

d) Y5 is mutated to K and linked to 4-(3-Indolyl)-butyric acid.

e) Mutation I219A in MHC α_1 domain

f) Mutation A245V in the MHC α3 domain.

g) Mutations Q65A, T69A, Q155A.

h) H2-K^{bm3} and H2-K^{bm8} are natural mutants of H2-K^{b}

i) WT with solubility mutations.

j) Solubility Mutations (F9Y, V12T, I23T)

k) Mutation T163A in the MHC binding pocket

m) Mutation W167V in th MHC binding pocket

n) F5: 3 4-difluoro-phenylalanine.

o) F5: 4-fluoro-phenylalanine

1. Giudicelli V, Chaume D, Lefranc M-P (2005) IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. Nucleic Acids Res 33: D256-D261.

S2 Table. Mutations within the framework region of the TCR structures. The sequences of the TCR structures were compared to the corresponding wild type (WT). Differences to the alleles provided in the IMGT Gene-DB [1] are provided as mutation pairs in brackets. Naming of the sheets and loops, and the residue indices follow the IMGT unique numbering [2].

Name	PDB-ID	Alpha Chain Mutations in Respect To WT Sequence Acc. To IMGT Unique Numbering (w/o CDR3/FG)									Beta Chain Mutations in Respect To WT Sequence Acc. To IMGT Unique Numbering (w/o CDR3/FG)							
		A	B	BC	C	C'	C'C''	C''	D	E	F	G	A	B	BC	C	C'	C'C''	C''	D	E	F	G
1G4 AV-wt	2F54	(E3Q)				(K48G)(G49K)			(S84A)	(R86S)													
1G4 c5c1	2PYF									(R86S)								(I65T)		(T84I)			
1G4 c5c1	2PYE									(R86S)								(I65T)		(T84I)			
1G4 c49c50	2F53						(Q57P)(S58F)(S59W)											(G58S)(A63V)(I65M)					
1G4 c58d62	2P5W	(E3Q)					(Q57T)(S58P)(S59W)			(R86S)								(G58S)(A63V)(I65M)		(T84I)			
1G4 c58c61	2P5E						(Q57T)(S58P)(S59W)			(R86S)								(G58A)(A63I)(G64Q)(I65T)		(T84I)			
3A6	1ZGL								(L76K)									(S58N)					
AS01	3O4L																						
B7	1BD2							(E67A)															
cf34	3FFC					(L53V)	(Q64G)																
DM1	3DXA	(P9T)																					
DM1	3DX9	(P9T)																					
Hy.1B1	3PL6												(A120T)(K123R)(K127I)										
MEL5	3HG1	(K2Q)																					
nkt 5E [H]	2CDF	(L4V)																					
OB.1A12	2WBJ												(C13S)										
OB.1A12	1YMM												(C13S)										
SB27 [K16Dq]	3KXF		(K16D)																				
TCR MS2-3C8	3O6F												(C13A)										
TK3 WT	3MV7											(Q120E)(I123Q)(H128N)											
TK3 Q55H	3MV8											(Q120E)(I123Q)(H128N)					(Q55H)						
TK3 Q55A	3MV9											(Q120E)(I123Q)(H128N)					(Q55A)						
1934.4	2PXY	(E8G)										(T125Q)(I126V)(K127V)		(G17E)			(H54Y)				(I90T)(L93S)		(S120P)
1F<u>LE</u>8	3MFF		(A19T)																				
2C T7	2ICW					(L50P)				(W96R)				(G17E)			(G49E)				(L93S)		
2C [T7-wt-s]	2OI9					(L50P)				(W96R)				(G17E)			(G49E)(H54Y)				(I90T)(L93S)		
2C m13 [T7-s]	3E3Q					(L50P)				(W96R)				(G17E)			(G49E)(H54Y)				(I90T)(L93S)		
2C m6 [T7-s]	2E7L					(L50P)				(W96R)				(G17E)			(G49E)(H54Y)				(I90T)(L93S)		
2C m67 [T7-s]	3E2H					(L50P)				(W96R)				(G17E)			(G49E)(H54Y)				(I90T)(L93S)		
B3K506	3C5Z	(Q10N)																					
BM3.3	1FO0											(I120T)											
BM3.3	2OL3											(I120T)											
d19	2Z31	(E8G)										(T125Q)(I126V)(K127V)		(G17E)			(H54Y)				(I90T)(L93S)		
D10	1D9K											(K120A)(V123R)(L125A)											
nkt [M.B]	2Q86			(F42Y)		(S52L)					(I103F)	(A120S)											
nkt [M.C]	3ARB		(C19S)																				
nkt [M.C]	3ARD		(C19S)																				
nkt [M.C]	3ARE		(C19S)																				
nkt [M.C]	3ARF		(C19S)																				
nkt [M.C]	3ARG		(C19S)																				
nkt [M.C]	3HE6		(C19S)																				
nkt [M.D]	3HE7		(C19S)																				
TCR 21.30	3MBE											(I120T)											
TCR172.10	1U3H													(G17E)			(H54Y)				(I90T)(L93S)		
YAe62	3C60	(I5T)																					

Remark: For localization of the secondary structure elements A, B, BC, . . . , F see Figure 1.3 on Page 7 in the main part of the thesis.

1. Giudicelli V, Chaume D, Lefranc M-P (2005) IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. Nucleic Acids Res 33: D256-D261.
2. Lefranc MP, Pommie C, Ruiz M, Giudicelli V, Foulquier E, et al. (2003) IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. Dev Comp Immunol 27: 55-77.

S3 Table. Multiple Sequence Alignments at the COR.

C	Name	PDB	COR α	COR β
1	1G4 c49c50	2f53	lqwfrQdpgkg	mswyrQdpgmg
	1G4	2bnr	lqwfrQdpgkg	mswyrQdpgmg
	1G4	2bnq	lqwfrQdpgkg	mswyrQdpgmg
	1G4 AV-wt	2f54	lqwfrQdpggk	mswyrQdpgmg
	E8	2ian	lqwfhQnpwgq	mywyrQdpgmg
	1G4 c58c61	2p5e	lqwfrQdpgkg	mswyrQdpgmg
	1G4 c58c62	2p5w	lqwfrQdpgkg	mswyrQdpgmg
	E8	2iam	lqwfhQnpwgq	mywyrQdpgmg
	1G4 c5c1	2pye	lqwfrQdpgkg	mswyrQdpgmg
2	TCR MS2-3C8	3o6f	ihwyrQlpsqg	mfwyrQfpkqs
	A6	1qse	ffwyrQysgks	mswyrQdpgmg
	A6	3h9s	ffwyrQysgks	mswyrQdpgmg
	A6	3d39	ffwyrQysgks	mswyrQdpgmg
	A6	3d3v	ffwyrQysgks	mswyrQdpgmg
	A6	3pwp	ffwyrQysgks	mswyrQdpgmg
	A6	1qsf	ffwyrQysgks	mswyrQdpgmg
	A6	2gj6	ffwyrQysgks	mswyrQdpgmg
	A6	1ao7	ffwyrQysgks	mswyrQdpgmg
	A6	1qrn	ffwyrQysgks	mswyrQdpgmg
3	JM22	2vlj	lqwyrQepgeg	mywyrQdpgqg
	JM22	2vlk	lqwyrQepgeg	mywyrQdpgqg
	JM22	1oga	lqwyrQepgeg	mywyrQdpgqg
	JM22 [S99βA]	2vlr	lqwyrQepgeg	mywyrQdpgqg
	SB27	2ak4	lfwykQppsge	mywyrQdpgmg
	SB27 [K16Dα]	3kxf	lfwykQppsge	mywyrQdpgmg
	TCR 21.30	3mbe	vqwfqQnhrgr	mywyrQdtghg
	OB.1A12	1ymm	lqwyrQnsgrg	mfwyrQfpkqs
	OB.1A12	2wbj	lqwyrQnsgrg	mfwyrQfpkqs
4	KK50.4	2esv	vywyrQihsqg	lywyrRvmgke
	LC13	3kpr	ihwyrQlpsqg	lfwyqQalgqg
	2B4	3qib	vqwfqQnsrgs	vfwyqQnknne
	LC13	3kps	ihwyrQlpsqg	lfwyqQalgqg
	226 TCR	3qiu	vqwfrKnsrgs	vfwyqQnknne
	226 TCR	3qiw	vqwfrKnsrgs	vfwyqQnknne
	2C m13 [T7-s]	3e3q	lfwyvQyprqg	mywyrQdtghe
	2C m6 [T7-s]	2e7l	lfwyvQyprqg	mywyrQdtghe
	2C [T7-wt-s]	2oi9	lfwyvQyprqg	mywyrQdtghe
	DM1	3dxa	vywyrQihsqg	lywyrQtlgqg
	LC13	1mi5	ihwyrQlpsqg	lfwyqQalgqg
	cf34	3ffc	lfwykQpssge	lywyqQilgqg
	BM3.3	2ol3	lfwykQtasge	mswyqQdlqkq
	BM3.3	1fo0	lfwykQtasge	mswyqQdlqkq
	BM3.3	1nam	lfwykQtasge	mswyqQdlqkq
5	D10	1d9k	fpwyrQfpgks	mywyrQdtghg
	2W20	3c6l	fpwyhQfpges	mywyrQdtghg
	AS01	3o4l	lywykQepgag	mfwyrQfpkqs
	KB5-C20	1kj2	fpwyqQfpgeg	mswyqQdlqkq
	HA1.7	1fyt	lfwyvQypnqg	mfwyrQdpglg
	HA1.7	1j8h	lfwyvQypnqg	mfwyrQdpglg
6	AHIII12.2	2jcc	lfwyvQhlnea	mywyrQdtghg
	AHIII12.2	2uwe	lfwyvQhlnea	mywyrQdtghg
	1934.4	2pxy	lfwyvQypgeg	mywyrQdtghg
	AHIII12.2	1lp9	lfwyvQhlnea	mywyrQdtghg

cl19	2z31	lfwyvQypgeg	mywyrQdtghg
B7	1bd2	flwykKypaeg	mswyrQdpgmg
TCR172.10	1u3h	fpwyqQfpgeg	mywyrQdtghg
B3K506	3c5z	lfwyvQypgeg	mywyrQdtghg
ELS4	2nx5	lfwyqQhagea	mywyrQdpghg
YAe62	3c60	lfwyvQysgeg	mywyrQdtghg
2C	2ckb	lfwyvQyprqg	mywyrQdtghg
2C	1mwa	lfwyvQyprqg	mywyrQdtghg
RA14	3gsn	lhwyrWetaks	mswyrQdpgmg
2C	1g6r	lfwyvQyprqg	mywyrQdtghg
Hy.1B1	3pl6	fpwykQelgkr	lywyrQslgqg
TK3 WT	3mv7	lfwyrQdpgkg	vywyqQsldqg
TK3 Q55H	3mv8	lfwyrQdpgkg	vywyqQsldqg
TK3 Q55A	3mv9	lfwyrQdpgkg	vywyqQsldqg
3A6	1zgl	lfwyvQypgeg	vswyqQtpgqg
2C m67 [T7-s]	3e2h	lfwyvQyprqg	mywyrQdtghe
2C T7	2icw	lfwyvQyprqg	mywyrQdtghe
MEL5	3hg1	ffwyrQysgks	lywyrQaagrg

S4 Table. Epitopes of the bound TCR structures.

C	Name	PDB	Peptide	MHC I/II α	MHC II β
1	1G4 c49c50	2f53	SLLMWITQC	HLA-A*0201	
	1G4	2bnr	SLLMWITQC	HLA-A*0201	
	1G4	2bnq	SLLMWITQV	HLA-A*0201	
	1G4 AV-wt	2f54	SLLMWITQC	HLA-A*0201	
	E8	2ian	GELIGTLNAAKVPAD	HLA-DRA*0101	HLA-DRB1*0101
	1G4 c58c61	2p5e	SLLMWITQC	HLA-A*0201	
	1G4 c58c62	2p5w	SLLMWITQC	HLA-A*0201	
	E8	2iam	GELIGILNAAKVPAD	HLA-DRA*0101	HLA-DRB1*0101
	1G4 c5c1	2pye	SLLMWITQC	HLA-A*0201	
2	TCR MS2-3C8	3o6f	FSWGAEGQRPGFG	HLA-DRA*0101	HLA-DRB1*0401
	A6	1qse	LLFGYPRYV	HLA-A*0201	
	A6	3h9s	MLWGYLQYV	HLA-A*0201	
	A6	3d39	LLFGFPVYV[k]	HLA-A*0201	
	A6	3d3v	LLFGFPVYV[l]	HLA-A*0201	
	A6	3pwp	LGYGFVNYI	HLA-A*0201	
	A6	1qsf	LLFGYPVAV	HLA-A*0201[a]	
	A6	2gj6	LLFGKPVYV[b]	HLA-A*0201	
	A6	1ao7	LLFGYPVYV	HLA-A*0201	
	A6	1qrn	LLFGYAVYV	HLA-A*0201	
3	JM22	2vlj	GILGFVFTL	HLA-A*0201	
	JM22	2vlk	GILGFVFTL	HLA-A*0201	
	JM22	1oga	GILGFVFTL	HLA-A*0201	
	JM22 [S99βA]	2vlr	GILGFVFTL	HLA-A*0201	
	SB27	2ak4	LPEPLPQGQLTAY	HLA-B*3508	
	SB27 [K16Dα]	3kxf	LPEPLPQGQLTAY	HLA-B*3508[c]	
	TCR 21.30	3mbe	GAMKRHGLDNYRGYSLGN	H2-Aa(d)	H2-Ab(NOD)
	OB.1A12	1ymm	ENPVVHFFKNIVTPR	HLA-DRA*0101	HLA-DRB1*1501
	OB.1A12	2wbj	FARVHFISALHGS	HLA-DRA*0101	HLA-DRB1*1501
4	KK50.4	2esv	VMAPRTLIL	HLA-E*0101	
	LC13	3kpr	EEYLKAWTF	HLA-B*4405	
	2B4	3qib	ADLIAYLKQATKG	H2-Ea(k)	H2-Eb(k)
	LC13	3kps	EEYLQAFTY	HLA-B*4405	
	226 TCR	3qiu	ADLIAYLKQATKG	H2-Ea(k)	H2-Eb(k)
	226 TCR	3qiw	ADLIAYLEQATKG	H2-Ea(k)	H2-Eb(k)
	2C m13 [T7-s]	3e3q	QLSPFPFDL	H2-L(d)[d]	
	2C m6 [T7-s]	2e7l	QLSPFPFDL	H2-L(d)	
	2C [T7-wt-s]	2oi9	QLSPFPFDL	H2-L(d)[d]	
	DM1	3dxa	EENLLDFVRF	HLA-B*4405	
	LC13	1mi5	FLRGRAYGL	HLA-B*0801	
	cf34	3ffc	FLRGRAYGL	HLA-B*0801	
	BM3.3	2ol3	SQYYYNSL	H2-K1(bm8)[e]	
	BM3.3	1fo0	INFDFNTI	H2-K1(b)	
	BM3.3	1nam	RGYVYQGL	H2-K1(b)	
5	D10	1d9k	GNSHRGAIEWEGIESG	H2-Aa(k)	H2-Ab(k)
	2W20	3c6l	FEAQKAKANKAVD	H2-Aa(b)	H2-Ab(b)[f]
	AS01	3o4l	GLCTLVAML	HLA-A*0201	
	KB5-C20	1kj2	KVITFIDL	H2-K1(b)	
	HA1.7	1fyt	PKYVKQNTLKLAT	HLA-DRA*0101	HLA-DRB1*0101
	HA1.7	1j8h	PKYVKQNTLKLAT	HLA-DRA*0101	HLA-DRB1*0401[j]
6	AHIII12.2	2jcc	ALWGFFPVL	HLA-A*0201[g]	
	AHIII12.2	2uwe	ALWGFFPVL	HLA-A*0201[h]	
	1934.4	2pxy	HSRGGASQYRPSQ	H2-Aa(u)	H2-Ab(u)
	AHIII12.2	1lp9	ALWGFFPVL	HLA-A*0201	
	cl19	2z31	RGGASQYRPSQ	H2-Aa(u)	H2-Ab(u)
	B7	1bd2	LLFGYPVYV	HLA-A*0201	
	TCR172.10	1u3h	SRGGASQYRPSQ	H2-Aa(u)	H2-Ab(u)
	B3K506	3c5z	FEAQKAKANKAVD	H2-Aa(b)	H2-Ab(b)[f]
	ELS4	2nx5	EPLPQGQLTAY	HLA-B*3501	

YAe62	3c60	FEAQKAKANKAVD	H2-Aa(b)	H2-Ab(b)[r]
2C	2ckb	EQYKFYSV	H2-K1(b)	
2C	1mwa	EQYKFYSV	H2-K1(bm3)[e]	
RA14	3gsn	NLVPMVATV	HLA-A*0201[i]	
2C	1g6r	SIYRYYGL	H2-K1(b)	
Hy.1B1	3pl6	NPVVHFFKNIVTPR	HLA-DQA1*0102	HLA-DQB1*0501
TK3 WT	3mv7	HPVGEADYFEY	HLA-B*3501	
TK3 Q55H	3mv8	HPVGEADYFEY	HLA-B*3501	
TK3 Q55A	3mv9	HPVGEADYFEY	HLA-B*3501	
3A6	1zgl	VHFFKNIVTPRTP	HLA-DRA*0101	HLA-DRB5*0101
2C m67	3e2h	QLSPFPFDL	H2-L(d)[d]	
(2C T7	2icw	PKYVKQNTLKLAT	HLA-DRA*0101	HLA-DRB1*0101)[s]
MEL5	3hg1	ELAGIGILTV	HLA-A*0201	
(JM22	2xn9	PKYVKQNTLKLAT	HLA-DRA*0101	HLA-DRB1*0101)[s]

Haplotypes of murine MHC H2 alleles are given in brackets to avoid confusion with footnotes.

a) Mutation I219A in MHC a_3 domain.

b) Y5 is mutated to K and linked to 4-(3-Indolyl)-butyric acid.

c) Mutation Q65A, T69A, Q155A

d) Solubility Mutations (F9Y, V12T, I23T)

e) H2-K1^{bm3} and H2-K1^{bm8} are natural mutants of H2-K1^{b}

f) Mutation R189K

g) Mutation W167V in the MHC binding pocket

h) Mutation T163A in the MHC binding pocket

i) Mutation A245V in the MHC a_3 domain

j) Mutation L209V in the MHC a_3 domain

k) Flouration: y5f^{4F}

l) Double-fluoration: y5f^{3,4FF}

s) Bound to super-antigen (Sag). The TCRs only directly contacts the SAgs and thus are not considered as MHC bound within the clustering process. However, the structures would associate with the cluster 6.

S5 Table. References to the TCR structures used for the analysis

Name	PDB	Ref.	Name	PDB	Ref.
1G4	2bnu	[1]	LC13	3kps	[2]
1G4	2bnq	[1]	MEL5	3hg1	[3]
1G4	2bnr	[1]	OB.1A12	2wbj	[4]
1G4 AV-wt	2f54	[5]	OB.1A12	1ymm	[6]
1G4 c5c1	2pyf	[7]	RA14	3gsn	[8]
1G4 c5c1	2pye	[7]	SB27	2ak4	[9]
1G4 c49c50	2f53	[5]	SB27[K16Dα]	3kxf	[10]
1G4 c58c62	2p5w	[7]	TCR MS2-3C8	3o6f	[11]
1G4 c58c61	2p5e	[7]	TK3 wt	3mv7	[12]
3A6	1zgl	[13]	TK3 Q55H	3mv8	[12]
A6	2gj6	[14]	TK3 Q55A	3mv9	[12]
A6	1qsf	[15]	1934,4	2pxy	[16]
A6	1qse	[15]	1F1E8	3mff	[17]
A6	3d3v	[18]	226 TCR	3qiu	[19]
A6	3d39	[18]	226 TCR	3qiw	[19]
A6	1qrn	[15]	2B4	3qib	[19]
A6	1ao7	[20]	2B4	3qjf	[19]
A6	3h9s	[21]	2C	1tcr	[22]
A6	3pwp	[23]	2C	1g6r	[24]
AS01	3o4l	[25]	2C	1mwa	[26]
B7	1bd2	[27]	2C	2ckb	[28]
cf34	3ffc	[29]	2C T7	2icw	[30]
DM1	3dxa	[31]	2C [T7-wt-s][a]	2oi9	[32]
DM1	3dx9	[31]	2C m13 [T7-s][a]	3e3q	[33]
E8	2ian	[34]	2C m6 [T7-s][a]	2e7l	[32]
E8	2iam	[34]	2C m67 [T7-s][a]	3e2h	[33]
E8	2ial	[34]	2W20	3c6l	[35]
ELS4	2nx5	[36]	5c.c7	3qjh	[19]
ELS4	2nw2	[36]	AHIII12.2	2uwe	[37]
HA1.7	1fyt	[38]	AHIII12.2	2jcc	[37]
HA1.7	1j8h	[39]	AHIII12.2	1lp9	[40]
Hy.1B1	3pl6	[41]	B3K506	3c5z	[35]
JM22	2vlj	[42]	BM3.3	1nam	[43]
JM22	2vlk	[42]	BM3.3	1fo0	[44]
JM22	1oga	[45]	BM3.3	2ol3	[46]
JM22	2vlm	[42]	cl19	2z31	[16]
JM22	2xn9	[47]	D10	1d9k	[48]
JM22	2xna	[47]	KB5-C20	1kj2	[49]
JM22 [S99βA]	2vlr	[42]	N15	1nfd	[50]
KK50.4	2esv	[51]	TCR 21.30	3mbe	[52]
LC13	1mi5	[53]	TCR172.10	1u3h	[54]
LC13	1kgc	[55]	YAe62	3c60	[35]
LC13	3kpr	[2]			

[a] WT with solubility mutations acc. to ref. [32]. More detailed information is available in Table S2.

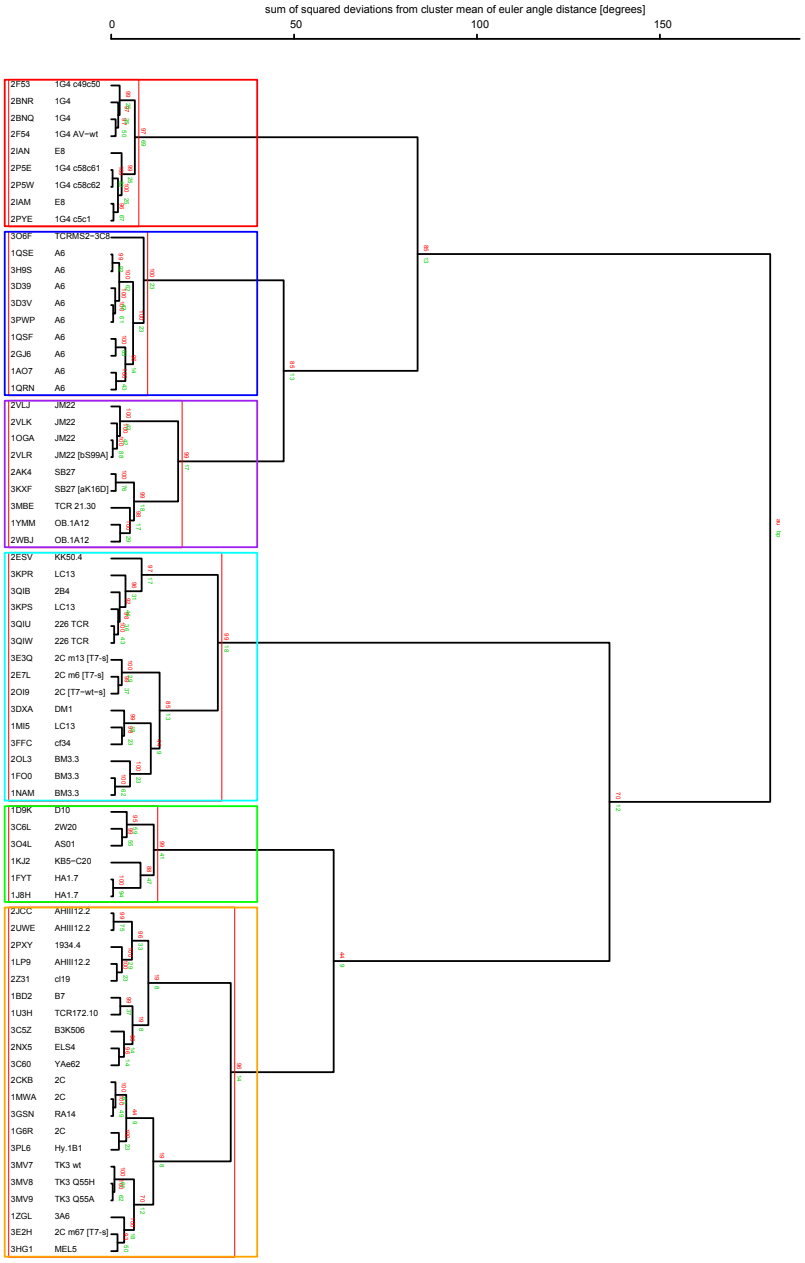
SI References:

1. Chen J-L, Stewart-Jones G, Bossi G, Lissin NM, Wooldridge L, et al. (2005) Structural and kinetic basis for heightened immunogenicity of T cell vaccines. J Exp Med 201: 1243-1255.
2. Macdonald WA, Chen Z, Gras S, Archbold JK, Tynan FE, et al. (2009) T cell allorecognition via molecular mimicry. Immunity 31: 897-908.
3. Cole DK, Yuan F, Rizkallah PJ, Miles JJ, Gostick E, et al. (2009) Germ line-governed recognition of a cancer epitope by an immunodominant human T-cell receptor. J Biol Chem 284: 27281-27289.
4. Harkiolaki M, Holmes SL, Svendsen P, Gregersen JW, Jensen LT, et al. (2009) T cell-mediated autoimmune disease due to low-affinity crossreactivity to common microbial peptides. Immunity 30: 348-357.
5. Dunn SM, Rizkallah PJ, Baston E, Mahon T, Cameron B, et al. (2006) Directed evolution of human T cell receptor CDR2 residues by phage display dramatically enhances affinity for cognate peptide-MHC without increasing apparent cross-reactivity. Protein Sci 15: 710-721.
6. Hahn M, Nicholson MJ, Pyrdol J, Wucherpfennig KW (2005) Unconventional topology of self peptide-major histocompatibility complex binding by a human autoimmune T cell receptor. Nat Immunol 6: 490-496.
7. Sami M, Rizkallah PJ, Dunn S, Molloy P, Moysey R, et al. (2007) Crystal structures of high affinity human T-cell receptors bound to peptide major histocompatibility complex reveal native diagonal binding geometry. Protein Eng Des Sel 20: 397-403.
8. Gras S, Saulquin X, Reiser J-B, Debeaupuis E, Echasserieau K, et al. (2009) Structural bases for the affinity-driven selection of a public TCR against a dominant human cytomegalovirus epitope. J Immunol 183: 430-437.
9. Tynan FE, Burrows SR, Buckle AM, Clements CS, Borg NA, et al. (2005) T cell receptor recognition of a 'super-bulged' major histocompatibility complex class I-bound peptide. Nat Immunol 6: 1114-1122.
10. Burrows SR, Chen Z, Archbold JK, Tynan FE, Beddoe T, et al. (2010) Hard wiring of T cell receptor specificity for the major histocompatibility complex is underpinned by TCR adaptability. Proc Natl Acad Sci U S A 107: 10608-10613.
11. Yin Y, Li Y, Kerzic MC, Martin R, Mariuzza RA (2011) Structure of a TCR with high affinity for self-antigen reveals basis for escape from negative selection. EMBO J 30: 1137-1148.
12. Gras S, Chen Z, Miles JJ, Liu YC, Bell MJ, et al. (2010) Allelic polymorphism in the T cell receptor and its impact on immune responses. J Exp Med 207: 1555-1567.
13. Li Y, Huang Y, Lue J, Quandt JA, Martin R, et al. (2005) Structure of a human autoimmune TCR bound to a myelin basic protein self-peptide and a multiple sclerosis-associated MHC class II molecule. EMBO J 24: 2968-2979.
14. Gagnon SJ, Borbulevych OY, Davis-Harrison RL, Turner RV, Damirjian M, et al. (2006) T cell receptor recognition via cooperative conformational plasticity. J Mol Biol 363: 228-243.
15. Ding YH, Baker BM, Garboczi DN, Biddison WE, Wiley DC (1999) Four A6-TCR/peptide/HLA-A2 structures that generate very different T cell signals are nearly identical. Immunity 11: 45-56.

16. Feng D, Bond CJ, Ely LK, Maynard J, Garcia KC (2007) Structural evidence for a germline-encoded T cell receptor-major histocompatibility complex interaction 'codon'. Nat Immunol 8: 975-983.
17. van Boxel GI, Holmes S, Fugger L, Jones EY (2010) An alternative conformation of the T-cell receptor alpha constant region. J Mol Biol 400: 828-837.
18. Piepenbrink KH, Borbulevych OY, Sommese RF, Clemens J, Armstrong KM, et al. (2009) Fluorine substitutions in an antigenic peptide selectively modulate T-cell receptor binding in a minimally perturbing manner. Biochem J 423: 353-361.
19. Newell EW, Ely LK, Kruse AC, Reay PA, Rodriguez SN, et al. (2011) Structural basis of specificity and cross-reactivity in T cell receptors specific for cytochrome c-I-E(k). J Immunol 186: 5823-5832.
20. Garboczi DN, Ghosh P, Utz U, Fan QR, Biddison WE, et al. (1996) Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. Nature 384: 134-141.
21. Borbulevych OY, Piepenbrink KH, Gloor BE, Scott DR, Sommese RF, et al. (2009) T cell receptor cross-reactivity directed by antigen-dependent tuning of peptide-MHC molecular flexibility. Immunity 31: 885-896.
22. Garcia KC, Degano M, Stanfield RL, Brunmark A, Jackson MR, et al. (1996) An alphabeta T cell receptor structure at 2.5 A and its orientation in the TCR-MHC complex. Science 274: 209-219.
23. Borbulevych OY, Piepenbrink KH, Baker BM (2011) Conformational melding permits a conserved binding geometry in TCR recognition of foreign and self molecular mimics. J Immunol 186: 2950-2958.
24. Degano M, Garcia KC, Apostolopoulos V, Rudolph MG, Teyton L, et al. (2000) A functional hot spot for antigen recognition in a superagonist TCR/MHC complex. Immunity 12: 251-261.
25. Miles JJ, Bulek AM, Cole DK, Gostick E, Schauenburg AJA, et al. (2010) Genetic and structural basis for selection of a ubiquitous T cell receptor deployed in Epstein-Barr virus infection. PLoS Pathog 6: e1001198.
26. Luz JG, Huang M, Garcia KC, Rudolph MG, Apostolopoulos V, et al. (2002) Structural comparison of allogeneic and syngeneic T cell receptor-peptide-major histocompatibility complex complexes: a buried alloreactive mutation subtly alters peptide presentation substantially increasing V(beta) Interactions. J Exp Med 195: 1175-1186.
27. Ding YH, Smith KJ, Garboczi DN, Utz U, Biddison WE, et al. (1998) Two human T cell receptors bind in a similar diagonal mode to the HLA-A2/Tax peptide complex using different TCR amino acids. Immunity 8: 403-411.
28. Garcia KC, Degano M, Pease LR, Huang M, Peterson PA, et al. (1998) Structural basis of plasticity in T cell receptor recognition of a self peptide-MHC antigen. Science 279: 1166-1172.
29. Gras S, Burrows SR, Kjer-Nielsen L, Clements CS, Liu YC, et al. (2009) The shaping of T cell receptor recognition by self-tolerance. Immunity 30: 193-203.
30. Wang L, Zhao Y, Li Z, Guo Y, Jones LL, et al. (2007) Crystal structure of a complete ternary complex of TCR, superantigen and peptide-MHC. Nat Struct Mol Biol 14: 169-171.
31. Archbold JK, Macdonald WA, Gras S, Ely LK, Miles JJ, et al. (2009) Natural micropolymorphism in human leukocyte antigens provides a basis for genetic control of antigen recognition. J Exp Med 206: 209-219.
32. Colf LA, Bankovich AJ, Hanick NA, Bowerman NA, Jones LL, et al. (2007) How a single T cell receptor recognizes both self and foreign MHC. Cell 129: 135-146.

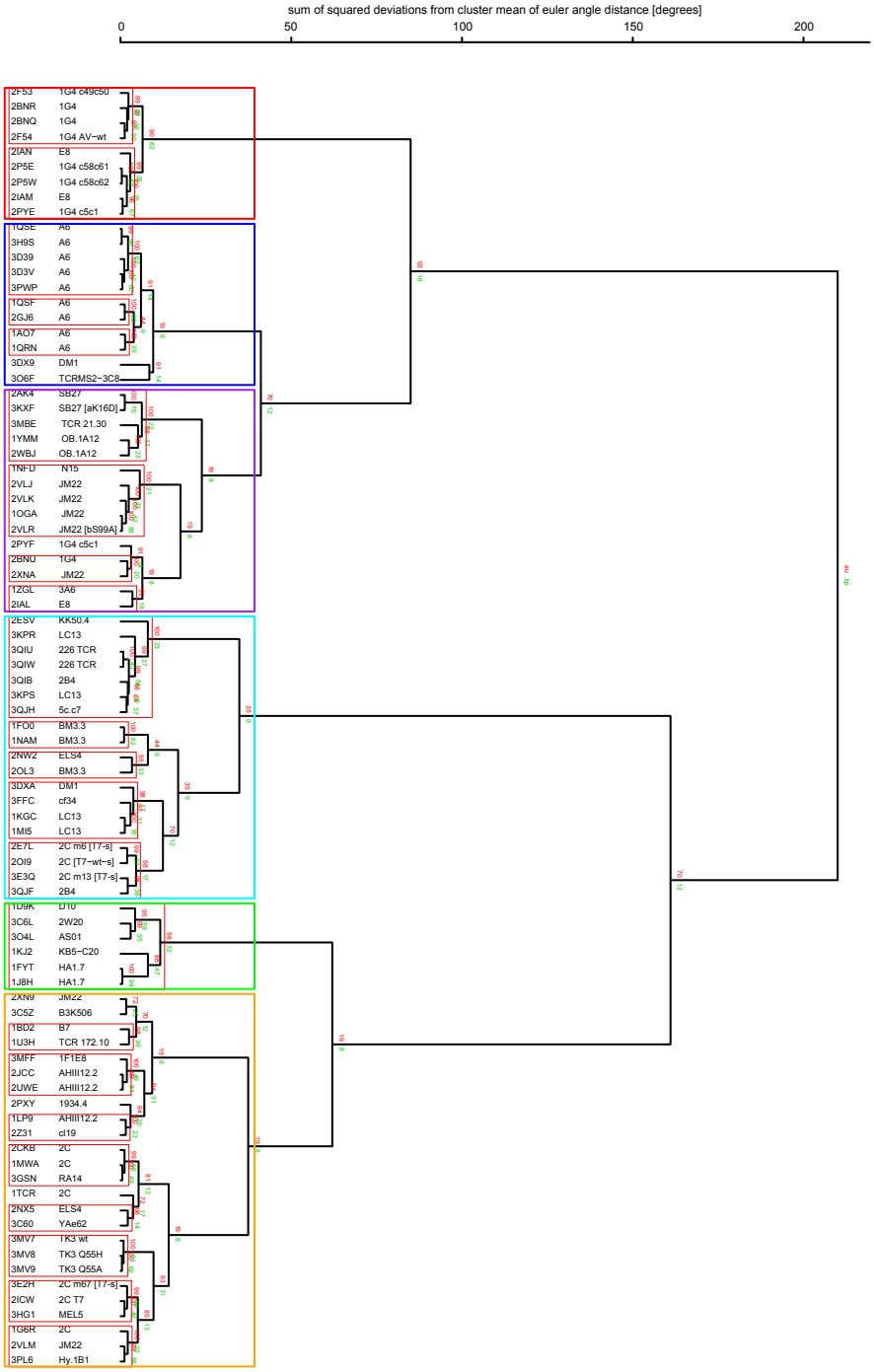
33. Jones LL, Colf LA, Stone JD, Garcia KC, Kranz DM (2008) Distinct CDR3 conformations in TCRs determine the level of cross-reactivity for diverse antigens, but not the docking orientation. J Immunol 181: 6255-6264.
34. Deng L, Langley RJ, Brown PH, Xu G, Teng L, et al. (2007) Structural basis for the recognition of mutant self by a tumor-specific, MHC class II-restricted T cell receptor. Nat Immunol 8: 398-408.
35. Dai S, Huseby ES, Rubtsova K, Scott-Browne J, Crawford F, et al. (2008) Crossreactive T Cells spotlight the germline rules for alphabeta T cell-receptor interactions with MHC molecules. Immunity 28: 324-334.
36. Tynan FE, Reid HH, Kjer-Nielsen L, Miles JJ, Wilce MCJ, et al. (2007) A T cell receptor flattens a bulged antigenic peptide presented by a major histocompatibility complex class I molecule. Nat Immunol 8: 268-276.
37. Miller PJ, Pazy Y, Conti B, Riddle D, Appella E, et al. (2007) Single MHC mutation eliminates enthalpy associated with T cell receptor binding. J Mol Biol 373: 315-327.
38. Hennecke J, Carfi A, Wiley DC (2000) Structure of a covalently stabilized complex of a human alphabeta T-cell receptor, influenza HA peptide and MHC class II molecule, HLA-DR1. EMBO J 19: 5611-5624.
39. Hennecke J, Wiley DC (2002) Structure of a complex of the human alpha/beta T cell receptor (TCR) HA1.7, influenza hemagglutinin peptide, and major histocompatibility complex class II molecule, HLA-DR4 (DRA*0101 and DRB1*0401): insight into TCR cross-restriction and alloreactivity. J Exp Med 195: 571-581.
40. Buslepp J, Wang H, Biddison WE, Appella E, Collins EJ (2003) A correlation between TCR Valpha docking on MHC and CD8 dependence: implications for T cell selection. Immunity 19: 595-606.
41. Sethi DK, Schubert DA, Anders A-K, Heroux A, Bonsor DA, et al. (2011) A highly tilted binding mode by a self-reactive T cell receptor results in altered engagement of peptide and MHC. J Exp Med 208: 91-102.
42. Ishizuka J, Stewart-Jones GBE, van der Merwe A, Bell JI, McMichael AJ, et al. (2008) The structural dynamics and energetics of an immunodominant T cell receptor are programmed by its Vbeta domain. Immunity 28: 171-182.
43. Reiser J-B, Darnault C, Grégoire C, Mosser T, Mazza G, et al. (2003) CDR3 loop flexibility contributes to the degeneracy of TCR recognition. Nat Immunol 4: 241-247.
44. Reiser JB, Darnault C, Guimezanes A, Grégoire C, Mosser T, et al. (2000) Crystal structure of a T cell receptor bound to an allogeneic MHC molecule. Nat Immunol 1: 291-297.
45. Stewart-Jones GBE, McMichael AJ, Bell JI, Stuart DI, Jones EY (2003) A structural basis for immunodominant human T cell receptor recognition. Nat Immunol 4: 657-663.
46. Mazza C, Auphan-Anezin N, Gregoire C, Guimezanes A, Kellenberger C, et al. (2007) How much can a T-cell antigen receptor adapt to structurally distinct antigenic peptides? EMBO J 26: 1972-1983.
47. Saline M, Rödström KEJ, Fischer G, Orekhov VY, Karlsson BG, et al. (2010) The structure of superantigen complexed with TCR and MHC reveals novel insights into superantigenic T cell activation. Nat Commun 1: 119.
48. Reinherz EL, Tan K, Tang L, Kern P, Liu J, et al. (1999) The crystal structure of a T cell receptor in complex with peptide and MHC class II. Science 286: 1913-1921.
49. Reiser JB, Grégoire C, Darnault C, Mosser T, Guimezanes A, et al. (2002) A T cell receptor CDR3beta loop undergoes conformational changes of unprecedented magnitude upon binding to a peptide/MHC class I complex. Immunity 16: 345-354.

50. Wang J, Lim K, Smolyar A, Teng M, Liu J, et al. (1998) Atomic structure of an alphabeta T cell receptor (TCR) heterodimer in complex with an anti-TCR fab fragment derived from a mitogenic antibody. EMBO J 17: 10-26.
51. Hoare HL, Sullivan LC, Pietra G, Clements CS, Lee EJ, et al. (2006) Structural basis for a major histocompatibility complex class Ib-restricted T cell response. Nat Immunol 7: 256-264.
52. Yoshida K, Corper AL, Herro R, Jabri B, Wilson IA, et al. (2010) The diabetogenic mouse MHC class II molecule I-Ag7 is endowed with a switch that modulates TCR affinity. J Clin Invest 120: 1578-1590.
53. Kjer-Nielsen L, Clements CS, Purcell AW, Brooks AG, Whisstock JC, et al. (2003) A structural basis for the selection of dominant alphabeta T cell receptors in antiviral immunity. Immunity 18: 53-64.
54. Maynard J, Petersson K, Wilson DH, Adams EJ, Blondelle SE, et al. (2005) Structure of an autoimmune T cell receptor complexed with class II peptide-MHC: insights into MHC bias and antigen specificity. Immunity 22: 81-92.
55. Kjer-Nielsen L, Clements CS, Brooks AG, Purcell AW, McCluskey J, et al. (2002) The 1.5 A crystal structure of a highly selected antiviral T cell receptor provides evidence for a structural basis of immunodominance. Structure 10: 1521-1532.

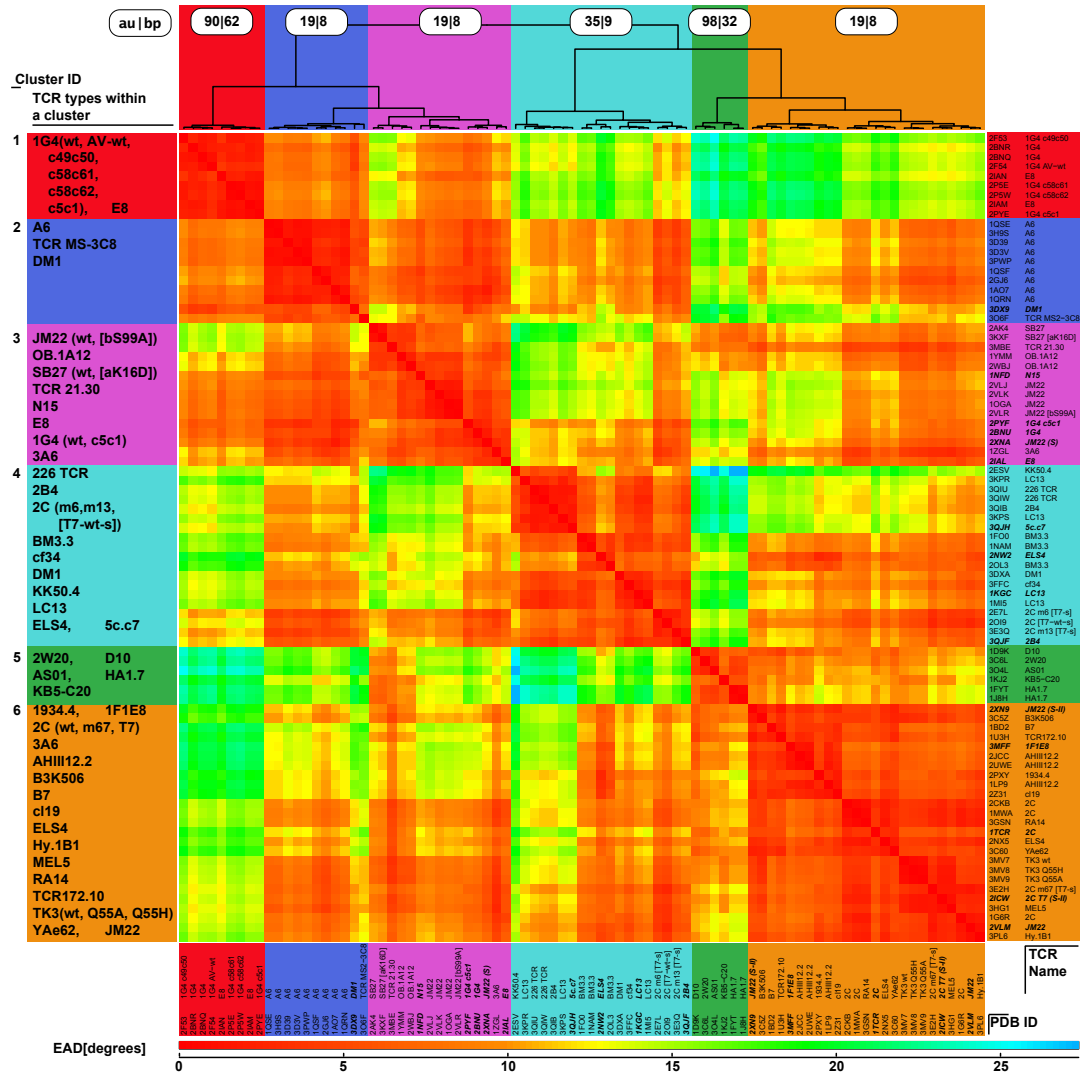


S1 Fig. Bootstrapping Dendrogramm of the Clustering of the MHC bound TCRs. Pairwise Euler-angle distances (EAD) were determined for all MHC-bound TCR structures and the free TCR structures according to Formula 1. Structures containing more than one biological unit were merged to one unique geometry. The distance matrix was hierarchically clustered using the Ward update formula. For each subtree of the dendrogram, the au (approximately biased) and the bp (bootstrapping probability) according to the bootstrapping method [1] are provided. We identified six siginificant clusters of an au-value greater than 95%. The clusters are marked by colored boxes.

1. Suzuki R, Shimodaira H (2006) Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22: 1540-1542.



S2 Fig. Bootstrapping Dendrogram of the Clustering of the free TCRs together with the MHC bound TCRs. For details see S1 Fig. The bootstrapping dendrogram was computed for the bound and free TCRs. The clusters of the unbound case are marked by colored boxes for comparison. Significant clusters are only found for smaller subtrees.



S3 Fig. Geometry Clusters of bound and unbound TCRs. Pairwise Euler-angle distances (EAD) were determined for all MHC-bound TCR structures and the free TCR structures according to Formula 2 (see Materials and Methods). Structures containing more than one biological unit were merged to one unique geometry. The distance matrix was hierarchically clustered using the Ward update formula. We identified six significant clusters, using a bootstrapping approach [1]. Notably, in most of the cases, TCRs of the same type occur in the same cluster. Upper panel: Clustering dendrogram with bootstrapping results (au=approximately unbiased, bp=bootstrapping probability) Left panel: TCR types occurring within a cluster. Right/lower panel: Structure PDB identifiers and corresponding TCR names. MHC unbound TCR structures are indicated by bold-italics fonts. Central panel: Pairwise Euler-angle distances (EAD). The color key is provided in the bottom of the figure.

1. Suzuki R, Shimodaira H (2006) Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22: 1540-1542.

Appendix C

Embedded Publication 3

C.1 Copy Permissions

The manuscript was submitted to the open access journal *BMC Structural Biology*. If accepted, the article will be published with the following copyright remark: *This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.*

C.2 Full Article

DynaDom: Structure-based prediction of T cell receptor inter-domain and T cell receptor-peptide-MHC association angles

Thomas Hoffmann, Antoine Marion, and Iris Antes*

Department of Biosciences and Center for Integrated Protein Science Munich

Technische Universität München

Emil-Erlenmeyer-Forum 8

85354 Freising

Germany

Email addresses:

thomas.hoffmann@tum.de

antoine.marion@tum.de

Corresponding author:

antes@tum.de

Keywords

T-cell recognition; TCR structural modeling; epitope prediction; glutamine side chain prediction; protein domain association angles; immunoinformatics; adoptive T-cell therapy; vaccine design

Abstract

Background

T cell receptor (TCR) molecules are involved in the adaptive immune response as they distinguish between self- and foreign-peptides, presented in major histocompatibility complex molecules (pMHC). Former studies showed that the association angles of the TCR variable domains (Vα/Vβ) can differ significantly and change upon binding to the pMHC complex. These changes can be described as a rotation of the domains around a general Center of Rotation, characterized by the interaction of two highly conserved glutamine residues.

Methods

We developed a computational method, DynaDom, for the prediction of TCR Vα/Vβ inter-domain and TCR/pMHC orientations in TCRpMHC complexes, which allows predicting the orientation of multiple protein-domains. In addition, we implemented a new approach to predict the correct orientation of the carboxamide endgroups in glutamine and asparagine residues, which can also be used as an external, independent tool.

Results

The approach was evaluated for the remodeling of 75 experimental structures of TCRpMHC complexes. We show that the DynaDom method predicts the correct orientation of the TCR Vα/Vβ angles in 96% and 89% of the cases, for the poses with the best RMSD and best interaction energy, respectively. For the concurrent prediction of the TCR Vα/Vβ and pMHC orientations, the respective rates reached 74% and 72%. Through an exhaustive analysis, we could show that the pMHC placement can be further improved by a straightforward, yet very time intensive extension of the current approach.

Conclusions

The results obtained in the present remodeling study prove the suitability of our approach for interdomain-angle optimization. In addition, the high prediction rate obtained specifically for the energetically highest ranked poses further demonstrates that our method is a powerful candidate for blind prediction. Therefore it should be well suited as part of any accurate atomistic modeling pipeline for TCRpMHC complexes and potentially other large molecular assemblies.

Background

An early event in the T cell mediated immune response is the recognition of pathogenic peptides contained in major histocompatibility complex (MHC) molecules. The capability of the vertebrate immune system to distinguish between a vast variety of pathogenic- and self-peptides is achieved by a tremendous population of different T cell variants (i.e., in a magnitude estimated from 10^6 to 10^7), which differ from each other in the T cell receptor (TCR) [1-3]. Such a diversity results from the combination of two membrane anchored TCR chains (α and β), which are encoded by gene segments joined in a process known as v(d)j recombination [4]. As depicted in Fig. 1, each chain consists of two immunoglobulin-like domains, the variable domain (further referred to as Vα and Vβ) and the constant domain. The v(d)j combination process occurs during the T cell maturation in the thymus, where variable (v) and joining (j) gene segments are combined while nucleotides are randomly introduced within the variable domains (V). In the case of the Vβ domain, an additional short segment is inserted in between the v and j segments, further increasing the TCR diversity (d). The binding interface of the TCR to the peptide-MHC molecule complex (pMHC) is formed by loops named as complementary determining regions (CDR), and each chain of TCR contains three CDRs. While the primary structure of CDR1 and CDR2 loops evolves together with the MHC molecules [5], the sequence of CDR3 loops is determined by the v(d)j recombination and thus exhibits a higher diversity [6].

The number of resolved bound and unbound TCR structures has drastically increased to 200 in the Protein Data Bank [7] during the past few years. Nevertheless, considering the vast variety of TCRs and the high polymorphism of the MHC molecules, the development of reliable structural methods is of crucial importance in

order to complement time consuming experimental structural techniques [8]. Such modeling approaches can help in the field of rational TCR design/optimization (*e.g.,* adoptive T cell cancer therapy) [9, 10], in the context of vaccine design [11, 12], and in the development of a consistent theory for T cell signal transduction, which is still not fully understood [13].

Over the past two decades, many theoretical methodologies have been developed and applied to model and predict TCRpMHC interactions.

The main focus in the area has been on the prediction of the peptide/MHC interaction without explicit consideration of the T-cell receptor as the experimental study of MHC-peptide binding has been a very active field since the mid-90s whereas the systematic investigation of the T-cell response started about a decade later. In addition, peptide-MHC binding is a necessary prerequisite for the T-cell response and thus has by itself already a highly predictive value. Therefore various sequence and structure based prediction tools have been developed of MHC-peptide binding in the past decades [14] [15]. Next to MHC-peptide specific structure-based prediction methods such as EpiDock, PREDEP, pDOCK, DynaPred, or DockTope [16-20], also general molecular docking approaches were applied [21, 22].

The first atomistic model of a TCRpMHC complex was built in 1995 by Almagro *et al.* using homology modeling and molecular dynamics techniques [23], before the first X-ray structures of a TCR (1tcr [24]) and of a TCRpMHC complex (1ao7[25]) were solved in 1996. Later, Michielin *et al.* realized a homology model of the T1 TCR structure bound to the photoreactive PbSC peptide and to the murin K^d MHC class I molecule, using the 1ao7 crystal structure of the TCRpMHC complex as a template [8]. The authors applied a methodology combining the MODELLER program with simulated annealing techniques [26], and suggested a rational homology model, which was refined based on previous mutation studies [27]. Further developments of the

approach led to the TCRep 3D method [28], which was recently applied in the context of rational TCR design [10]. In addition, Haidar *et al.* enhanced the affinity of the A6 TCR to TAX:HLA-A2 for about 100-fold using a structure-based model [29]. More recently, Pierce *et al.* [30] developed an approach based on their scoring function ZAFFI and on the Rosetta interface mutagenesis tool [31] to identify relevant point mutations that could increase the affinity of a TCR to a pMHC complex in the field of therapeutic immunology. The method allowed to optimize the DMF5 TCR to bind the ELAGIGILTV:HLA-A2 complex with a remarkable ~400-fold higher affinity. The same group also developed TCRFlexDock, a method to model a pMHC ligand onto a TCR that takes advantage of the Monte Carlo-based RosettaDock protocol [32, 33]. For a benchmark test set of twenty structure [33], the prediction of near native models was reached in 80% of the cases. The TCRFlexDock method was recently applied to predict models of TCRs bound to MHC like ligands such as CD1 and MR1 [34]. In that work, the authors showed that the use of multiple docking starting positions significantly improves the performance of the prediction.

In order to achieve an accurate molecular model of TCRpMHC complexes, it is necessary to consider several topological aspects of this sophisticated system. First, a precise description on an atomistic level is required, since small alterations in the TCR's or in the ligand's sequence can drastically affect the transduced signal [35]. As it was shown in other studies, mutations in the receptor or in the ligand can modify the binding affinity and thus the relative placement of the two units of the complex [36-39].

A second aspect to consider is the variation of the Vα/Vβ inter-domain angle within the TCR, as this is a system specific feature, and as it can adapt upon binding of the pMHC. The analysis of the inter-domain angle between the Vα and Vβ TCR domains as well as its influence on the binding of pMHC was analyzed in several

computational studies, which compared broad sets of TCR structures. Notably, by applying the pseudo-dyad method, McBeth *et al.* suggested that the resulting observed differences between the free and the MHC bound forms of TCRs constitute a feature of the receptor to adapt to different ligands, thus allowing cross reactivity [40]. Dunbar *et al.* analyzed a non-redundant set of TCRs with the ABangle methodology [41], which describes both the Vα- and the Vβ-orientation in an absolute manner, by considering a torsion angle, four bend angles and one distance as descriptors [42]. In the context of rational TCR-like antibody design, the authors found that antibodies adopt angles comprised in a different range than the one observed for TCRs. In our previous work [43], we analyzed the relative Vα- and Vβ-orientation by reducing the variable domains to cuboids, which served as basis for a distance based clustering. We observed that TCRs belonging to the same clonotype associate in the same angular cluster. Furthermore, we identified a Center of Rotation (further referred to as CoR_{β} and depicted in Fig. 1) and determined its location in the middle of a conserved interaction between two glutamine residues, one in the Vα and one in the Vβ domain. The various inter-domain angles in the evaluation set could be obtained through a rotation around this center. Recent studies, including ours, further emphasized the large range of values that the TCR Vα/Vβ inter-domain angle can adopt [40, 43, 44] and thus its influence on the positioning of the ligand binding CDR loops. These results suggest that next to the orientation of the pMHC ligand with respect to the TCR [24, 25, 36, 45, 46], also the Vα/Vβ inter-domain angle should be explicitly taken in account to assess an accurate homology modeling of TCRpMHC complexes. This last comment is in agreement with recent observations about the dynamics of the TCRpMHC system and the influence of the TCR on the pMHC structure [44, 47]. In addition, it was shown for antibodies that the consideration of the V_H-V_L angles for homology modeling can increase the accuracy considerably [41, 48, 49]. In this

context, Dunbar *et al.* identified key structural parameters, which provide a comprehensive description of the movement of the V_H and V_L domains with respect to each other [41]. Based on these features and on their respective values in the available antibody structures, Bujotzek *et al.* trained a predictor for the association of the two domains [48]. The authors further concluded that the consideration of the association angles is crucial for the prediction of highly accurate homology models of antibodies [49].

Along the course of the present study, we pointed out a third topological aspect that can have an impact on the success of TCRpMHC complexes modeling. The Vα/Vβ orientation directly depends on the proper interaction of two specific glutamine residues. During protein structure elucidation by X-ray crystallography, the ambiguous electron densities of nitrogen and oxygen atoms can hamper the correct assignment of these two elements. In the case of asparagine and glutamine residues, this often leads to misassigned atoms in the carboxamide group of the side chain. The detailed investigation of high-resolution structures shows that approximately 20% of these residues are assigned in a wrong flip state, leading to a non-optimal hydrogen bond network [50-53]. The respective orientation of asparagine and glutamine residues has a dramatic impact on most of molecular modeling techniques [53], and should be corrected by considering their direct environment. Due to this significance, several approaches have been developed in order to address this issue. Among those, the most popular ones are HBPLUS (X-PLOR package) [52], NETWORK (WHAT IF package) [53], Reduce (MolProbity package) [50, 54, 55], NQ Flipper [51, 56, 57], the Independent Cluster Decomposition Algorithm (ICDA) [58], Protonate 3D [59], Protoss [60, 61], and the Computational Titration method [62].

Despite the great improvements in TCRpMHC complex modeling achieved during the past decades, some of the critical aspects described above are still not taken into

account. To the best of our knowledge, none of the currently available methodologies explicitly include the adaption of the Vα/Vβ inter-domain angles, although these have a direct impact on the disposition of the CDR loops, and as a consequence, on the contact between the TCR and the pMHC ligand.

In what follows, we present a new method, DynaDom, for the prediction of TCR Vα/Vβ inter-domain and TCR/pMHC association angles. We implemented our approach into the DynaCell suite [63], a general force-field-based molecular modeling program developed in our group. Our new method uses an extendable multidimensional rigid body optimization approach based on the work by Mirzaei *et al.* [64]. The implementation is specifically designed in a way that allows for an arbitrary definition of rigid bodies and for the inclusion of local flexibility on different levels (e.g., from the domain to the residue level) into the modeling pipeline (Fig 2.). As a first application, we evaluate here the DynaDom method for the remodeling of a large set of TCR and TCRpMHC complexes. This evaluation intends to determine the general capability of a rotation-based algorithm and the relevance of our CoR-concept [43] for the successful prediction of association angles. Notably, we demonstrate here that it is possible to distinguish between correct and wrong models by solely using the force-field-based interaction energy computed between the different units of the complex. This is indeed very promising for future blind homology modeling of TCRpMHC complexes and others, especially if a sufficient amount of experimental data is not available for the training of an application-specific, knowledge–based scoring function.

Methods

Our procedure is intended to predict the relative orientation of the Vα and Vβ TCR domains as well as the binding of the pMHC ligand using a rigid body optimization

procedure. The new DynaDom prediction method is directly derived from our previous analysis of the structural features of TCRpMHC complexes [43] and uses the same theoretical framework as defined therein. In ref. [43], we performed a comprehensive, systematic analysis of the Vα/Vβ inter-domain angles in a set of 85 structures. Among this set, one structure was chosen as a coordinate frame reference (PDB-ID 2bnu [37]) and the Vα domain of the remaining TCRs in the set were aligned to this reference. The fixed Vα and the mobile Vβ domains were represented by unified cuboids based on the backbone atoms of the conserved residues that define the framework region. In this context, the Euler angle components associated with the positions of the Vβ cuboids were computed with respect to the Vα cuboid for each of the 85 structures in the set. The analysis of the obtained pairwise Euler angle distance matrix resulted in the identification of six structural clusters. In addition, we determined, in our previous study, a center of rotation for the Vβ domains (further referred to as CoR_{β}) by discretizing each Vβ cuboid into a grid and by identifying the grid point with the lowest deviation between the different Vβ cuboids. We also pointed out in ref. [43] that the identified center of rotation -CoR_{β}- is characterized by two highly conserved glutamine residues (Q) interacting with each other to form a stabilizing hydrogen bond network between Vα and Vβ. As we briefly mentioned in the Introduction section, the ambiguous character of the electron density of carboxamide groups in Q side chains can often lead to a misassignement of nitrogen and oxygen atoms. This issue requires further attention to ensure a proper modeling at an atomistic level, especially when considering the critical position of these residues in TCR complexes.

Based on our former, above-discussed observations, we designed the new DynaDom prediction method as a rotation-based algorithm. It uses the original concept of a center of rotation to explicitly account for the Vα/Vβ inter-domain angle in the

association process. In addition, due to the presence of glutamine residues at the CoR_β center of rotation, DynaDom includes a new strategy for the correction of misassigned Q residue side chains as an integral part of its prediction pipeline. In what follows, we present the general structural basis of DynaDom together with the extension of the Center of Rotation concept to the particular case of TCRpMHC association. Then, we detail the theoretical framework of our rigid body optimization algorithm and describe the corresponding algorithm pipeline. Finally, we define the particular data set used in the present test application of DynaDom for the remodeling of TCR and TCRpMHC complexes.

Structural basis for the prediction algorithm

As we previously introduced in ref. [43], we also define here the frame of the Vα domain in the 2bnu structure as the basic coordinate system (i.e., the Vα domain of each considered structure is first aligned to the one of 2bnu). In addition, each component of the TCRpMHC complex (i.e., Vα, Vβ, and pMHC) is described as a unified cuboid. The orientation of Vβ with respect to Vα is determined by the previously identified Center of Rotation, CoR_β (Fig. 1).

In the present application, we additionally evaluated the approach for the refinement of the association angles between the pMHC ligand and the TCR domains. Based on the analysis of the early structures of this complex [24, 25, 36], it has been suggested that the binding mode of the TCRpMHC complex occurs in a generally diagonal mode [46]. Rudolph *et al.* introduced a general unified method to measure the binding angle of TCRs with respect to their ligands and determined the angular range of 24 complexes [45], based on a general rotation axis. However, beside this main rotation, two more rotational operations as well as a translation of the pMHC are necessary to completely describe all the transformations in a three-dimensional space. To further

consider this feature, we adapted our previously introduced cuboid method [43] and measured the three Euler angle components of the TCR/pMHC orientation. In a similar way to the determination of CoR_β, we define a center of rotation for the orientation of the pMHC cuboid relative to the Vα domain, which we shall refer to as CoR_μ in what follows (see Fig. 1). Unlike CoR_β, CoR_μ does not correspond any conserved residue and lies in the middle of the peptide binding groove of the pMHC complex. As CoR_β and CoR_μ can be defined by the Vβ and pMHC coordinates alone, we further use their relative location within these domains as rotational centers for the rigid body optimization. These locations will be named as CoR_β- and CoR_μ-based rotational centers, respectively.

Prior to a prediction process, the Vα, Vβ, and pMHC domains together with their cuboids must be placed in a given starting conformation. Motivated by our former identification of the structural clusters representing the TCR Vα/Vβ association angles [43], we define here 11 different relative orientation of the Vβ domain with respect to Vα (for details see Additional file 1: Text S1). These 11 orientations are used to produce multiple starting conformations of the complex for each remodeling attempt, to ensure a proper sampling.

For the modeling of the TCR/pMHC association angles, only one single starting conformation is used for the placement of pMHC to lower the computational expense. To determine the starting orientation of the pMHC cuboid, we analyzed the Vα/pMHC angles associated with all structures considered in our set (see subsection Structural data sets). The crystal structure 3e3q [65] showed the lowest angular deviation with respect to the others and was chosen as a reference. In the present application, this reference orientation is used for the preplacement of pMHC with respect to Vα, for each of the 11 Vα:Vβ complexes (i.e., the complexes formed by the association of the Vα and Vβ domains considering the 11 starting orientations

described above and in the Supporting Information). Finally, to compensate for the use of one single starting orientation of the pMHC ligand, the corresponding cuboid is additionally translated away from the TCR, such that unphysical locking events due to side chain overlap are avoided. This translation is performed along an axis described by the normal vector to the plane defined by the MHC β-sheet backbone atoms of 3e3q in the frame of 2bnu (see sketches in Fig. 2).

Rigid body optimization

We implemented our method within the DynaCell suite [63] using a rigid body energy minimization approach based on the work by Mirzaei *et al.* [64] together with the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm as implemented in the GNU scientific library (libGSL; version 1.15 double) [66]. Mirzaei *et al.* introduced the original algorithm focusing on the RBEM problem for molecular docking [64]. The approach is specifically designed for an efficient rotation of the rigid bodies around a center of rotation and is particularly well suited for our application. However, the original method only allows for a simultaneous optimization of the relative position of only two rigid bodies. We therefore extended it such that the simultaneous optimization of the orientation of an arbitrary number of rigid bodies is possible. This was necessary, as the simultaneous optimization of at least three rigid bodies Vβ cuboids is Vβ cuboids required for the modeling of TCRpMHC complexes: i.e., the TCR Vα and Vβ domains as well as the pMHC. In addition, it allows the straightforward extension of the method towards the inclusion of local side chain flexibility.

We implemented the method in a generalized, modular way, allowing for the individual design of application specific optimization pipelines, based on a given combination of the different functions during runtime. Each pipeline step consists in

the assembly of sub-process operators (SOs), which evaluate an objective function and the corresponding gradient to further perform the resulting coordinates transformations. Each SO manipulates the coordinates of a subset of atoms, which form the rigid body, further referred to as selection $s_k \subseteq S$ of the entire system S (where k refers to an SO within a given pipeline step). The contribution e_k of the operator to the objective function (as well as the corresponding gradient) is computed for the selection s_k within a given context c_k defined as a subset of the entire system, which at least contains the selection. As an illustration, let us consider a TCRpMHC complex consisting of four chains, A, B, C, and D, respectively corresponding to Vα, Vβ, MHC and to the bound peptide. A rigid body operator transforming the Vβ position with respect to Vα but ignoring the pMHC ligand would operate on the selection s_β*="chain B"* and within the context c_β*="chain A and chain B"*. The objective function would be evaluated only for the atoms of chains A and B (context) and chain B (selection) would be transformed whereas the chains A, C, and D would remain in their original position. So far, we implemented three different families of SOs. We shall briefly describe them below, while a more detailed presentation can be found as Supporting Information (Additional file 2: Text S2).

The first family of SOs consists of the basic operators for the rigid body rotation and translation. The objective function of these operators is computed from the non-bonded interactions between the mobile rigid body (s_k) and the defined context c_k of the system. The operators modify the three parameters for the rotation and for the translation of the rigid body, either simultaneously or independently.

Next to these general operators, we used the modular implementation strategy to further define a specific carboxamide group rotation operator. The latter uses the above SOs in a tailored fashion to specifically adjust the orientation of the conserved glutamine residue in each domain of the TCR (i.e., Vα and Vβ). This operator is valid

for both asparagine and glutamine residues. The selection s_k contains all atoms of the amino acid carboxamide group and the context c_k is usually either both Vα and Vβ TCR domains or solely the domain that contains the carboxamide group of interest. The rotation axis is defined by the carbon-carbon bond next to the carboxamide group (i.e., Cβ-Cγ and Cγ-Cδ for an asparagine and a glutamine residue, respectively). The objective function accounts for all bonded energy terms within the corresponding side chain and non-bonded energies within the whole context (i.e., including the selection). This sub-process operator can be used within our prediction pipeline algorithm or independently, and we shall refer to as Q-flip correction tool in the following.

Finally, we added a third class of rigid body position restraint operators to prevent unrealistically large translational motions and hence to avoid irrelevant conformations. The objective function in this case consists of a harmonic potential applied on the distance between a given reference and a mobile point. The harmonic penalty is applied if the distance is greater than the defined threshold. The mobile point is usually defined as one atom or as the geometric center of a set of atoms, and belongs to the rigid body that has to be restrained. In the case of the present application, CoR_β and CoR_μ are used as reference points for restraining the positions of the Vβ and pMHC cuboids, respectively. The mobile point for Vβ is the beta carbon of the conserved glutamine residue side chain. In the case of pMHC, the mobile point is defined as the geometric center between the C- and N-terminal C_α atoms of the bound peptide.

TCRpMHC prediction pipeline

The standard modeling pipeline for the prediction of the Vβ orientation in Vα/Vβ complexes of TCRs as well as the orientation of both Vβ and pMHC in TCRpMHC complexes is summarized in Fig. 2. The steps corresponding to the former case are

shown in black circles while the steps of the latter case are circled in green color. In addition, an animation of the modeling process is available as Supporting Information (Additional file 3: Movie S1). A standard prediction run for a TCRpMHC complex consists of 7 steps, among which steps 1 and 2 are performed only once, while steps 3 to 7 are repeated for each considered starting conformation (i.e., 11 in the present application).

The first step consists in the preparation of the complex (either TCR or TCRpMHC), which is assembled from its individual components (Vα, Vβ, and pMHC) in a general zero position (see previous subsection for details).

In the second step, the orientation of the glutamine residues located at the TCR Center of Rotation (CoR_β) is optimized by applying our Q-flip correction tool. At this stage, the correction is performed separately for each single domain and the interaction with the other TCR domain is ignored. To escape a possible local minimum, an explicit sampling of the Cγ-Cδ dihedral angle by 180° is performed for each residue, resulting in two symmetric conformations. Then, the conformation with the lowest energy is further refined by performing 30 steps of BFGS energy minimization. The final geometries are then used as base structures for the assembly of the starting conformations in the following loop (i.e., step 3 to 7).

At step three the orientation of the Vβ domain is rotated around the CoR_β with respect to the fixed Vα domain to produce a specific starting conformation. In the present application, the starting conformations are based on the previously defined 11 Vα/Vβ relative orientations (as described in the previous subsection). The optimization process from this step to step 7 will thus be repeated 11 times, each time with another starting conformation. As illustration, we show in Fig. 3A the 11 starting orientations of Vβ with respect to the fixed Vα domain (depicted in blue in the figure) for the

remodeling of the Vα/Vβ complex of one particular structure (i.e., the crystal structure with the PDB-ID 2f53 [38]).

The fourth step of this pipeline performs the preplacement of the pMHC ligand (for TCR/pMHC modeling). To avoid strong initial forces due to steric hindrance, the pMHC unit is first translated away from the TCR by 13 Å along its preplacement axis (see subsection Structural basis for the prediction algorithm). Afterwards the position of the pMHC is preoptimized with respect to the fixed Vα and Vβ domains. For this purpose the position of the pMHC is optimized by simultaneous translation along its preplacement axis and rotation around its CoR_{μ}-based rotational center using 150 steps of minimization. Because the Vα/Vβ orientation angle is not optimized at this stage of the pipeline, strong forces can occur during the optimization process due to unrealistic charge repulsion and thus induce large translational motions of the pMHC unit. To avoid unrealistic conformations, a 27 Å rigid body distance restraint operator is added, which is centered on CoR_{μ}.

In the fifth step, both CoR_{β} Q residues of the Vα and Vβ domains are optimized in the assembled structures. In contrast to step 2, both glutamine residues are optimized together in the context of the two TCR domains ensuring a proper orientation of the Q residues with respect to each other for the current Vα/Vβ orientation. As in step 2 an explicit sampling of the Cγ-Cδ dihedral angle is performed for each residue prior to minimization. However, because the new context might drastically change the orientation of these two residues, the angular step is reduced to 18°, leading to 400 different orientations. The orientation with the lowest energy is then further minimized for 30 steps and used in the next step.

The sixth step constitutes the core of our prediction methodology. At this stage, the position of Vβ and pMHC as well as the orientation of the two glutamine residues are concurrently optimized. Thus, the Vβ TCR domain and the pMHC ligand are

simultaneously translated in space and rotated around their respective internal CoR_β- and CoR_μ-based rotational centers. During the optimization, the positions of the internal CoR_β- and CoR_μ-based rotational centers of the Vβ TCR domain and the pMHC ligand are restrained within a maximum distance of 7.5 and 13 Å, respectively, from the CoR_β and CoR_μ reference positions. Simultaneously, the Cγ-Cδ dihedral angles of both CoR_β glutamine residues are optimized using the same settings as in step 5. Step 5 and 6 are then repeated once in the present pipeline to ensure a proper orientation of the conserved glutamine residues. The optimization in step 6 is performed for 50 and 3000 steps during the first and the second run, respectively.

In the last step of this pipeline, the quality of the current model (i.e., originating from the i^{th} starting conformation out of 11 in the present application) is evaluated by computing the complex binding energy ($E_{i,bind}$) as:

$$E_{i,bind} = E_{i,complex} - \left(E_\alpha + E_\beta + E_\mu\right),$$

where $E_{i,complex}$ is the total energy of the complex and E_α, E_β, and E_μ are the energy terms of the individual complex components Vα, Vβ, and pMHC (notice that these last quantities are constant for each prediction run and are thus computed only once). The energy is evaluated using the OPLS-AA force field [67, 68]. In addition, the all-atom positional root mean square deviation (RMSD) with respect to the crystal structure is calculated for each of the 11 final models.

Finally, the ranking of the models (i.e., 11 per prediction run) is performed by two independent means, either according to an energy criterion (C_E) or to a structural one (C_R). With respect to C_E, the best model is the one associated with the lowest value of $E_{i,bind}$, while C_R ranks as first the structure with the best RMSD. For both of the two criteria, the remodeling attempt is considered successful if the selected model (i.e.,

either according to C_E or C_R) presents an RMSD value lower than 2 Å, with respect to the corresponding crystal structure.

Structural datasets

We selected 75 biological units (BUs) originating from 48 different crystal structures contained in the set that we previously analyzed in ref. [43].

In that study, we observed that the different BUs within a given crystal structure can slightly differ from each other (RMSD < 1 Å), especially in the exact location of side chain atoms. This is presumably due to the relatively high intrinsic flexibility of the complexes or to the limited resolution in some of the structures (differs from 1.5 to 3.5 Å). To evaluate the robustness of our method and its capability to tackle such inaccuracies, we included all BUs in our two main datasets. The inclusion of all BUs also results in a larger data set and, as no training of a scoring function is performed (DynaDom is a force-field based approach as described in the previous subsections), introduces no bias to the method itself. In addition, the current datasets only contain structures in which all atoms that are involved in the modeling process were experimentally resolved. Although these atoms or residues could be easily modeled, this would potentially introduce a bias in the set, which we prefer to avoid here. As summarized in Table 1, the TCRpMHC crystal structures selected for this work belong to different species (i.e., 17 murine and 31 human) and are representative for 22 different TCR types (mutations not accounted). The coordinates of each structure were aligned with respect to the conserved residues of the Vα domain, as described in ref. [43] and in the previous subsections. The TCR constant domains, the MHC α_3 domain, and the β-microglobulin were systematically removed from the structures as well as all non-protein atoms (the discarded domains are represented with transparent

colors in Fig. 1). Hydrogen atoms were added and topologies were created for the OPLS-AA force field [67, 68] using the pdb2gmx tool (Version 4.5.6) [98].

We further derived three different data sets. First, to evaluate the performance of our method for the remodeling of the association angle of the TCR Vα and Vβ domains, we removed the pMHC ligand in each structure. This resulted in a set of 75 Vα/Vβ complexes, which we shall refer to as DS_T in the following. In addition, we created a second data set, in which only the first BU in the PDB file of the corresponding structure was included (48 structures, further referred to as DS_T^*). Then, to perform the remodeling of TCRpMHC complexes, we selected among the 75 BUs only the structures containing an MHC class I molecule. The resulting third data set, named as DS_C, contains a total of 53 TCRpMHC complexes. We disregarded MHC class II molecules in the DS_C set to ensure a proper comparison between the samples. A third set could have been dedicated to MHC class II molecules. However, we sustained from remodeling also that set as the results obtained for the MHC class I complexes already showed that further optimization of the pipeline, beyond this publication, is necessary for accurate pMHC placement. In addition, the size of the MHC class II set (22 structures) would have been very small for a robust analysis. Further details about each data set are listed in Table S1 of the Supporting Information (Additional file 4: Table S1).

Results and Discussion

The structural prediction of immunologically relevant molecular assemblies has focused the interest of a wide range of methodological developments over the past decades, especially in the field of antibody-antigen interactions [41, 48, 49, 99]. Compared to the effort made so far in antibody modeling, the number of predicted TCRpMHC structures is still relatively small, as we discussed in the Introduction

section. In the case of antibodies, it was recently shown by some of us, that statistical learning techniques can efficiently predict the V_H-V_L association angles [49]. Such very appealing approaches are based on experimentally observed structural features and require a large amount of existing data. In the particular case of antibodies, over 2000 crystal structures are already available, thus allowing the application of such knowledge-based methodologies. Considering the relatively small amount of TCR structures referenced in the Protein Data Bank (i.e., about 200), such a road can unfortunately not be envisaged for the prediction of association angles in TCR complexes. As a consequence, we developed here a solely force-field based optimization strategy for TCR and TCRpMHC complexes modeling. Such a force-field based approach can potentially be applied to other similar systems, even if a sufficient amount of experimental data is not available for the training of a specific scoring function.

As we extensively described in the Introduction and in the Methods sections, this new algorithm, named as DynaDom, is derived from our previous comprehensive analysis of the Vα/Vβ TCR variable domain association angles [43]. The main conclusions that arose from this former work can be summarized as follows: i) TCR complexes can be classified into structural clusters, differing significantly in their Vα/Vβ inter-domain angles, ii) the angular differences between the structural clusters can be described by a simple rotation around a center of rotation (CoR_β, see the Methods section and Fig. 1 for details), and iii) the CoR_β is characterized by two highly conserved glutamine residues, which contribute to the interaction between the TCR Vα and Vβ domains *via* a stabilizing hydrogen bond network.

For the remodeling of TCRpMHC complexes, the DynaDom method uses a unified cuboid description of the three different units of this complex (i.e., Vα, Vβ, pMHC). The optimization of the total system is performed by a rotation-based algorithm,

which is based on our Center of Rotation concept (i.e., CoR_{β} and CoR_{μ}, as described in the Methods section). In our previous analysis study, we observed that the Vα/Vβ association angle spectrum is much larger in unbound TCRs than in structures bound to the pMHC [43]. Pierce *et al.* further emphasized that for the prediction of TCRpMHC complexes from unbound units [33] the side chains of the CDR loops must adapt to their environment in order to allow for a proper interaction between the different units of the complex. Therefore, the inclusion of local side chain flexibility at the domain interface would most likely be a necessary extension for the prediction of TCRpMHC structures from unbound or homologous TCR and pMHC structures by homology modeling techniques. Our generalized, modular implementation ensures that the additional inclusion of local flexibility is straightforward. However, the adaptation of the algorithm would require additional extensive evaluation efforts, which would go beyond the scope of the present work and will be part of future investigations. Nevertheless, in the present work we already tested such a feature by the inclusion of local side chain flexibility for the two Q-Q residues at the CoR_{β}, which we found to be crucial for the prediction success as we shall discuss in the following subsections.

The current version of our pipeline algorithm, results from an extensive series of evaluations intended to assess the effect of the different parameters. Hereafter, we present and discuss our main findings together with the actual evaluation of the method. We first discuss the optimization of the orientation of the glutamine residues (Q-flip correction) located at the interface between the two TCR variable domains (i.e., at the Center of Rotation CoR_{β} of Vβ with respect to Vα), based on the original experimental structures. Then, we analyze the effect of such a Q-flip correction together with the use of restraints on the remodeling of Vα/Vβ TCR and TCRpMHC complexes. Along this analysis, we compare the results obtained using either an

energy or a structure based selection criterion (i.e., C_E or C_R, respectively, as defined in the Methods section). This comparison intends to state if an atomistic force field energy based criterion could be used for future blind homology modeling of TCRpMHC complexes. We finally suggest further possible routes of improvement for our methodology, based on the analysis of the few cases in which the remodeling process did not lead to a satisfactory structure.

Glutamine orientation correction

The interface between the Vα and Vβ domains of TCRs is characterized by the interaction of two highly conserved glutamine (Q) residues [43]. While this Q-Q interaction appears to be of critical importance, the flip state of these residues is often wrongly assigned in experimental crystal structures, due to the ambiguous character of the carboxamide group electron density [50-53]. In the context of this work, we analyzed the flip state of the Q residues among the crystal structures contained in our set of 75 Vα/Vβ complexes (i.e., data set DS_T). Only 72.7% of the structures present a correct assignment of the Q-flip state. The details of this analysis are listed in Table S2 of the Supporting Information (Additional file 5: Table S2).

As discussed in the Introduction section, many modeling tools exist to correct the orientation of glutamine and asparagine residues in a given crystal structure. Among those, we tested Reduce [50, 54, 55] and Protoss [60, 61] on our DS_T set. The application of the Reduce and Protoss programs leads to an improvement of the glutamine flip state in our set, reaching 94.6% and 97.3% of correctly assigned Q-flip states, respectively. Analysis of the failed cases showed that they featured an interaction of the Q residues in an initial bifurcated orientation (i.e. associated in a perpendicular manner). Manual inspection showed that in these cases the perpendicular orientation allowed for optimal interactions with the rest of the domains

and should therefore be the most stable in the functional receptor (i.e. not a further artifact of the carboxamide assignments in the experimental structures). As the Reduce and Protoss programs only allow parallel orientations, these cannot successfully predict such particular interactions. Because of this limitation and the below discussed observation that the Q-flip state can change upon the association of the Vα/Vβ domains, we decided to implement an independent Q-flip correction approach using our already implemented rigid body operators, such that it can directly be included into our pipeline. This represents a first probing of the modular character of our implementation, which we shall follow towards the future inclusion of local flexibility.

We evaluated the performance of our method for the Q-flip correction in the crystal structure of the DS_T set. We present here the most relevant findings of our analysis, while a more detailed discussion can be found as Supporting Information (Additional file 6: Text S3), together with the entirety of our observations per structure (Additional files 5, 7, and 8: Tables S2, S3, and S4). Using the DynaDom correction module, 100.0% of the structures could be assigned in the good Q-flip state. An example of a successful Q-flip correction is depicted in Fig. 3B for the 2f53 crystal structure. This higher performance obtained by our method with respect to the other programs comes from the optimization-based methodology that we implemented. While standard tools only consider two possible parallel conformations per residue (i.e., the original and the flipped state), DynaDom performs an explicit sampling of the carboxamide group using adjustable angular step sizes, followed by an energy minimization step during which the atomistic environment of the residue is taken into account. Such a protocol allows the system to escape local minima in order to find the most favorable conformations of the Q residues in their environment. For this reason,

are ranked by their RMSD or energy score (see Additional file 9: Fig. S1 Fig. and the description in the Methods section). The remodeling experiment is then counted as successful if the RMSD of the selected model with respect to the original crystal structure is lower than 2 Å.

Our results are summarized in Table 2 and the different evaluation settings that we considered are generally labeled as M_T plus a bit string, which encodes for the use or not of the Q-flip correction and the restraint (e.g., M_T10 labels the remodeling of a TCR complex by applying the Q-flip correction but no distance restraint). The last M_T^*11 test corresponds to the M_T11 settings performed for the DS_T^* data set, which contains only the first BU of each experimental structure.

Considering the C_R criterion, the remodeling procedure already reaches a very high positive prediction rate of 94.7%, even if no Q-flip corrections or restraints are used (M_T00). This rate further increases to 96.0% if the Q-flip correction is switched on (M_T10), while no change is observed if the distance restraint is used alone (94.7% in the M_T01 case). As a consequence, the final prediction rate, with both parameters switched on, also reaches the remarkable rate of 96.0% (M_T11). Only three outliers are observed, originating from the 3dxa and from the 1mwa crystal structure. The relatively low resolution of the 3dxa structure (i.e. 3.5 Å) can partially explain this failure. Furthermore, our modeling process only considers the Vα and Vβ domains of the TCR complex. It is possible that the two constant domains of the complex play an important role in these three outlier cases. Regarding the additional experiment performed on the DS_T^* data set (M_T^*11), the results in Table 2 show that the differences in the achieved accuracies with respect to M_T11 are only marginal. This confirms that the inclusion of the BUs does not bias the overall results and demonstrates the robustness of our algorithm with respect to small variations in the

structures, thus highlighting the suitability of the approach in a future homology modeling pipeline.

In the perspective of a blind homology modeling experiment of TCR complexes, no reference crystal structure would be available and only an energy-based criterion could be considered for structure selection (i.e., C_E). Based on such a C_E criterion, our remodeling attempt reaches a prediction rate of 84.0% even if no Q-flip corrections or restraints are applied (M_T00). The prediction rate increases with both, the independent use of the Q-flip correction and the distance restraint to 85.3% and 88.0% for M_T10 and M_T01, respectively. If both parameters are used (M_T11), the prediction reaches the remarkable rate of 89.3% and even 89.6% for the M_T^*11 data set. This last result is very promising for the further applications of the DynaDom method in a real structure prediction setting.

It appears that the use of the distance restraint has a stronger impact on the prediction rate obtained according to the C_E criterion than it has for the C_R criterion. This could be attributed to the observation that without distance restraint, the algorithm can yield structures in which the two TCR domains are placed in an unrealistic conformation, which nevertheless has a lower interaction energy (see Additional file 10: Fig S2). Such unphysical associations are far from the original crystal structure and are intrinsically discriminated by an RMSD based selection criterion.

Next to the analysis of the best conformations according to the C_E and C_R criteria we also analyzed the overall performance of the algorithm regarding the quality of all predicted conformations. In Fig. 4, we present the percentage of structures having an RMSD value lower than 1, 2 and 3 Å, depending on the algorithm settings (i.e., M_T00, M_T01, M_T10, and M_T11) among all 75*11 models produced by our DynaDom procedure. In this context we also further analyzed the impact of the Q-flip correction by classifying the resulting models into two groups, with respect to their original Q-

flip state in the experimental structures as paired (51*11) and mispaired (20*11). Notice that 4*11 structures lack the presence of Q residues at CoR_β and were therefore not included in the respective analysis. The histograms (Fig. 4) show that an overall percentage of about 80% of the models feature an RMSD lower than 2 Å, thus fulfilling our success criterion. This demonstrates the robustness of the presented algorithm and thus its relevance as one step in a comprehensive structure prediction pipeline. By further analyzing the influence of the Q-flip correction on the prediction rates, it can be observed that the overall prediction success is higher for structures in which the Q-Q orientation is already correct in the X-ray structure (paired structures). For these structures 85% of the models have an RMSD value lower than 2 Å, whereas the rate drops to 75% for the mispaired structures. This might be due to the relatively smaller size of the latter data set, as an investigation of a possible correlation between the crystal structure resolution and the quality of the final models did not yield any significant outcome.

Regarding the percentage of structures having an RMSD value lower than 2 and 3 Å for both sets, paired and mispaired, the results are practically independent on the defined settings and only a slight trend towards an improved performance can be observed if the Q-flip correction is applied. This low impact on the overall structures is most likely due to the large surface area of the total TCR domain interface and thus the high number of other interactions, which drive the overall optimization of the domains orientations. The use of the Q-flip correction has, however, a remarkable effect on the quality of the resulting structures once the cutoff is lowered to 1 Å. The percentage of models featuring such a low RMSD indeed increases from 47% to 60% for the mispaired structure set, if the correction is switched on. These observations reveal the importance of a correct orientation of the conserved Q residues at the Center of Rotation CoR_β for an accurate modeling of the TCR variable domain

association and the need for their correction if they are wrongly assigned in the template structure.

Overall, this series of remodeling essays highlights the quality of our methodology. It also further emphasizes the applicability of a force field interaction energy-based criterion, which is very promising in the perspective of a homology modeling setting, as it shows that high-quality structures can be identified by this means.

Modeling of the pMHC position with simultaneous TCR variable domain placement

Regarding the successful results obtained for the remodeling of TCR Vα/Vβ assemblies discussed in the previous subsection, we shall now assess the performance of the DynaDom method to remodel TCRpMHC complexes. Here, the calculations were performed on the smaller set of structures DS_C, which contains a total of 53 biological units. Our results are listed in Table 3 and the labeling of the test settings follows the nomenclature introduced above (i.e., M_C label and a bit string for the use or not of Q-flip correction and distance restraint).

An example of successfully predicted complex is depicted in Fig. 3C for the structure 2f53. In the figure, the Vα, Vβ, and MHC units are respectively colored in blue, red, and green. The two images represent the complex before and after optimization (on the left and on the right hand side of the figure, respectively). The magnifying glass shows that the Q-flip state is efficiently corrected and one can observe that the final model successfully fits the reference crystal structure (depicted with gray color in the picture).

Regarding the prediction rates according to the C_R and C_E criteria in Table 3, the percentages obtained for the remodeling of TCRpMHC complexes reach a less striking prediction rate, though still relatively high (i.e., 73.6% and 71.7% according

to the C_R and C_E criteria, respectively). For the prediction based on the C_R criterion, the success rate appears to be independent on the use of Q-flip correction and distance restraints. A similar trend is observed with the C_E criterion. In this case, the use of one or both of the parameters only marginally increases the prediction rate. As for the modeling of TCR variable domains alone, the use of an energy based criterion yields very satisfactory results compared to a structure based one. This point also confirms the robustness and thus the suitability of our method for a blind homology modeling of TCRpMHC complexes.

Detailed performance analysis for the TCRpMHC prediction

Regarding the overall, nearly equal, performances of the different settings in Table 3, it clearly appears that the drop of the prediction rate for the remodeling of TCRpMHC complexes with respect to the modeling of only the TCR variable domains is barely dependent on the use of Q-flip correction and distance restraints. The former parameters only affect the relative orientation of the Vα and Vβ domains. This observation indicates that the lower performance observed for the remodeling of TCRpMHC complexes might originate from an incorrect placement of the MHC molecule. We thus performed additional analyses to gain more insights into the shortcomings of the current approach and to identify potential routes for future improvement of our algorithm. To further confirm that the issues encountered in the remodeling of TCRpMHC complexes are solely due to the prediction of the pMHC positions with respect to the TCR, we analyzed the impact of the initial placement of the pMHC ligand on the remodeling of TCRpMHC complexes. In this context, we shall only consider the models obtained according to the RMSD based criterion (C_R).

In the following series of test evaluations (T), we only consider the initial orientation of the TCR domains as found in their crystal structure (i.e., the remodeling procedure

does not start from the 11 starting conformations, but only one). This was done to eliminate any potential biasing errors originating from the TCR domain modeling. Next to that, we used the settings of the final modeling pipeline, i.e.Vβ optimization, Q-flip correction, and position restraints were systematically applied during these tests. As we described in the Methods section and depicted in Fig. 2, our modeling protocol includes a translation of the pMHC unit along a given axis to separate pMHC from the TCR, thus avoiding strong initial forces due to unphysical steric hindrance. This feature is one parameter that we shall analyze in the following tests (i.e., by switching it on or off). For each test setting, a rigid body optimization of the pMHC around its starting position was performed. Finally, for the first two test evaluations (T_1 and T_2), the MHC rigid body was initially placed in its crystal structure orientation, while for the last test (T_3), this unit was oriented according to the general zero conformation discussed in the Method section (i.e., the orientation used in the standard pipeline). The results and details of each test evaluation are presented in Table 4.

In the first test evaluation (T_1) in which the pMHC units are oriented according to their respective crystal structure orientation and no initial translation is performed, the prediction rate reaches 100.0%. Although such a result could be expected as we start from the experimental conformations, it proves that our algorithm does not lead to any conformational artifacts. The additional application of the initial translation step for the pMHC ligand (T_2) results in a drastic decrease of the prediction rate to 71.7%, slightly lower than the result obtained using our standard protocol (i.e., 73.6% in the M_C11 case). For the final test (T_3), in which the translated pMHC was placed according to our standard protocol, the prediction rate of our algorithm dramatically drops to 58.5%, which is significantly lower than for the final pipeline setting (73.6%).

These results show that the translation procedure and the preplacement of the pMHC ligand in a single general starting position constitute the accuracy limiting steps of our pipeline. In addition, we confirm here that the use of various starting positions for the Vβ domain clearly outperforms the case in which a single conformation is considered, even if the latter corresponds to the experimental crystal structure (i.e., 73.6% *versus* 58.6% for the M_C11 and T_3 cases, respectively). At first glance this is a surprising result. However, it clearly appears that the simultaneous optimization of both the TCR domains and the pMHC molecule is highly beneficial for the performance of the algorithm as it allows for an alternating adaption of the flexible units with respect to each other (Additional file 3: Movie S1). This leads to a smoother optimization path, thus lowering the probability for being trapped in a local minimum. Different starting positions further lower this probability as multiple paths are sampled.

Consequently, one straightforward way to improve our results should be to use multiple starting conformations for the pMHC ligand, in accordance with the 11 Vβ preplacement orientations. To evaluate this procedure, we chose one structure (PDB-ID 1oga) for which the modeling process failed in the last T_3 test settings. For this structure, the three Euler angle components defining the pMHCs CoR_μ-based rotational center were systematically varied by 5 degree and all 27 resulting starting poses were constructed. In accordance with the other test settings, the Vβ domain was here again placed in its crystal structure orientation. The results for the 27 resulting models are listed in the Supporting Information (Additional file 11: Table S5). The results improved considerably as this time five structures were obtained with an RMSD lower than 2Å, thus satisfying our success criterion. Notably, these five models also show the lowest interaction energy among the 27 predicted structures. This last test clearly confirms the necessity of more advanced sampling protocol for the MHC molecule orientation in our modeling strategy to avoid the complex

geometry to fall in an unfavorable local minimum. This is in agreement with the observations made by Pierce *et al.* [33] and demonstrates once again the importance of starting from multiple initial conformations. However, a straightforward combination of the 11 starting conformations of the TCR Vβ domain together with the 27 initial orientations of the pMHC unit would lead to a total of 297 structures to optimize per TCRpMHC complex, thus resulting in a dramatic increase of the computational cost.

Therefore, the presented algorithm provides excellent results and can readily be used for the optimization of the Vα/Vβ association angles. It also yields a fairly good prediction rate for the prediction of TCRpMHC complexes association. However, for the simultaneous optimization of both, the TCR domains and the placement of the pMHC in the latter case, further improvements and evaluations will be necessary prior to its practical use as one step in a real structure prediction pipeline. Considering the general, modular character of our implementation, also different approaches could be combined with the current method to tackle this issue. Among those, basin-hoping techniques [100] have proven to provide good results for the rigid body optimization of tryptophan zippers [101], and Monte Carlo-based rigid body sampling was recently applied by Pierce *et al.* for the placement of MHC like ligands alone [33, 34]. Despite the numerous tests that would be required for the combination of such techniques, this route represents a promising strategy for the future of our methodology.

Conclusions

In this work we presented a new procedure, DynaDom, for the optimization of protein domain-domain orientations, which was designed for and evaluated on the special case of remodeling T-cell-receptor-peptide-MHC complexes. The approach is based on several rigid body optimization and restraining routines, and uses atomistic force

field-based energy calculations. The individual optimization functions are combined in an application-specific pipeline. The method yields remarkable results for the remodeling of TCR Vα/Vβ association angles with prediction rates of 89-96% (RMSD < 2 Å) depending on the evaluation criterion.

The present study shows that it is possible to predict the TCR Vα/Vβ association angles on the basis of structural modeling only, without the need for a specially tailored experimental data dependent scoring function. It also demonstrates that the previously identified Center of Rotation concept [43] can be readily used for the structural prediction of the association angles.

Another striking result arising from this work is the observation that, by simply considering the best-energy conformation for each structure, high prediction rates of 89.3% for the Vα/Vβ association angles could be obtained. This is only marginally lower than the prediction rates obtained for the models with the smallest RMSD. This shows that ranking the modeled structures solely by their force field-based interaction energy allows the identification of high quality structures and demonstrates not only the robustness of the method, but also its suitability as part of a general structure prediction pipeline for TCRpMHC structures.

In a second step, we applied the concept to the simultaneous optimization of the TCR Vα/Vβ association angles and the pMHC positions on the TCR. However, due to efficiency considerations we used a simplified placement method for the pMHC, which resulted in lower prediction rates of 72-74%. This result is still in the predictive range, but not as high as for the TCR domain optimization. Additional preliminary investigations showed that the main reason lies indeed in the initial placement method of the pMHC ligand and that by simply using multiple initial conformations for the pMHC placement, already significant improvements in the placement accuracy are possible. However, a systematic optimization of the method for pMHC placement

would require further significant evaluation studies, which would go beyond the scope of this manuscript and which will be the topic of future studies together with the application of DynaDom to the blind homology modeling of TCRpMHC complexes. In general, the presented approach is very well suited to serve as basis for the development of such a method for the prediction of atomistic models of TCRs or TCRpMHC complexes taking inter-domain angles into account. Due to the modular design of our program, a straightforward combination and concurrent optimization of multiple features is possible, as already demonstrated in this work by the concurrent optimization of the Vβ orientation, the pMHC orientation, and the adaption of the glutamine residues connecting the two TCR chains. Thus, the future implementation of partial or full flexibility of side chains or protein backbone regions, which then could be simultaneously optimized while the rigid body positions are adapted, should be straightforward. In addition, including other domains of the complex, such as the TCR constant domains would also be possible. This could help to study e.g. scissoring effects observed for the constant domains [89] and to study TCR signaling, which was elsewhere discussed to be induced by conformational changes in the constant domains [102].

Finally, it is worth noting that the DynaDom strategy is not limited to TCRpMHC assemblies. The combination of the different modules can indeed be easily modified to fit the requirement of other rigid body based predictions of a large variety of biomolecular assemblies

Declarations

List of abbreviations

BFGS	Broyden-Fletcher-Goldfarb-Shanno

BU	Biological Unit; crystallographically independent molecule in the asymmetric unit.
C_E , C_R	ranking Criterion based on the Energy or the RMSD, respectively
CDR	Complementary Determining Region
CoR_β, CoR_μ	Center of Rotation in repect to the TCR Vβ and the pMHC, respectively
DS_C	Data Set TCRpMHC Complex
DS_T	Data Set TCR
DS_T*	reduced DS_T containing only one BU per structure
MHC	major histocompatibility complex
M_T, M_C	Modelling run for the TCR test set or the complex test set, respectively
PDB	Protein Data Bank
pMHC	peptide presented in a Major Histocompatibility Complex molecule
Q	glutamine
RMSD	Root Mean Square Deviation
SO	Sub-process Operator
T1,T2, and T3	Test evaluations with different conditions.
TCR	T Cell Receptor
Vα/Vβ	Variable domain of the TCR α- and the β-chain, respectively.
V_H, V_L	antibody Variable domains of the Heavy and the Light chain, respectively

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article (and its additional files).

Funding

TH was supported by the Deutsche Forschungsgemeinschft through the SFB1035 project A10, AM was supported by the Deutsche Forschungsgemeinschft through the SFB749, further financial support was provided by the CIPS^{M} cluster of excellence.

Competing interests

The authors do not have any competing interests.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Author's contribution

TH entirely programmed the code related to this work, performed all the evaluation studies, and contributed to the analysis, to the design of the study, to the development of the algorithms as well as to the writing of the manuscript. AM and IA contributed to the analysis, to the development of the algorithms and to the writing of the manuscript. IA also initiated and contributed to the general design of this work. All authors read and approved the final manuscript.

Acknowledgements

We thank Hanieh Mirzaei and Dima Kozakov for providing their rigid body optimization code, Stefan Bietz for the provision of a Protoss license, Angela Krackhardt, Atanas Patronov, and Ilke Ugur for helpful discussions and critical reading of the manuscript. We thank Antonia Stank for her preparative work on the used MHC complexes.

References

1. Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J, Kourilsky P: **A direct estimate of the human alphabeta T cell receptor diversity**. *Science* 1999, **286**(5441):958-961.
2. Casrouge A, Beaudoing E, Dalle S, Pannetier C, Kanellopoulos J, Kourilsky P: **Size estimate of the alpha beta TCR repertoire of naive mouse splenocytes**. *J Immunol* 2000, **164**(11):5782-5787.
3. Warren RL, Freeman JD, Zeng T, Choe G, Munro S, Moore R, Webb JR, Holt RA: **Exhaustive T-cell repertoire sequencing of human peripheral blood samples reveals signatures of antigen selection and a directly measured repertoire size of at least 1 million clonotypes**. *Genome research* 2011, **21**(5):790-797.
4. Tonegawa S: **Somatic generation of antibody diversity**. *Nature* 1983, **302**(5909):575-581.
5. Garcia KC: **Reconciling views on T cell receptor germline bias for MHC**. *Trends in immunology* 2012, **33**(9):429-436.
6. Davis MM, Bjorkman PJ: **T-cell antigen receptor genes and T-cell recognition**. *Nature* 1988, **334**(6181):395-402.
7. Bernstein FC, Koetzle TF, Williams GJ, Meyer EF, Jr., Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M: **The Protein Data Bank: a computer-based archival file for macromolecular structures**. *J Mol Biol* 1977, **112**(3):535-542.
8. Michielin O, Luescher I, Karplus M: **Modeling of the TCR-MHC-peptide complex**. *J Mol Biol* 2000, **300**(5):1205-1235.
9. Morgan RA, Dudley ME, Wunderlich JR, Hughes MS, Yang JC, Sherry RM, Royal RE, Topalian SL, Kammula US, Restifo NP *et al*: **Cancer regression in patients after transfer of genetically engineered lymphocytes**. *Science* 2006, **314**(5796):126-129.
10. Zoete V, Irving M, Ferber M, Cuendet MA, Michielin O: **Structure-Based, Rational Design of T Cell Receptors**. *Frontiers in immunology* 2013, **4**:268.
11. Rueckert C, Guzman CA: **Vaccines: from empirical development to rational design**. *PLoS Pathog* 2012, **8**(11):e1003001.
12. Flower DR, Macdonald IK, Ramakrishnan K, Davies MN, Doytchinova IA: **Computer aided selection of candidate vaccine antigens**. *Immunome Res* 2010, **6 Suppl 2**:S1.
13. Chakraborty AK, Weiss A: **Insights into the initiation of TCR signaling**. *Nat Immunol* 2014, **15**(9):798-807.
14. Backert L, Kohlbacher O: **Immunoinformatics and epitope prediction in the age of genomic medicine**. *Genome Med* 2015, **7**:119.
15. Lundegaard C, Lund O, Nielsen M: **Predictions versus high-throughput experiments in T-cell epitope discovery: competition or synergy?** *Expert Rev Vaccines* 2012, **11**(1):43-54.
16. Atanasova M, Patronov A, Dimitrov I, Flower DR, Doytchinova I: **EpiDOCK: a molecular docking-based tool for MHC class II binding prediction**. *Protein Eng Des Sel* 2013, **26**(10):631-634.
17. Schueler-Furman O, Altuvia Y, Sette A, Margalit H: **Structure-based prediction of binding peptides to MHC class I molecules: application to a broad range of MHC alleles**. *Protein Sci* 2000, **9**(9):1838-1846.

18. Khan JM, Ranganathan S: **pDOCK: a new technique for rapid and accurate docking of peptide ligands to Major Histocompatibility Complexes**. *Immunome Res* 2010, **6 Suppl 1**:S2.
19. Antes I, Siu SW, Lengauer T: **DynaPred: a structure and sequence based method for the prediction of MHC class I binding peptide sequences and conformations**. *Bioinformatics* 2006, **22**(14):e16-24.
20. Rigo MM, Antunes DA, Vaz de Freitas M, Fabiano de Almeida Mendes M, Meira L, Sinigaglia M, Vieira GF: **DockTope: a Web-based tool for automated pMHC-I modelling**. *Sci Rep* 2015, **5**:18413.
21. Liu T, Pan X, Chao L, Tan W, Qu S, Yang L, Wang B, Mei H: **Subangstrom accuracy in pHLA-I modeling by Rosetta FlexPepDock refinement protocol**. *J Chem Inf Model* 2014, **54**(8):2233-2242.
22. Liu Z, Dominy BN, Shakhnovich EI: **Structural mining: self-consistent design on flexible protein-peptide docking and transferable binding affinity potential**. *J Am Chem Soc* 2004, **126**(27):8515-8528.
23. Almagro JC, Vargas-Madrazo E, Lara-Ochoa F, Horjales E: **Molecular modeling of a T-cell receptor bound to a major histocompatibility complex molecule: implications for T-cell recognition**. *Protein Sci* 1995, **4**(9):1708-1717.
24. Garcia KC, Degano M, Stanfield RL, Brunmark A, Jackson MR, Peterson PA, Teyton L, Wilson IA: **An alphabeta T cell receptor structure at 2.5 A and its orientation in the TCR-MHC complex**. *Science* 1996, **274**(5285):209-219.
25. Garboczi DN, Ghosh P, Utz U, Fan QR, Biddison WE, Wiley DC: **Structure of the complex between human T-cell receptor, viral peptide and HLA-A2**. *Nature* 1996, **384**(6605):134-141.
26. Sali A: **Comparative protein modeling by satisfaction of spatial restraints**. *Mol Med Today* 1995, **1**(6):270-277.
27. Kessler B, Michielin O, Blanchard CL, Apostolou I, Delarbre C, Gachelin G, Gregoire C, Malissen B, Cerottini JC, Wurm F *et al*: **T cell recognition of hapten. Anatomy of T cell receptor binding of a H-2kd-associated photoreactive peptide derivative**. *J Biol Chem* 1999, **274**(6):3622-3631.
28. Leimgruber A, Ferber M, Irving M, Hussain-Kahn H, Wieckowski S, Derre L, Rufer N, Zoete V, Michielin O: **TCRep 3D: an automated in silico approach to study the structural properties of TCR repertoires**. *PloS one* 2011, **6**(10):e26301.
29. Haidar JN, Pierce B, Yu Y, Tong W, Li M, Weng Z: **Structure-based design of a T-cell receptor leads to nearly 100-fold improvement in binding affinity for pepMHC**. *Proteins* 2009, **74**(4):948-960.
30. Pierce BG, Hellman LM, Hossain M, Singh NK, Vander Kooi CW, Weng Z, Baker BM: **Computational design of the affinity and specificity of a therapeutic T cell receptor**. *PLoS computational biology* 2014, **10**(2):e1003478.
31. Simons KT, Kooperberg C, Huang E, Baker D: **Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions**. *J Mol Biol* 1997, **268**(1):209-225.
32. Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, Baker D: **Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations**. *J Mol Biol* 2003, **331**(1):281-299.

33. Pierce BG, Weng Z: **A flexible docking approach for prediction of T cell receptor-peptide-MHC complexes**. *Protein Sci* 2013, **22**(1):35-46.
34. Pierce BG, Vreven T, Weng Z: **Modeling T cell receptor recognition of CD1-lipid and MR1-metabolite complexes**. *BMC bioinformatics* 2014, **15**:319.
35. Ding YH, Baker BM, Garboczi DN, Biddison WE, Wiley DC: **Four A6-TCR/peptide/HLA-A2 structures that generate very different T cell signals are nearly identical**. *Immunity* 1999, **11**(1):45-56.
36. Ding YH, Smith KJ, Garboczi DN, Utz U, Biddison WE, Wiley DC: **Two human T cell receptors bind in a similar diagonal mode to the HLA-A2/Tax peptide complex using different TCR amino acids**. *Immunity* 1998, **8**(4):403-411.
37. Chen J-L, Stewart-Jones G, Bossi G, Lissin NM, Wooldridge L, Choi EML, Held G, Dunbar PR, Esnouf RM, Sami M *et al*: **Structural and kinetic basis for heightened immunogenicity of T cell vaccines**. *J Exp Med* 2005, **201**(8):1243-1255.
38. Dunn SM, Rizkallah PJ, Baston E, Mahon T, Cameron B, Moysey R, Gao F, Sami M, Boulter J, Li Y *et al*: **Directed evolution of human T cell receptor CDR2 residues by phage display dramatically enhances affinity for cognate peptide-MHC without increasing apparent cross-reactivity**. *Protein Sci* 2006, **15**(4):710-721.
39. Sami M, Rizkallah PJ, Dunn S, Molloy P, Moysey R, Vuidepot A, Baston E, Todorov P, Li Y, Gao F *et al*: **Crystal structures of high affinity human T-cell receptors bound to peptide major histocompatibility complex reveal native diagonal binding geometry**. *Protein Eng Des Sel* 2007, **20**(8):397-403.
40. McBeth C, Seamons A, Pizarro JC, Fleishman SJ, Baker D, Kortemme T, Goverman JM, Strong RK: **A new twist in TCR diversity revealed by a forbidden alphabeta TCR**. *J Mol Biol* 2008, **375**(5):1306-1319.
41. Dunbar J, Fuchs A, Shi J, Deane CM: **ABangle: characterising the VH-VL orientation in antibodies**. *Protein Eng Des Sel* 2013, **26**(10):611-620.
42. Dunbar J, Knapp B, Fuchs A, Shi J, Deane CM: **Examining variable domain orientations in antigen receptors gives insight into TCR-like antibody design**. *PLoS computational biology* 2014, **10**(9):e1003852.
43. Hoffmann T, Krackhardt AM, Antes I: **Quantitative Analysis of the Association Angle between T-cell Receptor Valpha/Vbeta Domains Reveals Important Features for Epitope Recognition**. *PLoS computational biology* 2015, **11**(7):e1004244.
44. Knapp B, Dunbar J, Deane CM: **Large Scale Characterization of the LC13 TCR and HLA-B8 Structural Landscape in Reaction to 172 Altered Peptide Ligands: A Molecular Dynamics Simulation Study**. *PLoS computational biology* 2014, **10**(8):e1003748.
45. Rudolph MG, Stanfield RL, Wilson IA: **How TCRs bind MHCs, peptides, and coreceptors**. *Annual review of immunology* 2006, **24**:419-466.
46. Sun R, Shepherd SE, Geier SS, Thomson CT, Sheil JM, Nathenson SG: **Evidence that the antigen receptors of cytotoxic T lymphocytes interact with a common recognition pattern on the H-2Kb molecule**. *Immunity* 1995, **3**(5):573-582.
47. Knapp B, Deane CM: **T-Cell Receptor Binding Affects the Dynamics of the Peptide/MHC-I Complex**. *Journal of chemical information and modeling* 2016, **56**(1):46-53.

48. Bujotzek A, Dunbar J, Lipsmeier F, Schafer W, Antes I, Deane CM, Georges G: **Prediction of VH-VL domain orientation for antibody variable domain modeling**. *Proteins* 2015, **83**(4):681-695.
49. Bujotzek A, Fuchs A, Qu C, Benz J, Klostermann S, Antes I, Georges G: **MoFvAb: Modeling the Fv region of antibodies**. *mAbs* 2015, **7**(5):838-852.
50. Word JM, Lovell SC, Richardson JS, Richardson DC: **Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation**. *J Mol Biol* 1999, **285**(4):1735-1747.
51. Weichenberger CX, Sippl MJ: **Self-consistent assignment of asparagine and glutamine amide rotamers in protein crystal structures**. *Structure* 2006, **14**(6):967-972.
52. McDonald IK, Thornton JM: **Satisfying hydrogen bonding potential in proteins**. *J Mol Biol* 1994, **238**(5):777-793.
53. Hooft RW, Sander C, Vriend G: **Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures**. *Proteins* 1996, **26**(4):363-376.
54. Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC: **MolProbity: all-atom structure validation for macromolecular crystallography**. *Acta crystallographica Section D, Biological crystallography* 2010, **66**(Pt 1):12-21.
55. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, 3rd, Snoeyink J, Richardson JS *et al*: **MolProbity: all-atom contacts and structure validation for proteins and nucleic acids**. *Nucleic Acids Res* 2007, **35**(Web Server issue):W375-383.
56. Weichenberger CX, Byzia P, Sippl MJ: **Visualization of unfavorable interactions in protein folds**. *Bioinformatics* 2008, **24**(9):1206-1207.
57. Weichenberger CX, Sippl MJ: **NQ-Flipper: recognition and correction of erroneous asparagine and glutamine side-chain rotamers in protein structures**. *Nucleic Acids Res* 2007, **35**(Web Server issue):W403-406.
58. Li X, Jacobson MP, Zhu K, Zhao S, Friesner RA: **Assignment of polar states for protein amino acid residues using an interaction cluster decomposition algorithm and its application to high resolution protein structure modeling**. *Proteins* 2007, **66**(4):824-837.
59. Labute P: **Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures**. *Proteins* 2009, **75**(1):187-205.
60. Lippert T, Rarey M: **Fast automated placement of polar hydrogen atoms in protein-ligand complexes**. *Journal of cheminformatics* 2009, **1**(1):13.
61. Bietz S, Urbaczek S, Schulz B, Rarey M: **Protoss: a holistic approach to predict tautomers and protonation states in protein-ligand complexes**. *Journal of cheminformatics* 2014, **6**:12.
62. Bayden AS, Fornabaio M, Scarsdale JN, Kellogg GE: **Web application for studying the free energy of binding and protonation states of protein-ligand complexes based on HINT**. *Journal of computer-aided molecular design* 2009, **23**(9):621-632.
63. Antes I: **DynaDock: A new molecular dynamics-based algorithm for protein-peptide docking including receptor flexibility**. *Proteins* 2010, **78**(5):1084-1104.
64. Mirzaei H, Beglov D, Paschalidis IC, Vajda S, Vakili P, Kozakov D: **Rigid Body Energy Minimization on Manifolds for Molecular Docking**. *J Chem Theory Comput* 2012, **8**(11):4374-4380.

65. Jones LL, Colf LA, Stone JD, Garcia KC, Kranz DM: **Distinct CDR3 conformations in TCRs determine the level of cross-reactivity for diverse antigens, but not the docking orientation**. *J Immunol* 2008, **181**(9):6255-6264.
66. Gough B: **GNU Scientific Library Reference Manual - Third Edition**. 2009.
67. Jorgensen W, Tirado-Rives J: **The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin**. *Journal of the American Chemical Society* 1988, **110**(6):1657-1666.
68. Jorgensen WL, Maxwell DS, Rives T: **Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids**. *J Am Chem Soc* 1996, **118**(45):11225-11236.
69. Reiser JB, Darnault C, Guimezanes A, Grégoire C, Mosser T, Schmitt-Verhulst AM, Fontecilla-Camps JC, Malissen B, Housset D, Mazza G: **Crystal structure of a T cell receptor bound to an allogeneic MHC molecule**. *Nat Immunol* 2000, **1**(4):291-297.
70. Feng D, Bond CJ, Ely LK, Maynard J, Garcia KC: **Structural evidence for a germline-encoded T cell receptor-major histocompatibility complex interaction 'codon'**. *Nat Immunol* 2007, **8**(9):975-983.
71. Hennecke J, Carfi A, Wiley DC: **Structure of a covalently stabilized complex of a human alphabeta T-cell receptor, influenza HA peptide and MHC class II molecule, HLA-DR1**. *EMBO J* 2000, **19**(21):5611-5624.
72. Hennecke J, Wiley DC: **Structure of a complex of the human alpha/beta T cell receptor (TCR) HA1.7, influenza hemagglutinin peptide, and major histocompatibility complex class II molecule, HLA-DR4 (DRA*0101 and DRB1*0401): insight into TCR cross-restriction and alloreactivity**. *J Exp Med* 2002, **195**(5):571-581.
73. Ishizuka J, Stewart-Jones GBE, van der Merwe A, Bell JI, McMichael AJ, Jones EY: **The structural dynamics and energetics of an immunodominant T cell receptor are programmed by its Vbeta domain**. *Immunity* 2008, **28**(2):171-182.
74. Reiser JB, Grégoire C, Darnault C, Mosser T, Guimezanes A, Schmitt-Verhulst AM, Fontecilla-Camps JC, Mazza G, Malissen B, Housset D: **A T cell receptor CDR3beta loop undergoes conformational changes of unprecedented magnitude upon binding to a peptide/MHC class I complex**. *Immunity* 2002, **16**(3):345-354.
75. Kjer-Nielsen L, Clements CS, Purcell AW, Brooks AG, Whisstock JC, Burrows SR, McCluskey J, Rossjohn J: **A structural basis for the selection of dominant alphabeta T cell receptors in antiviral immunity**. *Immunity* 2003, **18**(1):53-64.
76. Dai S, Huseby ES, Rubtsova K, Scott-Browne J, Crawford F, Macdonald WA, Marrack P, Kappler JW: **Crossreactive T Cells spotlight the germline rules for alphabeta T cell-receptor interactions with MHC molecules**. *Immunity* 2008, **28**(3):324-334.
77. Luz JG, Huang M, Garcia KC, Rudolph MG, Apostolopoulos V, Teyton L, Wilson IA: **Structural comparison of allogeneic and syngeneic T cell receptor-peptide-major histocompatibility complex complexes: a buried alloreactive mutation subtly alters peptide presentation substantially increasing V(beta) Interactions**. *J Exp Med* 2002, **195**(9):1175-1186.

78. Reiser J-B, Darnault C, Grégoire C, Mosser T, Mazza G, Kearney A, van der Merwe PA, Fontecilla-Camps JC, Housset D, Malissen B: **CDR3 loop flexibility contributes to the degeneracy of TCR recognition**. *Nat Immunol* 2003, **4**(3):241-247.
79. Stewart-Jones GBE, McMichael AJ, Bell JI, Stuart DI, Jones EY: **A structural basis for immunodominant human T cell receptor recognition**. *Nat Immunol* 2003, **4**(7):657-663.
80. Piepenbrink KH, Borbulevych OY, Sommese RF, Clemens J, Armstrong KM, Desmond C, Do P, Baker BM: **Fluorine substitutions in an antigenic peptide selectively modulate T-cell receptor binding in a minimally perturbing manner**. *Biochem J* 2009, **423**(3):353-361.
81. Maynard J, Petersson K, Wilson DH, Adams EJ, Blondelle SE, Boulanger MJ, Wilson DB, Garcia KC: **Structure of an autoimmune T cell receptor complexed with class II peptide-MHC: insights into MHC bias and antigen specificity**. *Immunity* 2005, **22**(1):81-92.
82. Archbold JK, Macdonald WA, Gras S, Ely LK, Miles JJ, Bell MJ, Brennan RM, Beddoe T, Wilce MCJ, Clements CS *et al*: **Natural micropolymorphism in human leukocyte antigens provides a basis for genetic control of antigen recognition**. *J Exp Med* 2009, **206**(1):209-219.
83. Colf LA, Bankovich AJ, Hanick NA, Bowerman NA, Jones LL, Kranz DM, Garcia KC: **How a single T cell receptor recognizes both self and foreign MHC**. *Cell* 2007, **129**(1):135-146.
84. Gras S, Burrows SR, Kjer-Nielsen L, Clements CS, Liu YC, Sullivan LC, Bell MJ, Brooks AG, Purcell AW, McCluskey J *et al*: **The shaping of T cell receptor recognition by self-tolerance**. *Immunity* 2009, **30**(2):193-203.
85. Hoare HL, Sullivan LC, Pietra G, Clements CS, Lee EJ, Ely LK, Beddoe T, Falco M, Kjer-Nielsen L, Reid HH *et al*: **Structural basis for a major histocompatibility complex class Ib-restricted T cell response**. *Nat Immunol* 2006, **7**(3):256-264.
86. Gras S, Saulquin X, Reiser J-B, Debeaupuis E, Echasserieau K, Kissenpfennig A, Legoux F, Chouquet A, Gorrec ML, Machillot P *et al*: **Structural bases for the affinity-driven selection of a public TCR against a dominant human cytomegalovirus epitope**. *J Immunol* 2009, **183**(1):430-437.
87. Borbulevych OY, Piepenbrink KH, Gloor BE, Scott DR, Sommese RF, Cole DK, Sewell AK, Baker BM: **T cell receptor cross-reactivity directed by antigen-dependent tuning of peptide-MHC molecular flexibility**. *Immunity* 2009, **31**(6):885-896.
88. Macdonald WA, Chen Z, Gras S, Archbold JK, Tynan FE, Clements CS, Bharadwaj M, Kjer-Nielsen L, Saunders PM, Wilce MCJ *et al*: **T cell allorecognition via molecular mimicry**. *Immunity* 2009, **31**(6):897-908.
89. Gagnon SJ, Borbulevych OY, Davis-Harrison RL, Turner RV, Damirjian M, Wojnarowicz A, Biddison WE, Baker BM: **T cell receptor recognition via cooperative conformational plasticity**. *J Mol Biol* 2006, **363**(1):228-243.
90. Deng L, Langley RJ, Brown PH, Xu G, Teng L, Wang Q, Gonzales MI, Callender GG, Nishimura MI, Topalian SL *et al*: **Structural basis for the recognition of mutant self by a tumor-specific, MHC class II-restricted T cell receptor**. *Nat Immunol* 2007, **8**(4):398-408.
91. Burrows SR, Chen Z, Archbold JK, Tynan FE, Beddoe T, Kjer-Nielsen L, Miles JJ, Khanna R, Moss DJ, Liu YC *et al*: **Hard wiring of T cell receptor specificity for the major histocompatibility complex is underpinned by TCR adaptability**. *Proc Natl Acad Sci U S A* 2010, **107**(23):10608-10613.

92. Yoshida K, Corper AL, Herro R, Jabri B, Wilson IA, Teyton L: **The diabetogenic mouse MHC class II molecule I-Ag7 is endowed with a switch that modulates TCR affinity**. *J Clin Invest* 2010, **120**(5):1578-1590.
93. Tynan FE, Reid HH, Kjer-Nielsen L, Miles JJ, Wilce MCJ, Kostenko L, Borg NA, Williamson NA, Beddoe T, Purcell AW *et al*: **A T cell receptor flattens a bulged antigenic peptide presented by a major histocompatibility complex class I molecule**. *Nat Immunol* 2007, **8**(3):268-276.
94. Gras S, Chen Z, Miles JJ, Liu YC, Bell MJ, Sullivan LC, Kjer-Nielsen L, Brennan RM, Burrows JM, Neller MA *et al*: **Allelic polymorphism in the T cell receptor and its impact on immune responses**. *J Exp Med* 2010, **207**(7):1555-1567.
95. Borbulevych OY, Piepenbrink KH, Baker BM: **Conformational melding permits a conserved binding geometry in TCR recognition of foreign and self molecular mimics**. *J Immunol* 2011, **186**(5):2950-2958.
96. Mazza C, Auphan-Anezin N, Gregoire C, Guimezanes A, Kellenberger C, Roussel A, Kearney A, van der Merwe PA, Schmitt-Verhulst A-M, Malissen B: **How much can a T-cell antigen receptor adapt to structurally distinct antigenic peptides?** *EMBO J* 2007, **26**(7):1972-1983.
97. Newell EW, Ely LK, Kruse AC, Reay PA, Rodriguez SN, Lin AE, Kuhns MS, Garcia KC, Davis MM: **Structural basis of specificity and cross-reactivity in T cell receptors specific for cytochrome c-I-E(k)**. *J Immunol* 2011, **186**(10):5823-5832.
98. Hess B, Kutzner C, van der Spoel D, Lindahl E: **GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation**. *Journal of Chemical Theory and Computation* 2008, **4**(3):435-447.
99. Almagro JC, Teplyakov A, Luo J, Sweet RW, Kodangattil S, Hernandez-Guzman F, Gilliland GL: **Second antibody modeling assessment (AMA-II)**. *Proteins* 2014, **82**(8):1553-1562.
100. Wales D, Doye J: **Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms**. *The Journal of Physical Chemistry A* 1998, **101**(28):5111-5116.
101. Kusumaatmaja H, Whittleston C, Wales D: **A Local Rigid Body Framework for Global Optimization of Biomolecules**. *Journal of Chemical Theory and Computation* 2012, **8**(12):5159-5165.
102. Beddoe T, Chen Z, Clements CS, Ely LK, Bushell SR, Vivian JP, Kjer-Nielsen L, Pang SS, Dunstone MA, Liu YC *et al*: **Antigen ligation triggers a conformational change within the constant domain of the alphabeta T cell receptor**. *Immunity* 2009, **30**(6):777-788.

Tables

Table 1. Description of the structural dataset DS_T and the subset DS_C.

PDB	DS[a]	TCR-Name	S[b]	L[c]	R[d]	PDB	DS[a]	TCR-Name	S[b]	L[c]	R[d]
1ao7	T/C	A6	H	I	[25]	2p5w	T/C	1G4 c58c62	H	I	[39]
1fo0	T/C	BM3.3	M	I	[69]	2pxy	T	1934.4	M	II	[70]
1fyt	T	HA1.7	H	II	[71]	2pye	T/C	1G4 c5c1	H	I	[39]
1j8h	T	HA1.7	H	II	[72]	2vlk	T/C	JM22	H	I	[73]
1kj2	T/C	KB5-C20	M	I	[74]	2vlr	T/C	JM22	H	I	[73]
1mi5	T	LC13	H	I	[75]	3c5z	T	B3K506	M	II	[76]
1mwa	T/C	2C	M	I	[77]	3c60	T	YAe62	M	II	[76]
1nam	T/C	BM3.3	M	I	[78]	3c6l	T	2W20	M	II	[76]
1oga	T/C	JM22	H	I	[79]	3d39	T/C	A6	H	I	[80]
1qse	T	A6	H	I	[35]	3d3v	T/C	A6	H	I	[80]
1u3h	T	TCR172.10	M	II	[81]	3dxa	T/C	DM1	H	I	[82]
2bnq	T/C	1G4	H	I	[37]	3e2h	T/C	2C m67 [T7]	M	I	[65]
2bnr	T/C	1G4	H	I	[37]	3e3q	T/C	2C m13 [T7]	M	I	[65]
2e7l	T/C	2C m6 [T7]	M	I	[83]	3ffc	T/C	cf34	H	I	[84]
2esv	T/C	KK50.4	H	I	[85]	3gsn	T/C	RA14	H	I	[86]
2f53	T/C	1G4 c49c50	H	I	[38]	3h9s	T/C	A6	H	I	[87]
2f54	T/C	1G4 AV-wt	H	I	[38]	3kpr	T/C	LC13	H	I	[88]
2gj6	T/C	A6	H	I	[89]	3kps	T/C	LC13	H	I	[88]
2iam	T	E8	H	II	[90]	3kxf	T/C	SB27	H	I	[91]
2ian	T	E8	H	II	[90]	3mbe	T	TCR 21.30	M	II	[92]
2nx5	T/C	ELS4	H	I	[93]	3mv8	T/C	TK3 Q55H	H	I	[94]
2oi9	T/C	2C [T7-wt]	M	I	[83]	3pwp	T/C	A6	H	I	[95]
2ol3	T/C	BM3.3	M	I	[96]	3qiu	T	226 TCR	M	II	[97]
2p5e	T/C	1G4 c58c61	H	I	[39]	3qiw	T	226 TCR	M	II	[97]

a) T: Structure only in dataset DS_T. T/C: Structure in both datasets, DS_T and DS_C.
b) Species (S): H=human, M=mouse.
c) Ligand type (L): MHC class I or II. See Table S1 of the Supporting Information for details about the MHC alleles and the different peptides.
d) References

Table 2. Prediction accuracy for the Vα/Vβ association angles modeled without pMHC.

ES[a]	Variants[b]		C_R[c]		C_E[d]	
	Q	R	#	(%)	#	(%)
M_T00	off	off	71	(94.7)	63	(84.0)
M_T01	off	on	71	(94.7)	66	(88.0)
M_T10	on	off	72	(96.0)	64	(85.3)
M_T11	on	on	72	(96.0)	67	**(89.3)**
M_T^*11	on	on	47	(97.9)	43	**(89.6)**

a) Evaluation setting label.
b) Variants: Q = glutamine carboxamide group orientation correction, R = rigid body position restraint.
c) Absolute and relative prediction rate according to the RMSD based criterion (i.e., C_R) in data set DS_T (75 structures). In the particular case of M_T^*11, the prediction was performed on the DS_T^* set (48 structures, without biological units). For each prediction run, the 11 models are ranked by RMSD and a success is counted if the selected structure has an RMSD value lower than 2 Å.
d) Same as c) using the energy criterion to rank the 11 structures and select the best. The prediction is considered as successful if the selected structure has an RMSD value lower than 2 Å.

Table 3. Prediction accuracy for the combined prediction of the Vα/Vβ and TCR/pMHC association angles.

ES[a]	Variants[b]		C_R[c]		C_E[d]	
	Q	R	#	(%)	#	(%)
M_C00	off	off	39	(73.6)	37	(69.8)
M_C01	off	on	39	(73.6)	38	(71.7)
M_C10	on	off	39	(73.6)	38	(71.7)
M_C11	on	on	39	(73.6)	38	**(71.7)**

a) Evaluation setting label.
b) Variants: Q = glutamine carboxamide group orientation correction, R = rigid body position restraint.
c) Absolute and relative prediction rate according to the RMSD based criterion (i.e., C_R) in data set DS_C (53 structures). For each prediction run, the 11 models are ranked by RMSD and a success is counted if the selected structure has an RMSD value lower than 2 Å.
d) Same as c) using the energy criterion to rank the 11 structures and select the best. The prediction is considered as successful if the selected structure has an RMSD value lower than 2 Å.

Table 4. Prediction rates of the test evaluations.

ES[a]	MHC initial orientation[b]	MHC translation[c]	C_R[d]	
			#	(%)
T_1	crystal	no	53	(100.0)
T_2	crystal	yes	38	(71.7)
T_3	general	yes	31	(58.5)

For each test the Q-flip correction as well as the use of distance restraint are systematically applied. The TCR Vβ domain is placed in its original crystal structure orientation and is optimized. The tests are performed for each of the 53 structures present in the DS_C data set and the MHC rigid body position is optimized in each case.
a) Evaluation setting label.
b) The initial orientation of the MHC unit is chosen either according to the original crystal structure or using the general zero orientation as for the standard version of our pipeline (see Material and Methods Section for more details).
c) Initial translation of the MHC unit to avoid steric hindrance, necessary if the MHC rigid body is not placed according to the crystal structure orientation (see Material and Methods Section for more details).
d) Absolute and relative prediction rate according to the RMSD based criterion (i.e., C_R) in data set DS_C (53 structures). For each prediction run, a success is counted if the resulting model has an RMSD value lower than 2 Å with respect to the crystal structure.

Additional Files:

Movie S1: Example for the prediction pipeline: Remodling of the structure with the PDB-ID 1ao7.

Text S1: Definition of the 11 starting conformation for the Vα/Vβ association angles.

Text S2: Detailed description of the operators.

Text S3: Glutamine correction and adaption (detailed Results and Discussion).

Fig S1: Discrimination of the models.

Fig S2: Influence of the restraint operator.

Table S1: Structural Dataset DS_T and the subset DS_C

Table S2: Performance of the Q-Q interaction optimization.

Table S3: Per residue flip states using Reduce, Protoss and DynaDom comparing single domains and TCR complexes.

Table S4: Angular deviations with respect to the crystal structures after DynaDom glutamine refinement.

Table S5: pMHC optimization for the structure 1oga with different pMHC start conformations.

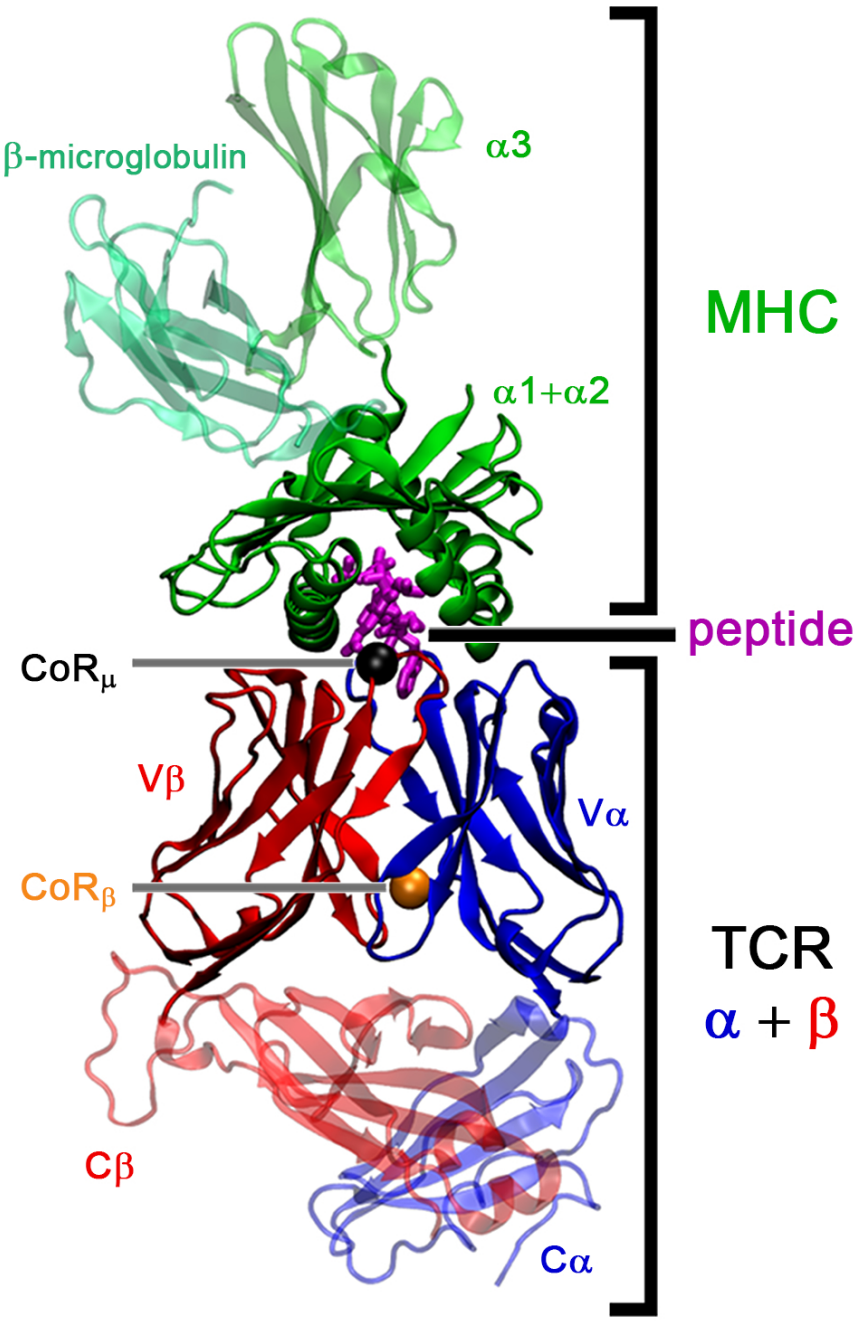


Figure C.1: **Representation of the TCRpMHC complex (PDB-ID 2bnq).** The MHC class I molecule is depicted in green (i.e., α_1, α_2, and α_3 domains). The β-microglobulin is colored in cyan and the peptide bound to MHC in magenta. The two chains of TCR, α and β, are represented in blue and red colors, respectively. In the present application, the domains shown as transparent are removed from the structure, and only the two variable domains of TCR (i.e., Vα and Vβ), the α_1 and α_2 domains of MHC as well as the peptide are modeled. In addition, the two centers of rotations CoR_β and CoR_μ, are respectively represented by an orange and a black colored ball.

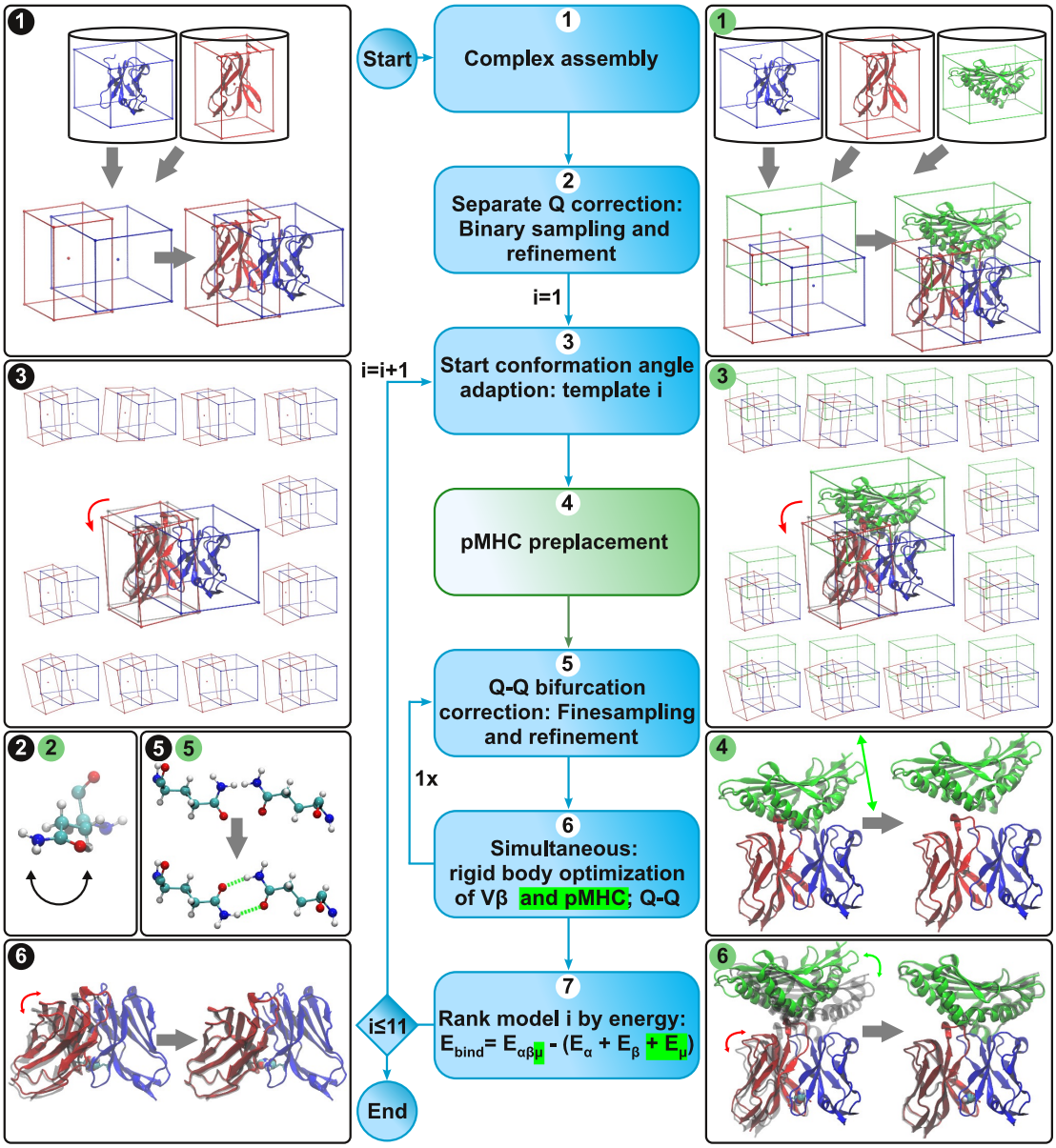


Figure C.2: **TCR and TCRpMHC complexes modeling pipeline.** Center column: standard pipeline (see Materials and Methods) for the remodeling of the TCR Vα/Vβ association angles and for the pMHC positioning with respect to the TCR. Blue highlighted steps are performed in both modeling pipelines: only Vβ and combined Vβ/pMHC placement. Green highlighted steps are performed only if the pMHC is included in the remodeling process. The left and the right columns illustrate the individual steps of the pipeline. Steps with numbers circled in black: TCR Vα/Vβ association angle modeling pipeline, steps with numbers circled in green: combined Vβ/pMHC modeling pipeline. Steps 3 to 7 are performed for each of the 11 starting conformations. The protein domains represented in blue, red, and green color correspond to the Vα, Vβ, and pMHC units, respectively..

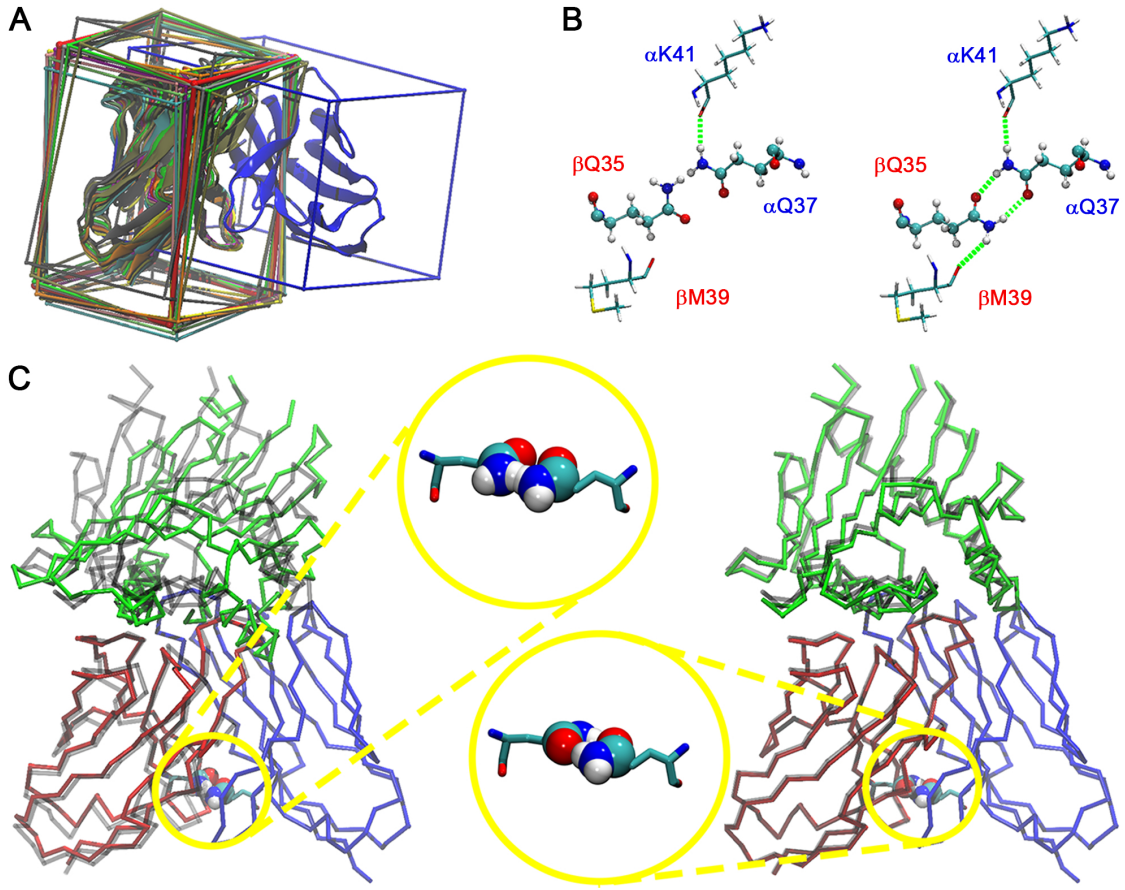


Figure C.3: **Remodeling of the 2f53 structure.** A) Superposition of the 11 Vβ starting orientations with respect to the Vα domain (represented in blue color). The average conformation of Vβ is shown in red color. B) Hydrogen bonds of the conserved Q-Q interaction at the CoR_{β} position. Left: misassigned conformation in the experimental crystal structure. Right: proper orientation of the Q residues after application of the Q-flip correction. The picture shows that the interaction between the two residues has been improved as well as the interaction of the residues with their respective environment. C) Modeling of the ternary TCRpMHC complex. The Vα, Vβ, and pMHC units are represented in blue, red, and green colors, respectively. The reference crystal structure is depicted in gray color. Left: initial assembly of the complex. Right: final model with an RMSD of 0.61 Å with respect to the crystal structure. Magnifications lenses: conformation of the conserved Q-Q interaction between the Vα and the Vβ domain.

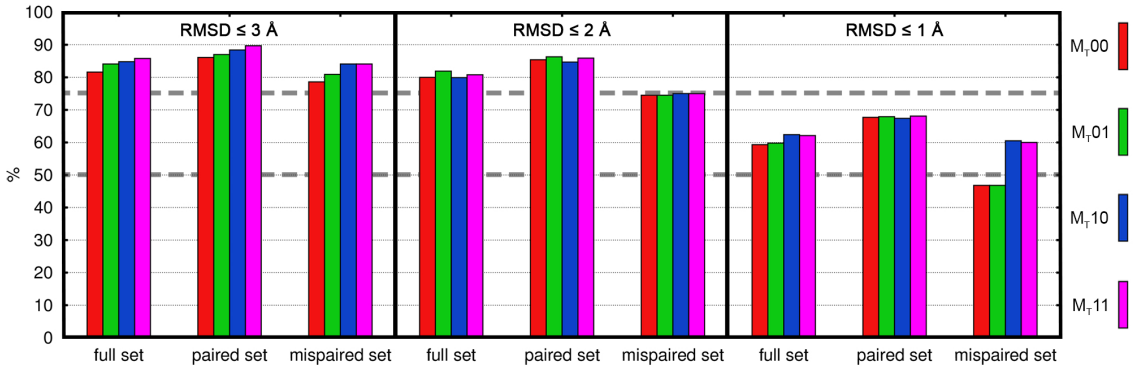


Figure C.4: **Percentage of structures with an RMSD value lower than 3, 2 and 1 Å among the 75*11 models.** The total set of 75*11 structures is separated into structures for which the Q residues were originally paired or mispaired within their corresponding crystal structure. Each histogram box corresponds to a different setting of the modeling procedure, i.e. with only distance restraint (M_T01), only Q-flip correction (M_T10), both (M_T11), or none them (M_T00). The percentage of structures with an RMSD value lower than 3, 2 and 1 Å are presented on the left, middle, and right plots, respectively. The right plot shows that for the structures presenting an originally wrong orientation of the Q residues, the Q-flip correction significantly improves the quality of the resulting model (i.e., M_T10 and M_T11).

C.3 Supporting Information

T. Hoffmann, A. Marion, and I. Antes - **DynaDom: Structure-based prediction of TCR inter-domain and TCR-pMHC association angles**

Additional File 1

Text S1: Definition of the 11 starting conformation for the Vα/Vβ association angles.

As we discussed in the main text, the DynaDom strategy uses multiple starting orientations of the Vβ domain with respect to Vα for each remodeling process attempt. For the present application, we defined 11 different orientations based on our former analysis of Vα/Vβ association angles [1]. This study was performed on a large set of 85 structures. In this former work, the analysis of the pairwise Euler angles distance matrix resulted in the identification of six structural clusters. These six cluster are labeled as cl1 to cl6 in ref. [1]. The 11 orientations used in the present work were derived from these clusters as follows:

- We included the representative structure of each cluster (6 orientations, labeled as cl1 to cl6 in Additional file 9: Fig. S1)
- We computed the average over the six representative structures (1 orientation, labeled as Av in Additional file 9: Fig. S1)
- The two main clusters (cl4 and cl6 in ref. [1]) were spitted into 2 sub-clusters each. The representative structure of each sub-cluster was then included here (4 orientations, labeled as cl4a, cl4b, cl6a, and cl6b in Additional file 9: Fig. S1).

References

1. Hoffmann T, Krackhardt AM, Antes I. Quantitative Analysis of the Association Angle between T-cell Receptor Valpha/Vbeta Domains Reveals Important Features for Epitope Recognition. PLoS computational biology. 2015;11(7):e1004244. doi: 10.1371/journal.pcbi.1004244. PubMed PMID: 26185983; PubMed Central PMCID: PMC4505886.

T. Hoffmann, A. Marion, and I. Antes - **DynaDom: Structure-based prediction of TCR inter-domain and TCR-pMHC association angles**
Additional File 2

Text S2: Detailed description of the operators.
The implementation of the DynaDom approach is based on the definition of three basic operator functions, which are combined in the final prediction pipeline, which is discussed in the main text. Her we provide a detailed description of these functions.

General objective function for the rigid body optimization operators: All rigid body operators (RMZ, RAT, and RAR) operate on a selection s_i, which is transformed relatively to an additional part of the system defined as context c_i. For the evaluation of the objective function only non-bonded interactions between atoms of the s_i and of the remaining context c_i are taken into account. Internal non-bonded interactions within the selected rigid body are not computed, as they remain constant during rigid body operations. The contribution of the objective function e_i is computed as follows:

$$e_{RXX,i} = \sum_{a \in s_i} \sum_{\substack{b \in c_i \\ b \notin s_i}} (E_{vdW}(a,b) + E_{Coul}(a,b)) \qquad (1)$$

In the context of the above described example: let w_1 and w_2 be atoms of the Vβ domain chain B, and w_3 and w_4 atoms of the Vα domain chain A. Then, $\{w_1, w_{2\}}\} \subseteq s_{TV\beta}$ is the selection and $\{w_1, w_2, w_3, w_4\} \subseteq c_{TV\beta}$ is the context, the intra-selection interactions (w_1, w_2) are ignored, whereas the inter-domain interactions (w_1, w_3), (w_1, w_4), (w_2, w_3), and (w_2, w_4) are taken into account.

The rigid body rotation and translation operator (RMZ) is based on the work of Mirzaei *et al.,* who recently introduced an algorithm focusing on the RBEM problem for molecular docking [2]. To minimize the interaction energy between two molecules, this algorithm rotates one molecule around a center (usually the center of mass), whereby this center is concurrently slightly translated to an optimum position. Their implementation makes use of the limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) Newton-Raphson method [3] and of a transformation of the force field gradient based on the Rodrigues formula. The approach was shown to outperform other RBEM methods by an order of magnitude in time [2].
The operator takes the non-bonded interactions between the flexible rigid body and the defined context c_i (without s_i) of the system into account, according to Equation 1. The operator simultaneously modifies three parameters for the rotation and three parameters for the translation of the rigid body. The gradient evaluation is discussed in detail in ref. [2]. We extended this method to allow for the simultaneous structural optimization of multiple rigid bodies. In the context of the TCRpMHC modeling, both the Vβ domain and the pMHC ligand are translated and rotated in one step.

The rigid body axis operators RAT and RAR: These operators define the separate translation along (RAT) or the rotation around (RAR) an axis according to the forces between the treated rigid body (s_i) and the remaining context c_i without s_i. The objective function e_i is computed according to Equation 1 of the main text.
To compute the gradient for the RAT operator, the non-bonded forces are projected on a predefined normalized axis ($\vec{A}$) as:

$$\overrightarrow{F_{RAT,i}} = \sum_{a \in s_i} \sum_{\substack{b \in c_i \\ b \notin s_i}} (\overrightarrow{F_{vdW}(a,b)} + \overrightarrow{F_{Coul}(a,b)}) \qquad \text{(S1)},$$

$$g_{RAT,i} = - \overrightarrow{F_{RAT,i}} \cdot \vec{A} \qquad \text{(S2)}$$

For the RAR operator the gradient is computed as the one-dimensional tangential force gradient:

$$F_t(s_i, c_i, \vec{A}, \vec{R}) = \sum_{a \in s_i} \sum_{\substack{b \in c_i \\ b \notin s_i}} \overrightarrow{T^r_{ab}} \cdot (1,0,0) \cdot \overrightarrow{D^r_a} \cdot (0,1,0) \qquad \text{(S3)}$$

$$g_{RAR,i} = -F_t(s_i, c_i, \vec{A}, \vec{R}) \quad \text{(S4)}$$

where s_i denotes the atoms of the treated rigid body, and where c_i is the remaining context to interact with. $\vec{R}$ is the coordinate vector of a predefined center of rotation and $\vec{A}$ defines the normalized rotation axis. F_t is the one dimensional tangential force vector with respect to the rotation axis $\vec{A}$ for all atoms a of the selection s_i interacting with all atoms b of the context c_i. The vector $\overrightarrow{T_{ab}^r}$ is the transformed projected force based on the objective function e_i, and $\overrightarrow{D_a^r}$ is the transformed distance vector of atom a with respect with respect to the origin.

The glutamine and asparagine carboxamide group rotation operator (AQR) defines the rotation of a glutamine or asparagine carboxamide group around an axis defined in the direction of the bond between two adjacent side chain atoms (*i.e.* Cβ-Cγ and Cγ-Cδ for an asparagine and a glutamine residue, respectively). The selection s_i ($s_i \subseteq c_i$) contains all atoms of the carboxamide group and the context is usually either the whole remaining protein or all atoms sharing the domain with the carboxamide group. The objective function contribution e_i is computed as follows:

$$e_{AQR,i} = \sum_{a \in s_i} \sum_{\substack{b \in c_i \\ b \neq a}} (E_{vdW}(a,b) + E_{Coul}(a,b) + E_{bonded}(a,b)) \quad \text{(S5)}$$

The evaluation of the tangential force gradient for the AQR operator is based on Equations S3 and S4, but differs from the RAR evaluation in two aspects: First, for the RAR operator, interactions within the selection are not taken into account, whereas for the AQR these interactions are computed and second, also bonded interactions are considered.

Rigid body position restraints (RPS and RPC): To avoid unfavorable large translational moves, we implemented two types of restraint operators, which can be combined with the rigid body operators RAT and RMZ. The restraint operators allow for a certain tolerance of deviation and add distance dependent restraints to the translational gradients and to the objective function. The first type of restraint operator (RPS) restrains the coordinates of one single atom contained in the rigid body to be around a defined position. The second type (RPC) restrains the geometric center of a set of atoms contained in the rigid body to be around a defined position.
For both cases harmonic restraints are used and are computed as follows:

$$e_{rest} = \begin{cases} 0, & if\ d \leq t \\ \frac{k}{2} \cdot (d-t)^2 & \end{cases} \quad \text{(S6)}$$

where t is the threshold, d is the distance between the desired reference position of the rigid body ($\vec{R}$) and the actual position of the rigid body ($\vec{P}$) and k an adjustable force constant (here, k = 100.000 kJ mol^{-1} nm^{-2}).
To restrain the position of the TCR Vβ domain, the RPS restraint operator is used. For this purpose, the reference position $\vec{R}$ is set to the Center of Rotation defined in ref. [1], the actual position $\vec{P}$ is defined as the coordinates of the C_β-atom of the CoR_β Q residue, and the threshold is set to 0.75 nm.
To restrain the pMHC position, the RPC operator is used. As stated above, a conserved residue at the CoR_μ could not be found and the CoR_μ is located close to the center of the bound peptide. Due to the lack of a conserved residue, we define the actual position $\vec{P}$ of the rigid body pMHC as the geometric center between the C- and N-terminal C_α-atoms of the bound peptide. The reference position $\vec{R}$ is set to the

CoR_{μ}, and a threshold of 2.7 nm and 1.3 nm is used for the preplacement pipeline step and the main pipeline step, respectively (see below for the pipeline).

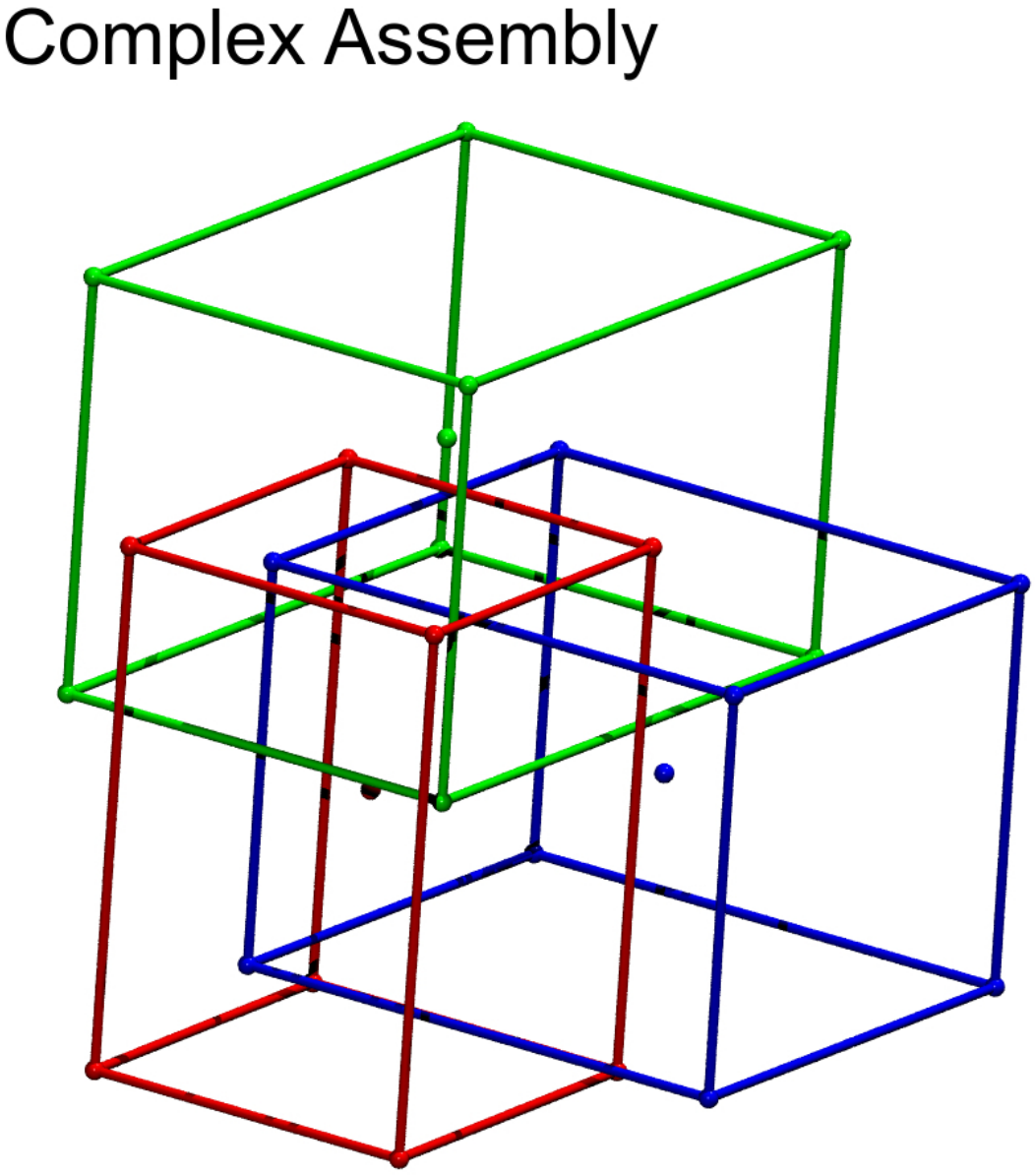
References

1. Hoffmann T, Krackhardt AM, Antes I. Quantitative Analysis of the Association Angle between T-cell Receptor Valpha/Vbeta Domains Reveals Important Features for Epitope Recognition. PLoS computational biology. 2015;11(7):e1004244. doi: 10.1371/journal.pcbi.1004244. PubMed PMID: 26185983; PubMed Central PMCID: PMC4505886.
2. Mirzaei H, Beglov D, Paschalidis IC, Vajda S, Vakili P, Kozakov D. Rigid Body Energy Minimization on Manifolds for Molecular Docking. Journal of chemical theory and computation. 2012;8(11):4374-80. doi: 10.1021/ct300272j. PubMed PMID: 23382659; PubMed Central PMCID: PMC3561712.
3. Liu D, Nocedal J. On the limited memory BFGS method for large scale optimization. Mathematical Programming. 1989;45(1-3):503-28. doi: 10.1007/BF01589116.

T. Hoffmann, A. Marion, and I. Antes - **DynaDom: Structure-based prediction of TCR inter-domain and TCR-pMHC association angles**

Additional File 3

Movie S1: Example for the prediction pipeline.
Remodeling of the structure with the PDB-ID 1ao7. The complete TCRpMHC modeling pipeline is shown for one of the eleven starting conformations (cl4a, for details see Additional file 1: Text S1) of structure 1ao7. After the preplacement of the pMHC ligand, the experimental crystal structure appears in transparent black as a reference. The final model converged to a conformation with an RMSD in respect to the experimental structure of 0.6 Å. Note: Both glutamine residues at the center of rotation are flipped in the model.



Remark: The movie will be available online with Publication 3. Playback of the movie embedded in the digital document requires an appropriate PDF viewer, such as Adobe Reader version 8 or higher.

T. Hoffmann, A. Marion, and I. Antes - **DynaDom: Structure-based prediction of TCR inter-domain and TCR-pMHC association angles**

Additional File 4

Table S1: Structural Dataset DS_T and the subset DS_C (part 1).

BU[a]	D[b]	Name	S[c]		Ligand		
				C[d]	Peptide	MHC I/II α	MHC II β
1AO7/DE	T/C	A6	H	I	LLFGYPVYV	HLA-A*0201	
1FO0/AB	T/C	BM3.3	M	I	INFDFNTI	H2-K1(b)	
1FYT/DE	T	HA1.7	H	II	PKYVKQNTLKLAT	HLA-DRA*0101	HLA-DRB1*010
1J8H/DE	T	HA1.7	H	II	PKYVKQNTLKLAT	HLA-DRA*0101	HLA-DRB1*040
1KJ2/AB	T/C	KB5-C20	M	I	KVITFIDL	H2-K1(b)	
1KJ2/DE	T/C	KB5-C20	M	I	KVITFIDL	H2-K1(b)	
1MI5/DE	T	LC13	H	I	FLRGRAYGL	HLA-B*0801	
1MWA/AB	T/C	2C	M	I	EQYKFYSV	H2-K1(bm3)	
1NAM/AB	T/C	BM3.3	M	I	RGYVYQGL	H2-K1(b)	
1OGA/DE	T/C	JM22	H	I	GILGFVFTL	HLA-A*0201	
1QSE/DE	T	A6	H	I	LLFGYPRYV	HLA-A*0201	
1U3H/AB	T	TCR172.10	M	II	SRGGASQYRPSQ	H2-Aa(u)	H2-Ab(u)
1U3H/EF	T	TCR172.10	M	II	SRGGASQYRPSQ	H2-Aa(u)	H2-Ab(u)
2BNQ/DE	T/C	1G4	H	I	SLLMWITQV	HLA-A*0201	
2BNR/DE	T/C	1G4	H	I	SLLMWITQC	HLA-A*0201	
2E7L/AD	T/C	2C m6 [T7]	M	I	QLSPFPFDL	H2-L(d)	
2E7L/BC	T/C	2C m6 [T7]	M	I	QLSPFPFDL	H2-L(d)	
2ESV/DE	T/C	KK50.4	H	I	VMAPRTLIL	HLA-E*0101	
2F53/DE	T/C	1G4 c49c50	H	I	SLLMWITQC	HLA-A*0201	
2F54/DE	T/C	1G4 AV-wt	H	I	SLLMWITQC	HLA-A*0201	
2F54/KL	T/C	1G4 AV-wt	H	I	SLLMWITQC	HLA-A*0201	
2GJ6/DE	T/C	A6	H	I	LLFGKPVYV	HLA-A*0201	
2IAM/CD	T	E8	H	II	GELIGILNAAKVPAD	HLA-DRA*0101	HLA-DRB1*010
2IAN/DE	T	E8	H	II	GELIGTLNAAKVPA	HLA-DRA*0101	HLA-DRB1*010
2IAN/IJ	T	E8	H	II	GELIGTLNAAKVPA	HLA-DRA*0101	HLA-DRB1*010
2IAN/NO	T	E8	H	II	GELIGTLNAAKVPA	HLA-DRA*0101	HLA-DRB1*010
2IAN/ST	T	E8	H	II	GELIGTLNAAKVPA	HLA-DRA*0101	HLA-DRB1*010
2NX5/IJ	T/C	ELS4	H	I	EPLPQGQLTAY	HLA-B*3501	
2NX5/NP	T/C	ELS4	H	I	EPLPQGQLTAY	HLA-B*3501	
2NX5/TU	T/C	ELS4	H	I	EPLPQGQLTAY	HLA-B*3501	
2OI9/BC	T/C	2C [T7-wt]	M	I	QLSPFPFDL	H2-L(d)	
2OL3/AB	T/C	BM3.3	M	I	SQYYYNSL	H2-K1(bm8)	
2P5E/DE	T/C	1G4 c58c61	H	I	SLLMWITQC	HLA-A*0201	
2P5W/DE	T/C	1G4 c58c62	H	I	SLLMWITQC	HLA-A*0201	
2PXY/AB	T	1934.4	M	II	HSRGGASQYRPSQ	H2-Aa(u)	H2-Ab(u)
2PYE/DE	T/C	1G4 c5c1	H	I	SLLMWITQC	HLA-A*0201	
2VLK/DE	T/C	JM22	H	I	GILGFVFTL	HLA-A*0201	
2VLR/DE	T/C	JM22	H	I	GILGFVFTL	HLA-A*0201	
2VLR/IJ	T/C	JM22	H	I	GILGFVFTL	HLA-A*0201	
3C5Z/AB	T	B3K506	M	II	FEAQKAKANKAVD	H2-Aa(b)	H2-Ab(b)
3C5Z/EF	T	B3K506	M	II	FEAQKAKANKAVD	H2-Aa(b)	H2-Ab(b)
3C60/AB	T	YAe62	M	II	FEAQKAKANKAVD	H2-Aa(b)	H2-Ab(b)
3C60/EF	T	YAe62	M	II	FEAQKAKANKAVD	H2-Aa(b)	H2-Ab(b)
3C6L/AB	T	2W20	M	II	FEAQKAKANKAVD	H2-Aa(b)	H2-Ab(b)
3C6L/EF	T	2W20	M	II	FEAQKAKANKAVD	H2-Aa(b)	H2-Ab(b)
3D39/DE	T/C	A6	H	I	LLFGFPVYV	HLA-A*0201	
3D3V/DE	T/C	A6	H	I	LLFGFPVYV	HLA-A*0201	

a) Biological unit (BU) given as PDB ID and the two TCR chains as indicator.

b) T: only in DS_T; T/C: contained in both datasets, DS_T & DS_C.

c) Species: either human (H) or mouse (M).

d) MHC class presents in the original crystal structure.

T. Hoffmann, A. Marion, and I. Antes - **DynaDom: Structure-based prediction of TCR inter-domain and TCR-pMHC association angles**

Additional File 4

Table S1: Structural Dataset DS_T and the subset DS_C (continued).

BU[a]	D[b]	Name	S[c]	C[d]	Ligand Peptide	MHC I/II α	MHC II β
3DXA/DE	T/C	DM1	H	I	EENLLDFVRF	HLA-B*4405	
3DXA/IJ	T/C	DM1	H	I	EENLLDFVRF	HLA-B*4405	
3DXA/NO	T/C	DM1	H	I	EENLLDFVRF	HLA-B*4405	
3E2H/BC	T/C	2C m67[T7]	M	I	QLSPFPFDL	H2-L(d)	
3E3Q/CF	T/C	2C m13[T7]	M	I	QLSPFPFDL	H2-L(d)	
3E3Q/de	T/C	2C m13[T7]	M	I	QLSPFPFDL	H2-L(d)	
3E3Q/DE	T/C	2C m13[T7]	M	I	QLSPFPFDL	H2-L(d)	
3E3Q/IJ	T/C	2C m13[T7]	M	I	QLSPFPFDL	H2-L(d)	
3E3Q/MN	T/C	2C m13[T7]	M	I	QLSPFPFDL	H2-L(d)	
3E3Q/RS	T/C	2C m13[T7]	M	I	QLSPFPFDL	H2-L(d)	
3E3Q/VW	T/C	2C m13[T7]	M	I	QLSPFPFDL	H2-L(d)	
3E3Q/Za	T/C	2C m13[T7]	M	I	QLSPFPFDL	H2-L(d)	
3FFC/DE	T/C	cf34	H	I	FLRGRAYGL	HLA-B*0801	
3FFC/IJ	T/C	cf34	H	I	FLRGRAYGL	HLA-B*0801	
3GSN/AB	T/C	RA14	H	I	NLVPMVATV	HLA-A*0201	
3H9S/DE	T/C	A6	H	I	MLWGYLQYV	HLA-A*0201	
3KPR/DE	T/C	LC13	H	I	EEYLKAWTF	HLA-B*4405	
3KPR/IJ	T/C	LC13	H	I	EEYLKAWTF	HLA-B*4405	
3KPS/DE	T/C	LC13	H	I	EEYLQAFTY	HLA-B*4405	
3KT/CF/D	T/C	SB27	H	I	LPEPLPQGQLTAY	HLA-B*3508	
3KXF/MO	T/C	SB27	H	I	LPEPLPQGQLTAY	HLA-B*3508	
3KXF/NP	T/C	SB27	H	I	LPEPLPQGQLTAY	HLA-B*3508	
3MBE/CD	T	TCR 21.30	M	II	GAMKRHGLDNY\ RGYSLGN	H2-Aa(d)	H2-Ab(NOD)
3MBE/GH	T	TCR 21.30	M	II	GAMKRHGLDNY\ RGYSLGN	H2-Aa(d)	H2-Ab(NOD)
3MV8/DE	T/C	TK3 Q55H	H	I	HPVGEADYFEY	HLA-B*3501	
3PWP/DE	T/C	A6	H	I	LGYGFVNYI	HLA-A*0201	
3QIU/CD	T	226 TCR	M	II	ADLIAYLKQATK	H2-Ea(k)	H2-Eb(k)
3QIW/CD	T	226 TCR	M	II	ADLIAYLEQATK	H2-Ea(k)	H2-Eb(k)

a) Biological unit (BU) given as PDB ID and the two TCR chains as indicator.

b) T: only in DS_T; T/C: contained in both datasets, DS_T & DS_C.

c) Species: either human (H) or mouse (M).

d) MHC class presents in the original crystal structure.

T. Hoffmann, A. Marion, and I. Antes - **DynaDom: Structure-based prediction of TCR inter-domain and TCR-pMHC association angles**
Additional File 5

S2 Table: Performance of the Q-Q interaction optimization.

PDB	CH[a]		Pairing state[b]				PDB	CH[a]		Pairing state[b]			
ID	α	β	ES	RD	PT	DD	ID	α	β	ES	RD	PT	DD
1ao7	D	E	p	p	p	p	2vlr	I	J	p	p	p	p
1fo0	A	B	p	p	p	p	3c5z	A	B	p	p	p	p
1fyt	D	E	m	p	p	p	3c5z	E	F	p	p	p	p
1j8h	D	E	p	p	p	p	3c60	A	B	p	p	p	p
1kj2	A	B	p	p	p	p	3c60	E	F	p	p	p	p
1kj2	D	E	p	p	p	p	3c6l	A	B	m	p	p	p
1mi5	D	E	p	p	p	p	3c6l	E	F	m	p	p	p
1mwa	A	B	p	p	p	p	3d39	D	E	p	p	p	p
1nam	A	B	p	p	p	p	3d3v	D	E	p	m	p	p
1oga	D	E	p	p	p	p	3dxa	D	E	m	p	p	p
1qse	D	E	p	p	p	p	3dxa	I	J	m	p	p	p
1u3h	A	B	m	p	p	p	3dxa	N	O	m	p	p	p
1u3h	E	F	m	p	p	p	3e2h	B	C	p	p	p	p
2bnq	D	E	p	p	p	p	3e3q	C	F	p	p	p	p
2bnr	D	E	p	p	p	p	3e3q	d	e	p	p	p	p
2e7l	A	D	m	p	p	p	3e3q	D	E	p	p	p	p
2e7l	B	C	p	p	p	p	3e3q	I	J	p	p	p	p
2esv	D	E	n/a	n/a	n/a	n/a	3e3q	M	N	p	p	p	p
2f53	D	E	m	p	p	p	3e3q	R	S	p	p	p	p
2f54	D	E	m	p	p	p	3e3q	V	W	p	p	p	p
2f54	K	L	m	p	p	p	3e3q	Z	a	p	p	p	p
2gj6	D	E	p	p	p	p	3ffc	D	E	p	p	p	p
2iam	C	D	p	p	p	p	3ffc	I	J	p	p	p	p
2ian	D	E	p	p	p	p	3gsn	A	B	n/a	n/a	n/a	n/a
2ian	I	J	p	p	p	p	3h9s	D	E	p	p	p	p
2ian	N	O	p	p	p	p	3kpr	D	E	p	p	p	p
2ian	S	T	p	p	p	p	3kpr	I	J	p	p	p	p
2nx5	I	J	p	p	p	p	3kps	D	E	p	p	p	p
2nx5	N	P	p	p	p	p	3kxf	D	E	p	p	p	p
2nx5	T	U	p	p	p	p	3kxf	M	O	p	p	p	p
2oi9	B	C	p	p	p	p	3kxf	N	P	m	p	p	p
2ol3	A	B	p	p	p	p	3mbe	C	D	x	x	x	p
2p5e	D	E	m	p	p	p	3mbe	G	H	x	x	x	p
2p5w	D	E	m	p	p	p	3mv8	D	E	m	p	p	p
2pxy	A	B	m	m	p	p	3pwp	D	E	p	p	p	p
2pye	D	E	m	p	p	p	3qiu	C	D	n/a	n/a	n/a	n/a
2vlk	D	E	p	p	p	p	3qiw	C	D	n/a	n/a	n/a	n/a
2vlr	D	E	p	p	p	p							

a) TCR chain ID contained in the same biological unit.
b) Pairing states in different models (ES: experimental crystal structure; RD/PT: crystal structure after application of the REDUCE/Protoss program. DD: final model after full application of the DynaDom approach on the crystal structure): p=paired, m=mispaired, x=crossing conformation, n/a=not applicable due to missing CoR_{β} glutamine in one of the chains.

T. Hoffmann, A. Marion, and I. Antes - **DynaDom: Structure-based prediction of TCR inter-domain and TCR-pMHC association angles**

Additional File 6

Text S3: Glutamine correction and adaption (detailed Results and Discussion).

One of the difficulties encountered in X-ray crystallography is to distinguish between neighboring oxygen and nitrogen atoms in electron density maps, since these two elements share a similar electron density. As a consequence, in many structures, the carboxamide atoms of glutamine and asparagine cannot be assigned to the correct coordinates with a high certainty and might by flipped by 180°. We investigated the conserved Q-Q interactions at the CoR_{β} position in the crystal structures of the dataset DS_T visually (Fig. 2B of the main text), with the MolProbity Reduce tool, and with the Protoss program. As listed in Table S2 (Additional file 5), we found several structures in which the glutamine residues were placed in a way that atoms of the same type (and thus of the same charge) were opposed to each other (1fyt/DE, 1u3h/AB and EF, 2e7l/AD, 2f53/DE, 2f54/DE and KL, 2p5e/DE, 2p5w/DE, 2pxy/AB, 2pye/DE, 3c6l/AB and EF, 3dxa/DE, IJ, and NO, 3kxf/NP, and 3mv8/DE). These presumed misplacements lead to unphysical repulsions between the two TCR chains, whereas in reality the glutamine residues form strong attractive electrostatic interactions induced by the interaction of the positively charged nitrogen and the negatively charged oxygen of the opposing residue. Furthermore, we found structural examples (1ao7/DE, 1fo0/DE, 1kj2/AB, 1mwa/AB, 1nam/AB, 1qse/DE, 2gj6/DE, 3kxf/DE) in which the glutamine residues indeed optimally interact with each other, but were both presumably flipped, as proposed by the Reduce tool and the Protoss tool (Additional files 7 and 8: Tables S3 and S4, columns CF). (Notably, the two different algorithms provided contradictory results for the structures 1nam, 2ol3, 2pxy, 3d3v, 3dxa, 3kxf, and 3mbe.) In these cases a rotation by 180° of both of the residues endgroups leads to a more favored interaction of these residues with the remaining protein. In most of these cases, the orientation of the Qα is determined by a hydrogen bond interaction between the amino terminal group of the glutamine and a backbone oxygen (Fig. 2C).

The application of Reduce[1-3] and Protoss[4,5] to the separated Vα and Vβ domains and to the Vα:Vβ complex showed that some TCRs adapt to different flip-states of the conserved CoR_{β} glutamine residues in the single chain state and in the complex state (see Additional Files 7 and 8: Tables S3 and S4, *e.g.* 1nam/AB, 1oga/DE, or 2gj6/DE). This result could be expected, as the glutamine residues at the CoR_{β} are surface residues in the individual chains and thus conformationally more flexible. Therefore a single preprocessing step involving only the individual chains is not suitable for an optimal placement of the conserved glutamine residues in the Vα:Vβ complex structure.

In order to address this issue, we developed a special procedure to correct the glutamine flip states concurrently before and during our rigid body optimization process. In this approach, both glutamine residues are first optimized only accounting for intra-domain interactions. Afterwards, the optimization is repeated by taking the inter-domain interactions into account as well. During the rigid body optimization both glutamine residues are free to rotate. However, a full rotation can be hampered by sterical clashes. To overcome this problem, we introduced a fast sampling step into our pipeline. For the correction of individual glutamine residues, in the sampling only the flipped and the unflipped states were taken into account. The sampling of both residues together was performed with smaller angular steps of 18°. After the sampling a short energy minimization rotating the amide groups was carried out.

In an intermediate pipeline step after 50 iterations of the main rigid body optimization, both glutamine residues were sampled and refined again. This readjustment of the two residues is used, since some initial rigid body orientations

prevent the formation of the bifurcation. In many cases a correct placement of the glutamine residues is only possible after some cycles of the rigid body optimization.
To validate our glutamine correction approach, we investigated the rotational states of the CoR_β glutamine residues before the rigid body optimization took place and compared the conformations of these residues with their counterpart found in the crystal structures (Additional file 5: Table S2). In all cases that show the assumed mispairing, at least one glutamine residue was flipped to allow the bifurcated Q-Q interaction. For all other structures in the DS_T dataset without an obvious mispairing, the Q-Q interaction was preserved. Our method performs similarly compared to the two methods used as a reference, namely Reduce and Protoss (Additional file 5: Table S2): With Protoss in all cases a pairing of the Q residues was reached. In contrast, the Reduce method did not correct the structure 2pxy/AB and introduced a mispairing in the structure 3d3v/DE. As a quantitative measure, we provide the pairing rates based on the number of structures containing two glutamine residues at the CoR_β: original crystal: 72.7%; Reduce: 94.6%; Protoss: 97.3%; DynaDom: 100%. Remarkably, our method was also able to transform the two orthogonal-like conformations into paired conformations.
Taking the correction of the bifurcated Q-Q interaction into account generally increases the prediction rate for the modeling of the TCR inter-domain angles as well as for the application on the TCRpMHC complex.

References

1. Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta crystallographica Section D, Biological crystallography. 2010;66(Pt 1):12-21. doi: 10.1107/S0907444909042073. PubMed PMID: 20057044; PubMed Central PMCID: PMC2803126.
2. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. 2007;35(Web Server issue):W375-83. doi: 10.1093/nar/gkm216. PubMed PMID: 17452350; PubMed Central PMCID: PMC1933162.
3. Word JM, Lovell SC, Richardson JS, Richardson DC. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. J Mol Biol. 1999;285(4):1735-47. doi: 10.1006/jmbi.1998.2401. PubMed PMID: 9917408.
4. Lippert T, Rarey M. Fast automated placement of polar hydrogen atoms in protein-ligand complexes. Journal of cheminformatics. 2009;1(1):13. doi: 10.1186/1758-2946-1-13. PubMed PMID: 20298519; PubMed Central PMCID: PMC3225823.
5. Bietz S, Urbaczek S, Schulz B, Rarey M. Protoss: a holistic approach to predict tautomers and protonation states in protein-ligand complexes. Journal of cheminformatics. 2014;6:12. doi: 10.1186/1758-2946-6-12. PubMed PMID: 24694216; PubMed Central PMCID: PMC4019353.

T. Hoffmann, A. Marion, and I. Antes - **DynaDom: Structure-based prediction of TCR inter-domain and TCR-pMHC association angles**

Additional File 7

Table S3: Per residue flip states using Reduce, Protoss and DynaDom comparing single domains and TCR complexes (part 1).

PDB	CH[a]		RI[b]		Reduce				Protoss				DynaDom[e]					
ID	α	β	α	β	*SF*[c]		CF[d]		*SF*[c]		CF[d]		*SF*[c]		SR[f]		*CR*[g]	
					α	*β*	α	β	*α*	*β*	α	β	*α*	*β*	α	β	*α*	*β*
1ao7	D	E	37	37	*K*	*F*	F	F	*F*	*F*	F	F	*F*	*F*	F	F	*F*	*F*
1fo0	A	B	37	37	*K*	*K*	F	F	*K*	*F*	F	F	*F*	*F*	F	F	*F*	*F*
1fyt	D	E	37	37	*F*	*K*	F	K	*F*	*K*	F	K	*K*	*K*	K	K	*F*	*K*
1j8h	D	E	37	37	*K*	*K*	K	K	*K*	*K*	K	K	*F*	*K*	F	K	*K*	*K*
1kj2	A	B	37	37	*F*	*F*	F	F	*F*	*F*	F	F	*F*	*F*	F	F	*F*	*F*
1kj2	D	E	37	37	*F*	*F*	F	F	*F*	*F*	F	F	*F*	*F*	F	F	*F*	*F*
1mi5	D	E	37	37	*K*	*K*	K	K	*K*	*K*	K	K	*F*	*K*	F	K	*K*	*K*
1mwa	A	B	37	37	*F*	*K*	F	F	*F*	*F*	F	F	*F*	*F*	F	F	*F*	*F*
1nam	A	B	37	37	*K*	*F*	K	K	*K*	*F*	F	F	*F*	*F*	F	F	*F*	*F*
1oga	D	E	38	39	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
1qse	D	E	37	37	*F*	*F*	F	F	*F*	*F*	F	F	*F*	*F*	F	F	*F*	*F*
1u3h	A	B	37	37	*K*	*F*	K	F	*K*	*F*	K	F	*K*	*F*	K	F	*K*	*F*
1u3h	E	F	37	37	*K*	*F*	K	F	*K*	*F*	K	F	*K*	*F*	K	F	*K*	*F*
2bnq	D	E	38	36	*F*	*F*	F	F	*F*	*F*	F	F	*F*	*F*	F	F	*F*	*F*
2bnr	D	E	38	36	*F*	*F*	F	F	*F*	*F*	F	F	*F*	*F*	F	F	*F*	*F*
2e7l	A	D	37	37	*K*	*F*	K	F	*K*	*F*	K	F	*K*	*F*	K	F	*K*	*F*
2e7l	B	C	37	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
2esv	D	E	36	n/a	*K*	*n/a*	K	n/a	*K*	*n/a*	K	n/a	*K*	*n/a*	K	n/a	*K*	*n/a*
2f53	D	E	37	35	*K*	*F*	K	F	*K*	*F*	K	F	*K*	*F*	K	F	*K*	*F*
2f54	D	E	37	35	*K*	*F*	K	F	*K*	*F*	K	F	*K*	*F*	K	F	*K*	*F*
2f54	K	L	37	35	*K*	*K*	K	F	*K*	*K*	K	F	*K*	*F*	K	F	*F*	*K*
2gj6	D	E	37	37	*F*	*K*	F	F	*F*	*K*	F	F	*F*	*F*	F	F	*F*	*F*
2iam	C	D	36	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
2ian	D	E	36	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
2ian	I	J	36	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
2ian	N	O	36	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
2ian	S	T	36	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
2nx5	I	J	37	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
2nx5	N	P	37	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
2nx5	T	U	37	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
2oi9	B	C	37	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
2ol3	A	B	37	37	*K*	*K*	F	F	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
2p5e	D	E	38	35	*K*	*F*	K	F	*K*	*F*	K	F	*K*	*F*	K	F	*K*	*F*
2p5w	D	E	38	35	*K*	*F*	K	F	*K*	*F*	K	F	*K*	*F*	K	F	*K*	*F*
2pxy	A	B	37	37	*F*	*K*	F	F	*F*	*K*	F	K	*F*	*K*	F	K	*F*	*K*
2pye	D	E	38	35	*K*	*F*	K	F	*K*	*F*	K	F	*K*	*F*	K	F	*K*	*F*
2vlk	D	E	38	39	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
2vlr	D	E	38	39	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
2vlr	I	J	38	39	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
3c5z	A	B	37	35	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
3c5z	E	F	37	35	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*

a) TCR chain pairs indicating the biological unit.
b) Residue IDs of the conserved glutamine residues at the center of rotation.
c) Flip state, after separate application of the tool to individual chains.
d) Flip state, after application of the tool to the complex.
e) Flip state before rigid body optimization.
f) Flip state, after separate application of the refinement to individual chains (Threshold 90°).
g) Flip state, after application of refinement tool to the complex (Threshold 90°).
Angular values of the refinements are listed in Additional file 8: Table S4. K=keep; F=flip.

T. Hoffmann, A. Marion, and I. Antes - **DynaDom: Structure-based prediction of TCR inter-domain and TCR-pMHC association angles**

Additional File 7

Table S3: Per residue flip states using Reduce, Protoss and DynaDom comparing single domains and TCR complexes (continued).

PDB	CH[a]		RI[b]		Reduce				Protoss				DynaDom[e]					
ID	α	β	α	β	*SF*[c]		CF[d]		*SF*[c]		CF[d]		*SF*[c]		SR[f]		*CR*[g]	
					α	*β*	α	β	*α*	*β*	α	β	*α*	*β*	α	β	*α*	*β*
3c60	A	B	37	35	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*F*	*K*
3c60	E	F	37	35	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*F*	*K*
3c6l	A	B	37	35	*K*	*F*	K	F	*K*	*F*	K	F	*K*	*F*	K	F	*K*	*F*
3c6l	E	F	37	35	*K*	*F*	K	F	*K*	*F*	K	F	*K*	*F*	K	F	*K*	*F*
3d39	D	E	37	37	*F*	*F*	F	F	*F*	*F*	F	F	*F*	*F*	F	F	*F*	*F*
3d3v	D	E	37	37	*F*	*F*	F	K	*F*	*F*	F	F	*F*	*F*	F	F	*F*	*F*
3dxa	D	E	44	44	*F*	*K*	F	K	*F*	*K*	K	F	*F*	*K*	F	K	*F*	*K*
3dxa	I	J	44	44	*F*	*K*	F	K	*F*	*K*	K	F	*F*	*K*	F	K	*F*	*K*
3dxa	N	O	44	44	*F*	*K*	F	K	*F*	*K*	F	K	*F*	*K*	F	K	*F*	*K*
3e2h	B	C	37	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
3e3q	C	F	37	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
3e3q	d	e	37	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
3e3q	D	E	37	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
3e3q	I	J	37	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
3e3q	M	N	37	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
3e3q	R	S	37	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
3e3q	V	W	37	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
3e3q	Z	a	37	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
3ffc	D	E	44	44	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
3ffc	I	J	44	44	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
3gsn	A	B	n/a	37	*n/a*	*K*	n/a	K	*n/a*	*K*	n/a	K	*n/a*	*F*	n/a	F	*n/a*	*F*
3h9s	D	E	37	37	*F*	*F*	F	F	*F*	*F*	F	F	*F*	*F*	F	F	*F*	*F*
3kpr	D	E	37	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
3kpr	I	J	37	37	*K*	*K*	K	K	*K*	*K*	K	K	*F*	*K*	F	K	*K*	*K*
3kps	D	E	37	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
3kxf	D	E	38	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
3kxf	M	O	38	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
3kxf	N	P	38	37	*K*	*K*	K	F	*F*	*K*	F	K	*F*	*K*	F	K	*F*	*K*
3mbe	C	D	44	44	*F*	*K*	K	K	*F*	*K*	F	K	*F*	*K*	F	K	*F*	*K*
3mbe	G	H	44	44	*F*	*K*	K	K	*F*	*K*	F	K	*F*	*K*	F	K	*F*	*K*
3mv8	D	E	44	44	*F*	*K*	F	K	*F*	*K*	F	K	*F*	*K*	F	K	*F*	*K*
3pwp	D	E	37	37	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*	K	K	*K*	*K*
3qiu	C	D	n/a	37	*n/a*	*K*	n/a	K	*n/a*	*K*	n/a	K	*n/a*	*K*	n/a	K	*n/a*	*K*
3qiw	C	D	n/a	37	*n/a*	*K*	n/a	K	*n/a*	*F*	n/a	K	*n/a*	*K*	n/a	K	*n/a*	*K*

a) TCR chain pairs indicating the biological unit.
b) Residue IDs of the conserved glutamine residues at the center of rotation.
c) Flip state, after separate application of the tool to individual chains.
d) Flip state, after application of the tool to the complex.
e) Flip state before rigid body optimization.
f) Flip state, after separate application of the refinement to individual chains (Threshold 90°).
g) Flip state, after application of refinement tool to the complex (Threshold 90°).
Angular values of the refinements are listed in Additional file 8: Table S4

T. Hoffmann, A. Marion, and I. Antes - **DynaDom: Structure-based prediction of TCR inter-domain and TCR-pMHC association angles**

Additional File 8

Table S4: Angular deviations with respect to the crystal structures after DynaDom glutamine refinement.

PDB	CH[a]		Angular Deviation[°]				PDB	CH[a]		Angular Deviation[°]			
			Single		Complex					Single		Complex	
ID	α	β	α	β	α	β	ID	α	β	α	β	α	β
1ao7	D	E	179.9	174.2	-168.3	-175.2	2vlr	I	J	11.4	7.9	7.7	5.7
1fo0	A	B	-179.5	179.9	-176.9	-176.0	3c5z	A	B	13.5	-5.8	9.3	2.5
1fyt	D	E	-11.5	-36.2	-135.3	-12.9	3c5z	E	F	13.7	-5.8	9.7	2.6
1j8h	D	E	174.2	-37.5	23.6	-10.5	3c60	A	B	-2.0	-18.5	96.0	-54.9
1kj2	A	B	-174.9	165.9	176.0	170.0	3c60	E	F	-3.2	-22.9	96.8	-56.9
1kj2	D	E	-161.2	167.8	-177.9	169.3	3c6l	A	B	33.5	157.2	69.2	121.4
1mi5	D	E	-168.6	5.6	8.0	6.4	3c6l	E	F	35.3	157.0	71.2	122.8
1mwa	A	B	-108.5	-178.1	176.1	-163.8	3d39	D	E	-154.7	175.1	-153.4	-177.4
1nam	A	B	179.9	158.8	168.1	174.0	3d3v	D	E	-165.6	174.2	-167.7	-172.0
1oga	D	E	7.8	16.8	3.4	14.8	3dxa	D	E	-147.1	17.7	-163.7	14.9
1qse	D	E	-173.3	157.0	-173.4	167.9	3dxa	I	J	-155.9	22.7	-172.6	19.9
1u3h	A	B	25.6	138.8	26.3	127.0	3dxa	N	O	-150.5	20.5	-162.9	12.7
1u3h	E	F	-2.2	134.1	-15.4	-149.7	3e2h	B	C	-2.8	-0.1	3.7	4.3
2bnq	D	E	-174.8	168.5	-176.5	-178.9	3e3q	C	F	5.4	-0.1	10.7	4.2
2bnr	D	E	-171.8	168.5	-173.5	-178.6	3e3q	d	e	2.3	-4.6	4.8	4.0
2e7l	A	D	20.9	157.0	24.3	162.1	3e3q	D	E	3.5	-3.7	7.1	2.8
2e7l	B	C	14.4	-4.3	22.1	-2.3	3e3q	I	J	5.7	-4.0	9.7	2.4
2esv	D	E	11.4	n/a	-1.5	n/a	3e3q	M	N	5.2	-0.1	10.5	3.9
2f53	D	E	0.8	-155.8	-9.0	-160.6	3e3q	R	S	1.2	-0.1	5.3	7.8
2f54	D	E	3.9	174.2	-20.9	-170.2	3e3q	V	W	3.7	-5.8	5.6	2.0
2f54	K	L	25.3	168.5	178.7	16.3	3e3q	Z	a	-5.8	0.2	-4.6	6.9
2gj6	D	E	-173.5	179.9	-172.7	-173.0	3ffc	D	E	-3.0	8.4	-0.5	19.2
2iam	C	D	-1.8	15.4	-9.0	17.4	3ffc	I	J	-5.2	17.2	-2.7	20.2
2ian	D	E	-5.8	14.8	-7.6	27.3	3gsn	A	B	n/a	154.8	n/a	157.7
2ian	I	J	-3.2	17.5	-3.7	18.8	3h9s	D	E	-180.0	174.2	-176.8	-177.8
2ian	N	O	-5.8	-5.8	-8.3	1.7	3kpr	D	E	-2.4	3.3	-7.3	1.4
2ian	S	T	-5.8	-5.8	4.1	-1.8	3kpr	I	J	180.0	-1.0	7.2	-2.7
2nx5	I	J	2.4	1.1	-10.9	8.2	3kps	D	E	12.8	-13.2	-7.8	-14.5
2nx5	N	P	21.1	11.4	12.9	7.3	3kxf	D	E	-2.0	-40.6	-5.1	-11.7
2nx5	T	U	10.5	16.4	-1.1	25.9	3kxf	M	O	-3.8	-44.2	3.4	-49.2
2oi9	B	C	5.3	-5.8	2.7	-3.5	3kxf	N	P	-166.3	-4.4	-166.0	14.7
2ol3	A	B	3.6	-11.5	10.9	-8.0	3mbe	C	D	180.0	-30.6	149.2	-43.1
2p5e	D	E	3.7	-157.6	-2.7	-159.2	3mbe	G	H	-180.0	-31.0	152.7	-38.2
2p5w	D	E	-3.0	-157.0	-14.9	-159.9	3mv8	D	E	-172.2	50.9	-174.7	22.9
2pxy	A	B	177.7	11.4	172.0	5.7	3pwp	D	E	5.6	2.3	3.7	9.4
2pye	D	E	-6.3	-157.0	-18.4	-160.4	3qiu	C	D	n/a	-0.2	n/a	14.1
2vlk	D	E	-8.8	23.1	-11.3	19.8	3qiw	C	D	n/a	4.2	n/a	-6.4
2vlr	D	E	3.7	9.4	0.8	7.6							

a) TCR chain ID contained in the same biological unit.

T. Hoffmann, A. Marion, and I. Antes - **DynaDom: Structure-based prediction of TCR inter-domain and TCR-pMHC association angles**

Additional File 9

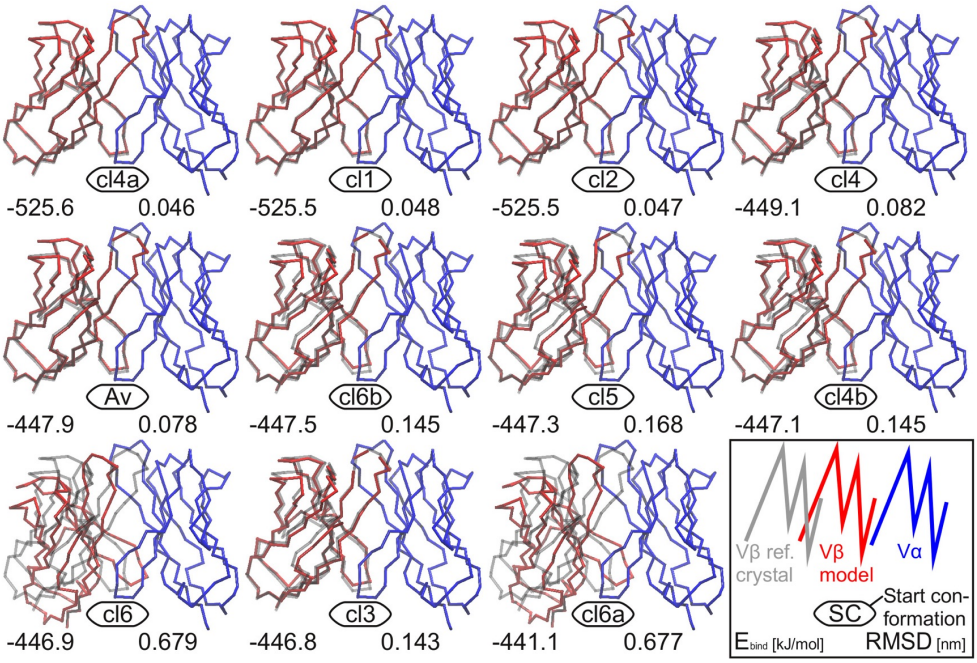


Figure S1: Discrimination of the models.
Eleven final models of structure 2p5e. Red: Vβ domain; blue: Vα domain; gray: crystal structure. Left side: interaction energy (kcal/mol); right side: RMSD (nm) (see also insertion above for explanation). The models are ordered according to their energy. Note that our method is able to discriminate accurate models of low RMSD with respect to the experimental structures (gray) based on the interaction energy between the Vα and the Vβ domains (first row). See Text S1 (Additional file 1) for the labeling of the 11 starting structures.

T. Hoffmann, A. Marion, and I. Antes - **DynaDom: Structure-based prediction of TCR inter-domain and TCR-pMHC association angles**

Additional File 10

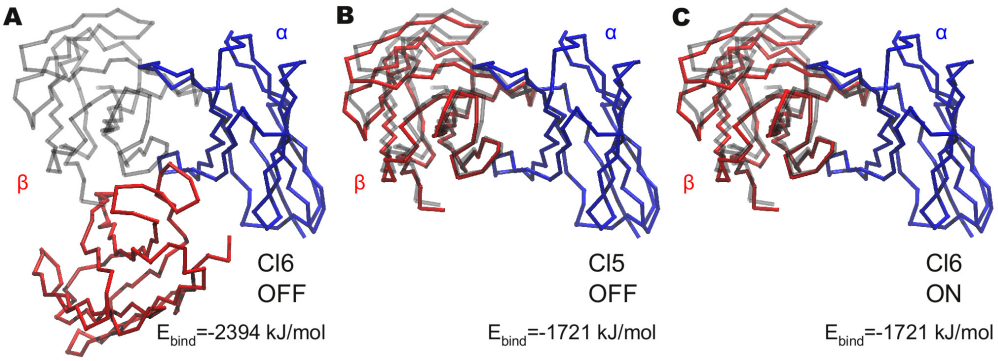


Figure S2: Influence of the restraint operator.
Final models of the structure 1kj2/DE after the optimization procedure (view: ligand binding site). The Vβ domain of the original experimental crystal structure is shown in transparent black as a reference. If the position of the Vβ domain is not constrained, the optimization from the starting conformation Cl6 leads to an unfavored model (A, RMSD 30.00 Å) with a better interaction energy than the preferred model starting from Cl5 (B, 1.54 Å), and thus the energy based prediction fails. C) By switching the restraints on, the model starting from Cl6 converges to the preferred conformation (RMSD 1.59 Å).

T. Hoffmann, A. Marion, and I. Antes - **DynaDom: Structure-based prediction of TCR inter-domain and TCR-pMHC association angles**

Additional File 11

Table S5: pMHC optimization for the structure 1oga with different pMHC start conformations.

pMHC Start Conf. Angle[a]			E_{bind}	$RMSD_{C\alpha}$ [nm]		
x [°]	**y** [°]	**z** [°]	[kJ/mol]	**Vβ**	**pMHC**	**Vβ&pMHC**
-5.0	-5.0	-5.0	-1435.71	0.0273	0.9121	0.7260
-5.0	-5.0	0.0	-1466.90	0.0257	1.0704	0.8520
-5.0	-5.0	5.0	-1518.47	0.0312	1.1560	0.9202
-5.0	0.0	-5.0	-1240.59	0.0366	0.8944	0.7121
-5.0	0.0	0.0	-1510.11	0.0217	1.0945	0.8712
-5.0	0.0	5.0	-1526.07	0.0334	1.1955	0.9516
-5.0	**5.0**	**-5.0**	**-1839.48**	**0.0259**	**0.0259**	**0.0259**
-5.0	**5.0**	**0.0**	**-1838.86**	**0.0227**	**0.0303**	**0.0278**
-5.0	5.0	5.0	-1246.40	0.0241	0.9172	0.7301
0.0	-5.0	-5.0	-1710.15	0.0183	0.8772	0.6982
0.0	-5.0	0.0	-1427.67	0.0331	1.0375	0.8259
0.0	-5.0	5.0	-1250.48	0.0210	0.9642	0.7674
0.0	0.0	-5.0	-1245.79	0.0307	0.8868	0.7060
0.0	0.0	0.0	-1706.83	0.0208	0.9177	0.7304
0.0	0.0	5.0	-1251.17	0.0246	0.9829	0.7824
0.0	**5.0**	**-5.0**	**-1839.87**	**0.0248**	**0.0205**	**0.0222**
0.0	5.0	0.0	-1248.86	0.0369	0.9318	0.7419
0.0	5.0	5.0	-1584.93	0.0280	0.8707	0.6931
5.0	-5.0	-5.0	-1708.17	0.0277	0.8988	0.7155
5.0	-5.0	0.0	-1706.07	0.0202	0.9106	0.7248
5.0	-5.0	5.0	-1245.24	0.0251	1.0039	0.7991
5.0	0.0	-5.0	-1577.50	0.0356	0.5658	0.4508
5.0	**0.0**	**0.0**	**-1840.44**	**0.0160**	**0.0191**	**0.0180**
5.0	0.0	5.0	-1248.18	0.0219	1.0019	0.7975
5.0	5.0	-5.0	-1707.68	0.0283	0.9075	0.7224
5.0	**5.0**	**0.0**	**-1839.45**	**0.0251**	**0.0229**	**0.0237**
5.0	5.0	5.0	-1708.90	0.0264	0.9007	0.7170

a) The TCR conformation is set to the crystal structure. The pMHC deviates from the zero-conformations by the values x, y, and z (Euler angle components).

Appendix D

Additional Data

D.1 Templates for Superpositioning

See Section 2.2.1.2 for explanations.

2bnu α Varibale Domain

The residues used for superpositioning are indicated with capital letters.

qevtqipaAL SVPegenLVL NCSFtdsaiY NLQWFRQdpg
kgLTSLLLIQ ssqrEQTSgr LNASLDkssg RSTLYIaasQ
PGDSATYLCA VRPtsggsyi ptfgrgtsli vhp

2bnu β Varibale Domain

The residues used for the placement of the β-cuboid are indicated with capital letters.

GVTQTpkFQV LKtgqsMTLQ CAQdmnheYM STRPYRQDpgmG
LRLIHYSVGa gITDQgevpn gyNVsrstte dfpLRLlsaA
PSQTSVYFCA SSyvgntgel ffgegSRLTV ledl

1k5n MHC class I α_1:α_2 templates T_μ and $T_\mu{}'$

The residues used for the placement of the β-cuboid are indicated with capital letters.

gSHSMRYFHT SVSrPGRgEPRFITVGYVDD TlFVRFdsda
aspreepraP WIEQEGPEYW DRETQICKAK AQTDREDLRT
LLRYYNQSEA GSHTLQNMYG CDVgpdgrll RGYHQHAYDG
KDYIALNEDL SSWTAADTAA QITQrKWEAa rvaEQLRAYL
EGECVEWLRR YlENG

D.2 Additional Figures and Table to Chapter 4

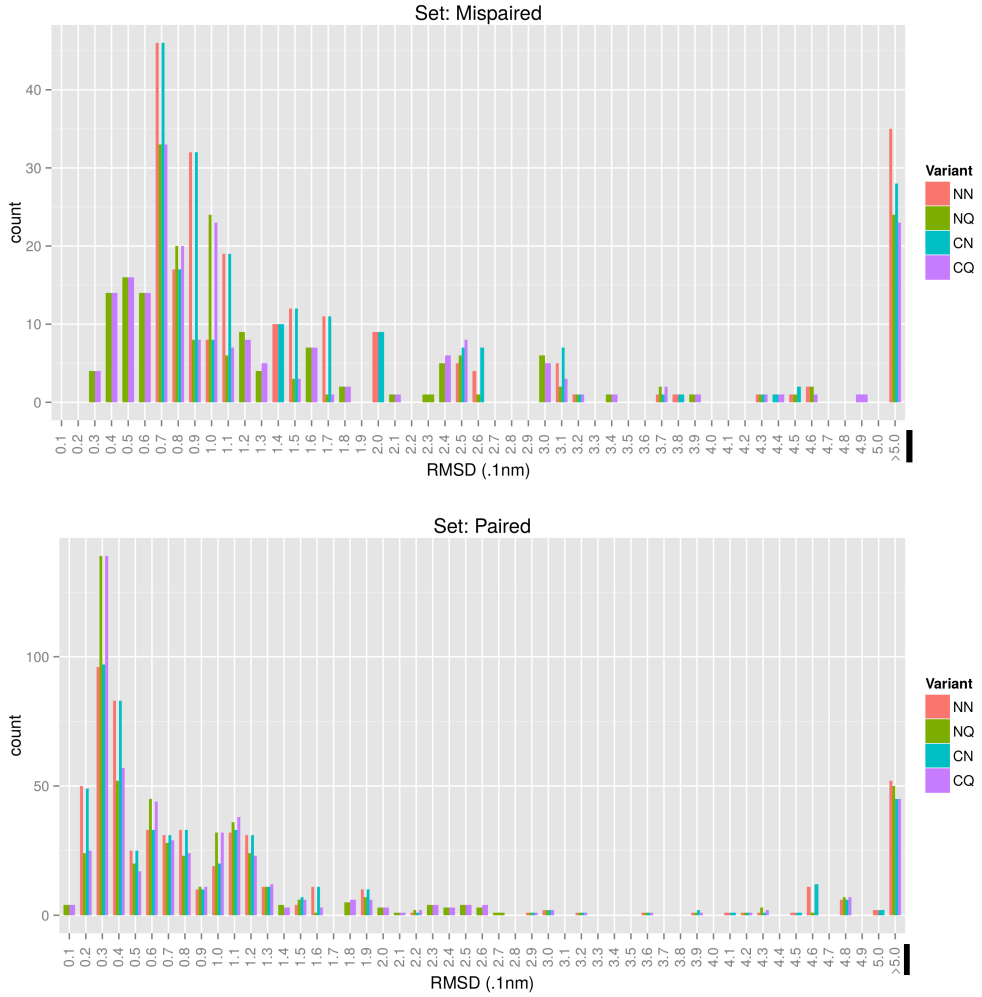


Figure D.1: **Influence of Glutamine Correction and of the Restraints - RMSD Distributions of all Models.** Two different subsets of the 75 TCR structures wer used, containing mispairings or bifurcated parings of the glutamine (Q) residues at the *Center of Rotation* of the variable β domain (CoR_β). For each structure in a set eleven models were generated with four different variants of the pipeline: NN) neither restraints nor Q optimization; NQ) no restraints, but Q optimization; RN) restraints but no Q optimization; RQ) restraints and Q optimization. The histograms show the *root mean square deviation* (RMSD) counts for bins of $0.1\,\text{Å}$ for all 11×75 models; values $> 5.0\,\text{Å}$ are summarized in on single bin. The usage of Q optimization significantly shifts the distribution to the better RMSD values by improving the models of the cases in which a mispairing of the glutamine-glutamine (Q-Q) interaction was observed in the experimental structure.

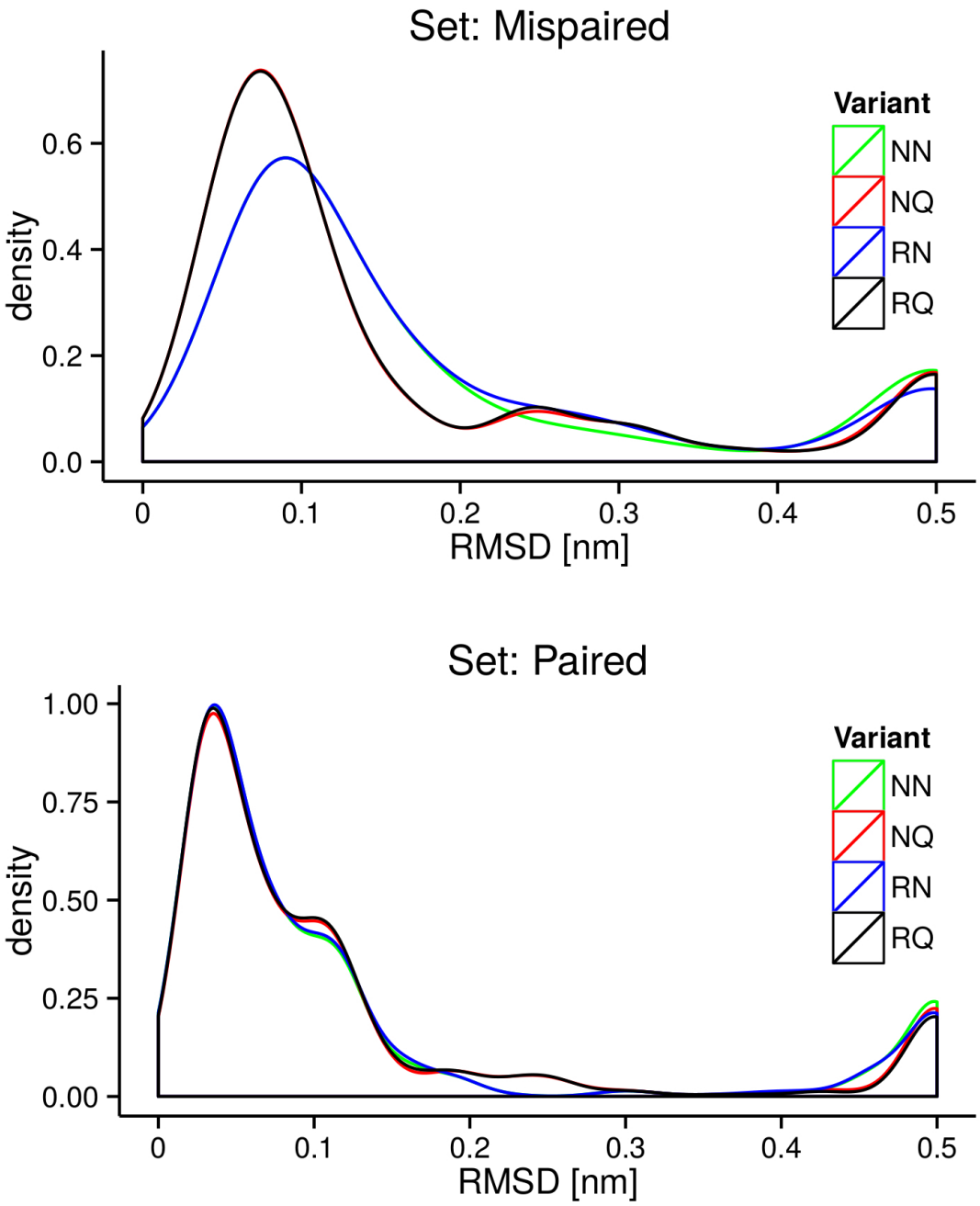


Figure D.2: **Influence of Glutamine Correction and of the Restraints - RMSD Densities of all Models.** Densities according to Fig. D.1 (Explanations and labels are found there).

Table D.1: **Prediction Accuracy for the Control Evaluation Studies**

ES[a]	**M**[b]	**V**[c]		**DS**[d]		**PS**[e]	
		Q	**R**	**N**	#	**Abs.**[a]	**Rel.**[%]
C_T	*TCR Control*						
C_T00	T	off	off	DS_{TC}	75	71	94.7
C_T10	T	on	off	DS_{TC}	75	71	94.7
C_T01	T	off	on	DS_{TC}	75	71	94.7
C_T11	T	on	on	DS_{TC}	75	71	94.7
C_{mMc}	*MHC optimization without axis pre-placement, crystal start conformation of both MHC and TCR, Vβ static*						
$C_{mMc}00$	M	n/a	off	DS_{CC}	53	52	94.7
$C_{mMc}10$	M	n/a	on	DS_{CC}	53	52	94.7
C_{cMc}	*MHC optimization without axis pre-placement, crystal start conformation of both MHC and TCR, Vβ dynamic)*						
$C_{cMc}00$	C	off	off	DS_{CC}	53	53	100.0
$C_{cMc}10$	C	on	off	DS_{CC}	53	53	100.0
$C_{cMc}01$	C	off	on	DS_{CC}	53	53	100.0
$C_{cMc}11$	C	on	on	DS_{CC}	53	53	100.0
C_{mpMc}	*MHC optimization with axis preplacement, crystal start conformation of both MHC and TCR, Vβ static*						
$C_{mpMc}01$	M	n/a	on	DS_{CC}	53	38	71.7
C_{cpMc}	*MHC optimization with axis preplacement, crystal start conformation of both MHC and TCR, Vβ dynamic*						
$C_{cpMc}01$	M	n/a	on	DS_{CC}	53	38	71.7
C_{cpM0}	*MHC optimization with axis preplacement, MHC in zero start conformation, TCR in crystal conformation, Vβ dynamic*						
$C_{cpM0}11$	C	on	on	DS_{M0}	53	31	58.5

[a] Evaluation setting.
[b] Method; T: TCR Vβ rigid body (RB); C: TCR Vα and pMHC RBs; M: pMHC RB.
[c] ES Variant: Q: CoR_β residue rotational optimization; R: restraints TCR/pMHC on/off; n/a: not available.
[d] Data set; N=Name (see Table 4.1); #=number of structures in the data set)
[e] Absolute and relative hit rates according to the RMSD based criterion $\mathcal{C}_\mathcal{R}$.

D.3 MHC Docking Directionality

The section of the appendix provides data from the Analysis of the pMHC docking directionality as described in Secs. 2.2.4.2 and 2.3.2.2. The TRAV-consensus sequence used for the PDB BLAST search was defined as:

```
gqsvtqsppg sltvqegesv tlnctystsy iaspylfwyv
qypgegpqll lkirsggekk egsgrftatf nkseksfslh
itssqpsdsa vyfcalsel
```

The superpositioning template was derived from the PDB structure 1ao7 using the residues MET138 to ASN174.

The template was used to align a vector, which is defined by to artificially added HETATM records with the coordinates: (25.541, -48.052, 9.805) and (60.462, -54.757, 36.313).

The result of the measurement is listed in Tab. D.2.

Table D.2: **MHC Docking Directionality**

PDB	α[a]	β[a]	I[a]	IIα[a]	IIβ[a]	CD[a]	Vα[b]	Vβ[b]	L[b]	Tx[c]	Ty[c]	Tz[c]	Mx[c]	My[c]	Mz[c]	Cz[c]
1ao7	D	E	A				H-TRAV12-2*02	H-TRBV6-5*01	I	-17.888	-15.133	2.679	-34.921	6.705	-26.508	-648.399
1bd2	D	E	A				H-TRAV29*01	H-TRBV6-5*01	I	-16.728	-15.142	3.835	-31.166	14.647	-27.951	-716.931
1d9k	A	B		C	D		M-TRAV14D-2*01	M-TRBV13-2*01	II	-17.095	-15.012	4.619	-27.192	21.531	-27.641	-776.279
1fo0	A	B	H				M-TRAV16*01	M-TRBV1*01	I	-18.461	-14.315	1.281	-37.032	20.374	-13.441	-906.237
1fyt	D	E		A	B		H-TRAV8-4*01	H-TRBV28*01	II	-16.088	-16.037	4.410	-34.242	13.671	-24.652	-769.078
1g6r	A	B	H				M-TRAV9-4*01	M-TRBV13-2*01	I	-17.171	-14.893	1.856	-42.185	3.660	-13.196	-691.107
1j8h	D	E		A	B		H-TRAV8-4*01	H-TRBV28*01	II	-16.286	-16.207	4.160	-33.193	12.895	-26.441	-747.967
1kj2	A	B	H				M-TRAV14-1*01	M-TRBV1*01	I	-17.634	-16.475	2.865	-40.237	6.308	-17.558	-774.140
1lp9	E	F	A				M-TRAV12D-2*01	M-TRBV13-3*01	I	-17.760	-14.285	3.334	-21.784	30.191	-24.106	-847.377
1mi5	D	E	A				H-TRAV26-2*01	H-TRBV7-8*01	I	-20.331	-12.213	2	-33.990	23.868	-15.559	-900.380
1mwa	A	B	H				M-TRAV9-4*01	M-TRBV13-2*01	I	-17.231	-14.861	2.292	-42.496	3.753	-12.128	-696.201
1nam	A	B	H				M-TRAV16*01	M-TRBV1*01	I	-18.288	-14.637	1.121	-37.041	18.159	-16.290	-874.261
1oga	D	E	A				H-TRAV27*01	H-TRBV19*01	I	-17.512	-14.334	2.879	-22.852	23.438	-29.925	-738.007
1qrn	D	E	A				H-TRAV12-2*02	H-TRBV6-5*01	I	-17.897	-14.725	2.592	-34.863	6.676	-26.591	-632.838
1qse	D	E	A				H-TRAV12-2*02	H-TRBV6-5*01	I	-18.031	-14.650	2.634	-33.702	7.858	-27.740	-635.422
1qsf	D	E	A				H-TRAV12-2*02	H-TRBV6-5*01	I	-17.527	-14.768	3.239	-34.298	5.479	-27.581	-602.543
1ymm	D	E		A	B		H-TRAV17*01	H-TRBV20-1*01	II	-16.374	-16.128	1.951	-7.706	32.745	-28.906	-660.449
2ak4	D	E	A				H-TRAV19*01	H-TRBV6-1*01	I	-15.760	-15.723	3.355	-23.892	30.763	-21.211	-860.479
2bnq	D	E	A				H-TRAV21*01	H-TRBV6-5*01	I	-17.764	-14.741	0.136	-21.029	32.174	-22.129	-881.527
2bnr	D	E	A				H-TRAV21*01	H-TRBV6-5*01	I	-17.658	-14.826	0.358	-21.128	31.862	-22.483	-875.863
2ckb	A	B	H				M-TRAV9-4*01	M-TRBV13-2*01	I	-17.195	-14.777	1.787	-42.064	4.277	-13.394	-695.123
2e7l	A	D	E				M-TRAV9-4*01	M-TRBV13-2*01	I	-18.251	-14.132	1.429	-34.561	19.360	-19.946	-841.755
2esv	D	E	A				H-TRAV26-1*01	H-TRBV14*01	I	-21.266	-11.119	0.294	-31.745	21.836	-21.968	-817.337
2f53	D	E	A				H-TRAV21*01	H-TRBV6-5*01	I	-17.872	-14.815	-0.225	-21.169	30.350	-24.452	-856.034
2f54	D	E	A				H-TRAV21*01	H-TRBV6-5*01	I	-18.229	-14.586	-0.129	-20.989	31.765	-22.750	-885.190

[a] chain IDs assigned to: TCR α, TCR β, MHC class Iα, MHC class IIα, MHC class IIβ, and CD1D1.
[b] TCR variable domain (Vα and Vβ) alleles and Ligand type (I=MHC class I, II=MHC class II, C=CD1D1). Human/murine TCR alleles are marked with H-/M-.
[c] TCR direction vector, MHC(like) lingand direction vector, and cross product Z component. [Å]

MHC Docking Directionality (continued)

PDB	α[a]	β[a]	I[a]	IIα[a]	IIβ[a]	CD[a]	Vα[b]	Vβ[b]	L[b]	Tx[c]	Ty[c]	Tz[c]	Mx[c]	My[c]	Mz[c]	Cz[c]
2gj6	D	E	A				H-TRAV12-2*02	H-TRBV6-5*01	I	-17.573	-14.928	3.241	-35.489	8.692	-25.142	-682.524
2iam	C	D		A	B		H-TRAV22*01	H-TRBV6-6*01	II	-17.313	-14.293	0.640	-25.052	17.164	-32.326	-655.229
2ian	D	E		A	B		H-TRAV22*01	H-TRBV6-6*01	II	-16.919	-14.317	1.245	-25.707	15.263	-32.761	-626.282
2j8u	E	F	A				M-TRAV12D-2*01	M-TRBV13-3*01	I	-17.812	-14.440	3.496	-21.559	29.931	-24.625	-844.443
2nx5	D	E	A				H-TRAV1-2*01	H-TRBV10-3*01	I	-17.495	-14.869	3.365	-26.746	22.846	-27.015	-797.377
2oi9	B	C	A				M-TRAV9-4*01	M-TRBV13-2*01	I	-18.068	-14.257	1.781	-34.024	21.688	-18.414	-876.939
2ol3	A	B	H				M-TRAV16*01	M-TRBV1*01	I	-18.436	-14.709	1.969	-38.078	19.096	-12.351	-912.143
2p5e	D	E	A				H-TRAV21*01	H-TRBV6-5*01	I	-17.733	-14.677	0.073	-21.403	31.063	-23.327	-864.972
2p5w	D	E	A				H-TRAV21*01	H-TRBV6-5*01	I	-17.958	-14.884	-0.020	-21.606	30.679	-23.645	-872.517
2po6	C	D	A				H-TRAV10*01	H-TRBV25-1*01	I	-19.629	-12.750	1.889	-43.521	4.876	7.022	-650.604
2pxy	A	B		C	D		M-TRAV6D-7*01	M-TRBV13-2*01	II	-17.398	-13.975	2.815	-29.833	22.418	-23.969	-806.945
2pye	D	E	A				H-TRAV21*01	H-TRBV6-5*01	I	-17.925	-15.105	0.243	-22.466	30.286	-23.349	-882.225
2uwe	E	F	A				M-TRAV12D-2*01	M-TRBV13-3*01	I	-17.917	-14.284	3.602	-21.784	30.370	-23.878	-855.302
2vlj	D	E	A				H-TRAV27*01	H-TRBV19*01	I	-17.363	-14.706	2.752	-23.865	22.380	-29.944	-739.543
2vlk	D	E	A				H-TRAV27*01	H-TRBV19*01	I	-17.610	-14.635	2.515	-22.966	24.605	-28.885	-769.401
2vlr	D	E	A				H-TRAV27*01	H-TRBV19*01	I	-17.847	-14.487	2.681	-23.260	24.213	-28.979	-769.097
2wbj	C	D		A	B		H-TRAV17*01	H-TRBV20-1*01	II	-16.316	-15.912	3.064	-9.712	30.827	-30.372	-657.511
2ypl	D	E	A				H-TRAV5*01	H-TRBV19*01	I	-18.795	-13.892	2.881	-39.564	10.889	-16.829	-754.282
2z31	A	B		C	D		M-TRAV6D-7*01	M-TRBV13-2*01	II	-17.292	-14.131	3.061	-30.717	21.110	-24.041	-799.096
3arb	C	D				A	M-TRAV11*02	M-TRBV13-2*01	C	-17.753	-14.860	3.412	-44.321	1.007	1.329	-676.487
3ard	C	D				A	M-TRAV11*02	M-TRBV13-2*01	C	-17.822	-14.757	3.331	-44.301	0.586	2.042	-664.194
3are	C	D				A	M-TRAV11*02	M-TRBV13-2*01	C	-17.758	-14.764	3.519	-44.314	0.447	1.803	-662.190
3arf	C	D				A	M-TRAV11*02	M-TRBV13-2*01	C	-17.825	-14.861	3.170	-44.230	1.505	2.922	-684.129
3arg	C	D				A	M-TRAV11*02	M-TRBV13-2*01	C	-17.702	-15.147	3.422	-44.310	0.688	1.825	-683.343
3d39	D	E	A				H-TRAV12-2*02	H-TRBV6-5*01	I	-18.062	-14.582	2.526	-33.864	8.605	-27.318	-649.228

[a] chain IDs assigned to: TCR α, TCR β, MHC class Iα, MHC class IIα, MHC class IIβ, and CD1D1.
[b] TCR variable domain (Vα and Vβ) alleles and Ligand type (I=MHC class I, II=MHC class II, C=CD1D1). Human/murine TCR alleles are marked with H-/M-.
[c] TCR direction vector, MHC(like) lingand direction vector, and cross product Z component. [Å]

MHC Docking Directionality (continued)

PDB	α[a]	β[a]	I[a]	IIα[a]	IIβ[a]	CD[a]	Vα[b]	Vβ[b]	L[b]	Tx[c]	Ty[c]	Tz[c]	Mx[c]	My[c]	Mz[c]	Cz[c]
3d3v	D	E	A				H-TRAV12-2*02	H-TRBV6-5*01	I	-17.955	-14.674	2.452	-33.920	8.377	-27.320	-648.151
3dxa	D	E	A				H-TRAV26-1*01	H-TRBV7-9*01	I	-20.864	-12.536	1.890	-22.401	27.905	-26.204	-863.029
3e2h	B	C	A				M-TRAV9-4*01	M-TRBV13-2*01	I	-18.048	-14.253	1.922	-34.737	20.618	-18.313	-867.220
3e3q	D	E	A				M-TRAV9-4*01	M-TRBV13-2*01	I	-18.299	-14.171	1.567	-34.930	20.036	-18.590	-861.632
3ffc	D	E	A				H-TRAV14*01	H-TRBV11-2*03	I	-18.375	-12.985	0.800	-37.007	9.507	-22.521	-655.227
3gsn	A	B	H				H-TRAV24*01	H-TRBV6-5*01	I	-16.821	-15.004	2.865	-37.181	12.413	-20.751	-766.663
3h9s	D	E	A				H-TRAV12-2*02	H-TRBV6-5*01	I	-17.790	-14.679	2.811	-33.723	8.532	-27.514	-646.804
3he6	C	D				A	M-TRAV11*02	M-TRBV13-2*01	C	-18.052	-14.581	3.293	-43.757	5.694	4.472	-740.809
3he7	C	D				A	M-TRAV11*02	M-TRBV29*01	C	-18.732	-13.613	1.509	-43.630	4.533	6.560	-678.847
3hg1	D	E	A				H-TRAV12-2*01	H-TRBV30*01	I	-17.169	-15.209	1.814	-28.714	16.339	-29.592	-717.236
3huj	G	H	A				H-TRAV10*01	H-TRBV25-1*01	I	-19.445	-12.974	2.893	-43.125	0.676	10.338	-572.649
3kpr	D	E	A				H-TRAV26-2*01	H-TRBV7-8*01	I	-21.396	-11.370	1.516	-32.842	26.216	-14.183	-934.331
3kps	D	E	A				H-TRAV26-2*01	H-TRBV7-8*01	I	-21.104	-11.668	1.586	-32.395	26.954	-13.826	-946.822
3kxf	D	E	A				H-TRAV19*01	H-TRBV6-1*01	I	-15.523	-15.919	3.707	-26.102	30.124	-19.451	-883.133
3mbe	C	D		A	B		M-TRAV4-2*01	M-TRBV13-3*01	II	-16.699	-15.572	1.652	-26.020	16.644	-31.829	-683.122
3mv7	D	E	A				H-TRAV20*01	H-TRBV9*01	I	-16.678	-14.904	2.736	-32.181	10.748	-28.566	-658.881
3mv8	D	E	A				H-TRAV20*01	H-TRBV9*02	I	-16.707	-14.876	2.704	-32.374	10.418	-28.470	-655.649
3mv9	D	E	A				H-TRAV20*01	H-TRBV9*01	I	-16.731	-15.017	2.676	-32.507	10.629	-28.239	-665.991
3o4l	D	E	A				H-TRAV5*01	H-TRBV20-1*01	I	-15.182	-16.500	5.300	-37.544	13.543	-19.341	-825.086
3o6f	C	D		A	B		H-TRAV26-2*01	H-TRBV20-1*01	II	-20.308	-12.955	2.510	-33.234	16.037	-24.605	-756.226
3o8x	C	D				A	M-TRAV11*02	M-TRBV13-2*01	C	-17.959	-15.118	2.634	-44.330	-0.372	1.353	-663.500
3o9w	C	D				A	M-TRAV11*02	M-TRBV13-2*01	C	-17.972	-15.060	2.741	-44.235	1.101	3.038	-685.966
3pl6	C	D		A	B		H-TRAV13-1*02	H-TRBV7-3*01	II	-16.987	-14.416	2.348	-41.655	5.520	-14.195	-694.267
3pqy	D	E	A				M-TRAV21*02	M-TRBV29*01	I	-20.861	-12.168	1.371	-31.088	28.597	-13.521	-974.841
3pwp	D	E	A				H-TRAV12-2*02	H-TRBV6-5*01	I	-18.042	-14.599	2.511	-33.507	8.753	-27.708	-647.090

[a] chain IDs assigned to: TCR α, TCR β, MHC class Iα, MHC class IIα, MHC class IIβ, and CD1D1.
[b] TCR variable domain (Vα and Vβ) alleles and Ligand type (I=MHC class I, II=MHC class II, C=CD1D1). Human/murine TCR alleles are marked with H-/M-.
[c] TCR direction vector, MHC(like) lingand direction vector, and cross product Z component. [Å]

MHC Docking Directionality (continued)

PDB	α[a]	β[a]	I[a]	IIα[a]	IIβ[a]	CD[a]	Vα[b]	Vβ[b]	L[b]	Tx[c]	Ty[c]	Tz[c]	Mx[c]	My[c]	Mz[c]	Cz[c]
3qdg	D	E	A				H-TRAV12-2*01	H-TRBV6-4*01	I	-15.835	-15.407	2.282	-36.504	7.389	-24.082	-679.422
3qdj	D	E	A				H-TRAV12-2*01	H-TRBV6-4*01	I	-15.711	-15.467	2.281	-36.846	7.550	-23.505	-688.515
3qdm	D	E	A				H-TRAV35*02	H-TRBV10-3*01	I	-17.226	-14.972	3.524	-39.646	9.347	-17.547	-754.591
3qeq	D	E	A				H-TRAV35*02	H-TRBV10-3*01	I	-17.244	-14.585	2.758	-33.195	19.292	-22.204	-816.820
3qfj	D	E	A				H-TRAV12-2*02	H-TRBV6-5*01	I	-17.867	-14.638	2.671	-34.289	8.506	-26.815	-653.899
3qi9	C	D				A	M-TRAV11*02	M-TRBV19*01	C	-18.243	-13.992	2.927	-44.203	3.428	-1.219	-681.025
3qib	C	D		A	B		M-TRAV4D-4*02	M-TRBV26*01	II	-17.771	-13.919	0.737	-29.650	13.782	-29.968	-657.618
3qiu	C	D		A	B		M-TRAV4D-4*03	M-TRBV26*01	II	-17.060	-14.245	0.172	-28.108	14.876	-30.915	-654.183
3qiw	C	D		A	B		M-TRAV4D-4*03	M-TRBV26*01	II	-17.353	-14.267	0.158	-28.282	15.048	-30.672	-664.627
3qux	C	D				A	M-TRAV11*02	M-TRBV13-2*01	C	-17.604	-15.441	2.872	-44.298	-0.571	2.122	-673.954
3quy	C	D				A	M-TRAV11*02	M-TRBV13-2*01	C	-17.851	-15.111	2.644	-44.323	-0.223	1.586	-665.784
3quz	C	D				A	M-TRAV11*02	M-TRBV13-2*01	C	-17.934	-15.150	2.621	-44.345	-0.368	0.648	-665.227
3rdt	A	B		C	D		M-TRAV14D-3*02	M-TRBV13-2*01	II	-17.206	-14.688	4.687	-35.366	13.865	-22.894	-758.017
3rgv	A	B	C				M-TRAV6-3*01	M-TRBV13-2*01	I	-19.043	-13.821	2.978	-29.015	22.036	-25.290	-820.648
3rtq	C	D				A	M-TRAV11*02	M-TRBV13-2*01	C	-17.715	-15.278	2.764	-44.265	-0.376	2.745	-669.620
3rug	E	F				A	M-TRAV13D-3*01	M-TRBV13-3*01	C	-17.580	-15.135	0.313	-44.053	-3.825	3.436	-599.499
3rzc	C	D				A	M-TRAV11*02	M-TRBV13-2*01	C	-17.758	-15.271	2.845	-44.298	-0.303	2.180	-671.094
3scm	C	D				A	M-TRAV11*02	M-TRBV19*01	C	-18.314	-13.908	3.132	-44.143	4.175	-1.027	-690.402
3sda	C	D				A	M-TRAV11*02	M-TRBV19*01	C	-18.338	-13.937	2.885	-44.236	3.214	0.111	-675.455
3sdc	C	D				A	M-TRAV11*02	M-TRBV19*01	C	-18.347	-13.884	2.974	-44.143	4.239	-0.747	-690.654
3sdd	C	D				A	M-TRAV11*02	M-TRBV19*01	C	-18.270	-14.014	3.048	-44.082	4.839	0.658	-706.174
3sdx	E	F	A				H-TRAV10*01	H-TRBV25-1*01	I	-19.408	-13.173	2.248	-43.422	6.725	6.032	-702.517
3sjv	D	E	A				H-TRAV12-1*01	H-TRBV6-2*01	I	-19.079	-14.176	1.119	-34.492	16.549	-22.439	-804.697
3t0e	C	D		A	B		H-TRAV26-2*01	H-TRBV20-1*01	II	-20.323	-13.129	2.900	-33.644	16.595	-23.660	-778.972
3ta3	C	D				A	M-TRAV11*02	M-TRBV13-2*01	C	-17.731	-15.318	2.820	-44.300	-0.527	2.086	-669.243

[a] chain IDs assigned to: TCR α, TCR β, MHC class Iα, MHC class IIα, MHC class IIβ, and CD1D1.
[b] TCR variable domain (Vα and Vβ) alleles and Ligand type (I=MHC class I, II=MHC class II, C=CD1D1). Human/murine TCR alleles are marked with H-/M-.
[c] TCR direction vector, MHC(like) lingand direction vector, and cross product Z component. [Å]

MHC Docking Directionality (continued)

PDB	α[a]	β[a]	I[a]	IIα[a]	IIβ[a]	CD[a]	Vα[b]	Vβ[b]	L[b]	Tx[c]	Ty[c]	Tz[c]	Mx[c]	My[c]	Mz[c]	Cz[c]
3tf7	C	D	A				M-TRAV9D-3*01	M-TRBV13-1*02	I	-18.798	-13.358	1.427	-34.060	27.326	-7.770	-968.648
3tfk	C	D	A				M-TRAV9D-3*01	M-TRBV13-1*02	I	-17.620	-14.526	3.285	-38.688	21.361	-3.750	-938.363
3tjh	C	D	A				M-TRAV9D-3*01	M-TRBV13-1*02	I	-17.587	-14.294	4.041	-43.152	-5.670	8.540	-517.096
3tn0	C	D				A	M-TRAV11*02	M-TRBV13-2*03	C	-17.692	-14.931	3.438	-44.027	3.894	3.685	-726.260
3to4	C	D				A	M-TRAV11*02	M-TRBV1*01	C	-17.985	-15.772	1.281	-44.333	1.275	0.075	-722.151
3tpu	A	B	I				M-TRAV9D-3*01	M-TRBV13-1*02	I	-19.324	-13.291	2.812	-36.990	18.973	-15.453	-858.268
3tvm	C	D				A	M-TRAV11*02	M-TRBV13-2*01	C	-17.811	-15.438	2.556	-44.258	-1.695	2.344	-653.065
3tzv	G	H				C	H-TRAV10*01	H-TRBV25-1*01	C	-19.365	-12.986	1.320	-43.027	7.198	-7.996	-698.138
3uts	D	E	A				H-TRAV12-3*01	H-TRBV12-4*01	I	-17.922	-14.984	1.873	-35.122	11.284	-24.621	-728.500
3utt	D	E	A				H-TRAV12-3*01	H-TRBV12-4*01	I	-17.035	-15.597	3.712	-34.952	17.075	-21.305	-836.019
3vwj	C	D				A	H-TRAV10*01	H-TRBV25-1*01	C	-19.066	-13.228	3.004	-43.311	1.733	9.394	-605.959
3vwk	C	D				A	H-TRAV10*01	H-TRBV25-1*01	C	-19.161	-13.463	3.230	-43.345	0.844	9.361	-599.726
3vxm	D	E	A				H-TRAV8-3*01	H-TRBV4-1*01	I	-18.297	-14.052	0.754	-24.417	14.484	-34.076	-608.121
3vxr	D	E	A				H-TRAV21*01	H-TRBV7-9*01	I	-16.929	-14.413	0.616	-27.510	18.768	-29.292	-714.225
3vxs	D	E	A				H-TRAV21*01	H-TRBV7-9*01	I	-17.119	-14.534	0.338	-26.768	20.638	-28.717	-742.348
3vxu	D	E	A				H-TRAV12-2*02	H-TRBV27*01	I	-17.588	-14.979	1.602	-23.455	24.347	-28.708	-779.547
3w0w	D	E	A				H-TRAV12-2*02	H-TRBV27*01	I	-16.771	-15.101	1.047	-27.062	22.322	-27.139	-783.026
4apq	C	D				A	M-TRAV11*02	M-TRBV19*01	C	-18.268	-13.946	2.922	-43.608	7.846	1.989	-751.488
4e41	D	E		A	B		H-TRAV22*01	H-TRBV5-8*01	II	-17.377	-14.376	1.064	-24.692	16.798	-32.790	-646.871
4ei5	C	D				A	M-TRAV7-4*02	M-TRBV3*01	C	-18.844	-13.964	1.789	-14.792	39.518	-13.661	-951.233
4elm	G	H				A	M-TRAV7-4*02	M-TRBV3*01	C	-19.549	-13.529	2.112	-14.719	40.491	-10.531	-990.692
4en3	A	B				C	H-TRAV17*01	H-TRBV25-1*01	C	-17.808	-13.782	3.947	-44.343	0.865	0.002	-626.539
4eup	G	H	D				H-TRAV12-2*02	H-TRBV28*01	I	-17.731	-14.871	3.054	-40.003	12.649	-14.383	-819.164
4ftv	D	E	A				H-TRAV12-2*02	H-TRBV6-5*01	I	-17.949	-14.935	2.553	-34.309	7.504	-27.087	-647.094
4g8g	D	E	A				H-TRAV14*02	H-TRBV6-5*01	I	-18.272	-13.370	-0.297	-37.558	22.148	-8.122	-906.839

[a] chain IDs assigned to: TCR α, TCR β, MHC class Iα, MHC class IIα, MHC class IIβ, and CD1D1.
[b] TCR variable domain (Vα and Vβ) alleles and Ligand type (I=MHC class I, II=MHC class II, C=CD1D1). Human/murine TCR alleles are marked with H-/M-.
[c] TCR direction vector, MHC(like) lingand direction vector, and cross product Z component. [Å]

MHC Docking Directionality (continued)

PDB	α[a]	β[a]	I[a]	IIα[a]	IIβ[a]	CD[a]	Vα[b]	Vβ[b]	L[b]	Tx[c]	Ty[c]	Tz[c]	Mx[c]	My[c]	Mz[c]	Cz[c]
4g9f	D	E	A				H-TRAV14*02	H-TRBV6-5*01	I	-18.629	-13.519	-0.234	-37.690	22.096	-7.635	-921.157
4gg6	G	H		A	B		H-TRAV26-2*01	H-TRBV9*01	II	-20.537	-11.512	1.412	-30.499	23.707	-21.794	-837.975
4grl	C	D		A	B		H-TRAV13-1*02	H-TRBV7-3*01	II	-16.979	-14.340	2.299	-41.377	5.332	-15.055	-683.878
4h1l	I	J		A	B		H-TRAV8-3*01	H-TRBV19*01	II	-17.752	-14.011	2.091	-23.097	22.505	-30.449	-723.121
4irj	C	D				A	M-TRAV11*02	M-TRBV13-2*01	C	-17.929	-15.183	2.656	-44.320	-0.571	1.580	-662.673
4irs	C	D				A	M-TRAV11*02	M-TRBV13-2*01	C	-17.942	-15.120	2.723	-44.316	-0.186	1.777	-666.721
4jfd	D	E	A				H-TRAV12-2*01	H-TRBV30*01	I	-17.153	-14.711	0.728	-30.515	14.006	-28.979	-689.151
4jfe	D	E	A				H-TRAV12-2*01	H-TRBV30*01	I	-17.122	-14.746	0.601	-30.490	14.122	-28.950	-691.402
4jff	D	E	A				H-TRAV12-2*01	H-TRBV30*01	I	-17.159	-14.581	0.650	-30.250	14.147	-29.188	-683.824
4jrx	D	E	A				H-TRAV19*01	H-TRBV6-1*01	I	-15.672	-15.690	3.053	-24.888	28.070	-23.659	-830.406
4jry	D	E	A				H-TRAV39*01	H-TRBV5-6*01	I	-17.065	-14.417	-0.230	-14.472	22.449	-35.408	-591.735
4l3e	D	E	A				H-TRAV12-2*01	H-TRBV6-4*01	I	-15.977	-15.448	1.971	-37.081	7.520	-23.140	-692.974
4l4t	G	H	A				H-TRAV1-2*01	H-TRBV6-1*01	I	-18.275	-14.280	2.234	-22.911	25.596	-28.054	-794.936
4l4v	G	H	A				H-TRAV1-2*01	H-TRBV6-1*01	I	-18.339	-14.286	2.412	-22.684	25.923	-27.939	-799.466
4l9l	A	B	C				H-TRAV1-2*01	H-TRBV6-2*01	I	-18.695	-13.981	2.822	-22.441	31.067	-22.323	-894.545
4lcw	G	H	A				H-TRAV1-2*01	H-TRBV6-1*01	I	-18.134	-14.310	2.262	-23.176	25.037	-28.340	-785.670
4may	C	D		A	B		H-TRAV13-1*02	H-TRBV7-3*01	II	-16.940	-14.380	2.407	-41.684	5.583	-14.084	-693.992
4mji	D	E	A				H-TRAV17*01	H-TRBV7-3*01	I	-17.040	-14.412	1.611	-34.416	18.340	-21.126	-808.517
4mnq	D	E	A				H-TRAV22*01	H-TRBV6-5*01	I	-16.836	-14.606	-0.226	-30.627	16.293	-27.634	-721.647
4ms8	C	D	A				M-TRAV9D-3*01	M-TRBV13-1*02	I	-17.659	-14.094	2.430	-28.222	31.770	-12.700	-958.787
4nhu	A	B	G				M-TRAV9-4*01	M-TRBV13-2*01	I	-16.826	-15.181	2.744	-1.437	27.669	-34.634	-487.374
4nqc	D	E	C				H-TRAV1-2*01	H-TRBV6-1*01	I	-18.311	-14.172	2.361	-22.797	25.805	-27.955	-795.594
4nqd	D	E	C				H-TRAV1-2*01	H-TRBV6-1*01	I	-17.894	-14.359	2.830	-22.805	26.583	-27.209	-803.133
4nqe	D	E	C				H-TRAV1-2*01	H-TRBV6-1*01	I	-18.123	-14.270	2.586	-22.494	27.124	-26.934	-812.558
4ozf	G	H		A	B		H-TRAV26-1*01	H-TRBV7-2*01	II	-21.910	-11.150	-1.519	-25.941	30.353	-19.311	-954.276

[a] chain IDs assigned to: TCR α, TCR β, MHC class Iα, MHC class IIα, MHC class IIβ, and CD1D1.
[b] TCR variable domain (Vα and Vβ) alleles and Ligand type (I=MHC class I, II=MHC class II, C=CD1D1). Human/murine TCR alleles are marked with H-/M-.
[c] TCR direction vector, MHC(like) lingand direction vector, and cross product Z component. [Å]

MHC Docking Directionality (continued)

PDB	α[a]	β[a]	I[a]	IIα[a]	IIβ[a]	CD[a]	Vα[b]	Vβ[b]	L[b]	Tx[c]	Ty[c]	Tz[c]	Mx[c]	My[c]	Mz[c]	Cz[c]
4ozg	G	H		A	B		H-TRAV26-1*01	H-TRBV7-2*01	II	-21.084	-11.938	0.237	-31.201	20.989	-23.518	-815.010
4ozh	G	H		A	B		H-TRAV26-1*01	H-TRBV7-2*01	II	-22.236	-10.992	-0.480	-29.061	22.130	-25.156	-811.521
4ozi	G	H		A	B		H-TRAV4*01	H-TRBV20-1*01	II	-20.927	-11.881	0.681	-28.548	31.034	-13.746	-988.627
4p23	A	B		C	D		M-TRAV14D-3*02	M-TRBV13-2*01	II	-17.195	-14.743	4.777	-34.996	14.324	-23.177	-762.247
4p2o	C	D		A	B		M-TRAV4D-4*02	M-TRBV26*01	II	-17.450	-14.080	0.826	-28.357	14.263	-30.977	-648.156
4p2q	D	E		A	B		M-TRAV4D-4*03	M-TRBV26*01	II	-17.992	-14.095	1.128	-31.634	10.384	-29.300	-632.710
4p2r	D	E		A	B		M-TRAV4D-4*03	M-TRBV26*01	II	-18.080	-13.953	1.266	-31.711	10.528	-29.167	-632.810
4p46	A	B		C	D		M-TRAV14D-3*02	M-TRBV13-2*01	II	-17.230	-14.772	4.699	-34.180	12.256	-25.467	-716.078
4p4k	C	D		A	B		H-TRAV9-2*01	H-TRBV5-1*01	II	-17.789	-13.781	2.650	-31.447	24.253	-19.749	-864.808
4p5t	A	B		C	D		M-TRAV14D-3*02	M-TRBV13-2*01	II	-16.773	-15.306	3.799	-31.076	14.341	-28.209	-716.191
4pj5	G	H	A				H-TRAV1-2*01	H-TRBV6-1*01	I	-18.179	-14.399	2.278	-23.128	25.553	-27.915	-797.548
4pj7	G	H	A				H-TRAV1-2*01	H-TRBV6-4*01	I	-18.105	-14.563	2.316	-23.390	24.327	-28.779	-781.069
4pj8	C	D	A				H-TRAV1-2*01	H-TRBV20-1*01	I	-19.222	-13.675	3.131	-26.262	23.698	-26.755	-814.656
4pj9	C	D	A				H-TRAV1-2*01	H-TRBV6-4*01	I	-17.409	-14.770	2.770	-25.616	24.385	-26.764	-802.867
4pja	E	F	A				H-TRAV1-2*01	H-TRBV6-1*01	I	-18.119	-14.338	2.318	-23.203	25.636	-27.777	-797.183
4pjb	E	F	A				H-TRAV1-2*01	H-TRBV6-1*01	I	-18.233	-14.174	2.050	-23.189	24.896	-28.453	-782.610
4pjc	E	F	A				H-TRAV1-2*01	H-TRBV6-1*01	I	-17.748	-14.596	2.555	-23.604	25.168	-27.867	-791.206
4pjd	E	F	A				H-TRAV1-2*01	H-TRBV6-1*01	I	-18.080	-14.433	2.445	-23.343	25.278	-27.987	-793.936
4pje	E	F	A				H-TRAV1-2*01	H-TRBV6-1*01	I	-18.392	-14.150	2.301	-22.222	26.292	-27.965	-798.004
4pjf	G	H	A				H-TRAV1-2*01	H-TRBV6-1*01	I	-18.168	-14.307	2.388	-23.380	26.032	-27.254	-807.447
4pjg	G	H	A				H-TRAV1-2*01	H-TRBV6-1*01	I	-18.418	-14.180	2.032	-23.491	24.903	-28.198	-791.766
4pjh	E	F	A				H-TRAV1-2*01	H-TRBV6-1*01	I	-18.182	-14.270	2.125	-22.818	26.249	-27.522	-802.872
4pji	G	H	A				H-TRAV1-2*01	H-TRBV6-1*01	I	-18.166	-14.528	2.719	-22.712	25.225	-28.548	-788.197
4pjx	E	F	A				H-TRAV1-2*01	H-TRBV6-1*01	I	-17.913	-14.377	2.063	-23.884	24.312	-28.383	-778.881
4prh	D	E	A				H-TRAV20*01	H-TRBV9*01	I	-17.092	-14.967	2.259	-31.066	11.718	-29.406	-665.249

[a] chain IDs assigned to: TCR α, TCR β, MHC class Iα, MHC class IIα, MHC class IIβ, and CD1D1.
[b] TCR variable domain (Vα and Vβ) alleles and Ligand type (I=MHC class I, II=MHC class II, C=CD1D1). Human/murine TCR alleles are marked with H-/M-.
[c] TCR direction vector, MHC(like) lingand direction vector, and cross product Z component. [Å]

MHC Docking Directionality (continued)

PDB	α[a]	β[a]	I[a]	IIα[a]	IIβ[a]	CD[a]	Vα[b]	Vβ[b]	L[b]	Tx[c]	Ty[c]	Tz[c]	Mx[c]	My[c]	Mz[c]	Cz[c]
4pri	D	E	A				H-TRAV20*01	H-TRBV9*01	I	-17.024	-14.804	2.530	-31.649	10.610	-29.204	-649.156
4prp	D	E	A				H-TRAV20*01	H-TRBV9*01	I	-16.881	-14.913	2.769	-32.248	10.572	-28.554	-659.380
4qok	D	E	A				H-TRAV12-2*01	H-TRBV30*01	I	-17.356	-15.086	1.787	-28.611	16.547	-29.575	-718.815
4qrp	D	E	A				H-TRAV9-2*02	H-TRBV11-2*01	I	-18.468	-14.058	2.143	-26.346	13.076	-33.197	-611.860
4qrr	D	E	A				H-TRDV1*01	H-TRBV5-1*01	I	-17.872	-14.134	1.659	-37.661	19.565	-12.882	-881.966
4wo4	C	D	A				H-TRDV1*01	H-TRBV20-1*02	I	-17.267	-15.753	2.429	-29.214	22.082	-25.021	-841.498
4x6c	G	H	A				H-TRAV12-3*01	H-TRBV6-2*01	I	-17.560	-14.631	3.260	-7.331	34.126	-27.363	-706.512
4x6d	G	H	A				H-TRAV12-3*01	H-TRBV6-2*01	I	-17.269	-14.869	3.358	-9.673	37.652	-21.352	-794.040

[a] chain IDs assigned to: TCR α, TCR β, MHC class Iα, MHC class IIα, MHC class IIβ, and CD1D1.
[b] TCR variable domain (Vα and Vβ) alleles and Ligand type (I=MHC class I, II=MHC class II, C=CD1D1). Human/murine TCR alleles are marked with H-/M-.
[c] TCR direction vector, MHC(like) lingand direction vector, and cross product Z component. [Å]