TECHNISCHE UNIVERSITÄT MÜNCHEN

Department Chemie

Lehrstuhl für Biotechnologie

Folding and association of antibody domains

Eva-Maria Herold

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. Stephan A. Sieber
Prüfer der Dissertation:	1. Univ.-Prof. Dr. Johannes Buchner
	2. Univ.-Prof. Dr. Bernd Reif

Die Dissertation wurde am 19. Juni 2013 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 28. August 2013 angenommen.

Contents

Contents

Summary

Immunoglobulins (Ig) are the key players of the adaptive immune response. They directly neutralise pathogens and pathogen-derived products and recruit molecular and cellular immune effectors to eradicate infections and tumour cells. Immunoglobulins either exist in a membrane-bound form (B-cell receptors) or as secreted effector molecules (antibodies). In higher vertebrates five different classes, so-called isotypes, of immunoglobulins exist (IgA, IgD, IgE, IgG and IgM). IgGs form the most abundant isotype family. Antibodies and their derivatives play an important role in therapeutic applications and in diagnostics. However, there is often a need to improve certain properties, such as stability, specificity or immunogenicity. Engineering of single qualities often influences others in a negative manner. Therefore the more we know about the relationship of the different attributes the more successful we will be in modifying them. This thesis focuses on the folding and stability but also on association of and functionality the variable domains.

Antibodies are ideal model system for folding analysis. They consist of variable and constant domains which share a common structural motif, the immunoglobulin fold. As single domains and various fragments are stable in isolation, the complexity of the system can be varied. One special property of antibodies, fragments and some isolated domains is their ability to adopt a stable structural state with defined secondary and non-defined tertiary interactions at acidic pH conditions. As this state depicts non-native secondary structure it was termed alternatively folded state (AFS). In some aspects it is similar to the well-described molten globule state but differs due to its high thermodynamic stability and oligomer formation. In this thesis the abilities to gain an AFS for the isolated Fab domains and the second constant domain of the heavy chain (C_H2), which is glycosylated in intact antibodies, were characterised. Although they share the same Ig fold under physiological conditions surprising differences between the domains could be observed at acidic pH. All isolated domains except the constant domain of the light chain (C_L) were able to adopt an AFS at acidic conditions and moderate ionic strength. The ability to form oligomers seemed to correlate with hydrophobic intermolecular interactions for Fab (fragment, antigen binding) domains but not for C_H2. The thermodynamic stability differed between the domains, too. Interestingly, particularly those domains which were quite unstable in the native state or even intrinsically disordered, adopted very stable structures at pH 2.0. Furthermore, the AFS is a reversible state, although for some domains only in a limited concentration range. The glycan of C_H2 lost the stabilising effects observed in physiological conditions when the AFS is adopted. The pathway to the AFS is comparable for all domains. The rearrangements on the secondary structure level are enabled by a proceeding partially of completely unfolded intermediate. In addition, we were able to gain first insight in the structure of the AFS of C_H2 by NMR revealing a dynamic system instead of a fixed state with partially unstructured regions.

Summary

The analysis of the folding behaviour of isolated antibody domains had shown that they differ. One domain which was not yet analysed is the variable domain of the heavy chain (V_H). In general, this domain is less stable than its counterpart, the variable domain of the light chain (V_L), and often tends to aggregate in isolation. V_H of MAK33, an IgG1 antibody which is well characterised, is very unstable. An R44A mutation can stabilise the domain and grants reversibility of unfolding. It was not possible to elucidate a complete folding pathway within this work. However, some insights in the differences between V_H and other antibody domains as well as between different V_H domains could be gained. The folding of V_H is a complex process with multiexponential phases within the observed kinetic traces. Some of the apparent rate constants showed no dependency on the concentration of denaturant indicating slow isomerisation processes. Prolyl isomerisation plays an important role for all antibody domains characterised so far. Yet, it could not definitely be shown that this process appears in the pathway of V_H. In addition, V_H domains of different antibodies were examined and interesting differences could be observed in the complexity of their folding behaviour.

Another field of interest within this thesis was the interaction mechanism between the two variable domains. Previous covariation analysis identified conserved amino acids at or adjacent to the interface but their functions remained unclear. These residues were exchanged in V_L and V_H against alanine and the mutants analysed. Here, for the first time isolated domains were addressed and not mutations in Fab or scFv constructs. The focus was set on the influence of amino acid exchange on structure, stability and interaction between V_H and V_L as well as on functionality (i.e., antigen recognition). In terms of structure and stability, V_H was much more sensitive compared to V_L when conserved amino acids were exchanged. The often presumed correlation between antigen binding and interaction between V_H and V_L could not be confirmed on the single domain level. Therefore the studies on the interaction based on scFvs or Fab fragments should be reconsidered. The third complementarity determining region of V_H (CDR-H3) is most flexible in sequence and length. Introduction of single charged positive residues in this loop had positive effects on the stability and did not disturb the interaction or functionality. However, combinations were non-favourable. Hence, a single point mutation can be sufficient to improve the properties of V_H.

In summary, in this thesis new insights on the folding behaviour of single antibody domains were gained as well as on the interaction mechanism between the variable domains. These findings will be helpful for antibody engineering and are a thought-provoking impulse to rethink interaction analysis performed so far. However, if the effects observed on the level of isolated domains of MAK33 are also present in other antibodies or if they can be compensated within Fab of scFv fragments needs to be tested.

Zusammenfassung

Immunoglobuline nehmen eine Schlüsselrolle in der adaptiven Immunantwort ein. Ihre Aufgabe besteht darin, Pathogene und deren schädliche Produkte zu neutralisieren um Infektionen sowie Tumorzellen zu beseitigen. Immunoglobuline treten in zwei verschiedenen Formen auf. Zum einen können sie an die Zellmembran gebunden werden (B-Zellen Rezeptor), zum anderen können sie als Effektormoleküle sekretiert werden (Antikörper). Höhere Wirbeltiere besitzen fünf verschiedene Immunoglobulin Typen (sogenannte Isotypen): IgA, IgD, IgE, IgG und IgM. IgGs machen den größten Teil aller Immunglobuline aus. Antikörper und deren Derivate spielen auch eine wichtige Rolle in der therapeutischen Anwendung und Diagnostik. Es müssen jedoch immer wieder bestimmte Eigenschaften für die gewünschte Anwendung verbessert werden. Solche Veränderungen betreffen zumeist Stabilität, Spezifität oder die immunogene Aktivität. Eine Herausforderung bei der Optimierung einer bestimmten Eigenschaft ist die komplexe Beziehung dieser untereinander, welche häufig dazu führt, dass sie sich gegenseitig negativ beeinflussen. Aufgrund dessen ist es wichtig, das Verständnis dieser komplexen Beziehungen auszuweiten und damit erfolgreichere Strategien entwickeln zu können. Der Schwerpunkt dieser Arbeit liegt auf der Analyse der wichtigsten Eigenschaften für den Einsatz von Antikörpern: Faltung, Stabilität, wechselseitige Beziehungen zwischen variablen Domänen und deren Funktionalität.

Antikörper stellen hervorragende Modellsysteme dar, weil sie aus verschiedenen Domänen bestehen. Es gibt variable und konstante Domänen welche ein gemeinsames Faltungsmotiv teilen, den sogenannten *immunoglobulin fold*. Sowohl die einzelnen Domänen, als auch verschiedene Antikörperfragmente sind stabil und erlauben das komplexe System Antikörper stufenweise zu vereinfachen. Eine besondere Eigenschaft von Antikörpern, deren Fragmente und einiger isolierter Domänen ist die Fähigkeit bei sauren pH Bedingungen einen stabilen Zustand einzunehmen. Dieser zeichnet sich durch einen definierten Sekundärstrukturanteil aus, wohingegen die Tertiärstruktur meist nicht definiert ist. Da die Sekundärstruktur gewöhnlich von der nativen Struktur abweicht, wurde dieser „alternativer Faltungszustand“ (*alternatively folded state*, AFS) benannt. Dieser Faltungszustand ähnelt in einigen Eigenschaften dem bereits charakterisierten *molten-globule*-Zustand. Deutliche Unterschiede zwischen den beiden Zuständen bestehen in deren Stabilitäten und ihrer Fähigkeit höhere Molekülstrukturen zu bilden. Der AFS ist ein sehr stabiler Zustand und führt häufig zur Bildung von Oligomeren. In dieser Arbeit wurde der AFS von den Domänen des Fab Fragmentes sowie der C_H2-Domäne (zweite konstante Domäne der schweren Kette) untersucht. C_H2 ist eine besondere Domäne, da sie in Antikörpern glykosyliert vorliegt. Obwohl alle Domänen im nativen Zustand dasselbe Faltungsmuster aufweisen, unterschied sich deren Verhalten bei saurem pH. Alle Domänen, mit Ausnahme der konstanten Domäne der leichten Kette (C_L), waren in der Lage, einen AFS bei sauren Bedingungen und moderater Ionenstärke zu bilden. Bei den Fab Domänen schien die Fähigkeit, Oligomere zu bilden, mit

der Möglichkeit hydrophobe Wechselwirkungen auszubilden, zusammenzuhängen. Im Fall von C_H2 ließ sich diese Beobachtung nicht bestätigen. Zudem gab es Unterschiede in den Stabilitäten der einzelnen Domänen. Interessanterweise waren solche Domänen im AFS besonders stabil, welche im nativen Zustand instabil oder sogar ungefaltet sind. Der AFS ist ein reversibler Zustand, jedoch gilt das für manche Domänen nur in einem eingeschränkten Konzentrationsbereich. Die gebundenen Zucker im Fall von C_H2 hatten einen positiven Einfluss auf die Stabilität der Domänen. Änderten sich allerdings die Bedingungen von physiologischem zu saurem pH, so ging dieser Effekt verloren. Der Mechanismus der AFS-Bildung ist für alle Domänen vergleichbar. Um die Umstrukturierung der Sekundärstruktur zu ermöglichen, passieren die Domänen ein teilweise oder vollständig entfaltetes Intermediat. Bisher sind keine Strukturdetails über den AFS bekannt. In dieser Arbeit konnte zum ersten Mal mittels NMR ein Einblick in den AFS gewonnen werden. Der Zustand zeichnet sich durch eine hohe Flexibilität aus, da einige Bereiche unstrukturiert bleiben und ist allgemein sehr dynamisch.

Durch die Untersuchungen der Faltungsmechanismen von isolierten Antikörperdomänen weiß man heute, dass sich diese trotz des identischen Faltungsmotivs unterscheiden. Die variable Domäne der schweren Kette (V_H) wurde bisher nicht in die Untersuchungen eingeschlossen. V_H-Domänen zeichnen sich zumeist durch eine geringere Stabilität als die variable Domäne der leichten Kette (V_L) sowie durch ihre Aggregationsanfälligkeit aus. In dieser Arbeit ist es zwar nicht gelungen, den vollständigen Faltungsweg aufzuklären, jedoch konnten einige Unterschiede sowohl zu den Faltungswegen der anderen Domänen als auch zwischen verschiedenen V_H-Domänen beobachtet werden. Der Faltungsmechanismus der V_H-Domänen ist komplex und es lassen sich mehrere Geschwindigkeitskonstanten für Entfaltung und Rückfaltung beobachten. Einige langsame Geschwindigkeitskonstanten zeigen keine Abhängigkeit von der Konzentration des Denaturierungsmittels. Dies ist häufig ein Indiz für langsame Isomerisierungsprozesse. Am häufigsten wurde *cis-trans* Isomerisierung von Prolinen beobachtet, welche in allen Antikörperdomänen eine wichtige Stellung einnimmt. Es konnte im Fall von V_H jedoch nicht eindeutig gezeigt werden, dass auch diese Domäne davon beeinflusst wird. Zusätzlich hat die Untersuchung an verschiedenen V_H-Domänen gezeigt, dass sie sich in ihren Faltungswegen unterscheiden.

Ein weiterer Fokus dieser Arbeit lag auf der Interaktion zwischen den variablen Domänen. Mittels bioinformatischer Analysen konnten konservierte Aminosäuren im Bereich im und in der Nähe der Interaktionsflächen identifiziert werden. Deren tatsächliche Funktion ist jedoch noch nicht bekannt. Die konservierten Reste wurden vereinzelt gegen Alanin ausgetauscht und die Folgen auf Struktur, Stabilität, Wechselwirkung und Funktionalität (Antigen-erkennung) untersucht. Die vorliegende Arbeit unterscheidet sich von früheren Studien, da hier die variablen Domänen einzeln und nicht in einem Fab oder scFv Fragment analysiert wurden. Die Struktur und Stabilität von V_H wurde häufiger durch den Alaninaustausch beeinflusst als V_L. In früheren Studien wurde von einer Korrelation zwischen der Fähigkeit der variablen Domänen miteinander zu interagieren und der Antigenbindung ausgegangen. Diese Annahme konnte auf Ebene der isolierten Domänen nicht bestätigt werden. Aufgrund

dessen sollte man den Ansatz dieser Studien überdenken. Der dritte variable Bereich in V_H, welcher unter anderem für Antigenerkennung verantwortlich ist (*complementarity determining region;* CDR-H3), weist die höchste Diversität auf. Durch eine Punktmutation, welche eine zusätzliche positive Ladung in CDR-H3 einbrachten, konnte die Stabilität von V_H verbessert werden, ohne die anderen Eigenschaften zu beeinträchtigen. Kombinationen dieser Mutationen führten hingegen zu einem gegenteiligen Effekt.

Diese Arbeit ermöglichte neue Einblicke in das Faltungsverhalten einzelner Antikörperdomänen. Zudem konnte das Verständnis von Wechselwirkungen zwischen V_H und V_L erweitert werden. Die Beobachtungen können dabei helfen, die Optimierung von Antikörpern effizienter zu gestalten. Es sollte zudem überdacht werden, ob Aussagen über die Interaktion zwischen den variablen Domänen wirklich zutreffend sind, wenn diese nur indirekt über Antigenbindung bestimmt wurden. Es muss aber noch gezeigt werden, ob die Beobachtungen für MAK33 auch für andere Antikörper gelten. Zudem besteht die Möglichkeit, dass die Effekte, welche bei den isolierten Domänen beobachtet wurden, in komplexeren Gefügen kompensiert werden können.

Chapter 1

1. Introduction

1.1 Proteins in general

Proteins are multifunctional molecules and abundant in all forms of life. There are key players in nearly all biological processes. With an increasing understanding of cellular processes hundreds of thousands of different proteins could be identified. They play a role as catalysts, in transport and storage of other molecules, and they provide immunity. They can also enable movements, transfer nerve impulses and control growth and differentiation. Although the functions of proteins are manifold, proteins consist of just 20 naturally occurring amino acids. In order to be able to perform all their tasks correctly, proteins must fold into their native structures [1].

The cellular environment is extraordinarily crowded and proteins as well as other macromolecules can be present in concentrations of up to 300-400 mg/ml [2]. These conditions can cause misfolding of proteins [1]. To cope with this situation, mechanisms were developed during evolution to assist correct protein folding. Molecular chaperones play an important role *in vivo* in this process as they can protect unfolded or partially folded intermediates from non-productive interactions and correct mistakes [3, 4]. Chaperones increase the efficiency of folding but not the specificity of the fold itself [3]. In addition, folding catalysts are present, for example protein disulphide isomerases (PDI) and peptidyl-prolyl *cis trans* isomerases (PPIase) [5] which promote the formation of the correct structure as well. For the case that a protein cannot be reconstituted into its native structure it will be removed before it might cause harm to the cell by proteasome-dependent degradation pathway or by autophagy [6, 7].

During the past few years, considerable progress has been made in our quest for a better understanding of the relationship between amino acid sequence and structure. For this knowledge of the structure of proteins is essential. Nuclear magnetic resonance (NMR) and X-ray crystallography enable protein structure determination in great detail [8-11]. Although proteins vary in their amino acid sequences, many homologous structures have been identified and categorised into protein families [12]. Large proteins are often organised in domain structures. Therefore many different proteins can be ensembled by a smaller set of folding motifs (i.e. domains). Protein structural domains are contiguous stretches of the polypeptide chain that fold independently into compact globular structures and comprise the building blocks of a protein's overall structure. The ordered series of these domains is called a protein's multi-domain architecture [13, 14].

Understanding the relationship between protein structure and sequence would permit to predict the three dimensional structure adapted by the primary amino acid sequence or the *de novo* design of new proteins. As many human diseases have been discovered which are caused by protein misfolding, the topic of protein folding mechanisms has recently gained more and more importance [15]. Additionally, the advent of protein engineering and the ability to produce any protein, often in an insoluble and inactive form, granted the topic of folding great practical importance [16]. Knowing the mechanisms enables the development of new therapeutics [17]. The first model systems to study folding *in vitro* were small proteins which did not need assistance of chaperones for adopting their native structure [18, 19] simplifying the analysis. By now many elements of protein structure are largely understood and seem to be related to function [20, 21]. However, it is still not yet possible to make reliable predictions of protein structure starting from the amino acid sequence for all proteins [22]. Additionally, a general mechanism of protein folding could not be identified up to date [23]. However, there is an on-going discussion that at least small proteins share a hierarchic folding mechanism [24, 25] including of an ensemble of different states [26].

In the last 50 years of research on the folding pathways of proteins, experimental data and computer simulations complemented one another. Consequently, our understanding of protein folding and structure increased although it is not yet completed. We have learned that proteins fold rapidly due to random thermal motions that cause conformational changes which lead energetically downhill toward the native structure. This principle is captured in funnel-shaped energy landscapes [27]. Additionally, the large set of known structures in the Protein Data Bank (www.pdb.org) played an important role for predicting protein structures. Nowadays the prediction is far more successful than it was thought possible in the early days [27].

1.2 The protein folding problem

The genetic information of proteins contains only the primary structure of a protein, the amino acid sequence. This information is sufficient for the protein to adopt its defined three-dimensional, tertiary structure. Many proteins refold spontaneously into their native structure what was first shown by Anfinsen and co-workers in 1961 [28, 29]. It was also Anfinsen who distilled in 1973 the two main considerations: I) How is secondary and tertiary structure of a protein encoded within the primary sequence and II) what is the mechanism by which the polypeptide chain is able to reversibly resemble into its three dimensional native structure [30].

If a linear protein of 100 amino acids would fold in a random manner, it would take more than a billion years to adopt the native structure throughout more than 10^{30} possible conformations [31]. Protein folding occurs in a much faster timescale in the range of milliseconds to minutes [32] depending on the complexity of the molecule. Levinthal proposed in 1968 that proteins cannot fold in a random manner sampling all the

conformational space. Protein folding follows rather kinetic pathways that restrict the conformational search and allow the polypeptide chain to fold rapidly and consistently to the same low energy structure. This became known as “Levinthal`s paradox”. It indicates that the time for a protein to randomly find the most energetically stable structure is not sufficient and therefore proteins must fold by a directed process [30, 33].

Significant progress has been made for the prediction of the structure of small proteins or domains of databases from known structural elements, however it is still not possible to predict structures of all proteins *de novo* and the prediction of folding pathways is even more limited [30]. The nature of the folding landscape was determined to be rugged for many proteins. Besides, there is experimental evidence that related proteins from the same family can fold via apparently different mechanisms [34]. The same can be observed for representatives of the same protein domain as shown for example for antibody domains [35]. On the other hand, the elucidation of folding principles has enabled the *de novo* design of ideal protein structures [36].

1.3 Protein folding vs. misfolding

Understanding the mechanisms underlying protein misfolding and aggregation has become a central issue in biology and medicine. There is evidence that amyloid aggregate formation has a negative impact in cell function and that it causes the most prevalent human degenerative disorders: Alzheimer`s, Huntington`s and Parkinson`s diseases [37] as well as spongiform encephalopathies such as Creutzfeldt-Jakob disease [38] and amyotrophic lateral sclerosis [39]. Less frequent but fatal are light chain amyloidosis and the even more rarely heavy chain amyloidosis [40].

All amyloid diseases are highly heterogeneous but have one thing in common. They are all in general based on a specific soluble protein that can under certain circumstances misfold and aggregate as insoluble amyloids. The latter form deposits predominantly in the extracellular or intracellular space or in both forms [40]. Up to date 27 proteins are characterised which have the propensity to cause human amyloid diseases [41].

Table 1 gives an overview of some amyloid diseases. Besides the main protein the final deposits can contain additional components like other proteins and carbohydrates [42]. Clinically two different forms of amyloidosis are differentiated. In the systemic forms different organs are affected, in local amyloidosis one specific organ (which is in general the organ or tissue where the precursor protein is synthesised) [40].

Table 1: A selection of human amyloid diseases and their precursor proteins*

Disease	Precursor protein	Organs affected	Systemic / Local
Light chain amyloidosis (AL)	Immunoglobulin light chain	Heart, kidney, liver	systemic
Heavy chain amyloidosis (AH)	Immunoglobulin heavy chain	Kidney, liver	systemic
Alzheimer`s disease	Amyloid β-peptide tau	Brain	local
Parkinson`s disease	A-synuclein	Brain	local
Creutzfeldt-Jakob disease	Prion protein	Brain and others	local and systemic
Huntington`s disease	Huntingtin	Brain	local
Amyotrophic lateral sclerosis	Superoxide dismutase (dominant)	Spinal motor neurons and motor cortex	local

*Adopted from [39, 40, 42]

In vitro studies of amyloid formation indicate that the process of amyloid fibril formation is a nucleated self-assembly process. Fibril formation is favoured under destabilising conditions that promote formation or accumulation of non-native folding intermediates or both [40]. Three phases of fibril formation describe the self-polymerization process. The process starts with a lag or nucleation phase. The formation of the amyloid fibril nucleus is slow and starts with the oligomerisation of non-native protein. After a critical concentration is exceeded, the elongation phase starts where either native protein or more non-native conformation is incorporated. At some point a steady-state (or plateau phase) is reached and no further polymerisation will occur [40, 43, 44]. The final amyloid structure was found to be similar for several proteins despite different sequences and structures of the precursors [45, 46].

The fibrillar structures are long and unbranched, and typically exhibit a twisted shape a few nm in diameter. They share a characteristic cross-β X-ray fibre diffraction pattern and a core structure composed of β-sheets with strands perpendicular to the fibril axis [42]. *In vitro* experiments of purified aggregation-prone proteins resulted in the same fibril characteristics. Therefore self-assembly is possible without the need of other components [42]. There is an on-going discussion about the toxicity of fibrils. It is now considered that some intermediates formed on the way from the disease protein to the plaques are probably the toxic species and that the deposits might even have a protective function [39]. Figure 1 shows a model of fibril formation and structure.

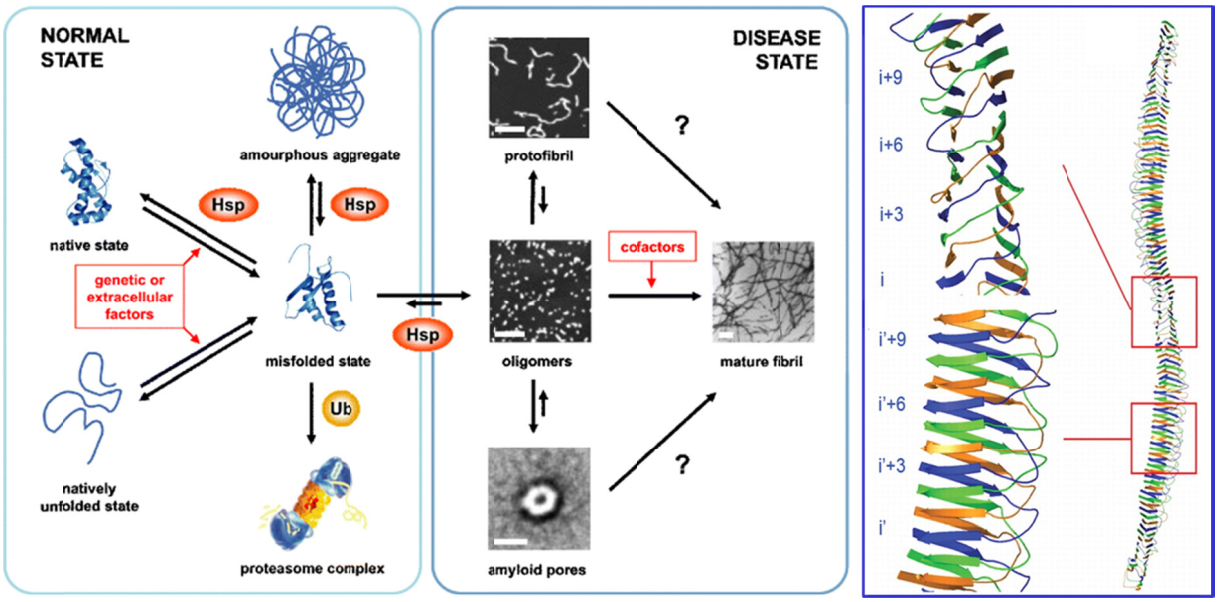

Figure 1: Model for fibril formation and structure

Left side: In the normal state quality control mechanisms are active that avoid misfolding of proteins. Middle: In the disease state misfolded proteins are able to form oligomers, protofibrils and amyloid aggregates or mature fibrils; adopted from [47]. Right side: Model for the fibrillar structure of a human islet polypeptide [48].

Why do some proteins misfold? In different diseases several mechanisms were observed. Here, just three will be emphasised. In prion diseases, for example, conformational changes are propagated that promote amyloid aggregation. As neither overproduction nor degradation could be observed, these new conformations have to serve as a molecular template and influence the native state to change their conformations, too [49]. In Alzheimer`s disease a failure or dysregulation of proteolytic processes of unfolded or misfolded proteins is the reason for the accumulation of aggregation-prone proteins. In this case also no overproduction of the precursor is observed but often supporting mutations, which can either destabilise the precursor protein or influence protein degradation components [50]. A third example is light chain amyloidosis. Here, the overproduction of the amyloid protein results in high levels of circulating protein that can overcome extracellular protein quality control and degradation systems [43, 51].

The increasing knowledge of the mechanisms of these diseases enables to determine sequences with high amyloidogenic potential [52, 53]. For both directions of design – on the one hand induction of fibrillisation and on the other hand preventing fibril formation- successful examples are published. Induction was shown for designed peptides with the ability to fibrilise [54] and the induction of fibrilogenic propensity into non-fibrilogenic proteins by introduction of small sequences [55]. Amyloidogenisity could be reduced by either substitution of a few residues [55] or insertion of small loop regions [56]. The ability to change the propensity of amyloid and non-amyloid forming proteins indicates that the formation of amyloid structures might be a common part of the folding pathway with most proteins evolved not to follow the pathway of amyloids [6]. Drug development is the next challenge after we gained a good understanding –although by no means complete- in fibril

formation. In some cases amyloid formation could already be inhibited or at least reduced [44, 57, 58].

1.4 Driving forces of protein folding

Protein folding is influenced by two different kinds of interactions. First, so called weak interactions and second, auxiliary factors such as electrostatic interactions.

1.4.1 Weak interactions in protein folding

Weak interactions include hydrophobic free energy, van der Waals interactions, peptide hydrogen bonds and peptide solvation. The term “weak interactions” is misleading as the single contributions are indeed small but there are so many interactions within a protein or peptide that they are considered to be the main force of protein folding [59].

Van der Waals attractions results from interactions amongst fixed or induced dipoles and hydrogen bonds are formed when a hydrogen atom is shared between generally two electronegative dipoles. Peptide solvation describes the interaction between peptides and water and it has an electrostatic character. Hydrogen bonds and peptide solvation depend on each other. Peptide hydrogen bonds and peptide solvation are closely connected because hydrogen bonds between water and the peptide N-H and C=O groups must be broken before peptide hydrogen bonds can be formed [60]. In the past they were discussed to be the driving forces of folding but they cannot fulfil the requirements to be so. The dominant driving force must be able to explain why the folded state is favoured compared to the denatured state [61]. In the case of hydrogen bonds this criterion is not fulfilled as intra-chain hydrogen bonds in the folded state have the same energy compared to those in the unfolded chain to water [61]. Nevertheless, hydrogen bonds play an important role in protein folding and stability (for example for secondary structure formation) as well as van der Waals interactions.

For a few years the general agreement has been that hydrophobic free energy is the major driving force of protein folding, although a dispute over its proper definition earlier made this issue controversial [60, 62]. Hydrophobic free energy is described as the transfer free energy which occurs when a hydrocarbon solute is transferred from water to a non-aqueous solvent, or when a nonpolar side chain is buried in its hydrophobic core (i.e., the burial of non-polar surface area) [62]. The denatured protein has many options to interact with water, but the hydrophobic side chains prefer to be in a nonpolar environment. Hence, burying these residues into the core of the protein is favourable. In addition, the entropy of surrounding water molecules increases. Walter Kauzmann already predicted in 1959 that the hydrophobic free energy is the main factor in protein folding. Back then, he termed them hydrophobic bonds [63]. Recent studies and models confirm the dominant role of the hydrophobic effect for protein folding [64].

Folding is not only an exclusively favourable enthalpic effect. During the folding process the entropy from conformational movements of the backbone is decreased. As both effects are in a comparable dimension of app. 200 kJ mol^{-1} they can mostly cancel each other out. In the end the majority of proteins are marginally stable in a range of app. 50 kJ mol^{-1} [61, 65]. The low stability of proteins is discussed to be crucial for protein`s turnover and function [66].

The discussion about the driving forces of protein folding can be summarised as the following: hydrophobic interactions are the driving force for protein folding, but there are many other auxiliary interactions that stabilise the native state of a protein [59].

1.4.2 Auxiliary interactions

Besides the influence of the van der Waals interactions and hydrogen bonds (see 1.4.1) additional factors are influencing protein stability. Electrostatic interactions between ionised/charged side-chains play an important role in folding and stabilisation, as well as α-helix and β-sheet formation [60, 61]. Secondary structure formation depends strongly on hydrogen bond formation. Specific pairwise interactions between side chains are not only a result of hydrogen bonds; salt bridges and cation-pi interactions influence pairing as well [60]. Furthermore, disulphide bonds are covalent bonds between two cysteine residues which influence folding by restricting entropy [61].

In the resolved structures of proteins nonpolar amino acid side chains are predominantly directed into the interior of the protein whereas polar and charged amino acids are exposed to the surface. Some polar groups are located in the core of the protein and can be stabilised by hydrogen bonds and support secondary structure formation. These residues can interact with the solvent upon exposure. Hence, hydrophobic interactions exist in general in the protein core, hydrogen bonds in the core and at the surface and electrostatic interactions predominantly on the surface.

Coulombic interactions and salt bridges are long-range, whereas hydrophobic and van der Waals interactions are short-range interactions concerning the distance of the interacting side chains. Charge-charge interactions are able to influence folding in an early stage of the process [67] what makes them interesting to study. Acidic (aspartic acid, and glutamic acid) and basic amino acids (lysine, arginine, and histidine) can be partly or fully exposed to the surface and can react according to Coulomb`s law. Whether the charge-charge interaction energy becomes unfavourable (means large and positive) depends on the pH value. A low as well as a high pH have a disadvantageous influence and disfavour these interactions [60]. Coulombic interactions can increase in some cases the stability of proteins e.g. *via* introduction of salt bridges as shown by mutational experiments [68, 69]. On the other hand the stability is in some cases just little affected or not affected at all. This case was observed for example in studies on barnase [70] or T4 lysozyme [71]. Removal of a native salt bridge or hydrogen bonds within a protein in general decreases stability [61].

1.4.3 The activation barrier

An activation barrier separates the native and the unfolded state from each other. Whenever a protein wants to change between both states it has to overcome this barrier. The barrier is influenced by entropic and enthalpic contributions. The activation barrier can also be described as Gibbs free energy of activation ($\Delta G^{\neq} = \Delta H^{\neq} - T\Delta S^{\neq}$; $\Delta H^{\neq}$ is the enthalpy and $\Delta S^{\neq}$ the entropy term) [72]. Nowadays examples are known where the activation barrier is so small that folding is possible without any barriers. These proteins follow the downhill folding principle and fold in general ultrafast with folding rates in the range of 10^6–10^4 s^{-1} [73].

The increasing order of the protein backbone decreases the entropy as chain configurations are restricted. Enthalpic contributions have often been attributed to desolvation effects [74]. The activation barrier and folding rates seems to be connected to each other. Fast folding proteins can form few local contacts early in the folding process and are influenced by their native state topology [75]. Plaxco and co-workers introduced a relative contact order parameter to be able to determine the average distance between residues that interact with each other within a sequence. Their findings are that mainly long-range interactions result in high contact order and slow folding rates, whereas more local interactions lead to a low contact order and fast folding rates [75].

1.5 Protein folding mechanisms

As protein folding is a complex process in which several interactions and forces participate, it is an on-going search to find a simple model to explain all folding processes observed in experimental studies.

1.5.1 "Two-dimensional" models of protein folding

The first model proposed was the "nucleation-growth" mechanism. This model assumes that a folding nucleus consisting of secondary structure is slowly formed and afterwards the processes accelerates and the remaining structure is rapidly formed [76]. A drawback of this model is that folding intermediates cannot be explained. Therefore this model is not suitable for all proteins.

Subsequently, hierarchic models were introduced. Here, folding starts from the primary amino acid sequence and the degree of complexity increases during the process from secondary to tertiary structure formation [30]. The "framework" model is an alternative to the "nucleation growth" model as protein folding starts with the formation of an intermediate which is rich in secondary structure. In the next step the tertiary structure is formed [77]. Related but still different to the "framework" is the "diffusion-collision" model. Here, also secondary structure is accumulated before tertiary interactions but the secondary structure is unstable and just transiently formed. The collision between these "micro-domains" provokes folding [78, 79]. This model can describe very fast folding proteins where

intermediates are within the dead time of the experimental setup as no stable intermediates are formed.

Beside a hierarchic point of view, models were proposed that differ from the previous mechanisms. In the "hydrophobic collapse" model, collapsed hydrophobic regions and tertiary structure are formed. The formation is favourable as the compaction of the polypeptide chain increases the efficiency of conformational search for native-like interactions [80]. This model can describe the formation of molten globule (or collapsed) states [81, 82]. Some of these states can be formed very fast even in a range within submilliseconds [83]. A more recent model is the "nucleation-condensation" mechanism. This model is efficient for folding of small proteins that consist of a single domain and combines some features of the models described so far [84, 85]. As a first step a nucleus is formed. From this state secondary and tertiary structure formation are propagated in parallel. Thus, inherently unstable and short-range interactions from secondary structure formation can be stabilised by native-like tertiary interactions within the folding nucleus, which act as long-range interactions. This model fits well for small two-state folding proteins [86]. Figure 2 shows an overview of three of these folding models.

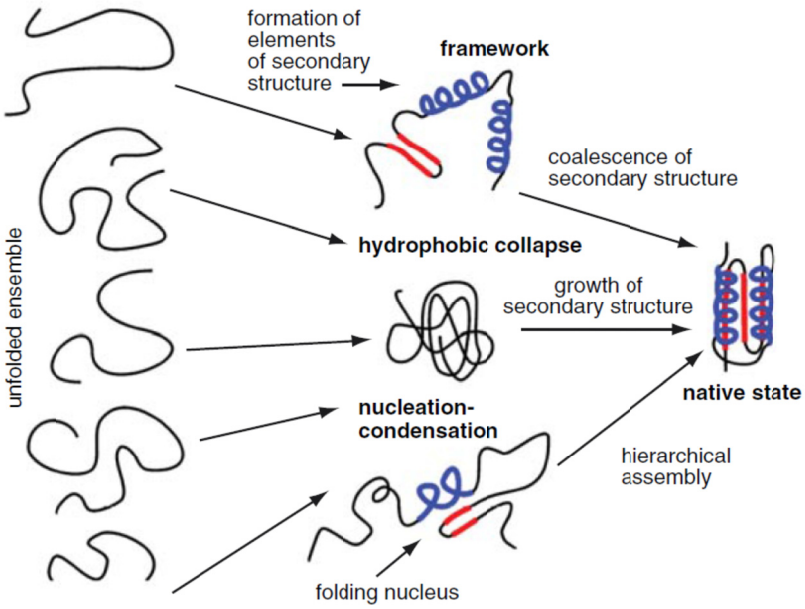

Figure 2: Different two-dimensional mechanisms of protein folding
Key differences in the assembly of the native state are shown between three selected folding models: framework, hydrophobic collapse and nucleation-condensation model. The entropy decreases with the increase of structural complexity (from left to right). Figure taken from [30].

1.5.2 The protein folding landscape

None of the two-dimensional models can cope with the complex behaviour of some proteins. Hence, an alternative view on protein folding mechanisms was developed. In a new approach, folding is considered as a three-dimensional energy landscape. This concept was developed as a statistical approach and includes the energetics of protein conformation. It accounts for diverse folding pathways and the observed unique behaviour of some proteins

[87] in contrast to the previously discussed models. Nevertheless, this approach has some limitations, too.

Proteins are predicted to fold downwards in a folding funnel thereby reaching their energetic minimum and consequently the most stable state (i.e., in general the native state). The funnel surface is best described as a complex rugged structure as there are many possible routes on the way from the denatured to the native state [30]. Figure 3 shows an overview of the development of protein folding models from Levinthal to rugged folding funnels.

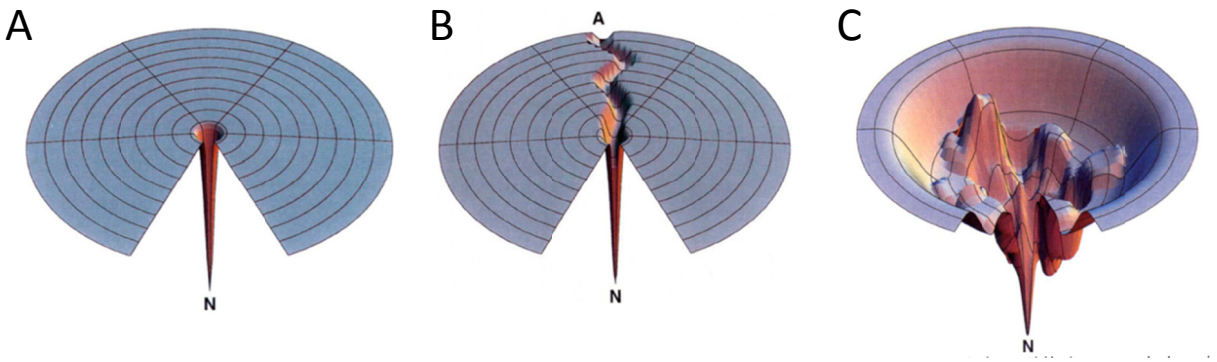

Figure 3: Development of protein folding models
From left to right the different point of views on protein folding are shown: A) Levinthal`s "golf-course" landscape with N as the native state; the search for N is random B) "pathway" solution of folding; starting from A (denatured state) is the native state N the result from a direct conformational search what increases the speed of folding C) In the new view the energy landscape is rugged and contains kinetic traps, energy barriers, and some narrow paths to the native state; folding can be a multi-state process. Adapted from [88].

Within this funnelled and rugged landscape fast-folding pathways can be present to avoid the formation of trapped-species or bottlenecks [30]. The funnelled shape suggests that the folding process is robust if mutations appear as long as the stability is maintained [89]. Single point mutations have confirmed that they can influence the profile of the folding landscape and in some cases they even changed the folding route [89-91]. Nevertheless, the native state will be adopted, as long as the overall stability is not influenced [89]. One challenge of the folding model is the denatured state – the origin of folding. In the first attempts to analyse protein folding it was assumed that the denatured state is completely unstructured and will explore the conformational space in a random manner [92]. More recent results indicate that the denatured state is not completely unstructured [93, 94] and contains residual, fluctuating structure what can be analysed by combining NMR and computational modelling [95].

1.6 Structural states of proteins

Protein folding studies start with denatured proteins independent of the complexity of the folding mechanism. As mentioned in 1.5.2 the denatured state is more complex than a true

random-coil structure. Different methods can be used to disrupt the noncovalent interactions that stabilise the native state and to denature a protein. Frequently, chemical denaturants such as chaotropic agents like urea or guanidinium chloride (GdmCl) are used [96]. Thermal, pressure or pH dependent denaturation is possible as well [97]. In folding studies mostly the native state is considered, but this is only half the truth. To be able to determine the mechanism properly one has to consider besides the native state all other structures and stabilities adopted during folding: the denatured state, intermediate(s), the transition state and native state [98].

1.6.1 The native state of a protein

The native state was always seen as the most stable and best defined state. It is possible to gain detailed structural information about the native structure of a protein using X-ray crystallography or NMR-spectroscopy. Structures can be resolved with resolution values as low as app. 1 Å what might confirm the idea of the native state as one specific conformation. However, new methods are beginning to prove what has already been assumed for several years [99, 100]: the native state is dynamic and flexibility is often crucial for the functionality of the protein [101, 102]. Additionally, a protein class was discovered that shows hardly any defined tertiary structure called “intrinsically disordered” proteins. Although this protein class was not paid attention to for years, it was recently assumed to make up about 40% of the human proteome and to play an important role in signalling and regulatory processes [103, 104]. Many representatives of this class can undergo conformational changes upon ligand binding and increase their structural order [105].

1.6.2 The denatured state(s) of a protein

The terms “denatured state” and “unfolded state” are often used as synonyms. If one wants to be precise, both terms specify not exactly the same. The idea behind “denatured state” was an operational definition. It should describe a noncovalent, cooperative and reversible major change from the original native structure [106]. The “unfolded state” on the contrary describes a specific subset of denatured states; this means highly open and solvent-exposed conformations with little or no residual structure (usually obtained under strongly denaturing conditions) [28]. The unfolded state implies the state of a protein most similar to a random coil as possible. In an ideal random coil the rotation angles of all backbone bonds and side chains are independent of the angles of distant bonds and therefore all conformations share a comparable free energy [107]. Limitations are steric repulsion between atoms close in the covalent structure and the excluded volume effect (i.e., spatial overlap between atoms distant in the covalent structure) [107].

The denatured state of a protein is a less considered object for study than the native state. On the one hand, the denatured state possess less structure than the native state or stable intermediates, on the other hand this state is also more challenging to study [94]. However,

the importance of this topic was recognised and explored. After all, at least three important phenomena are influenced by the denatured state: I) Protein folding and stability, II) Transport across membranes and III) Proteolysis and protein turnover [106].

In some publications the denatured state is also called “denatured state ensemble” to demonstrate that it is not one specific conformation [108]. There is evidence that every molecule of a population in the unfolded state of a protein is likely to have a unique conformation at each instant of time [109]. One has to consider that not all methods for denaturation result in the same denatured state, too. Comparison of thermally and chemically denatured ribonuclease A showed that in the first case still residual structure can be detected, whereas after chemical denaturation this is not the case [97]. In addition, the properties of denatured states depend on the solvent conditions [106].

While the exact mechanism of chaotropic agents is still discussed it is clear that urea and GdmCl are more effective to disrupted noncovalent interactions which define native conformations than acidic or thermal unfolding [97]. Chaotropic agents and the degree of denaturation are concentration dependent. In general, denaturants are considered to weaken hydrophobic interactions and to improve the solvation of nonpolar residues in water [106, 110]. Two main models try to explain the denaturing effect in more detail. The transfer model includes a correlation between the effect of denaturants on protein stability and free energies of transfer of the polypeptide chain from water to denaturant solutions [111, 112]. The solubility of most parts of a protein is increased in GdmCl or urea compared to water and hence the unfolded state, which is more solvent-exposed than the native state, is stabilised. The second model is based upon the assumption that denaturants bind weakly to the polypeptide chain. The unfolded state has more binding sites than the native state and thus will be favoured in denaturing solvents [113, 114]. Both models can explain the empirically determined linear dependency between the free energy of unfolding and denaturant concentration within the region of the unfolding transition [115].

Many methods which are commonly used in protein folding analysis cannot handle the flexibility and dynamic character of the denatured state. Additionally, many of these methods are only applicable for soluble samples what is not always the case for denatured proteins [94]. Spectroscopic methods like intrinsic tryptophan fluorescence or circular dichroism spectroscopy cannot resolve these structures. In contrast, small-angle X-ray scattering is a suitable method to gain at least quantitative information on global features such as size and shape, whereas high-resolution NMR can be applied to detect and characterise local structure. One drawback of NMR-spectroscopy is that the state must be monomeric and soluble [94] and the data is often difficult to interpret. However, these studies could show that besides random coil characteristics weak local interactions and some tendencies to form α-helical structures exist [109].

1.6.3 Transition state and intermediates

1.6.3.1 Transition state and folding intermediates

The transition states share the same primary amino acid sequence as the native and denatured state because during folding no covalent bonds are formed or broken [24] (exceptions are covalent disulphide bonds). Only indirect observation of the transition state is possible, as it appears only in low amounts. The analysis of the rates of folding and unfolding in the presence of various concentrations of denaturant can be used to determine the apparent free energies and therefore enables conclusions on the transition state of the observed reaction [109]. The free energy of the transition state depends linearly on temperature or denaturant concentration although some minor deviations are possible. Thus, the nature of the transition state is not varying substantially [32, 116]. Studies of the transition state have shown that its nature is very similar to the native state but some parts of the protein may be distorted [85, 117].

In folding reactions, the transition state is not exactly the same as in ordinary chemical processes. In a folding reaction, an ensemble of different species can appear. In addition, the rate of product formation in chemical reactions is dependent upon the frequency of vibration of a critical bond what is not true for folding [24]. Nevertheless, the transition state theory is an approximation which is valid for protein folding [24].

Hence, a model for the role of transition states can be determined from ordinary chemical reactions as a basis (see Figure 4). The model assumes that under refolding conditions the unfolded state U can rapidly fold in parts or undergoes compaction. This rapidly appearing intermediate can be a denatured state or an early intermediate depending on the structural amount formed (D). If other intermediates appear they will exhibit energy between D and N (i.e., native state). For the case that the free-energy barrier between D and $I^{\neq}$ (i.e., transition state) is large enough, D and $I^{\neq}$ equilibrate before N is formed. This assumption leads to the conclusion that the formation of N is proportional to $[I^{\neq}]$ and that folding kinetics behaviour is exponential.

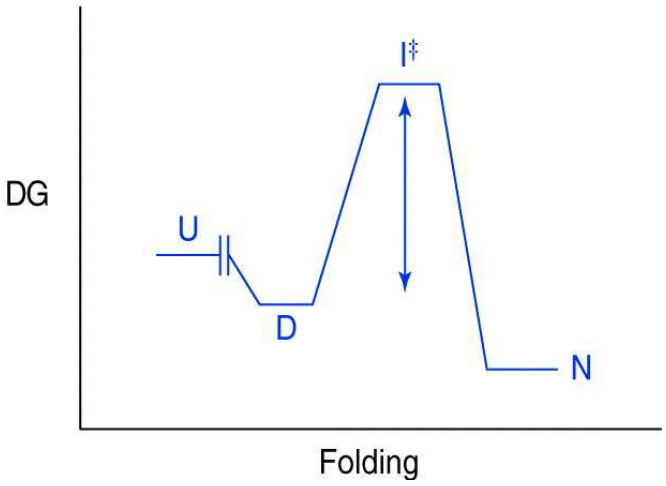

Figure 4: Free-energy diagram of folding modelled on an ordinary chemical reaction

The free energy of folding is plotted against the reaction coordinate. U is the unfolded state, D the denatured state or early intermediate, $I^{\neq}$ is the transition state and N the native state; figure taken from [24].

Folding intermediates have different characteristics. They can often be populated for a significant amount of time [118-120] long enough to be studied. Early folding intermediates can adopt some secondary and tertiary structures and have free energies comparable to or even more stable than the unfolded state [97]. Early intermediates are very rapidly formed (within first few milliseconds) and often magnitudes faster than the subsequent processes, thus a burst phase can be observed. The amplitude of the burst phase and the population of intermediates are directly proportional [97]. The structure and properties of these intermediates appear rather common and could be observed for several examples like α-lactalbumin [121], hen lysozyme [121] and dihydropholate reductase [122]. Nevertheless, at least two different types of early intermediates differing in their responses seem to exist. In class I the dependence of the burst-phase amplitude for the far-UV ellipticity signal is in a sigmoidal manner dependent on the final denaturant concentration. In class II the decrease in the burst phase signal is similar to a saturation curve behaviour as the final denaturant concentration is increased [97].

1.6.3.2 Intermediates in extreme solvent conditions

Extreme solvent conditions can be useful to study intermediates. Under conditions in which the native state is not stable, residual structures can be present long enough to be studied in detail. Low pH or high temperatures were frequently used and lead to the detection of intermediates which represent minima on the conformational energy hypersurface including the molten globule state, the alternatively folded state (AFS), and amyloid structures (see 1.3).

One of the best described intermediates is the molten globule state. The term “molten globule state” [123, 124] was introduced to describe partially folded proteins that showed a significant amount of secondary structure (often already native-like), but only very few nonnative or native-like tertiary interactions. Since the secondary structure elements are only weakly packed against each other (i.e., the tertiary interactions are *molten*), the hydrodynamic radius of the protein is usually 10-20% larger than in the native state [123, 125] but smaller compared to the unfolded state (i.e., still a compact state and thus *globule*). The stability of a molten globule is often only marginal, and its uncooperative unfolding reflects its structural characteristics [124]. Under appropriate solvent conditions (often acidic pH), this state could be populated for a large variety of proteins at equilibrium and, hence, conveniently studied [123, 124]. This was of particular interest as its structural similarity to intermediates formed on many protein folding pathways could be shown [126-128]. Thus, analysis of the equilibrium molten globule state allowed the intrinsic problem of studying transiently populated kinetic species to be circumvented and helped to approach the question about how proteins fold to their native state (admittedly its significance has been questioned [129]). In general, two different types of molten globules are distinguished. Wet molten globules feature a hydrated core and considerably decreased packing compared to the native state [130]. Otherwise, dry molten globules are slightly expanded forms of the native state and show increased conformational flexibility, native-like main-chain hydrogen

bonding and dry interiors [130]. More recently, a distinction between the molten globule state and the pre-molten globule state has been made, where the latter does not display a globular structure but contains considerable secondary structure and a compacted polypeptide chain, but both to a lesser extent than is typically associated with a molten globule [126, 131].

In many respects similar to but nevertheless distinct from the molten globule state is the so-called alternatively folded state (AFS) formed at low pH for some proteins [132]. Similar to the molten globule state, is its richness in secondary structure, but often, as its name implies, the structure adopted is clearly distinct from the one found in the native state or in folding intermediates. Furthermore, the AFS often possesses a defined quaternary structure and, strikingly, it usually shows a rather high stability and cooperative unfolding transitions [132-134].

A species in between the molten globule and the AFS structure is the so-called “acid state” or A-state. It has been reported for a large number of proteins and is thought to form under acidic conditions for the case that the charge repulsion of the protonated side chains is compensated for by the presence of anions [135-137]. It differs from the AFS as its structure is often native like but only marginally stable [138, 139] and it does not adopt the globular shape that is a prerequisite for molten globule structures [123].

An IgG antibody was one of the first proteins shown to form an AFS [132]. Not only the whole antibody also its antigen binding fragment (Fab), adopt a stable conformation at low pH and low ionic strength, clearly distinct from the native state [134]. Importantly, this structure is stabilised by quaternary interactions conveyed by the disulphide bridge between the light chain and the Fd fragment present in the Fab fragment [134] (Fd and Fab fragments will be described in the next section). The notion that significant conformational changes have to take place for the formation of the AFS was strengthened by the fact that reduction of intradomain disulphide bridges facilitated its formation [140]. In other monoclonal antibodies an AFS was also observed [141, 142], as well as in smaller isolated domains such as the IgG C_H3 fragment (i.e., the third constant domain of the antibody heavy chain) [133]. Lately, AFS formation of human IgG Fc fragments induced by thermal treatment instead of acidic conditions was described [143].

1.7 Immunoglobulins and their role as model substrates

1.7.1 Immunoglobulins *in vivo*

Immunoglobulins belong to the eponymous immunoglobulin superfamily [144, 145] and are the key players of the adaptive immune system. The immunoglobulin superfamily is a large group of functionally diverse proteins that share the immunoglobulin fold as a common structural feature [144]. Immunoglobulins directly neutralise pathogens and pathogen-derived products and recruit molecular and cellular immune effectors to eradicate infections

and tumour cells [146]. They can exist in a membrane bound form (i.e., cell surface bound B-cell receptors) or as secreted effector molecules (antibodies). In higher vertebrates five different classes, so called isotypes, of immunoglobulins exist: IgA, IgD, IgE, IgG and IgM. The main difference lies within their heavy chains (α, δ, ε, γ, and μ chain) although two different types of light chains exist as well (λ and κ) which can be combined with all types of heavy chains. IgG is the predominant isotype with the longest serum half-life. In humans four IgG subclasses (IgG1-IgG4) are distinguished according to differences in structure, antigens and function [147] of which IgG1 is the most frequently appearing class.

An organism can generate a huge set of antibodies. Although the number of gene loci encoding for single antibody domains is limited, several mechanisms developed to increase the amount of possible immunoglobulin sequences. Somatic hypermutation within the hypervariable regions, combinatorial rearrangement between different gene segments (VDJ gene segments, i.e., variable, diversity and joining), and combinatorial association between different heavy and light chains create a preimmune antibody repertoire of more than 10^{16} different immunoglobulins [147] and enable a response to the various environmental influences. There is a high risk for the formation of antibodies deficient in folding, assembly, transport or interaction with signalling molecules. Mechanisms have evolved to control antibody development and to select functional combinations. Heavy and light chains of antibodies are co-translationally translocated into the endoplasmic reticulum (ER), coupled with an intricate system to examine and test their fitness [148]. Typically, those HCs and LCs that do not find a suitable partner will be retained in the ER and can be degraded (see light chain amyloidosis for exceptions). Thus, B-cells will express functional antigen receptors and upon antigen binding induce differentiation into plasma cells, which produce a tremendous amount of effector antibodies. The actual antigen repertoire is smaller than the preimmune antibody repertoire, within 10^{7}-10^{9} different Ig molecules [148].

Besides assembly of HCs and LCs other posttranslational modifications are performed in the ER. Disulphide bonds are formed and at the conserved motifs glycosylation is induced. Afterwards, the glycosylation process is completed in the Golgi apparatus. The extension of the glycosylation pattern of an immunoglobulin depends on the isotype [149]. A conserved Asn-297 residue in the second constant domain of the heavy chain (C_H2) is the glycosylation site for IgGs (i.e., N-linked glycosylation). A heptasaccharide consisting of N-acetylglucosamine and mannose form the core of the biantennary sugar complexes. Variations of glycosylation between IgG molecules as well as within different sites in the same molecule result in more than 30 different possible glycosylation patterns [147].

Glycans have a strong influence on immunoglobulins in nearly all fields; they are important for structure, binding, maintaining solubility and conformation as well as for subcellular transport, secretion, clearance and effector functions [149]. In Figure 5 the influence of sugar moieties on the structure of the Fc fragment is shown as well as a selection of typical glycosylation patterns in human organisms. For IgG, it was observed that in case the glycosylation site is mutated and glycosylation is no longer possible there is only a reduced

or no binding to the Fcγ receptor (FcγR) [150]. Besides FcγR binding, the antibody-dependent cellular cytotoxicity (ADCC) is influenced by the glycan composition, too. In the ADCC antibody-coated antigens activate effector cells (i.e., natural killer cells or monocytes) to destroy the antibody-coated target by binding the complex to the FcγR. Glycoengineering experiments have shown, that a reduction of the fucose content increases the ADCC activity [151, 152]. Another important defence mechanism, the complement-dependent cytotoxicity, is also glycan dependent as there is no C1q binding in the absence of glycans [147, 153]. Besides immune defense, false glycosylation patterns can cause autoimmune diseases. Reasons for the induced inflammation are abnormal sugars at the N- termini when sialic acid or galactose is absent.

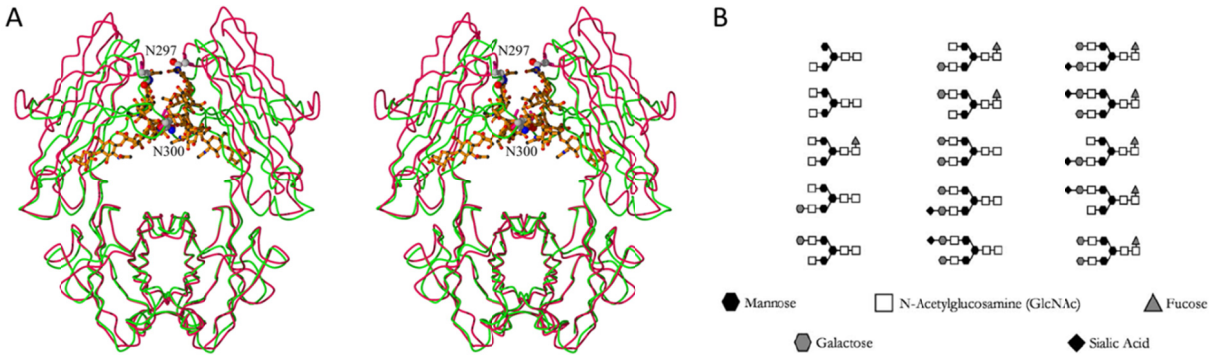

Figure 5: The influence of glycans on the Fc structure

In (A) a comparison between an unglycosylated murine Fc fragment with its glycosylated human counterpart is shown. Glycosylation influences mostly the structure C_H2 domains. The conserved asparagine residues are highlighted and the sugar moiety of the glycosylated Fc is depicted in a ball-and-stick representation; figure taken from [154]. In (B) a selection of typical glycosylation patterns of human organisms is shown. They all have two N-acetylglucosamines and three mannose sugars in common; figure adopted from [155]

The resulting N-acetylglucosamine at the termini can activate the complement cascade when binding to mannose-binding lectin [156]. In addition to the Fc region, the variable domains of antibodies can also be glycosylated thereby modulating antigen binding [149].

1.7.2 Immunoglobulins as model substrates

Antibodies are perfect model systems for studying protein folding, dynamics and association due to the presence of homo- and heterodimeric interactions, conserved proline residues, disulphide bonds and population of intermediate states. The most common antibodies, IgGs, are comprised of two light chains (LCs, consisting of the V_L and C_L domains) and two heavy chains (HCs, with the V_H and C_H1-3 domains). The complete IgG can be subdivided in an Fc fragment (i.e., fragment crystallisable) composed of the homo-dimer of the C_H2-C_H3 domains and two Fab fragments (i.e., fragment antigen binding) composed of the heterodimer of the LC and the V_H and C_H1 domains of the HC (see Figure 6). The Fab, Fc and Fv fragments can be expressed, purified and are stable in isolation. The same holds true for the isolated domains.

Therefore it is possible to reduce the level of complexity subdividing the molecule into fragments or single domains.

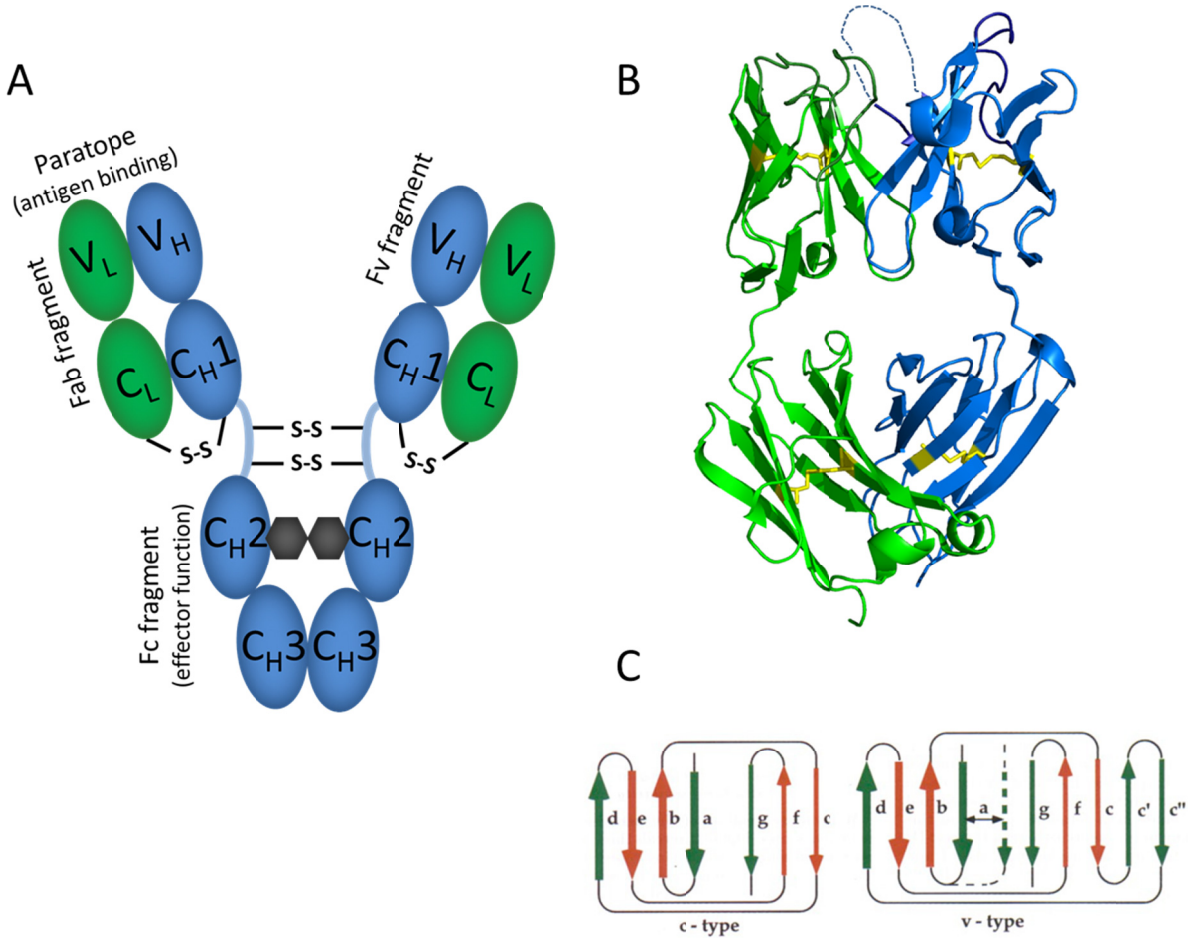

Figure 6: Domain arrangement of an IgG antibody molecule and structure of the single domains
LCs are depicted in green, HCs in blue. (A) Domain organisation of an IgG antibody; S-S indicates a disulphide bridge and the grey hexagons represent oligosaccharides. Functional elements of the antibody are indicated; figure adapted from [35]. (B) Cartoon of the Fab fragment to demonstrate that all domains share the Ig fold and a conserved internal disulphide bridge (yellow sticks); the variable domains have additionally hypervariable regions (long loops) highlighted in dark green and dark blue for V_L and V_H, respectively. The structure is based on PDB ID 1FH5 and CDR-H3 is indicated as a dashed line as it is not resolved in the structure. (C) Two-dimensional diagrams of observed hydrogen bonding patterns; Immunoglobulin constant domains have seven strands in general in c-type topology, variable domains have in total nine strands in general in v-type topology, respectively; figure taken from [157].

The variable domains, V_H and V_L, together constitute the antigen binding site which is crucial for antibody specificity and antigen binding. Within each variable domain, three hypervariable regions (i.e., complementarity determining regions or CDRs) constitute the residues interacting with antigens. The association of the two variable domains is important for the specificity and activity of antibodies [158]. The framework region of the interface is formed by the packing of two β-sheets [159] and this geometry is preserved in Fv fragments consisting of the two variable domains only [160, 161]. Interestingly, the most conserved residues in the framework of the Fv fragment turned out to be those positioned in the interface between the two domains [162, 163]. Computational approaches, such as

covariation analysis [160, 164-167] identified correlated amino acid changes [164]. Covariation analysis has been the most frequently used method and determines if there is a correlation between single amino acids within a sequence by analysing whether the presence (or absence) of a particular amino acid at one position correlates with the presence (or absence) of a particular amino acid at a second position within a multiple sequence alignment [160, 164-167]. The covariation analysis of Wang and co-workers (published 2009) [164] is the largest analysis of variable domain sequences that was performed up to date. In their study, 2432 naturally occurring individual sequences of various species were analysed. With this approach, they aimed to identify naturally occurring amino acid networks that are generally important for antibody structure and function. In Figure 7 the conserved residues within the interface identified by Wang and co-workers are shown. Computational methods alone are not sufficient to determine important residues. Experimental studies addressed the influence of the exchange of particular amino acids on Fv stability [168-172]. While these studies lay the groundwork for deciphering this important aspect of antibody structure, we are still far from a detailed and comprehensive picture of the organization of the Fv interface.

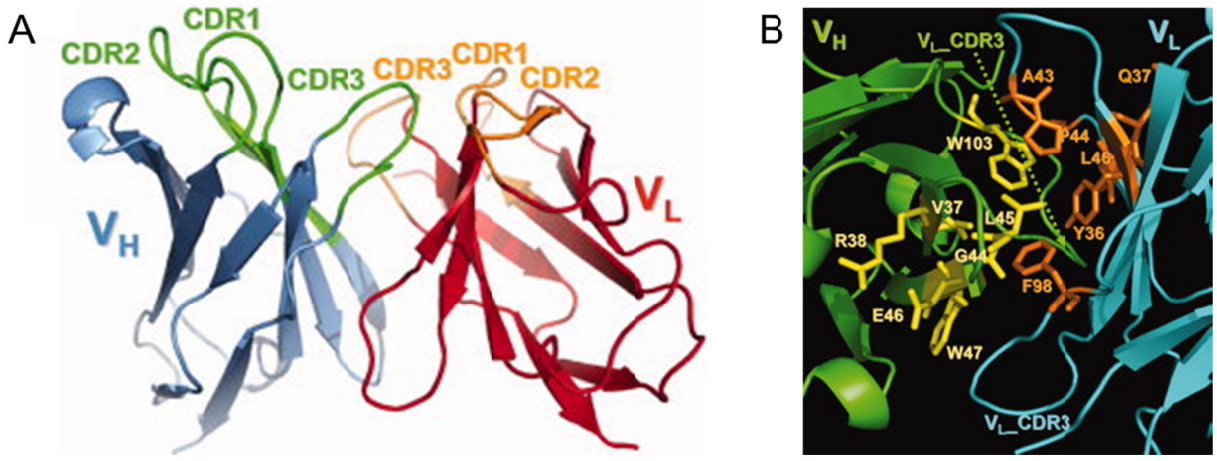

Figure 7: Conserved amino acids of the interface between V_H and V_L

In (A) a ribbon diagram of the Fv region of an antibody is shown. A V_H (blue) and a V_L (red) domain form the Fv fragment. The complementarity determining regions (CDRs) are highlighted in green and orange for V_H and V_L, respectively, and comprise the antigen binding site. In (B) those residues within the interface of V_H (green, residues in yellow) and V_L (blue, residues in orange) are shown as sticks which were determined to be highly conserved. Figure is adopted from [164].

Despite diverse amino acid sequences, the antibody domain structure is highly conserved. This so-called immunoglobulin (Ig) fold, is a β-barrel structure of approximately 110 amino acids in which the two halves of the barrel are linked by a conserved disulphide bond [173] (see Figure 6). Dissecting antibodies into individual domains or fragments allowed the detection of differences in the folding of these structurally highly similar domains. Pioneering studies from Goto and co-workers on immunoglobulin folding focused on denaturing and refolding LCs [174]. The first isolated domains examined have been C_L [175, 176] and C_H3 [177] domains. Their analysis revealed that the domains can fold autonomously

and that peptidyl-prolyl bond *cis-trans* isomerisation plays an important and rate-determining role [176]. This reaction can be catalysed and accelerated by peptidyl-prolyl *cis-trans* isomerases (PPIases) [178, 179]. For the C_H3 domain, it could be shown that prolyl isomerisation has to take place before the association to the homo-dimer [180]. In the folding pathway of the C_L domain, a small helix connecting two strands is the first element that adopts its native structure upon refolding and seems to guide the orientation of the rest of the protein [56]. Despite their structural similarity, recent results showed that the folding pathways of the individual domains can be quite different. One special observation was made for the C_H1 domain, which is in the native antibody the partner of C_L. C_H1 cannot fold autonomously but only in the context of the C_L domain [181].

The V_L domain of the murine monoclonal antibody MAK33 shows a triangular folding pathway in contrast to other antibody domains studied so far with two intermediates in two different pathways [119]. The first intermediate identified is obligatory for folding and shows early molten globule-like characteristics. Two possible pathways were determined from this intermediate to the native state. One pathway is directly leading to the native structure, whereas an alternative pathway includes a more native-like intermediate [119]. It is known that some V_L variants are associated with misfolding diseases and can form pathological fibrillar structures [182, 183]. V_L of MAK33 was not previously known to be linked to fibrillation but it was demonstrated that this domain is able to form fibrils under certain conditions [119]. This finding leads to the assumption that one of these folding pathways could be the junction to fibril formation in agreement with previous observation that amyloid formation starts at partially folded or intermediate states [184-186]. In this fatal protein deposition disease, antibody LCs with mutated V_L domains are produced in large amounts and form fibrillar structures in the organs of patients [182] (see also 1.3). A comparison of amyloidogenic with non-amyloidogenic V_L variants from patients has identified residues that occur with a higher frequency in the former [184, 187-190]. Some of the potential amyloidogenic residues have been identified within the framework region of the β-barrel fold. However, it is not clear how these mutations affect folding and why of all antibody domains specifically the V_L domain is most susceptible to amyloidosis.

1.8 Objective

Within this thesis three objectives are addressed:

I) The alternatively folded state formation of isolated antibody domains.
II) The mechanism of the interaction between V_H and V_L.
III) The folding behaviour of V_H domains.

The murine monoclonal antibody MAK33 which is directed against human creatine kinase was chosen as the main model system. It is a well characterised antibody [56, 119, 132, 140, 191, 192] (see also 1.7.2). It is a representative of the IgG1 κ type and therefore of one of the most frequently appearing antibody types. Another advantage of this antibody is that crystal

structures of its Fab and Fc part as well as of the C_H3 dimer are available in the PDB (entry numbers 1FH5, 3HKF and 1CQK).

It was previously shown that besides intact antibodies, the Fab fragment and the isolated C_H3 domain are also able to adopt an alternatively folded state at acidic pH conditions (see 1.7.2). Within this thesis the view of the AFS formation of isolated domains should be expanded focusing on the Fab domains and C_H2. All immunoglobulin domains share the same folding motif under physiological condition. Thus, it is interesting to see whether the domains behave the same at acidic conditions as well. As differences in the folding pathways from the unfolded to the native state between the isolated domains were observed [35], it is possible to observe variations also at acidic pH. C_H2 is of special interest because this domain has a conserved glycosylation site (see 1.7.1). As glycosylation influences the stability of proteins, the influence of glycans in an acidic environment should be addressed.

Antibodies as therapeutics are a promising and profitable market [193]. Antibody engineering often faces the problem that binding and stability are mutually exclusive. Up to date some conserved residues within the otherwise variable V_H and V_L domains are known. Most results are based on computational studies. Thus, within this thesis the influence of these residues on folding, stability and interaction between both domains should be elucidated experimentally. The role of CDR-H3 was also included as it is the most flexible hypervariable region. This study should help to understand the mechanism of interaction between V_H and V_L and should improve antibody engineering approaches.

Nowadays a lot is known about the folding behaviour of the isolated antibody domains. The folding pathways of C_L, C_H2 and C_H3 could be resolved in great detail [35]. The folding pathway of MAK33 V_L was of special interest and could be elucidated in 2009 [119]. Some V_L domains are associated with diseases (see 1.3) and their ability to misfold might be explained due to their folding mechanism. V_L was shown to be able to form fibrils at certain buffer conditions and shows a triangular mechanism of which one might be the entry point to misfolding disease. The variable domains of LC and of HC have different biophysical properties. Isolated V_L domains are usually not prone to aggregate [194, 195], except for those variants that can form fibrils [196] whereas many V_H domains are not stable in isolation. The comparison of the two pathways will allow to determine differences between V_L and V_H domains and to correlate sequence and structural features with biophysical properties. Besides insight into basic aspects of protein structure, these results may also be useful for diagnostic and therapeutic applications of antibodies [197] and allow insights into the mechanisms leading to amyloid formation when compared to the folding pathway of amyloidogenic V_L domains. V_H domains of different species -for example between mice and humans- can be clustered according to their CDR and framework sequences into different germline families [198]. Besides differences on sequence level also variations in the properties between germline families were observed [199]. Hence, if one wants to determine the folding pathway of V_H domains more than one germline family member should be addressed for a complete study.

Chapter 2

2. Materials and Methods

2.1 Materials

2.1.1 Devices

Autoclave Varioclav EP-Z	H+P
Biacore X100	GE Healthcare
Cell Disruption Apparatus Basic Z	Constant Systems
Centrifuges	
Avanti J25 and J26 XP	Beckman Coulter
Optima XL-A (equipped with FDS)	Beckman Coulter (Aviv)
Optima XL-I	Beckman Coulter
Rotina 46R	Hettich
Rotina 420R	Hettich
Universal 320R	Hettich
Tabletop centrifuge 5418	Eppendorf
Tabletop centrifuge Mikro R200	Hettich
Chromatography systems	
ÄKTA Prime	GE Healthcare
ÄKTA Purifier + REC 112	GE Healthcare
FPLC + REC 112	GE Amersham
Frac-900/950 fraction collectors	GE Healthcare
Superloops (various volumes)	GE Healthcare
Circular dichroism spectropolarimeters	
J710 (with PFD-350S Peltier device)	Jasco
J715 (with PTC 348 WI Peltier device)	Jasco
Thermoblocks	
Digital heat block	VWR
Eppendorf-Thermomixer	Eppendorf
TB1 Thermoblock	Biometra
Fluorescence spectrophotometers	
FluoroMax-2	Jobin Yvon Spex
FluoroMax-4	Horiba Jobin Yvon
Gel documentation system	
Biodoc II	Biometra
ImageQuant 300	GE Healthcare
Image Scanner III	GE Healthcare
Gel electrophoresis and blotting devices	Hoefer
Homogeniser Ultra Turrax DIAX900	Heidolph
HPLC systems	Jasco

Ice maker	Zieger
Incubator	New Brunswick Scientific
Magnetic stirrer	
MR2000	Heidolph
MR3001	Heidolph
MR80	Heidolph
Ultraflex II MALDI ToF/ToF	Bruker Daltonics
Microscopes	
Atomic force microscope	Veeco
Transmission electron microscope JEM100CX	Joel
Axiolab re	Zeiss
DM IL HC	Leica
Membrane vacuum pump	Sartorius
pH meter	WTW
Power amplifiers EPS 3500, 3501 and 1001	GE Healthcare
MARKII (refractometer)	Leica
Scales	
BP 121 S	Sartorius
BL 310	Sartorius
Thermal cycler Primus 25	MWG
SX18-MV stopped-flow apparatus with pdp spectra kinetic monochromator and fluorescence detector	Applied Photophysics
GENios plate reader	Tecan
Thermal cycler MJ Mini 48 well	Biorad
Typhoon 9400	GE Healthcare
Ultra filtration cell 8050	Amicon
UV-Vis spectrophotometers	
Helios γ	Thermo Fisher
UltroSpec 1100 pro	Amersham Biosciences
Nanodrop	Peqlab
Vortex MS2	IKA
Water bath F6-K	Haake
X-ray film processor Optimax TR	MS Laborgeräte

2.1.2 Chemicals

$^{15}NH_4Cl$	Cambridge Isotope Laboratories
^{13}C-glucose	Cambridge Isotope Laboratories
2-Mercaptoethanol	Sigma
5,5’-Dithiobis(2-nitrobenzoic acid) (DTNB)	Sigma
8-Anilino-1-naphthalenesulfonic acid (ANS)	Sigma
ABTS tablets	Roche

Acetic acid	Roth
Acrylamid/Bisacrylamide solution 38:2 (40% w:v)	Serva Electrophoresis
Agar Agar	Serva Electrophoresis
Agarose	Serva Electrophoresis
Ammonium persulphate (APS)	Roth
Ampicillin sodium salt	Roth
Bacto-pepton	BD Biosciences
Bacto-trypton	BD Biosciences
Blocking reagent	Roche
Bromphenol blue	Serva Electrophoresis
Coomassie Blue R	Serva Electrophoresis
Coomassie Brilliant Blue R-250	Serva Electrophoresis
Deoxynucleoside triphosphates (dNTPs)	Roche
Dimethyl sulfoxide (DMSO)	Sigma
Dithiothreitol (DTT)	Roth
EDC	Biacore
EDTA	Merck
EMCH	Thermo scientific
Ethanol	Merck
Ethidium bromide	Sigma
Glucose	Merck
Glutathione, oxidized (GSSG)	Sigma
Glutathione, reduced (GSH)	Sigma
Glycerol	Roth
Guanidinium chloride (GdmCl)	Sigma
HCl 32%	Merck
Imidazole	Sigma
Isopropyl β-d-1-thiogalaktopyranoside (IPTG)	Serva Electrophoresis
Kanamycin sulphate	Roth
KCl	Carl Roth
KH_2PO_4	Merck
L-Arginine	Sigma-Aldrich
LB medium	Serva Electrophoresis
Milk powder	Roth
NaCl	Merck
Na_2HPO_4*2 H_2O	Merck
NaH_2PO_4* H_2O	Merck
NHS	Biacore
Protease inhibitor Mix G, HP	Serva Electrophoresis
Sodium dodecylsulphate (SDS)	Serva Electrophoresis
Stain G	Sigma
TCEP	Pierce
Tetraethylethylendiamin (TEMED)	Roth
Tris	Roth

Triton X-100	Merck
Tween-20	Merck
Urea	Merck

2.1.3 Consumables

Amicon Ultra-15 Centrifugal Filter Units	Millipore
Amicon Ultra-4 Centrifugal Filter Units	Millipore
Biacore Sensor Chip CM5	GE Healthcare
Blotting paper	Whatman
Cuvettes, plastic, 1 mL	Brand
Dialysis membranes Spectra/Por (various MWCOs)	Spectrum Laboratories
Immobilon-P membrane (PVDF)	Roth
Membrane discs	Sartorius
PCR tubes	BioRad Laboratories
PE tubes, 15 and 50 mL	Greiner & Söhne
Petri dishes, PS, 94 mm	Greiner & Söhne
pH indicator	Merck
Reaction tubes, various volumes	Sarstedt
Streptawell, streptavidin-coated microplate	Roche Diagnostics
X-ray film X-OMAT AR	Kodak
Zip Tip	Millipore

2.1.4 Enzymes, Standards and Kits

Antarctic phosphatase	NEB
Cellfectin Reagent II	Invitrogen
ECL plus Western Blotting Detection System	GE Healthcare
Endoglycosidases	NEB
SDS-PAGE Standard Low Range	Biorad
peqGold 1 kb ladder Orange G	Peqlab
Pfu DNA polymerase	Promega
Phusion HF DNA polymerase	NEB
Pro-Q Emerald 488 Glycoprotein Stain Kit	Invitrogen
Restriction enzymes	NEB and Promega
T4 ligase	Promega
Wizard Plus SV Minipreps DNA Purification System Protocol	Promega
Wizard SV Gel and PCR clean-up system	Promega

2.1.5 Biacore kits

Biacore HBS-P+ Buffer 10x	GE Healthcare
Biacore Immobilization Reagents	GE Healthcare
Biacore Maintenance Kit, type 2	GE Healthcare
Biacore Regeneration Scouting Kit	GE Healthcare

2.1.6 Oligonucleotides

QC-PCR	Name	Sequence (5`-3`)
MAK33 V_L	Y36A_f	AACAACCTACACTGGGCGCAACAAAAATCACATG
	Y36A_r	CATGTGATTTTTGTTGCGCCCAGTGTAGGTTGTT
	Q37A_f	CAACCTACACTGGTATGCGCAAAAATCACATGAGTCTCC
	Q37A_r	GGAGACTCATGTGATTTTTGCGCATACCAGTGTAGGTTG
	S43A_f	CAAAAATCACATGAGGCGCCAAGGCTTCTCATC
	S43A_r	GATGAGAAGCCTTGGCGCCTCATGTGATTTTTG
	P44A_f	CAAAAATCACATGAGTCTGCGAGGCTTCTCATCAAATATGCTTCC
	P44A_r	GGAAGCATATTTGATGAGAAGCCTCGCAGACTCATGTGATTTTTG
	L46A_f	CTCCAAGGGCGCTCATCAAATATG
	L46A_r	CATATTTGATGAGCGCCCTTGGAG
	F98A_f	AGCTGGCCTCTCACGGCGGGTGCTGGGACC
	F98A_r	GGTCCCAGCACCCGCCGTGAGAGGCCAGCT
MAK33 V_H	V37A_f	TATTACATGTATTGGGCGCGCCAGACTCCGG
	V37A_r	CCGGAGTCTGGCGCGCCCAATACATGTAATA
	R38A_f	GTATTGGGTTGCGCAGACTCCG
	R38A_r	CGGAGTCTGCGCAACCCAATAC
	R44A_f	CAGACTCCGGAAAAGGCGCTGGAGTGGGTCGC
	R44A_r	GCGACCCACTCCAGCGCCTTTTCCGGAGTCTG
	L45A_f	ACTCCGGAAAAGAGGGCGGAGTGGGTCGCAACC
	L45A_r	GGTTGCGACCCACTCCGCCCTCTTTTCCGGAGT
	E46A_f	CCGGAAAAGAGGCTGGCGTGGGTCGCAACCATTAG
	E46A_r	CTAATGGTTGCGACCCACGCCAGCCTCTTTTCCGG
	W47A_f	AAGAGGCTGGAGGCGGTCGCAACCATTAG
	W47A_r	CTAATGGTTGCGACCGCCTCCAGCCTCTT
	W103A_f	GATGCTATGGACTACGCGGGTCAAGGAACCTC
	W103A_r	GAGGTTCCTTGACCCGCGTAGTCCATAGCATC

	VH_W103Y_f	GCTATGGACTACTATGGTCAAGGAACC
	VH_W103Y_r	GGTTCCTTGACCATAGTACATAGC

CDR / FW	Name	Sequence (5`-3`)
	MAK_MAK FW1 begin (FW)	CATGCCATGGAAGTTCAAGGTGTGGAGTCTGGGGGA
	MAK_MAK FW3 begin/ CDR3 DHU_MAK(RV)	GCATAAAAACCATCACCACCCCATCTTGCACAGTAATACATGGC
	DHU_MAK CDR3 begin (FW)	TGGGGTGGTGATGGTTTTTATGCTATGGATTATTGGGG
	DHU_MAK FW4 end (RV)	CCCAAGCTTGGGTTATTAGGTAACGCTGGTGCCCTG
QC-PCR	MAK VH WT A103E_FW	CTGTGCAAGAGATAAGGAATACTATGGTAACTACGGC
	MAK VH WT A103E_RV	GCCGTAGTTACCATAGTATTCCTTATCTCTTGCACAG
	MAK VH WT A106R_FW	GAGATAAGGACTACTATCGTAACTACGGCGATGC
	MAK VH WT A106R_RV	GCATCGCCGTAGTTACGATAGTAGTCCTTATCTC
	MAK_DHU WT A103E_FW	GTGCACGTGATAAAGAATATTATGGCAACTATGG
	MAK_DHU WT A103E_RV	CCATAGTTGCCATAATATTCTTTATCACGTGCAC
	MAK_DHU WT A106R_FW	CGTGATAAAGCCTATTATCGCAACTATGGTGATGC
	MAK_DHU WT A106R_RV	GCATCACCATAGTTGCGATAATAGGCTTTATCACG

ELISA and Biacore	Name	Sequence (5`-3`)
MAK V_L	VL_ctermFLAG_f	CATGCCATGGATATTGTGCTAACT
	VL_ctermFLAG_r	CCCAAGCTTGGGCTATTACTTGTCATCGTCGTCCTTGTAGTCT CTTTTCAGCTCCAGCTTG
MAK V_H	VH_ctermFLAG_f	CATGCCATGGATATTGTGCTAACT
	VH_ctermFLAG_r	CCCAAGCTTGGGCTATTACTTGTCATCGTCGTCCTTGTAGTCT CTTTTCAGCTCCAGCTTG
MAK V_L	VL_Nterm_Cys_f	CATGCCATGGGTTGCTCTGGTGATATTGTGCTAACTCAG
	VL_Nterm_Cys_r	CTGAGTTAGCACAATATCACCAGAGCAACCCATGGCATG

MAK V_H	VH_Nterm_Cys_f	CATGCCATGGGTTGCTCTGGTGAAGTTCAAGGTGTGGAG
	VH_Nterm_Cys_r	CTCCACACCTTGAACTTCACCAGAGCAACCCATGGCATG
	VL_Nterm_Hind_r	CCCAAGCTTGGGTTATTATCTTTTCAG
	VH_Nterm_Hind_r	CCCAAGCTTGGGTTATTAGGCTGAGGA

C_H2	**Name**	**Sequence (5`-3`)**
pSVL-Fc-TEV-His_6	MAK_Fc_BamHI_for	GCGGGATCCGCGTCATCTGTCTTC
	MAK_CH3-His-linker_rev	GCTGCCACTACCGTGATGGTGATG
	His-linker_for	CATCACCATCACGGTAGTGGCAGC
	TEV-His_NcoI_rev	TATACCATGGTTATTAGTGATGGT
pSVL-C_H2-TEV-His_6	CH2-linker_rev	GCTGCCACTACCTTTGGAGATGGT
	CH2-linker_for	ACCATCTCCAAAGGTAGTGGCAGC
	CH2_TEV_His_rev_NcoI	TATACCATGGTTATTAGTGATGGTGATGGTGATGATCG TAATCGCCCTGAAAATACAGGTTTTCGCTGCCACTACCT TTGGAGATGGTTTTCTCG
	Fc_TEV_His_rev_NcoI	TATACCATGGTTATTAGTGATGGTGATGGTGATGATCG TAATCGCCCTGAAAATACAGGTTTTCGCTGCCACTACCG TGATGGTGATGGTGATG
pAcGP67-Fc-TEV-His_6	MAK_Fc_BamHI_for	GCGGGATCCGCGTCATCTGTCTTC
	MAK_CH3-His-linker_rev	GCTGCCACTACCGTGATGGTGATG
	His-linker_for	CATCACCATCACGGTAGTGGCAGC
	TEV-His_NcoI_rev	TATACCATGGTTATTAGTGATGGT
pAcGP67-C_H2-TEV-His_6	CH2-linker_rev	GCTGCCACTACCTTTGGAGATGGT
	CH2-linker_for	ACCATCTCCAAAGGTAGTGGCAGC
C_H2 WT pET28	CH2_pET_Xba_f	GCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGAT ATACCATGTCATCTGTCTTCATC
	CH2_pET_Hind_r	CCCAAGCTTGGGCTATTAGATGGTTTTCTC

All primers were purchased from Eurofins MWG Operon.

2.1.7 Bacterial strains, insect cells and Plasmids

All constructs for expression in *Escherichia coli* were cloned into pET28-b (Novagen), for *Spodoptera frugiperda* Sf9 insect cell expression into pAcGP67-B (BD Biosciences).

Strain	Genotype	Origin
E. coli BL21 (DE3) codon plus	F– *ompT hsdS*(rB– mB–) *dcm*+ Tetr *gal endA* Hte (*argU proL* Camr)	Stratagene
E. coli BL21 star (DE3)	F-*ompT hsdS*$_B$($r_B^-m_B^-$) *gal dcm rne131* (DE3)	Invitrogen
E.coli Mach1	ΔrecA1398 endA1 tonA Φ80ΔlacM15 ΔlacX74 hsdR (r_k^- m_k^+)	Invitrogen
Sf9		Invitrogen (Cat.no. B825-01)

2.1.8 Media for expression in *E. coli* and *Sf9* insect cells

LB_0	LB medium for plates: Agar Agar	20 g/l 15 g/l
SOB	yeast extract tryptone NaCl KCl MgCl2 MgSO4	0.5% 2% 10 mM 2.5 mM 10 Mm 10 mM
TB solution (for supercompetent cells)	HEPES $CaCl_2$ KCl pH 6.7 with KOH or HCl $MnCl_2$	10 mM 15 mM 250 mM 55 mM
M9	Na_2HPO_4 KH_2PO_4 NaCl pH 7.3 (NaOH) add before use: D-(+)-glucose (^{13}C or ^{12}C) $^{15}NH_4Cl$ $MgSO_4$ $CaCl_2$ Thiamine HCl 100x trace elements	6 g/L 3 g/L 0.5 g/L 0.2 g/mL; 10 mL / L 0.1 g/L; 10 ml/L 1 M; 1 mL/L 1 M; 0.3 mL/L 1 mg/mL; 1 mL/L
Insect-XPRESS™ Protein-free Insect Cell Medium (Lonza Ltd.)	Sf9 medium	n.d.
Antibiotics (1000x stocks)	Ampicillin Kanamycin	100-200 µg/l (in ddH_2O) 35-50 µg/l (in ddH_2O)

100x trace elements contain per 1L: 5 g EDTA, 0.83 g $FeCl_2$*6 H_2O, 84 mg $ZnCl_2$, 10.25 g $CuCl_2$*2 H_2O, 10 mg $CoCl_2$*6H_2O, 10 mg H_3BO_3, and 1.35 mg $MnCl_2$*4H_2O. Solution was sterile filtered before use.

All media except Insect-XPRESS™ Protein-free Insect Cell Medium and TB solution were sterilised using an autoclave at 121°C for 20 min. Antibiotic stocks were passed through a sterile filter (0.22 µm) and stored at -20°C.

2.1.9 Antibodies

Anti-His_6-Peroxidase	Roche
Anti-FLAG M2-peroxidase (HRP)	Sigma-Aldrich

2.1.10 Chromatography materials and columns

Superdex 75 Prep Grade	GE Healthcare
HisTrap Fast Flow	GE Healthcare
Q Sepharose Fast Flow	GE Healthcare
SP Sepharose Fast Flow	GE Healthcare
Concanavalin A column	GE Healthcare

2.1.11 Buffers

2.1.11.1 Protein Purification

Inclusion Body preparation buffer	Tris / HCl pH 7.5	50 mM
	EDTA	5 mM
	NaCl	5 mM
Low salt buffer	Tris/HCl pH 7.8-8.0	25 mM
	EDTA	5 mM
	urea	5 M
High salt buffer	Tris/HCl pH 7.8-8.0	25 mM
	EDTA	5 mM
	Urea	5 M
	NaCl	1 M
IB dissolving buffer	Tris/HCl pH 7.8-8.0	25 mM
	EDTA	5 mM
	urea	8 M
	B-mercaptoethanol	10 mM
Refolding buffer V_L, C_L, C_H2	Tris/HCl pH 8.0	250 mM
	EDTA	5 mM
	L-Arginine	100 mM
	GSH	0.5 mM
	GSSG	1 mM

Refolding buffer C_H1	Tris/HCl pH 8.2	250 mM
	EDTA	5 mM
	L-Arginine	100 mM
	GSH	0.5 mM
	GSSG	1 mM
Refolding buffer V_H	Tris/HCl pH 8.0	250 mM
	EDTA	5 mM
	L-Arginine	400 mM
	GSH	0.5 mM
	GSSG	1 mM
Phosphate-buffered saline (PBS)	NaCl	137 mM
	KCl	2.7 mM
	$Na_2HPO_4 \cdot H_2O$	10 mM
	KH_2PO_4	2.0 mM
	pH 7.4	
C_L Ni-NTA buffer A (load)	Na_2HPO_4	100 mM
	GdmCl	2.5 M
	pH 7.5 (o-phosphoric acid)	
C_L Ni-NTA buffer B (elution)	Na_2HPO_4	100 mM
	GdmCl	2.5 M
	pH 3 (o-phosphoric acid)	
HisTrap FF binding / wash buffer	NaH_2PO_4	50 mM
	NaCl	150 mM
	Imidazol	10 mM
	pH 8.0	
HisTrap FF elution buffer	NaH_2PO_4	50 mM
	NaCl	150 mM
	Imidazol	300 mM
	pH 8.0	
TEV restriction buffer	NaH_2PO_4	50 mM
	NaCl	70 mM
	Imidazol	10 mM
	pH 8.0	
Concanavalin A binding / wash buffer	Tris / HCl pH 7.4	20 mM
	NaCl	0.5 M
	$CaCl_2$	2 mM
Concanavalin A elution buffer	Tris / HCl pH 7.4	20 mM
	NaCl	0.5 M
	D-glucose	0.5 M
Concanavalin A regeneration buffer	Tris / HCl pH 8.5	20 mM
	NaCl	0.5 M
Concanavalin A storage buffer	Acetate buffer	0.1 M
	NaCl	1 M
	$CaCl_2$	1 mM
	$MnCl_2$	1 mM
	$MgCl_2$	1 mM
	Ethanol	20%

2.1.11.2 Buffers for SDS Polyacrylamide Gelelectrophoresis

Fairbanks A (Coomassie staining)	2-Propanol	25% (v/v)
	Acetic acid	10% (v/v)
	Coomassie Blue R	0.05% (w/v)
Fairbanks D (destaining solution)	Acetic acid	10% (v/v)
Laemmli sample buffer (5x)	Tris/HCl, pH 6.8	0.3 M
	SDS	10% (w/v)
	Glycerol	50% (v/v)
	2-Mercaptoethanol	5% (v/v)
	Bromophenol blue	0.05% (w/v)
SDS running buffer (10x)	Tris/HCl, pH 8	0.25 M
	Glycine	2 M
	SDS	1% (w/v)
Separation gel buffer (2x)	Tris/HCl, pH8.8	1.5 M
	SDS	0.8% (w/v)
Stacking gel buffer (4x)	Tris/HCl, pH 6.8	0.25 M
	SDS	0.4% (w/v)

2.1.11.3 Buffers for Western Blotting

Transfer buffer	Tris/HCl, pH 7.5	50 mM
	Glycine	40 mM
	SDS	0.04% (w/v)
	Methanol	20% (v/v)
PBS-T	1x PBS	
	Tween	0.1% (w/v)

2.1.11.4 Buffers for Glycostaining

Fix solution	Methanol	50% (v/v)
	Glacial acetic acid	5% (v/v)
Wash solution	Glacial acetic acid	3% (v/v)

2.1.11.5 Buffers for Molecular Biology

TAE (50x)	Tris/Acetate pH 8.0	2 M
	EDTA	50 mM
Loading buffer (10x)	Glycerol	50% (v/v)
	Tris/HCl pH 8.0	10 mM
	Orange G	0.2% (w/v)
	Xylencyanol	0.2% (w/v)

2.1.11.6 Buffers for Biacore X100

HBS-P+ (10 x)	HEPES	0.1 M
	NaCl	1.5 M
	Surfactant P20	1.5% (v/v)
	pH 7.4	

2.1.11.7 Buffers for ELISA

Reaction Mix I	human creatine kinase (biotinylated)	25 µg pre-dissolved in 2 ml H_2O
	blocking reagent in H_2O	5%
Reaction Mix II	Na_2HPO_4 / NaH_2PO_4 pH 7.5	100 mM
	NaCl	50 mM
	EDTA	0.5 mM
	Tween-20	0.1%
	Blocking Reagent	0.5 g/50 ml
Reaction Mix III	ABTS-Pufferin	0.835 g/50 ml
	APTS	1 tablet

Blocking reagent for ELISA, ABTS Tablets, ABTS Buffer Powder, as well as the Streptavidine-coated microwell plates were purchased from Roche Diagnostics.

2.1.11.8 Buffers for AFS formation

pH 2.0, NaCl	NaH_2PO_4	50 mM
	NaCl	100 mM
	pH 2.0 (o-phosphoric acid)	
pH 7.5, NaCl	Na_2HPO_4	50 mM
	NaCl	100 mM
	pH 7.5 (o-phosphoric acid)	

2.2 Software, Databases and Web-based Tools

Software

Adobe Illustrator CS5	Adobe Inc.
Adobe Photoshop CS5	Adobe Inc.
BioEdit	Ibis Biosciences
DynaFit	BioKin Ltd.
EndNote X6	Thomson Reuters
Epson TT8 Launch Silver Fast	Epson
ImageQuant	GE Healthcare

Microsoft Office 2010	Microsoft
MMass V2.4.0	Open Source
OriginPro 8.6	OriginLab Corp.
Pymol V1.3	Schrödinger
SedFit	Peter Schuck

Databases

PDB	www.rcsb.org/pdb
PubMed	www.ncbi.nlm.nih.gov/pubmed
UniProt	www.uniprot.org

Web-based Tools

Denaturant Calculator	http://mcb.berkeley.edu/labs/krantz/tools/gdmcl.html
Kabat numbering	http://www.bioinf.org.uk/abs/
NCBI Blast	http://blast.ncbi.nlm.nih.gov/
NEBcutter V 2.0	http://tools.neb.com/NEBcutter2/
OligoAnalyzer 3.1	http://eu.idtdna.com/analyzer/applications/oligoanalyzer/
ProtParam	http://web.expasy.org/protparam/
Reverse complement	http://www.bioinformatics.org/sms/rev_comp.html

2.3 Molecular Biological Methods

All solutions were autoclaved or sterile filtered. If not otherwise stated all experiments were performed at room temperature. Polymerase chain reaction (PCR), DNA electrophoreses, restriction digestion, cultivation and *E. coli* transformation were conducted according to standard protocols described elsewhere (see Sambrook and Russell, Molecular Cloning, 2001). In the following the most frequently used methods are briefly described.

2.3.1 *E. coli* cultivation

Cultivation on LB plates or in liquid medium of *E. coli* was performed in an incubator at 37°C. Antibiotics addition depended on the plasmid transformed. Liquid cultures were inoculated 1:60 from fresh overnight cultures or with a single colony from plates. Bacterial growth was controlled by Vis spectroscopy at 600 nm. For cryo-stocks 1 ml of fresh overnight culture was mixed with 500 µl undiluted glycerol and snap-frozen. Cultures were stored at -80°C.

2.3.2 Supercompetent *E. coli* cell preparation

10-12 large colonies were transferred from a fresh overnight cultured plate into 250 mL SOB medium and incubated at 19°C with vigorous shaking to an OD_{600} of 0.5 (approx. 24 h). After a cold shock of 10 minutes on ice, cells were spinned down at 4000 rpm for 10 min at 4°C. Cells were gently resuspended in 80 ml ice-cold TB solution and stored on ice for 10 min.

After spinning the cells down, the pellet was resuspended in 20 ml ice-cold TB solution containing 1.4 ml DMSO. Afterwards, cells were aliquoted to 100 µL, frozen in liquid nitrogen and stored at -80 °C.

2.3.3 Agarose gel electrophoresis

The method was used for isolation of PCR products or digested fragments by gel extraction. Depending on fragment size, gels between 0.8 and 1.5% agarose in TAE were prepared. For size determination was a 1 kb DNA ladder (Peqlab) additionally loaded. Gel runs were done with 120 mV voltage. For staining of the DNA was either ethidium bromide or Stain G (Serva Electrophoresis) used and DNA detected at UV light.

2.3.4 PCR

Quick Change mutagenesis or target amplification was done with PCR. Pfu polymerase (Promega) or Phusion Polymerase (NEB) was used depending on the template. Primers were designed by hand and controlled with the OligoAnalyzer. Primers were ordered from Eurofins MWG.

For Pfu Polymerase are approximately two minutes/1000 bp recommended and for the Phusion polymerase 15-30 seconds per kb. Two typical cycler programs for most PCRs are:

Pfu	**cycles**	**T (°C)**	**time (min)**
Initial Denaturation	1	95	0.5
Melting		95	0.5
Annealing	x 29	50-65	0.5
Amplification		72	2-8
	30	72	5-15

Phusion	**cycles**	**T (°C)**	**time (min)**
	1	98	0.5
		98	10 sec
	x 29	50-65	20 sec
		72	1-6
	30	72	4-12

Pipetting schema

1 µL template DNA (50 ng/µL)
1 µL primer_f (10 pmol/µL)
1 µL primer_r (10 pmol/µL)
1 µL dNTPs (10 mM per dNTP)
5 µL 10x Pfu-Polymerase Puffer
1 µL Pfu-Polymerase
40 µL nuclease free ddH_2O
Σ50 µL

Pipetting schema

1 µL template (50 ng/µL)
2.5 µLprimer_f (10 pmol/µL)
2.5 µL primer_r (10 pmol/µL)
1 µL dNTPs (10 mM per dNTP)
10 µL 5x Phusion HF Buffer
1.5 µL DMSO (3%)
0.5 µL Phusion HF DNA-Polymerase
31 µL Nuclease free ddH_2O
Σ 50 µL

2.3.5 Cloning strategies

2.3.5.1 Quick Change PCR

For the insertion of single point mutations forward and reverse primers carrying the target mutation were designed. Wild type or single point mutant inserts within the target plasmid (usually pET28-b) were used as a PCR template. Template DNA was afterward digested with DpnI (NEB) for 1 hour at 37°C. After heat inactivation for 20 minutes at 65°C the PCR product was transformed into supercompetent Mach 1 *E. coli* cells (heat dependent transformation). Cells were plated onto LB plates with appropriate antibiotics and incubated overnight at 37°C. Single colonies were picked, incubated in liquid LB with antibiotic overnight at 37°C and the plasmid DNA isolated with the Wizard Plus SV Miniprep kit. Selected samples were sent for sequencing to Eurofins MWG Operon or GATC to verify the desired mutation using the T7 forward or pET-RP sequencing primer.

2.3.5.2 Overlapping PCR

For introduction of sequences in between a template sequence overlapping PCR strategies were used. For this method three PCR reactions were performed: amplification of insert part 1 with template DNA, amplification of insert part 2 with same template and a third PCR using the overlaps between insert 1 and 2 and forward primer of PCR 1 and reverse primer of PCR 2 to combine both parts. PCR product formation was controlled by agarose gel electrophoresis and the correct band purified out of the gel (Wizard DNA gel purification kit). Insert and vector were digested with the appropriate restriction enzymes for usually 2 hours at 37°C. After heat inactivation was the vector DNA dephosphorylated with Antarctic phosphatase (NEB). Insert and target were again purified and in a 1:1 or 1:2 ratio ligated in a total volume of 10 µl. As a standard ligase the T4 DNA ligase from Promega was used and samples were either incubated overnight at 4°C in a water bath or for 30 minutes at room temperature (Quick Ligation). Half of the reaction was transformed into supercompetent Mach 1 *E. coli* cells and selected on LB plates with antibiotics. Sequences were verified by sequencing. This strategy was successful for Fc-linker-TEV-His_6 cloning and some CDR mutants of V_H.

2.3.5.3 Standard PCR

For cases where new restriction sites or additional sequence information at the 5`or 3`ends had to be introduced standard PCRs were performed. Primers containing the additional sequences were designed and PCR with target template conducted. PCR product control protocol, digest and ligation were same as in 2.3.5.2. Sequences were finally confirmed by sequencing. The strategy was successful for subcloning applications with new restriction enzymes, C_H2ctr and C_H2-linker-TEV-His_6 as well as additional cysteine and FLAG-tagged V_H and V_L mutants.

2.3.5.4 Subcloning strategies

For the transfer of an insert between two plasmids the target was digested with suitable restriction enzymes and ligated (see 2.3.5.2 for details).

2.4 Protein chemical methods

2.4.1 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

General procedure was according to Laemmli [200]. Polyacrylamide gels were cast to final concentrations of 5% (w/v) acrylamide/bisacrylamide 19:1 (40% w/v) for stacking and 15-18% (w/v) acrylamide/bisacrylamide 19:1 (40% w/v) for separation gels. Polymerization of the solution was induced by adding ammonium persulfate (APS, 10% w/v in H_2O) and tetramethylethylendiamin (TEMED). All samples were mixed with 5x Laemmli (+5% β-mercaptoethanol) sample buffer and incubated at 95°C for 5 min before loading. Electrophoresis was carried out at a constant conductant currant of 30 mA/gel for 45-60 minutes. Gels were stained with Coomassie according to a modified protocol of Fairbanks and co-workers [201].

2.4.2 Protein Expression and Purification

2.4.2.1 Proteins refolded from inclusion bodies (E. coli)

Purification of wild type C_L, C_H1, and V_L was already published by Feige and Simpson [119, 191, 202]. Wild type V_H was purified according to Feige et al. [192] with the exception that besides a Q-sepharose column also a SP-sepharose was applied for pre-purification. C_H2 was subloned from the pAkF-T5/CH2 plasmid which was provided by Michael J. W. Thies [203] into pET 28b for expression in *E. coli* BL21 star. Expression and purification was according to V_L.

Proteins were expressed in 5 L flasks with LB medium and antibiotic. Inoculation occurred with a 1:60 dilution of overnight culture and incubated at 37°C. At an OD_{600}= 0.6-0.8 cells were induced with 1 mM IPTG (final concentration) and incubated overnight at 37°C. After cell harvest and disruption in IB preparation buffer with Protease inhibitor Mix G (or HP for His-Tag labelled C_L) and DNase I the lysate was incubated with 1% Triton X-100 at 4°C for at least 2 hours. Inclusion bodies were harvested (20.000 rpm, 20 min, JA 25.50) and washed twice IB preparation buffer.

Inclusion bodies were dissolved in IB dissolving buffer with 10 mM β-mercaptoethanol for 2 hours at room temperature. After centrifugation (20.000 rpm, 25 min, JA 25.50) supernatant was loaded on a Q-sepharose FF (or Q- and SP-sepharose in parallel) column (GE). Flow through, as well as eluate, were collected separately and the fraction containing the target protein refolded in an appropriate refolding buffer overnight at 4°C. For $C_H2$ctr the pH of the

refolding buffer was adjusted at room temperature to pH 7.5 and refolded for 48 hours at 4°C. pH of Tris buffer is temperature dependent, thus after 48 hours at 4°C the pH will change from pH 7.5 to pH 8.0.

After refolding, proteins were concentrated to 10 mL and size exclusion chromatography with High Load Superdex 26/69 75 pg column (GE Healthcare) was performed. Protein purity was confirmed by SDS-PAGE and the correct mass by Matrix-assisted laser desorption ionization time-of-flight (MALDI-ToF) mass spectrometry.

His-tagged labelled C_L was pre-purified under denaturing conditions with a self-packed HisTrap column (Ni-NTA from GE Healthcare). Elution of the target protein was done by pH shift. Afterwards the procedure was the same as for the untagged single domains.

Expression of ^{15}N- and ^{13}C-^{15}N-labelled C_H2 was performed in *E. coli* BL21 star in M9 minimal medium containing $^{15}NH_4Cl$ and glucose or $^{15}NH_4Cl$ and ^{13}C-glucose (Cambridge Isotope Laboratories), respectively. Before inoculation of the main culture, the overnight pre-culture in LB was harvested and washed once in PBS to remove LB traces. Purification protocol is according to unlabelled domains.

2.4.2.2 Glycosylated C_H2 from Sf9 expression

For the expression of glycosylated C_H2 *Sf9* cells were used. C_H2 was subcloned into a pAcGP67-B vector (BD Bioscience) with a C-terminal Gly-Ser-Gly-Ser linker and a TEV restriction site as well as a His_6-tag. For Baculo virus transfection $9.0*10^5$ cells were transfected with 2 µg C_H2-pAcGP67-B plasmid and 0.5 µl linearized DNA (BD Bioscience) with Cellfectin Reagent II (Invitrogen) according to the manufacturer`s protocol.

Sf9 cells were transfected with active virus (1:75 ratio) and incubated for 72 h at 27°C in Insect-XPRESS™ Protein-free Insect Cell Medium with L-glutamine (Lonza Ltd). Protein purification was subsequently performed at 8°C. The medium was dialysed after cell harvest (5.000 rpm, 15 min, JA 10) twice against 50 mM NaH_2PO_4, 150 mM NaCl, 10 mM imidazole, pH 8.0 for each 2 h. Afterwards C_H2-His_6 was bound to a HisTrap Fast Flow column (GE Healthcare) and after washing eluted with 300 mM imidazole in the presence of protease-inhibitor Mix HP (Serva Electrophoresis). Then, the protein was dialysed over night against 50 mM NaH_2PO_4, 70 mM NaCl, 10 mM imidazol, pH 8.0 and the His_6-tag was cleaved with TEV-protease in the presence of 3 mM GSSG and 0.3 mM GSH overnight. Free His-tag and uncleaved protein was removed by Ni-NTA (GE Healthcare). For the separation of glycosylated from nonglycosylated C_H2 a HiTrap Con A 4B column (GE Healthcare) was used according to the manufacturer`s recommendations. Finally, the both variants of C_H2 were purified by gel filtration chromatography with a High Load Superdex 26/69 75 pg (GE Healthcare) in phosphate buffer.

2.4.3 Western Blotting for His-tagged domains

The blotting of proteins from SDS-PAGE gels onto PVDF membranes (BioRad) was carried out using a FastBlots system (Biometra). For blotting, the following components were stacked from bottom to top and placed between the plate electrodes: three layers of blotting paper (Whatman), PVDF membrane, gel and three layers of blotting paper (Whatman). Prior to blotting, the membrane was activated by rinsing with methanol. The activated membrane, the blotting paper and the gel were incubated for 5 min in western blot transfer buffer. The blotting was performed applying 72 mA per gel for 60 min. After blotting, the membrane was incubated in blocking buffer (PBS-T with 5% milk powder) for 60 min or overnight at 4°C. The Anti-His_6-HRP antibody (horse reddish peroxidase coupled) was 1:10.000 diluted in 10 mL PBS-T/blot and 1 hour incubated at room temperature. After three washing steps for each 15 minutes the blot was analysed by chemiluminescent detection using ECL Western Blotting Detection Regents (GE Healthcare) and photographic film (Kodak), which was developed after exposure using an Optimax TR (MS-Laborgeräte).

2.5 Glycobiology

2.5.1 Staining of Glycoproteins

Staining of glycosylated C_H2 was performed with the Pro-Q Emerald 488 Glycoprotein Stain Kit (Invitrogen) according to the manufacturer`s recommendations.

2.5.2 PNGase F and Endo H digest

PNGase F and Endo H (both NEB) are endoglycosidases. The recognition sites of PNGase F and Endo H are shown in Figure 8. Both enzymes were used to test the glycosylation pattern of C_H2. Protocols according to manufacturer`s recommendations were used.

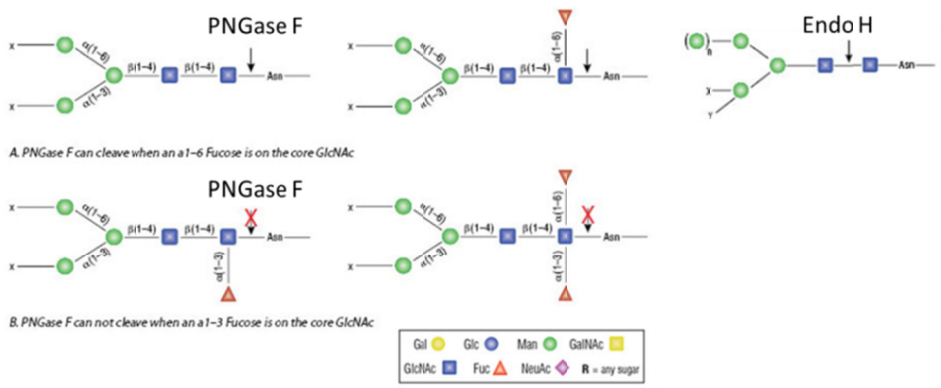

Figure 8: Cleavage sites for PNGase F and Endo H
From NEB homepage

2.6 Spectroscopy

2.6.1 UV-Vis spectroscopy

Absorption of a molecule is possible when it contains a chromophore (i.e., a system of conjugated double bonds with excitable electrons). When an incoming photon with a defined energy is absorbed electrons are excited from the ground to an excited state. Protein concentrations can be determined by absorption in the near-UV range (280 nm). In proteins the aromatic amino acids tryptophan and tyrosine are mainly responsible for absorption, to a small amount also phenylalanine and disulphide bonds. Calculation of the protein concentration from UV absorbance was carried out according to the Beer-Lambert law (Equation 1).

$$A = \varepsilon \cdot c \cdot d \leftrightarrow c = \frac{A}{\varepsilon \cdot d}$$

Equation 1: Beer-Lambert law.
A = Absorbance, ε = molar extinction coefficient (M^{-1} cm^{-1}), c = molar protein concentration (M), path length (cm)

Theoretical molar extinction coefficients were determined using the ProtParam tool. All UV spectra were recorded with a Helios γ (Thermo Fischer) or Ultrospec 1100 (Amersham Biosciences) UV/Vis spectrophotometer at room temperature and baseline corrected for buffer absorbance.

2.6.2 Circular Dichroism (CD) Spectroscopy

Circular dichroism is a powerful method for studying chiral molecules in solution as it is uniquely sensitive to the asymmetry of the system. Proteins are chemically asymmetric structures and exhibit optical activity. Thus, CD spectroscopy is an attractive method to study protein structure. The optical activity of proteins can have two reasons: one lies within the difference of the refractive index the second within the difference of the absorption of the two types of circularly polarised light. The latter phenomenon is called circular dichroism. Beer-Lambert law is also valid for CD spectroscopy and applies to both right- and left-handed circularly polarised light. CD is defined as the difference in extinction coefficients for each:

$$\Delta A(\lambda) = A_L(\lambda) - A_R(\lambda) = [\varepsilon_L(\lambda) - \varepsilon_R(\lambda)]lc = \Delta\varepsilon lc$$

Equation 2: Absorption between left and right handed circularly polarised light
A= absorption, λ= wavelength, ε= extinction coefficient, l= path length, c= concentration, R and L subscripts describe left- and right-handed circular polarised

CD spectroscopy can be used in the far-UV range to determine secondary structure of a protein. The far UV region lies within a range between 260 and 170 nm. Distinctive signals

are caused predominantly due to the differential absorption of backbone amide groups in asymmetric secondary structures. Typical secondary structure motifs like α-helices or β-sheets can be identified according to their characteristic shape. α-helical structures exhibit a negative band at about 222 nm and a negative and positive couplet at about 208 nm and 190 nm. The CD for β-sheet structures has a negative band at about 217 nm and a positive band at about 198 nm [204]. Random coil structures show one characteristic minimum at 200 nm.

CD is measured as ellipticity (Θ) in degrees. The conversion of the measurement signal to mean residue ellipticity is carried out by:

$$\Theta_{MRW} = \frac{\Theta \cdot 100 \cdot M}{d \cdot c \cdot N_{aa}}$$

Equation 3: Determination of the mean residue ellipticity
Θ_{MRW} = mean residue ellipticity (deg cm^2 dmol^{-1}), Θ= ellipticity (deg), M= molecular mass (g/mol), d= path length (cm), c = concentration (M), N_{aa}= number of amino acid residues

Besides secondary structure formation CD spectroscopy is a tool to determine a characteristic fingerprint of defined tertiary structure within a protein. In the near UV-range (320-260 nm) aromatic amino acids and to a small amount disulphide bonds are responsible for the observable CD bands. Aromatic amino acids have planar chromophores and are intrinsically symmetric. The signal in the near UV range is zero when the aromatic amino acids are mobile. In an ordered structure the environment of aromatic amino acids becomes asymmetric and therefore CD bands can be observed [205].

2.6.2.1 AFS of Fab domains

Unless otherwise stated, all experiments for the AFS analysis were carried out at 25°C. Measurements were performed in 50 mM sodium phosphate buffer at pH 7.5 or pH 2.0, in the presence or in the absence of sodium chloride. Urea concentrations were determined refractrometrically. Except for kinetic experiments, where proteins were diluted, proteins were dialysed against the appropriate buffer overnight or for 3 days, where indicated.

CD measurements were performed in a J-720 spectropolarimeter (Jasco) equipped with a Peltier element. FUV-CD spectra were recorded from 194 nm to 250 nm at a protein concentration of 50 µM in 0.2-mm quartz cuvettes; for NUV-CD spectra, a protein concentration of 50 µM in 5-mm quartz cuvettes was used. Spectra were accumulated 16 times and buffer corrected. Temperature-induced equilibrium unfolding transitions were monitored by recording the change in ellipticity at 205 nm in the case of C_H1 and C_L and at 215 nm in the case of V_H and V_L. For urea-induced equilibrium unfolding transitions, structural changes in C_H1 and C_L were monitored at 220 nm, and structural changes in V_H and V_L were monitored at 215 nm. The only exception was V_L under physiological conditions, where unfolding was observed at 212 nm. All measurements were performed at protein concentrations of 20 µM for temperature-induced unfolding transitions and 10 µM for urea-

induced unfolding transitions using a 1-mm quartz cuvette. In temperature-induced unfolding transitions, heating and cooling rates were set to 20°C h^{-1}. For denaturant-induced unfolding transitions, samples were incubated overnight at 25°C at the different urea concentrations prior to measurements. Kinetics of AFS formation were measured at 25°C in 1-mm quartz cuvettes with a final protein concentration of 10 µM. Kinetics for V_H and V_L were followed at 220 nm; kinetics for C_H1 at 205 nm. Formation of the AFS after kinetic experiments was verified by measurement of an FUV-CD spectrum for each domain.

Data were evaluated with Origin (OriginLab); for urea-induced unfolding transitions, a two-state model was applied [206].

$$y(D) = y_N^0 + m_N \cdot [D] - \frac{(y_N^0 + m_N \cdot [D]) - (y_U^0 + m_U \cdot T)}{1 + exp\left(-\frac{\Delta G_{stab} + m_c \cdot [D]}{R \cdot T}\right)}$$

Equation 4: Two-state model with the assumption of linear dependency of emission from native and unfolded protein
y(D) = fluorescence or CD signal; y = y-intercepts; N = data of native protein; U = data of unfolded protein; m = slope of the best fit straight line; m_c = cooperativity; [D] = denaturant concentration; T = temperature in Kelvin (20 °C (RT) ~ 293.1 K); ΔG_{stab} = free enthalpy of denaturant induced unfolding; R = universal gas constant (8.314 J mol^{-1} K^{-1})

2.6.2.2 AFS of C_H2

Unless otherwise stated, all experiments for the AFS analysis were carried out at 25°C. Measurements were performed in 50 mM sodium phosphate buffer at pH 7.5 or pH 2.0, in the presence or in the absence of sodium chloride. Sample preparation was done as denoted above.

CD measurements were carried out using a Jasco J-720 spectropolarimeter (Jasco) equipped with a Peltier element. Far-UV CD spectra were measured using 20 µM protein in 0.5-mm quartz cuvettes between 260 nm and 195 nm and near-UV CD spectra between 320 nm and 260 nm using 50 µM protein in 5-mm quartz cuvettes. The spectra were accumulated 16 times and buffer-corrected. Temperature-induced equilibrium unfolding transitions were monitored by recording the change in ellipticity at 205 nm. All measurements were performed at protein concentrations of 20 µM in 0.5-mm quartz cuvettes. Heating and cooling rates were set to 20°C h^{-1}.

2.6.2.3 Folding of V_H and Interaction between V_H and V_L

Unless otherwise stated, all experiments for the interaction study were carried out at 20°C. Measurements were all performed in PBS. GdmCl stock solution in PBS was pH-corrected (as GdmCl is acidic) and concentrations were determined refractrometrically.

CD measurements were performed in a J-720 or J-715 spectropolarimeter (Jasco) equipped with a Peltier element. FUV-CD spectra were recorded from 195 nm to 250 nm at a protein concentration of 20 µM in 0.5-mm quartz cuvettes; for NUV-CD spectra, a protein

concentration of 50 µM in 5-mm quartz cuvettes was used. Domains where only very low concentrations could be obtained where measured in a 1-cm quartz cuvette with 10 µM concentration. Spectra were accumulated 16 times and buffer-corrected. Temperature induced equilibrium unfolding transitions were monitored by recording the change in ellipticity at 215 nm for V_H and V_L mutants. For GdmCl-induced equilibrium unfolding transitions, structural changes in V_H domains were monitored at 215 nm. For GdmCl-induced equilibrium refolding were V_H domains in 1.5 M GdmCl denatured and refolded by dilution in PBS. All measurements were performed at protein concentrations of 20 µM for temperature-induced unfolding transitions and 10 µM or 7.5 µM (V_L and V_H, respectively) for GdmCl-induced unfolding transitions using a 1-mm quartz cuvette. In temperature-induced unfolding transitions, heating and cooling rates were set to 20°C h^{-1}. For denaturant-induced unfolding transitions, samples were incubated overnight at 20°C at the different GdmCl concentrations prior to measurements. For 1HEZ V_H a two-state unfolding model for dimeric proteins was additionally used (adopted from [154]).

$$N_2 = 2\,U$$

$$S_{obs} = Y_N(1 - F_U) + y_u\,F_U$$

$$F_U = \frac{[U]}{c_P} = \frac{\sqrt{K_U^2 + 8K_U c_p} - K_U}{4c_P}$$

$$K_U = \frac{[U]^2}{[N_2]} = e^{-\frac{\Delta G_U}{RT}} = e^{-\frac{(\Delta G_{H2O} - m\,[GdmCl])}{RT}}$$

Equation 5: Two-state model for dimeric proteins.
Adopted from [154]. N denotes to the native state and U to the unfolded state; S_{obs} is the observed fluorescence or CD signal at a certain GdmCl concentration, y_N is the signal of native protein and y_U of the unfolded protein (both assumed to be linearly dependent on the GdmCl concentration); F_U denotes the fraction of unfolded protein and c_p the total concentration of protein; ΔG describes the protein stability and was assumed to be linearly dependent with the cooperativity parameter m. R is the universal gas constant and T the absolute temperature.

2.6.3 Fluorescence spectroscopy

After excitation a molecule can emit fluorescence when an electron returns from the first excited state (S_1) back to the ground state (S_0). The lifetime of the excited state is long compared to the time for excitation and an excited molecule has ample time to relax into the lowest accessible vibrational and rotational states of S_1 before it returns to the ground state. Consequently, the energy of emitted light is always smaller than the one from the absorbed light.

Following excitation, a molecule loses energy as heat by cascading through the vibronic levels of all excited singlet states (i.e., internal conversion). From the S_1 the molecule returns to S_0 by different mechanisms. The return can either be driven by internal conversion, conversion to the triplet state (intersystem crossing) or emission of a photon (fluorescence).

Fluorescence of proteins originates from phenylalanine, tyrosine, and tryptophan residues. The fluorescence intensity of a particular chromophore depends on both, its absorption at the excitation wavelength and its quantum yield at the emission wavelength. Tryptophan has the highest quantum yield and is able to quench tyrosine fluorescence by resonance energy transfer (FRET). Fluorescence of tryptophan is often dominant in proteins [205].

All antibody domains have at least one conserved tryptophan residue. In the folded state this residue is buried beside the conserved disulphide bond and quenched. Upon unfolding this residue is exposed and the fluorescence signal increases. This conserved residue is a well suited probe to determine protein folding and unfolding.

Besides intrinsic tryptophan fluorescence numerous extrinsic fluorescence probes can be used. To determine the hydrophobicity at the surface of a protein or to observe folding the environment-sensitive fluorescent dye 1-anilinonaphthalene-8-sulfonic acid (ANS, Sigma-Aldrich) can be used. The fluorescence of ANS is increased when the environment becomes more hydrophobic and less exposed to water. Therefore structural changes within a protein can be determined as the native state is in general polar at the surface and hydrophobic residues are exposed during unfolding. Molten globule intermediates or partially unfolded regions can be characterised due to their high ANS fluorescence intensities caused by the exposure of hydrophobic regions.

2.6.3.1 AFS of Fab domains

Fluorescence measurements were carried out in a Spex FluoroMax II fluorimeter (Horiba-Yvon) or a Spex FluoroMax IV fluorimeter (Horiba Jobin-Yvon). Intrinsic tryptophan fluorescence spectra were recorded from 300 nm to 450 nm, exciting the samples at 280 nm with excitation and emission slit widths set to 3 nm and 5 nm, respectively. All measurements were performed at 25°C using a protein concentration of 10 µM in 1-cm quartz cuvettes. Spectra were accumulated three times and buffer-corrected. ANS binding was measured by fluorescence emission recorded from 400 nm to 650 nm, with excitation at 380 nm (slits as denoted above). Experiments were performed at a protein concentration of 10 µM and an ANS concentration of 100 µM in 1-cm quartz cuvettes. All spectra were accumulated three times and buffer-corrected.

2.6.3.2 AFS of C_H2

Fluorescence measurements were carried out in a Spex FluoroMax IV fluorimeter (Horiba Jobin-Yvon). Intrinsic tryptophan fluorescence spectra were recorded as in 2.6.3.1 with excitation and emission slit widths set to 2 nm and 5 nm, respectively. All measurements were performed at 25°C using a protein concentration of 1 µM. ANS binding was measured as denoted above.

For urea-induced equilibrium unfolding transitions, structural changes in C_H2 samples were excited with 280 nm and signal was monitored at 350 nm (pH 7.5) or 370 nm (pH 2.0) with

excitation and emission slits set to 2 nm and 5 nm, respectively. All measurements were performed at protein concentrations of 1 µM using a 1-cm quartz cuvette. In denaturant-induced unfolding transitions, samples were incubated overnight at 25°C at the different urea concentrations prior to measurements. Data analysis was according to 2.6.2.1.

2.6.3.3 Folding of V_H and Interaction between V_H and V_L

Fluorescence measurements were carried out in a Spex FluoroMax IV fluorimeter (Horiba Jobin-Yvon). All experiments were performed at 20°C if not otherwise stated. Intrinsic tryptophan fluorescence spectra were recorded as in 2.6.3.1 with excitation and emission slit widths set to 2 nm and 5 nm for V_L and 1 nm and 3 nm for V_H domains, respectively. ANS binding was measured as denoted above.

For GdmCl-induced equilibrium unfolding transitions, structural changes in V_L and V_H samples were excited with 280 nm and signal was monitored at the wavelength with the largest signal difference (in general 355 nm, determined *via* spectra of native and unfolded state) with excitation and emission slits as denoted above. For refolding transitions of V_L and V_H were proteins denatures in 2.5 M and 1.5 M GdmCl, respectively. Unfolding was induced by dilution with PBS. All measurements were performed at protein concentrations of 1 µM using a 1-cm quartz cuvette. In denaturant-induced unfolding transitions, samples were incubated overnight at 20°C at the different urea concentrations prior to measurements. Data analysis was according to 2.6.2.1.

2.7 Kinetic experiments

2.7.1 Manual mixing experiments for V_H folding

For refolding experiments, either 10 or 50 µM protein was denatured in 1.5 M or 2.0 M GdmCl (MAK and 1OPG, respectively) overnight and then diluted into a refolding buffer in a 1-cm quartz cuvette. For unfolding kinetics, native protein was diluted 25-fold or 50-fold into an unfolding buffer. The final protein concentration was 1 µM or 5 µM. The change in fluorescence intensity over time was monitored at the unfolded maximum of 355 nm, after excitation at 280 nm (excitation and emission slits set to 1 nm and 3 nm).

In order to assess the effect of prolyl isomerases, Cpr6 (kindly provided by Priyanka Sahasrabudhe and Oliver Lorenz) was added to the refolding buffer to a final concentration of 0–200 nM and the kinetics were observed at 355 nm for 600 seconds. Tree traces were averaged.

2.7.2 Stopped-flow fluorescence spectroscopy

Stopped-flow is a spectroscopic technique used for studying fast reaction mechanisms in solution over timescales of about 1 ms up to a few 100 seconds. In general, two reagents are

rapidly mixed together and then 'stopped' in an observation cell. Proceeding of the reaction can be visualised as the change in the recorded signal as a function of time observable when the sample cell is irradiated with monochromatic light. Besides fluorescence also absorption or circular dichroism can be detected (dependent on instrument set-up). Analysis of the resulting kinetic transient can determine reaction rates, complexity of the reaction mechanism, information on short-lived reaction intermediates etc. A series of stopped-flow experiments can be used to show the effect of parameters such as temperature, pH and reagent concentration on the kinetics of the reaction.

2.7.2.1 AFS formation of C_H2

Rapid mixing experiments were carried out using an SX18-MV stopped-flow apparatus (Applied Photophysics) in single mixing mode with a mixing ratio of 1:10 at 25°C. 10 µM protein in NaP buffer pH 7.5 (+NaCl) was mixed with NaP buffer pH 1.92 (+NaCl) to observe AFS formation. To determine if AFS formation is reversible, 10 µM protein in NaP buffer pH 2.0 (+NaCl) were mixed with NaP buffer pH 10.76 (+NaCl). Excitation and emission slits were set to 0.34 mm and 0.44 mm, respectively. Data were evaluated with Origin (OriginLab) using single, double and triple exponential equations to describe the traces. Rate constants were determined from t (sec).

2.7.2.2 Unfolding and refolding of V_H domains

Rapid mixing experiments were carried out using an SX18-MV stopped-flow apparatus (Applied Photophysics) in single mixing mode with a mixing ratio of 1:10 at 20°C. Protein (5 µM, 10 µM or 50 µM) was diluted in various concentrations of refolding or unfolding buffer. Final protein concentrations were in the range between 0.5 µM and 5 µM. Fluorescence was excited at 280 nm, and a cut-off filter at 305 nm was used. Excitation and emission slits were set to 0.4 mm each. Five traces were averaged in each case.

2.7.3 Double jump experiments for V_H folding

Native protein of MAK33 V_H and 1OPG V_H was diluted into 1.5 M GdmCl or 2.0 M GdmCl, respectively, to initiate unfolding. After an aging time between 5 and 600 s, the sample was further diluted 30 or 40 times into 0.05 M GdmCl, resulting in a final protein concentration of 1 µM. Fluorescence changes in the second step were monitored under the conditions described for manual mixing experiments.

2.7.4 Interrupted refolding experiments for V_H folding

The V_H domain was fully unfolded overnight in 1.5 M GdmCl. The experiment was initiated by diluting the protein 25 times in 0.06 M GdmCl. After an aging time varying between 5 and 600 s, the protein was further diluted in 1.0 M GdmCl to interrupt the refolding process.

Fluorescence changes in this step were monitored under the conditions described above. The final protein concentration was 1 µM.

2.7.5 Data analysis of kinetic data

Exponential changes in fluorescence were fitted using Origin (OriginLab). Different exponential functions (single, double and triple exponential) were tested and the residual compared. The fit with the best residual was chosen. More complex unfolding or refolding kinetics were fitted to more complex exponential functions created in Origin. Global fitting was performed with DynaFit.

2.8 Quaternary structure analysis

2.8.1 Analytical ultracentrifugation

Analytical ultracentrifugation (AUC) is a versatile and powerful method for the quantitative analysis of macromolecules in solution. AUC has broad applications for the study of biomacromolecules in a wide range of solvents and over a wide range of solute concentrations. Three optical systems are available for the analytical ultracentrifuge (absorbance, interference, and fluorescence) that permit precise and selective observation of sedimentation in real time [207].

AUC makes use of the application of centrifugal force together with the simultaneous real-time observation of the sedimentation of macromolecules in the centrifugal field. A benefit of AUC experiments is that proteins can be studied in solution and their hydrodynamic and thermodynamic characterisation is not influenced by any interaction with a matrix or a surface. There are two basic types of ultracentrifugation experiments for the characterisation of proteins: sedimentation velocity (SV) and sedimentation equilibrium (SE) methods, which are described in the following sections.

In SV experiments, a sufficiently large centrifugal force field is applied that leads to a movement of molecules toward the bottom of the centrifuge cell. The sedimentation process is determined by three factors: the gravitational force, the buoyancy and the hydrodynamic friction. The gravitational force is proportional to the square of the rotor speed, thus adjustment of the rotor speed allows to study of a wide range of particle sizes, ranging from kDa to GDa molecular weights. The Svedberg equation accounts for the balance of all three forces:

$$s = \frac{v}{\omega^2 r} = \frac{MD(1-\bar{V}\rho)}{RT}$$

Equation 6: Svedberg equation.
s= sedimentation coefficient (s), v= observed radial velocity (m/s), ω= angular velocity of the rotor (m/s²), ω²r= centrifugal field, M= molar mass (g/mol), $\bar{V}$= partial specific volume (cm³/g), ρ= density of the solvent (g/cm³), D= diffusion coefficient (m²/s), R= gas constant (8.314472 J/K mol), T= absolute temperature (K)

The sedimentation coefficient describes the sedimentation velocity *v* in relation to the centrifugal field $\omega^2 r$. The s-values are commonly reported in Svedberg (S) units, which correspond to 10^{-13} s.

In sedimentation equilibrium experiments, the centrifugal force is adjusted sufficiently small so that the process of diffusion significantly opposes the process of sedimentation, and an equilibrium concentration distribution of macromolecules is obtained throughout the cell. For an ideal non-interacting single component system, the equilibrium distribution is an exponential function of the buoyant mass of the molecule, $M(1-\bar{V}\rho)$, as described by:

$$c(r) = c_0 \cdot e^{\frac{M(1-\bar{V}\rho)\omega^2(r^2-r_0^2)}{2RT}}$$

Equation 7: Equilibrium within an ideal non-interacting single component system.
c(r)= sample concentration at radial position r, c_0= sample concentration at reference radial distance r_0, $\bar{V}$= partial specific volume, ω= angular velocity of the rotor, R= gas constant, T= absolute temperature

2.8.1.1 AFS of Fab domains and C_H2

For the determination of the quaternary structure of the Fab domains and C_H2 variants under physiological and AFS conditions sedimentation velocity experiments were performed at 25°C. AUC was carried out with a ProteomLab XL-I (Beckman) equipped with absorbance and interference optics. 400 µl of the 10 µM samples and 410 µl of the phosphate buffer were loaded into assembled cells with quartz windows and 12 mm pathlength charcoal-filled epon double-sector centrepieces and centrifuged at 42.000 rpm in an eight-hole Beckman-Coulter AN50-Ti rotor. Sedimentation was monitored at 280 nm or 230 and continuously scanned with a radial resolution of 30 µm, and three replicates were taken. Data analysis was carried out with SEDFIT using the continuous c(S) distribution mode of SEDFIT (Peter Schuck, University of Maryland). For interference scans, the cells were assembled with sapphire windows, and data were analysed with Ultrascan and the USLIMS platform (Borries Demeler, Department of Health Sciences, University of Texas). In brief, after two-dimensional grid analysis and Monte Carlo simulation (50 iterations), the data fit was further refined by a genetic algorithm and finally tested for robustness by another Monte Carlo simulation (50 iterations).

Additionally, AUC was used to determine the quaternary structure of the V_H and V_L mutants. One challenge to determine the complex formation between V_L and V_H by AUC is their similar size (11.9 kDa and 13.9 kDa, respectively). Thus, a prerequisite for the binding studies is the monomeric state of both interaction partners to be sure to observe exclusively the

heterodimeric complex which could not be distinguished from the homo-dimer complex by size differences. Therefore all domains were tested for their ability to be monomeric within the applied concentration range (0.5 µM to 7.5 µM). As a standard method AUV sedimentation velocity experiments were performed, for some cases additionally sedimentation equilibrium experiments to confirm the observations of the standard experiments fitted to an ideal non-interacting single component system.

2.8.1.2 Folding of V_H and Interaction between V_L and V_H

AUC was performed using a Beckman ProteomLab XL-A centrifuge (Beckman-Coulter). 500 µl of the protein samples and 510 µl of PBS were loaded into assembled cells with quartz windows of 12-mm pathlength and were centrifuged at 42.000 rpm in a three- or eight-hole Beckman AN50-Ti rotor. Sedimentation was monitored at 230 nm or 280 nm to reach an absorption signal between 0.1 and 0.8. Data analysis was done with SEDFIT (Peter Schuck), using a non-model based continuous Svedberg distribution methods (c(S)) with time and radial invariant noise.

2.8.2 Transmission Electron Microscopy and Atomic Force Microscopy

The transmission electron microscope (TEM) operates on the same basic principles as the light microscope but uses electrons instead of light. Light microscopes have a limited resolution due to the wavelength character of light. As TEMs use electrons as a "light source" and their much lower wavelength it is possible to get a resolution a thousand times better than with a light microscope (resolution limit is approximately 10^{-10} m). The possibility for high magnifications has made the TEM a valuable tool in both medical, biological and materials research.

Atomic force microscopy (AFM) is a high-resolution imaging technique that can resolve features on molecular and atomic level. AFM works by bringing a cantilever tip in contact with the surface to be imaged. An ionic repulsive force from the surface applied to the tip bends the cantilever upwards and the amount of bending, which is measured by a laser spot reflected on to a split photo detector, can be used to calculate the force. By keeping the force constant while scanning the tip across the surface, the vertical movement of the tip follows the surface profile and is recorded as the surface topography by the AFM. The method can be used for imaging any conducting or non-conducting surface.

To determine if the Fab domains can form fibril under AFS condition two independent microscopic techniques were chosen. For fibrillisation experiments, 10 µM protein samples were incubated at 25°C for 20 days. Afterwards, samples for TEM measurements were fixed on a 200-mesh copper grid. 10 µl of the samples were loaded on the activated grid and incubated for 1 minute. After a washing step with 10 µl H_2O, the samples can be stained with 10 µl of a 1.5% uranyl-acetate solution. Samples were recorded with 100 kV voltage and with a 33.000x magnification. Bettina Richter performed all TEM measurements.

For AFM samples with a protein concentration of 10 µM were incubated at 25°C, with mild agitation. After 20 days, 20 µl of each sample was transferred to a freshly cleaved mica disc. The disc was washed three times with water and allowed to dry. AFM measurements were carried out with a Digital Instruments multimode scanning probe microscope (Veeco) in contact mode using DNP-S20 tips. Experimental setup was as published previously [56].

2.8.3 Nuclear Magnetic Resonance

NMR experiments were performed in collaboration with Professor Bernd Reif (TUM). Elke Prade performed all measurements. All samples were recorded in a NMR- spectrometer from Bruker at 25°C if not stated otherwise.

For assignments, 750 MHz were applied; titration experiments of C_H2 were measured with 750 MHz or 500 MHz. For the spectrum at pH 2.0 with NaCl, HNCA, HNCACB and HCNCO (all best versions; according to [208]) as well as (H)CC(CO)NH were assigned. In (H)CC(CO)NH measurements, in which water signals had to be suppressed, watergate was applied. All ^{1}H-^{15}N 2D spectra are SOFAST-HMQC [209].

2.9 Interaction analysis

2.9.1 Sedimentation equilibrium experiments

Samples of V_H or V_L wild type together with the binding partner were prepared in a 1:1 and for some additionally in a 1:2 ratio within a concentration range of 0.5 µM- 7.5 µM and incubated on ice during AUC cell preparation. Afterwards samples were loaded into six channel epon centrepieces (12 mm window) and equilibrium at three different velocities (25.000, 30.000 and 35.000 rpm) measured. Scans were performed at 280 nm, 250 nm or 230 nm to be in absorption range within 0.1-0.8. 25 replicates were recorded. According to the results of the AUV sedimentation velocity experiments the domains are monomeric and observed complexes a result of hetero-association. Data was globally fitted to determine the association constant with a heterogeneous interaction model (1:1 complex) which is based on the mass action law $C_{AB} = K_a c_A c_B$ (Equation 8) with the programme Origin 8.6. Three different velocities and concentrations were included in the fit.

$$a\,(r) = C_A(r_0)\varepsilon_A d\; exp\left[M_{b,A}\;\frac{\omega^2}{2RT}\;(r^2 - r_0^2)\right] + C_B(r_0)\varepsilon_B d\; exp\left[M_{b,B}\;\frac{\omega^2}{2RT}\;(r^2 - r_0^2)\right] + K_a C_A(r_0) C_B(r_0)(\varepsilon_A + \varepsilon_B) d\; exp\left[(M_{b,A} + M_{b,B})\frac{\omega^2}{2RT}(r^2 - r_0^2)\right]$$

Equation 8: Model for heterogeneous interaction between two components.
c(A)/c(B)= sample concentration of interaction partner A and partner B, r_0= reference radius, $M_{b,A}/M_{b,B}$= buoyant masses of component A and B, ω= angular velocity of the rotor, R= gas constant, T= absolute temperature, d= optical path length, K_a= association constant

The K_D values can be determined from the K_A values with Equation 9.

$$K_D = \frac{1}{K_A}$$

Equation 9: K_D determination.
K_D = dissociation constant, K_A = association constant

2.9.2 Surface Plasmon Resonance

Surface Plasmon Resonance (SPR) is a suitable method to determine protein interactions. On- and off-rates, as well as binding strength and specificity, can be characterised. In the experimental setup the ligand is immobilised on a sensor chip and the analyte is constantly flushed through the sample cell. The SPR phenomenon occurs when polarised light strikes an electrically conducting gold layer under conditions of total internal reflection at the interface between media of different refractive index: the glass of a sensor surface (high refractive index) and a buffer (low refractive index). A wedge of polarised light, covering a range of incident angles, is directed toward the glass face of the sensor surface. Reflected light is detected within the system. Electric field intensity (i.e., the evanescent wave) is generated when the light strikes the glass. This evanescent wave interacts with, and is absorbed by, free electron clouds in the gold layer. This consequently generates electron charge density waves called plasmons and causes a reduction in the intensity of the reflected light. The resonance angle at which this intensity minimum occurs is a function of the refractive index of the solution close to the gold layer on the opposing face of the sensor surface.

The ligand (V_L or V_H wild type with N-terminal Gly-Ser-Cys) was immobilized through the thiol group of the additional cysteine residue on the carboxymethylated dextran (CM) surface of a Biacore Sensor Chip CM5 using a Biacore X100 instrument. In an immobilisation pH-scouting experiment, the ideal buffer conditions (acetate buffer pH 4.0/4.5/5.0) with a protein concentration of 40 µg/ml were determined. Afterwards, the protein had to be coupled on the chip using the maleimide coupling protocol from the Biacore control software. The N-terminal cysteine must be a free reduced thiol group for a successful coupling reaction thus the protein was incubated with 1 mM TCEP for 15 min at RT. Afterwards the immobilisation protocol from the Biacore X100 software was started. Therefore the surface was activated with EDC/NHS. Then, EMCH was injected to introduce maleimide groups on the chip surface. To deactivate excess reactive groups Ethanolamine (pH 7.0) was added. The N-terminal free thiol group on the injected V_L or V_H (40 µg/ ml) domain in 10 mM acetate buffer with pH 4.5 then reacted with the maleimide group to form a stable thioether linkage. Deactivation of excess maleimide groups was done by injecting a cysteine/NaCl solution. The final immobilisation RU was set to 500.

For the actual SRP measurements, different V_H or V_L domains were added as analytes. HBS-P+ buffer was used as running buffer for measurements at 20°C with a flow rate of 10 µL/min. Multi-cycle kinetic assays were performed ranging from 0.2 µM to 25 µM of analyte

concentrations. Regeneration was done with 1 M NaCl. The experimental data were analysed with the Biacore X100 Evaluation Software.

2.10 Enzyme-linked Immunosorbent Assay (ELISA)

To test the functionality of the mutated V_H and V_L domains an ELISA assay with isolated variable domains was established. The experimental setup is depicted in Figure 9.

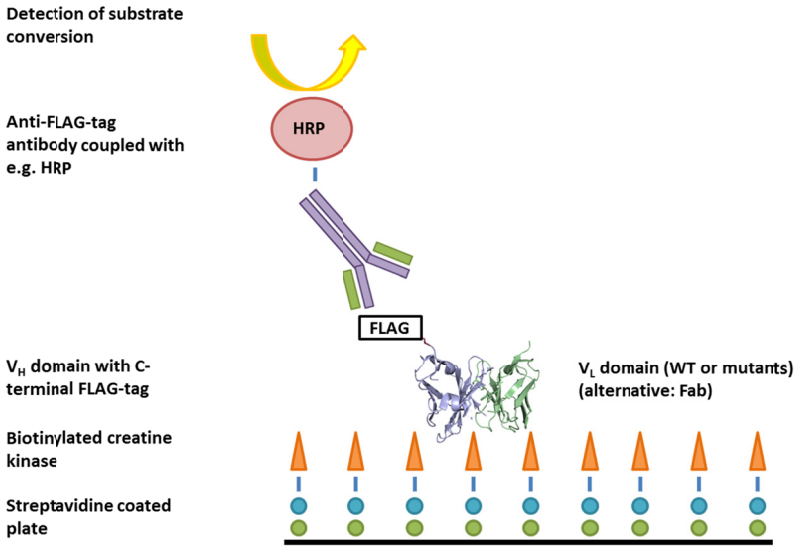

Figure 9: ELISA set-up to test binding affinities of isolated domains
Biotinylated human creatine kinase is coupled to a streptavidine coated microwell plate; V_H and V_L wild type with C-terminal Flag-tag domains and the mutant interaction partner are incubate and detection is enabled via a Horseradish peroxidase (HRP) coupled Anti-Flag-tag antibody.

Assay components and microwell plates were purchased from Roche. Assay was performed according to the manufacturer`s recommendations. Shortly, samples were prepared in 10 µl volume. Different mutants were tested within a concentration range of 100 nM up to 25 µM against wild type V_L or V_H with a C-terminal Flag-tag for detection. After the addition of 90 µl reaction mix I the sample was incubated on a streptavidine coated microwell plate to immobilise human biotinylated creatine kinase. Incubation was performed for 45 minutes with constant agitation at 20°C or 10°C for V_L and V_H mutants, respectively. 100 µl/well of reaction mix II were prepared in between and the detection antibody against Flag-tagged antibody in a 1:15.000 dilution added. Before reaction mix II and the Anti-Flag-tag antibody were added and the samples for 45 minutes at room temperature incubated, the samples were washed with sterile pure water for three times. Afterwards, the samples were washed three times with water for a second time and then 100 µl of reaction mix III was added and the measurement in the GENios plate reader (Tecan) started. Absorption was measured at 405 nm without active pre-mix setup (i.e., agitation of the sample for 8 seconds before start). Measurement time was 0.5-3 hours with data intervals of 1-3 minutes until a plateau was reached.

2.11 MALDI-ToF MS

The molecular weight analysis of purified proteins was performed on a Bruker Ultraflex-2 MALDI-TOF/TOF mass spectrometer.

To remove potassium ions (from PBS buffer) a Zip Tip preparation was performed for all samples following the instructions of the manufacturer (Millipore). The HCCA matrix was dissolved in 0.1% TCA, 80% acetonitrile and 20% ddH2O. The full-length proteins and peptides were spotted in the matrix solution and analysed after the evaporation of the solvent. For each spectrum, the MS analysis at different points of the preparation was averaged and the evaluation was done with the program m/z (Moverz).

Chapter 3

3. Results and Discussion

3.1 The alternatively folded state of antibody domains

3.1.1 Dissecting the AFS of Fab domains[1]

3.1.1.1 Different propensities of individual Fab fragment domains to form an AFS

In previous studies, MAK33 [210] had been shown to adopt an AFS at low pH [132]. Later on, the concept was extended to the Fab fragment, the isolated light chain and C_H3 [133, 134, 140]. To assess the propensity of the individual Fab fragment domains (V_H and C_H1 of the heavy chain, and V_L and C_L of the light chain) (Figure 10 A) to form an AFS, we expressed and purified each of the isolated domains. As previously described [119, 191], all these domains, except for C_H1, which is intrinsically disordered [181], show a predominantly β-sheet structure under physiological conditions (100 mM NaCl, pH 7.5) (Figure 10 B; in green).

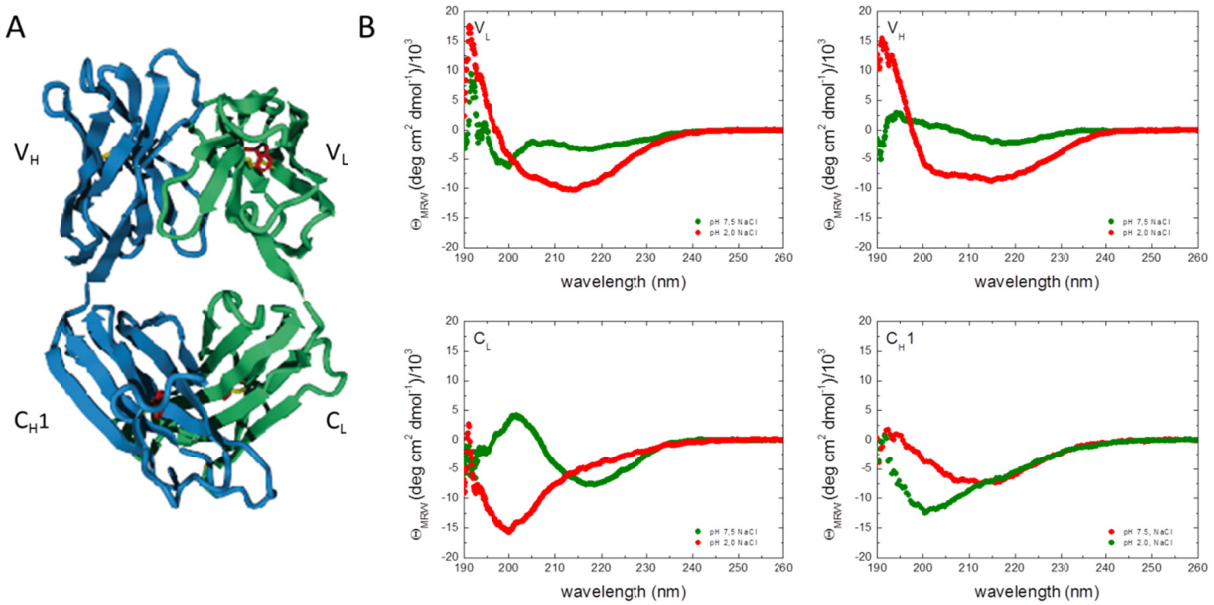

Figure 10: Structure of the Fab fragment and secondary structure of its constituting domains under physiological and AFS conditions

(A) The structure of an IgG Fab fragment (PDB code 12E8). The two N-terminal heavy chain domains, V_H and C_H1, are shown in blue. The light chain is composed of V_L and C_L (depicted in green). In (B) the FUV CD spectra of the single domains measured under physiological conditions [50 mM sodium phosphate (pH 7.5) and 100 mM

[1] Published in 192. Feige, M.J., et al., *Dissecting the alternatively folded state of the antibody Fab fragment.* J Mol Biol, 2010. **399**(5): p. 719-30.

NaCl] (in green), as well as under AFS conditions [50 mM sodium phosphate (pH 2.0) and 100 mM NaCl] (in red) are shown. All measurements were performed at a protein concentration of 50 µM at 25°C

This type of secondary structure is characterised by a minimum around 218 nm and a maximum around 200 nm in far-UV (FUV) CD spectra (Figure 10 A). As the variable domains (V_H and V_L) possess a significant amount of unstructured amino acids in their antigen binding regions [complementarity-determining regions (CDRs)], the β-sheet characteristics of the FUV-CD spectra for these domains are less pronounced than for the C_L domain (Figure 10 B). To induce the formation of an AFS in individual domains, we applied conditions established for the IgG C_H3 domain [133] (100 mM NaCl, pH 2; in red; denoted as AFS conditions in the following). It is evident from Figure 10 B that all domains, except for C_L, which globally unfolds under the chosen solvent conditions, form a significant amount of secondary structure at pH 2.0.

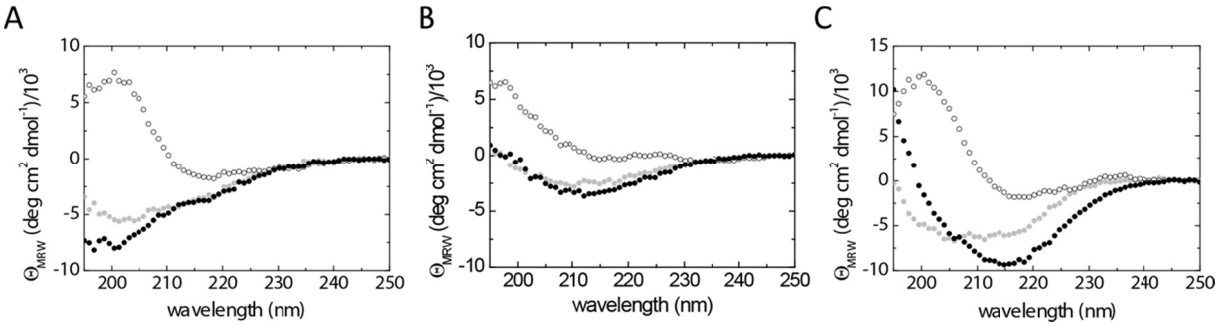

Figure 11: AFS formation for different variable IgG domains.

To analyse if the AFS can be generally adopted by isolated variable IgG domains and to assess a potential influence of the complementarity determining regions, three variable domains of different antibodies were examined: The V_L domain of an anti-HIV1 antibody (pdb code: 2FX7) (A) and an anti-IL1β antibody (pdb code: 2KH2) (B) as well as a V_H domain of an anti-Her2 antibody (pdb code: 1FVC) (C). FUV-CD spectra were recorded under native conditions (50 mM sodium phosphate, pH 7.5, 100 mM NaCl) (open circles) and AFS conditions (50 mM sodium phosphate, pH 2.0, 100 mM NaCl) (closed circles). Additionally, spectra were recorded under low salt conditions (50 mM sodium phosphate, pH 2.0) (light grey circles). The protein concentration was 50 µM and the spectra were recorded at 25°C

The formation of an AFS is not restricted to the MAK33 Fab domains, as we could observe this effect also for another murine anti-IL1β V_L domain [211] and a humanised anti-Her2 V_H domain [212] (Figure 11). In contrast, a human V_L domain directed against human immunodeficiency virus-1 (HIV-1) [213] did not form an AFS under the tested conditions (Figure 11 A). In general, a minimum in FUV-CD spectra around 215 nm (Figure 10 B; Figure 11) seems to be typical of this alternative structure at low pH. Notably, the spectra are different in shape and intensity from those of the native and unfolded states (Figure 10 B; Figure 11; Figure 12).

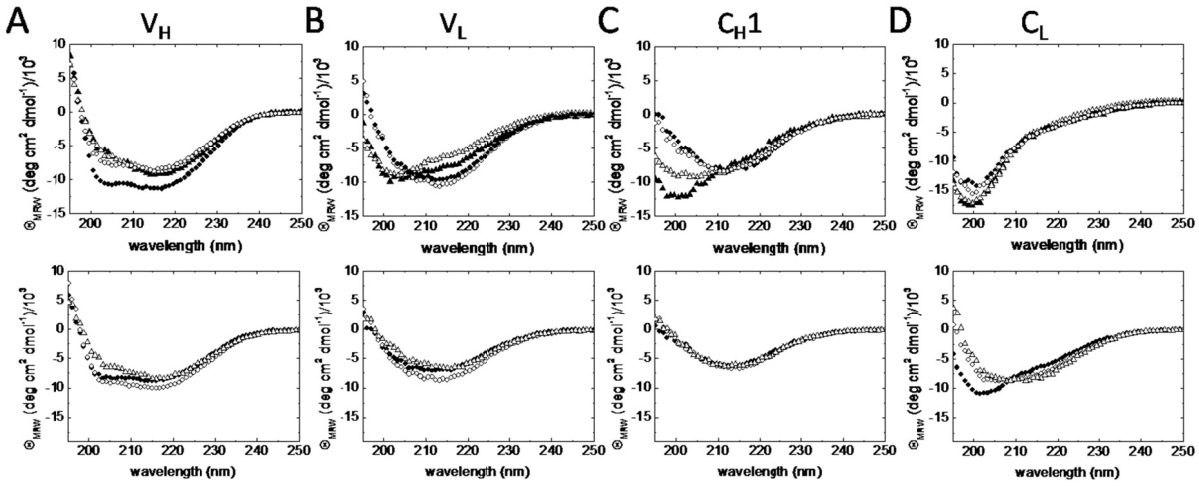

Figure 12: Influence of a reducing agent (TCEP) and salt concentrations on AFS formation.

To study potential effects of a reducing agent on the formation of the AFS, FUV-CD spectra of V_H (A), V_L (B), C_H1 (C) and C_L (D) (upper panels) were recorded under AFS conditions (50 mM sodium phosphate, pH 2.0, 100 mM NaCl) in the presence (closed circles) and absence (open circles) of 1 mM TCEP. In addition, all four domains of the Fab fragment were analysed under acidic low salt conditions (50 mM sodium phosphate, pH 2.0) in the presence (closed triangles) and absence (open triangles) of 1 mM TCEP. The effects of higher salt concentrations (> 100 mM) on AFS formation are shown in the lower panels. Three different salt concentrations were used: 150 mM (closed circles), 175 mM (open circles) and 200 mM (open triangles). FUV-CD spectra were recorded at protein concentrations of 50 µM at 25°C

The ionic strength of the solvent and the internal disulphide bridge in immunoglobulin domains and fragments had been reported to influence the formation of the AFS [132-134, 140]. Indeed, for C_H1 and V_L, the FUV-CD spectra at low pH and low ionic strength (0 mM NaCl) were typical for globally unfolded proteins (Figure 12). For V_H, in contrast, identical FUV-CD spectra were obtained in the absence and in the presence of NaCl at pH 2 (Figure 12). At higher ionic strengths up to 200 mM NaCl, no significant additional changes in the FUV-CD spectra of the AFS of V_H, V_L, and C_H1 could be observed (Figure 12). In contrast, for C_L, formation of secondary structure above 175 mM NaCl was detectable (Figure 12). The presence of a reducing agent [1 mM tris(2-carboxyethyl)phosphine (TCEP)] did not have an effect on the formation of the AFS for any of the domains; only for V_H slightly higher signal intensities were observed under reducing conditions (Figure 12). Taken together, the individual Fab fragment domains exhibit strikingly different propensities to form secondary structures distinct from the native state under conditions known to induce the AFS. C_L seems to be completely unfolded under conditions that induce an AFS in all other domains and only gains structure with increasing ionic strength. V_H even forms the AFS in the absence of salt. For all domains forming an AFS, the structural changes induced at low pH were found to be almost completely reversible at low protein concentrations (10 µM, Figure 13), with some aggregation occurring at higher protein concentrations (70 µM, not shown).

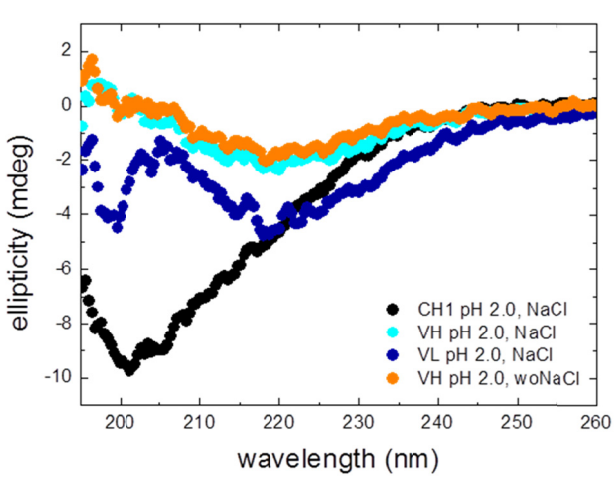

Figure 13: The AFS for the Fab domains is reversible at low concentrations.
To test if AFS formation is reversible, the Fab domains were dialysed overnight against 50 mM sodium phosphate (pH 2.0) and 100 mM NaCl (or in addition without NaCl for V_H) and afterwards transferred into 50 mM sodium phosphate (pH 7.5) and 100 mM NaCl and dialysed overnight. The FUV-CD spectra for C_H1 (black circles), V_H (pale cyan), V_L (dark blue circles), and V_H (from low salt to physiological condition; orange circles) are shown.

3.1.1.2 Structural characterisation of the AFS

To obtain a more detailed picture of structural changes that the Fab domains undergo upon formation of their AFS, we used near-UV (NUV) CD and fluorescence spectroscopy. These spectroscopic techniques allow tertiary structural changes around the aromatic amino acids Phe, Tyr, and predominantly Trp to be monitored. Under physiological conditions, all domains, except for C_H1, show well-defined NUV-CD spectra indicative of a defined structure around the aromatic reporter groups (Figure 14, top, in green). Under AFS conditions, V_L completely loses its NUV-CD signal, giving rise to the typical featureless spectra of an unfolded protein comparable to the spectra of C_L, which does not adopt an alternative structure under the applied conditions (Figure 10 B and Figure 14 B and D, top). In contrast, V_H shows a well-defined NUV-CD spectrum under AFS conditions that is strikingly different from that of the native protein (Figure 14 A, top). C_H1 does not exhibit a defined NUV-CD spectrum at pH 7.5 or pH 2.0 (Figure 14 C, top).

The fluorescence spectra of the different domains corroborate the findings from the NUV-CD measurements. For V_H, the maximum of the fluorescence spectra is at 342 nm in the native state and in the AFS, indicative of Trp residues being partially shielded from the solvent in both states (Figure 14 A, bottom). For all other domains, a small shift of the emission maximum is observed upon transfer to low pH. For C_L, which unfolds under AFS conditions, a redshift from 350 nm to 355 nm appears, whereas for the AFS-forming domains V_L and C_H1, a blueshift from 355 nm to 350 nm is observed (Figure 14 B-D, bottom). Hence, if an AFS is formed, Trp residues are not completely solvent-exposed in this state. As the Trp fluorescence is usually quenched in the native state of folded antibody domains due to the local proximity of the buried intrinsic disulphide bridge to an adjacent buried Trp residue of the hydrophobic core (Figure 10 A), changes in the fluorescence intensity provide another means to follow underlying structural changes.

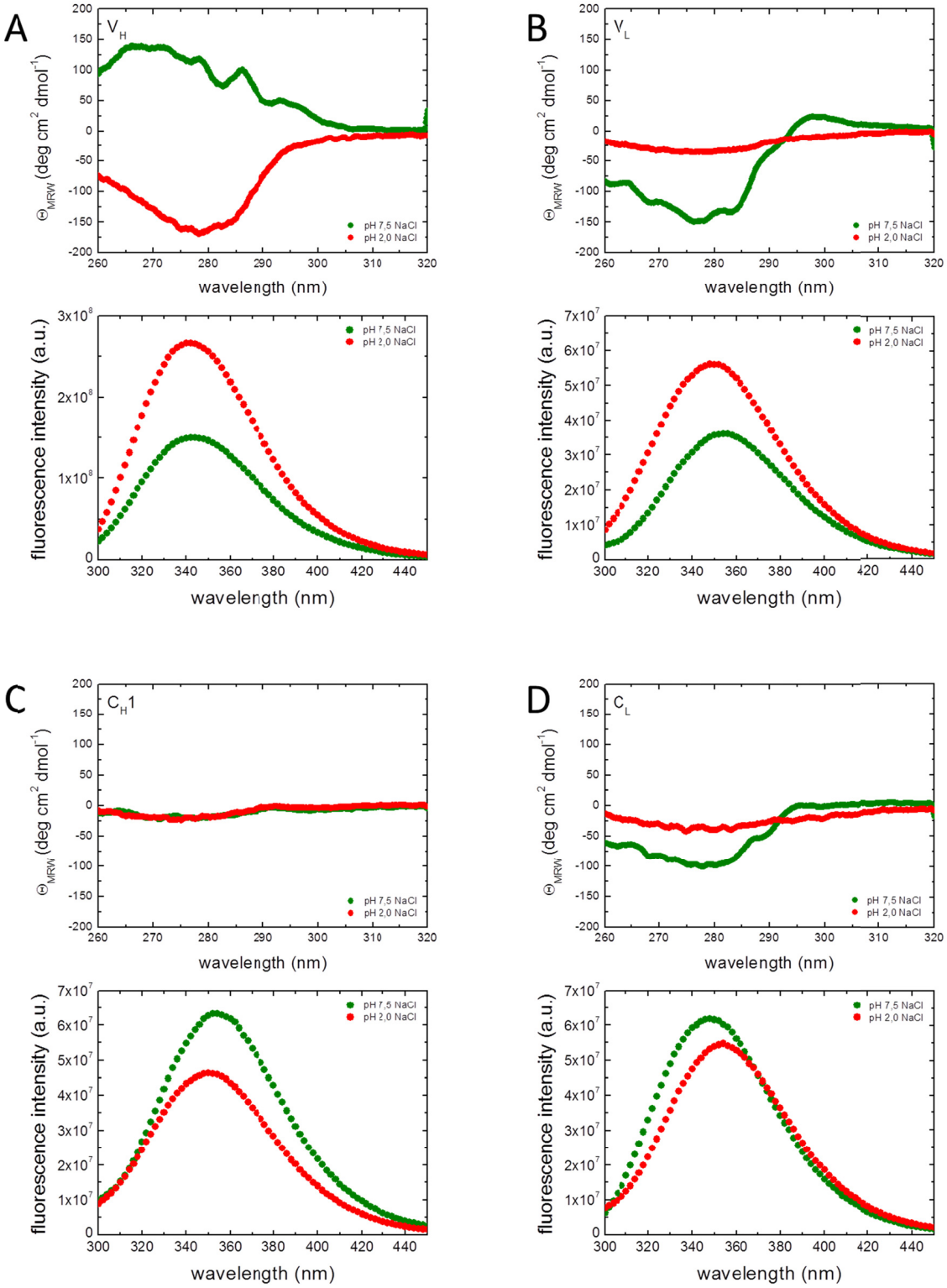

Figure 14: Tertiary structure of individual Fab fragment domains under physiological and AFS conditions.

NUV-CD spectra (top) and fluorescence spectra (bottom) of V_H (a), V_L (b), C_H1 (c), and C_L (d) were measured. Spectra were obtained under physiological conditions [50 mM sodium phosphate (pH 7.5) and 100 mM NaCl] (green), as well as under AFS conditions [50 mM sodium phosphate (pH 2.0) and 100 mM NaCl] (red). For NUV-CD measurements, a protein concentration of 50 µM was used; for fluorescence measurements, a protein concentration of 10 µM was used. Protein fluorescence was excited at 280 nm. All measurements were performed at 25°C.

For V_L and V_H, the fluorescence intensity of the AFS is significantly higher than in the native state (Figure 14 A and B, bottom). This argues for structural rearrangements moving Trp residues away from the internal disulphide bridge. A decrease in fluorescence intensity upon AFS formation is observed for C_H1 (Figure 14 C, bottom). As C_H1 is not folded under physiological conditions, an interpretation is less straightforward in this case. For C_L, surprisingly, a slight decrease in fluorescence intensity is observed upon unfolding, which might argue for other quenching reactions (e.g., due to solvent exposure) in the unfolded state or in a misfolded state at low pH (Figure 14 D, bottom).

Formation of exposed hydrophobic patches can be monitored by the fluorescence increase of the hydrophobic dye 8-anilino-1-naphthalene sulfonic acid (ANS) upon binding to these patches [214, 215]. For all proteins populating an AFS (i.e., V_H, V_L, and C_H1), but not for C_L, under the applied conditions, a significant increase in exposed hydrophobic surface was observed upon AFS formation, as assessed by ANS binding (Figure 15). This was particularly pronounced for the variable domains (Figure 15 A and B).

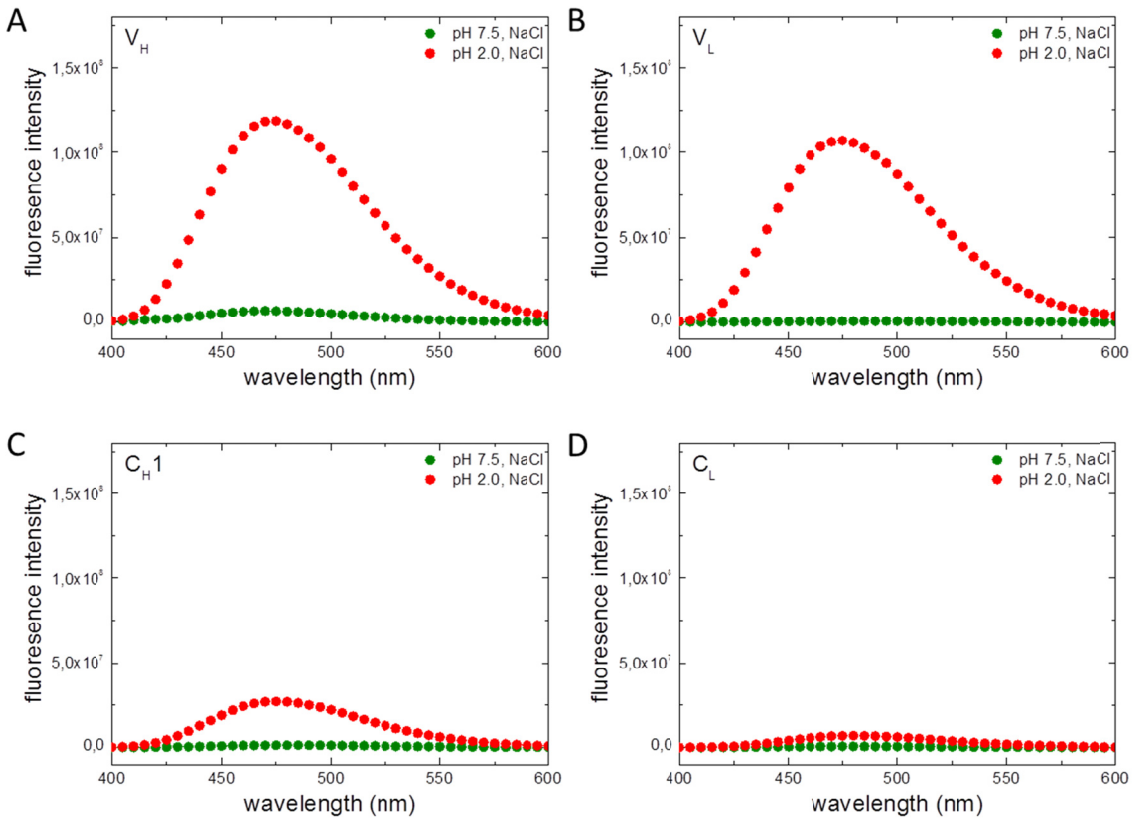

Figure 15: Exposure of hydrophobic surface upon AFS formation.

To analyse whether AFS formation leads to an increase of exposed hydrophobic surface, ANS binding assays were performed. ANS fluorescence spectra of the four Fab domains - V_H (A), V_L (B), C_H1 (C) and C_L (D) - under physiological conditions (50 mM sodium phosphate, pH 7.5, 100 mM NaCl) (green) and AFS conditions (50 mM sodium phosphate, pH 2.0, 100 mM NaCl) (red) in the presence of 100 µM ANS were measured. Protein concentrations were 10 µM for each domain. Excitation was set to 380 nm. Measurements were performed at 25°C

A typical feature of the AFS is the formation of higher-order protein assemblies [133, 140]. Fibril formation at low pH, a state some antibody domains can adopt [216, 217], could not be detected for any of the proteins by atomic force microscopy or transmission electron microscopy under the conditions tested (data not shown). Sedimentation velocity analytical ultracentrifugation analysis of the Fab domains revealed that they were monomeric at pH 7.5 (Table 2). Under AFS conditions, however, different oligomeric states were observed (Table 2; Figure 16). V_H and V_L predominantly formed oligomers in the range from 200 kDa to 500 kDa, corresponding to roughly 14–36 and 17–42 monomers, respectively. Additionally, for both domains, oligomers in the ranges between 50–200 kDa and 500–1000 kDa were detected. C_H1 populated a mixture of monomers and predominantly smaller oligomers (50–200 kDa, corresponding to 5–20 monomers), and C_L remained monomeric. Thus, the differences observed in tertiary structural changes and ANS binding of the individual Fab domains also partially transfer to quaternary structural changes.

Table 2: Quaternary structure of individual Fab domains.

Domain	Buffer condition	Monomer [%]	Oligomers [%]		
			50-200 kDa	200-500 kDa	500-1000 kDa
V_L	pH 7.5, NaCl	~ 100	-	-	-
	pH 2.0, NaCl	-	~ 25	~ 55	~ 20
V_H	pH 7.5, NaCl	~ 100	-	-	-
	pH 2.0, NaCl	-	~ 5	~ 70	~ 25
C_L	pH 7.5, NaCl	~ 100	-	-	-
	pH 2.0, NaCl	-	~ 70	~ 30	-
C_H1	pH 7.5, NaCl	~ 100	-	-	-
	pH 2.0, NaCl	~ 100	-	-	-

Sedimentation velocity analytical ultracentrifugation was used to determine the oligomeric state of the Fab domains under physiological conditions [50 mM sodium phosphate (pH 7.5) and 100 mM NaCl] and AFS conditions [50 mM sodium phosphate (pH 2.0) and 100 mM NaCl]. The observed quaternary structure was grouped into four categories.

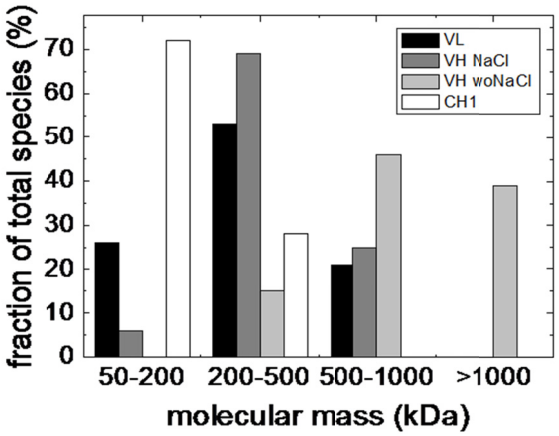

Figure 16: Analysis of the higher order protein assemblies for the different Fab domains that form an AFS.

Sedimentation AUC velocity analysis was performed to dissect the higher order oligomers under AFS conditions. The fractions of total species in % for AFS conditions (50 mM sodium phosphate, pH 2.0, 100 mM NaCl) are blotted for V_L (black), V_H (dark grey) and V_H (without NaCl) (light grey) and C_H1 (white). The protein concentration was 10 µM and the measurements were performed at 25°C.

3.1.1.3 Stability of the AFS of Fab domains

One defining characteristic of the AFS in comparison to molten globule states, which are also characterised by spectroscopic signatures different from the native state [218], is its high stability against thermal and chemical unfolding [132, 133, 140].

Thus, to further characterise the AFS, its stability against elevated temperatures and the chemical denaturant urea in comparison to the native state were analysed. Structural changes were monitored by FUV-CD spectroscopy. As under AFS conditions none of the transitions was found to be reversible, thermodynamic stabilities could not be determined. At pH 7.5, no data were obtained for the intrinsically disordered C_H1 domain. V_L and C_L show a similar transition midpoint of ~50°C in temperature-induced unfolding experiments (Figure 17 B and D, top; Table 3). V_H is the least stable of the structured domains and unfolds uncooperatively, with significant aggregation occurring already slightly above 30°C (Figure 17 A, top; Table 3). Therefore its transitions could not be evaluated. Under AFS conditions, all domains populating the AFS show a transition midpoint higher than that under physiological conditions (Figure 17, top; Table 3). In particular, V_H is highly stable against elevated temperatures in its AFS, with a melting temperature of ∼77 °C (Figure 17 A, top; Table 3). For V_L and C_H1, the transitions are quite uncooperative. No changes in stability were observed when proteins were incubated under AFS conditions for extended times (3 days) prior to thermal unfolding. In contrast, an increase in the thermal midpoint was observed with increasing protein concentrations (data not shown). An overall similar trend is observed in urea-induced unfolding experiments under AFS conditions. V_L and C_H1 unfold at rather low urea concentrations (Figure 17 B and C, bottom; Table 3), whereas V_H, even though unfolding is not very cooperative, shows residual structure up to ∼6 M urea (Figure 17 A, bottom). C_L, being unfolded under AFS conditions, does not show any defined transitions (data not shown). Taken together, all domains of the Fab fragment that adopts an AFS show significant thermodynamic stability in this state. In particular, V_H is more stable in the AFS than in the native state.

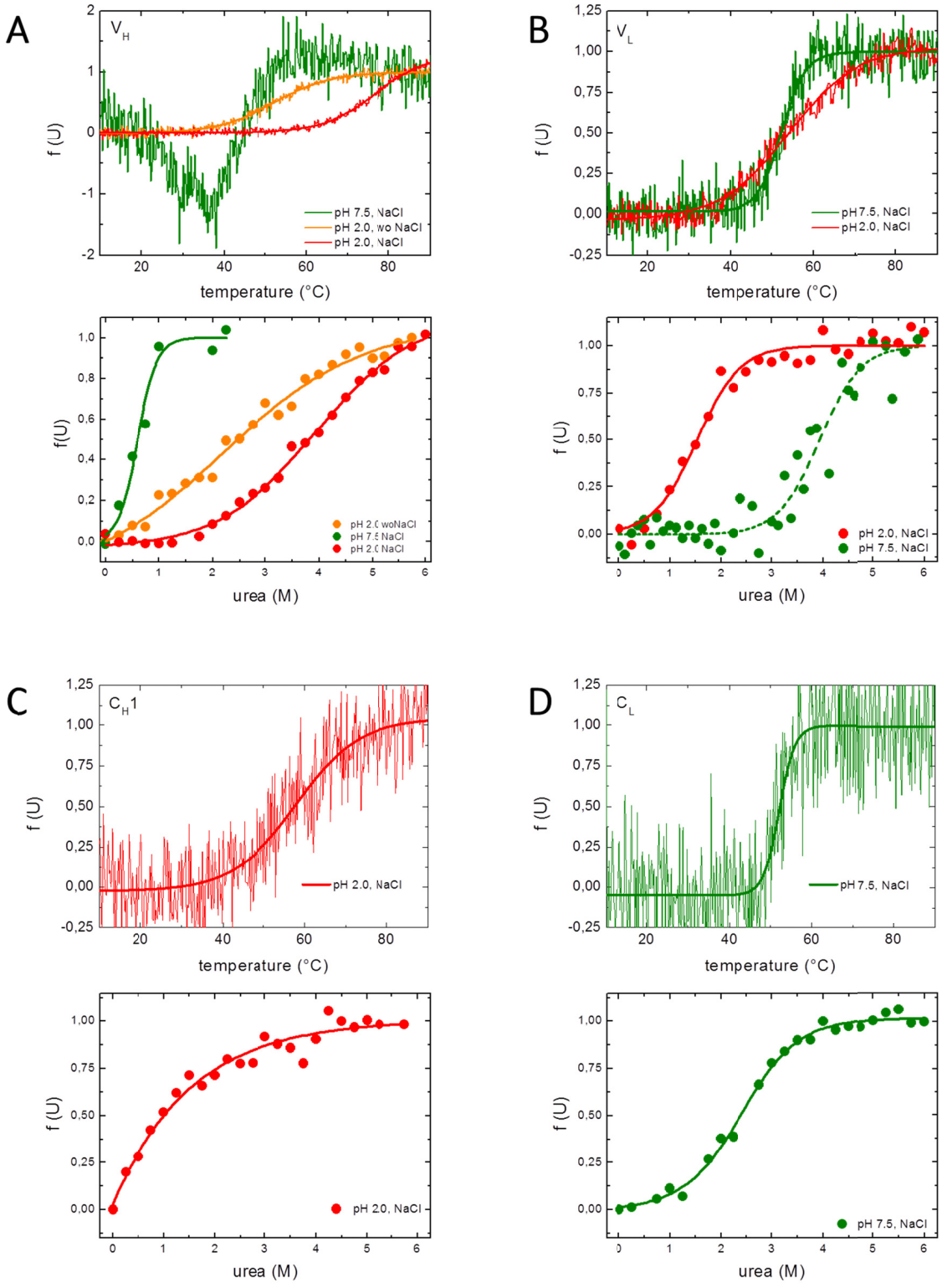

Figure 17: Stability of individual Fab domains in their native state and in the AFS.

To assess the stability of the different Fab domains under physiological and AFS solvent conditions, temperature-induced (top) and urea-induced (bottom) unfolding experiments were performed. Data for V_H (A), V_L (B), C_H1 (C), and C_L (D) are shown. Temperature-induced unfolding transitions are shown in green/ green lines for physiological conditions and in red/red lines for AFS conditions. For urea melts, green lines/green circles denote data under physiological conditions [50 mM sodium phosphate (pH7.5) and 100 mM NaCl], and red lines/red circles denote data under AFS conditions [50 mM sodium phosphate (pH 2.0) and 100 mM NaCl]. For V_H additionally the thermal and urea-induced transitions in orange lines/orange circles are shown.

Temperature-induced unfolding was monitored by FUV-CD spectroscopy at a fixed wavelength with a heating rate of 20°C h^{-1}. Data were fitted to a Boltzmann function to obtain transition midpoints. Chemical denaturation of the proteins was induced by urea, and structural changes were monitored by FUV-CD spectroscopy at a fixed wavelength. Data were evaluated according to a two-state unfolding model to obtain midpoints and cooperativity parameters of the transitions. Measurements were performed at 25°C at a protein concentration of 20 µM (temperature-induced transitions) or 10 µM (urea-induced transitions). As C_H1 is intrinsically disordered under physiological conditions and as C_L does not populate an AFS, no data could be obtained for these domains under the respective conditions.

Table 3: Thermal and chemical stabilities of different Fab domains.

	pH 7.5, 100 mM NaCl			pH 2.0, 100 mM NaCl		
Domain	**T_{melt} [°C]**	**$D_{1/2}$ [M]**	**m [kJ mol^{-1} M^{-1}]**	**T_{melt} [°C]**	**$D_{1/2}$ [M]**	**m [kJ mol^{-1} M^{-1}]**
C_L	52.1 ± 0.4	2.4 ± 0.3	4.5 ± 0.5	NA	NA	NA
V_L	52.3 ± 0.2	3.9 ± 3.4	6.0 ± 2.0	53.9 ± 0.3	1.5 ± 0.02	6.1 ± 0.4
C_H1	NA	NA	NA	57.8 ± 0.9	~ 1.0	NA
V_H	NA	0.6 ± 0.08	13.8 ± 4.5	76.5 ± 0.3 (52.0 ± 0.3)	3.3 ± 2.1 (2.2 ± 0.7)	2.0 ± 0.4 (1.9 ± 0.4)

Stabilities against the thermal and chemical denaturation (urea) of the Fab domains under physiological conditions [50 mM sodium phosphate (pH 7.5) and 100 mM NaCl] and AFS conditions [50 mM sodium phosphate (pH 2.0) and 100 mM NaCl] are shown. Midpoints of thermal transitions are shown as T_{melt}. Even though none of the urea-induced unfolding transitions was reversible, data were evaluated according to a two-state equilibrium unfolding model to derive the midpoint of transitions ($D_{1/2}$), as well as the cooperativity parameter (m), for a qualitative comparison of the data. When aggregation occurred or when proteins were not structured under a particular solvent condition, data were not evaluated (NA). Values in brackets are for low salt conditions [50 mM sodium phosphate (pH 2.0)]

3.1.1.4 The Pathway to the AFS

Different mechanistic scenarios are conceivable for how the AFS is reached. The proteins could either directly reorganise their structure upon changes under the solvent conditions, completely unfold before the AFS is established, or populate other states *en route* to the AFS. To discriminate between these alternatives, kinetic FUV-CD experiments for the Fab domains that populate an AFS were performed.

For V_H, two kinetic phases were observed; one fast phase with an observable rate of $k_{obs}1$= 1.0 ± 0.2×10^{-2} s^{-1} and one slow phase with an observable rate of $k_{obs}2$= 1.1 ± 0.2×10^{-4} s^{-1} at 25°C (Figure 18 A, right). As FUV-CD spectra could only be measured after the fast phase had already been completed (Figure 18 A, left), no unambiguous statement can be made as to how V_H forms its AFS. In the absence of salt, however, formation of the AFS was much slower for V_H (k_{obs}= 4.0 ± 0.1×10^{-5} s^{-1}), proceeding through the unfolded state of the domain (Figure 19). For V_L, upon changing solvent conditions from pH 7.5 to pH 2, fast global unfolding was observed (Figure 18 B, left). Only subsequently in a slow process (k_{obs}= 3.9 ± 0.1×10^{-5} s^{-1} at 25°C) the AFS was formed (Figure 18 B, right). As C_H1 is intrinsically disordered, formation of the AFS has to start from an unfolded state (Figure 18 C, left). The

AFS forms with an observable rate of k_{obs}= 5.2 ± 0.5×10 s^{-1} at 25°C for C_H1, very similar to the value obtained for V_L and V_H in the absence of salt. Thus, surprisingly small differences in the overall slow rates of conversion were observed for the Fab domains populating an AFS. The only exception is V_H, which quite rapidly establishes an AFS in the presence of salt. The underlying mechanism seems to be the same for all domains where spectra could be obtained: the proteins first unfold partially or completely and then establish the AFS.

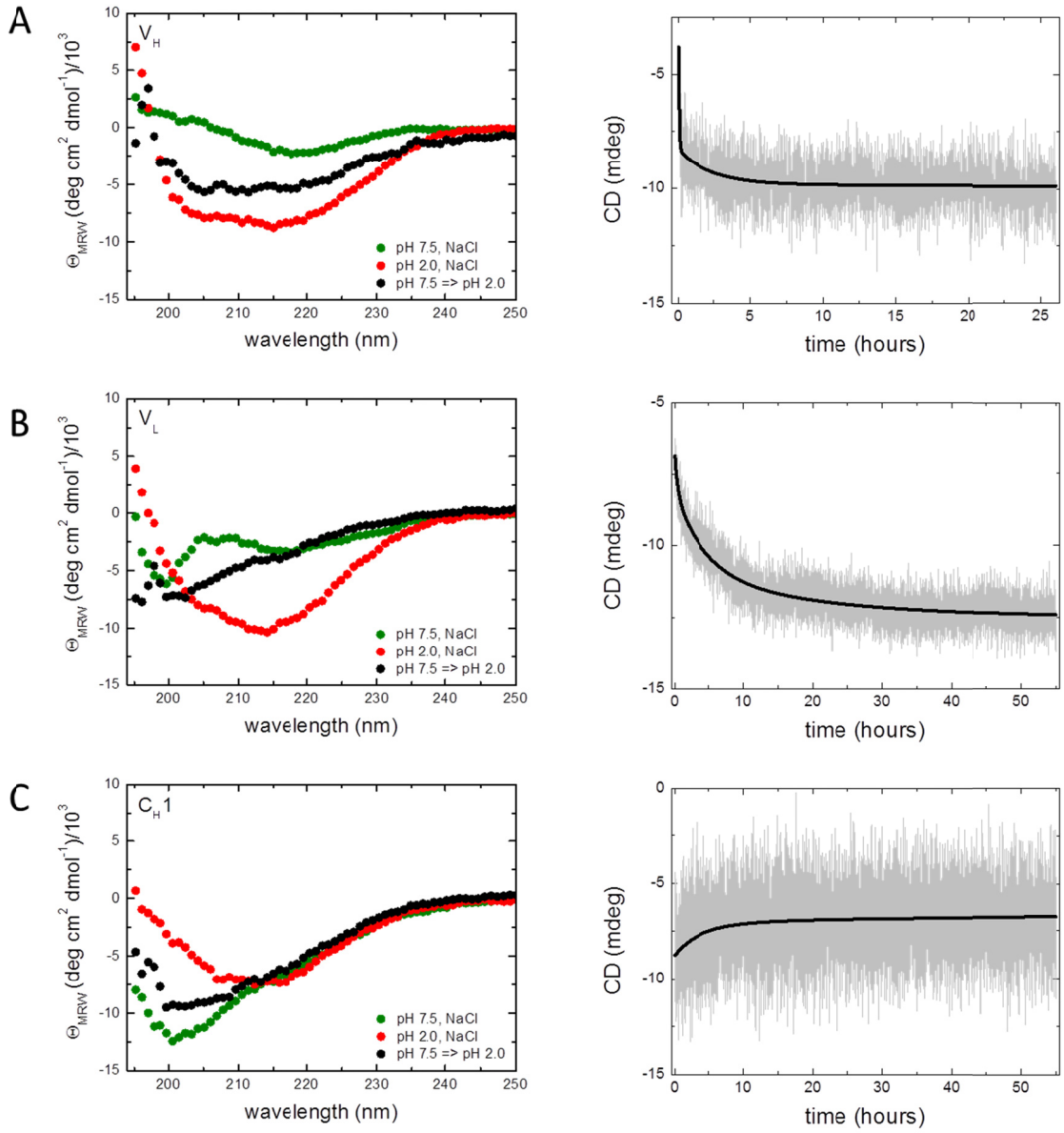

Figure 18: The pathway to the AFS for individual Fab domains.

To obtain structural insights into the pathway leading to the AFS, kinetic FUV-CD experiments for all domains that are able to form an AFS were performed: V_H (A), V_L (B), and C_H1 (C). Left: The FUV-CD spectra of the domains directly after a single jump from physiological conditions [50 mM sodium phosphate (pH 7.5) and 100 mM NaCl] to AFS conditions [50 mM sodium phosphate (pH 2.0) and 100 mM NaCl] are shown (black circles). The jumps were recorded with 20 nm s^{-1} scanning speed (1 accumulation), repeated 12 times and averaged. For comparison, the FUV-CD spectra under physiological conditions (green circles) and AFS conditions (red circles) are shown. Right: Changes in ellipticity observed at a constant wavelength (C_H1 at 205 nm; V_L and V_H at 220 nm)

after the transfer of the protein from physiological conditions to AFS conditions are shown. All measurements were performed in a 1-mm quartz cuvette at 25°C at a protein concentration of 10 µM.

The AFS formation of V_H was slower when only low ionic strength is present. One phase could be observed (k_{obs}= 1.5 ± 6.8x10^{-3} s^{-1} at 25°C; Figure 19, right) whereas for V_H in presence of 100 mM NaCl two phases were determined (Figure 18 A, right). However, the FUV-CD spectra after the pH jump indicate that the pathway is similar to V_H under AFS conditions sharing the presence of an unstructured intermediate on the way from the native state to the AFS (Figure 18 A and Figure 19, left).

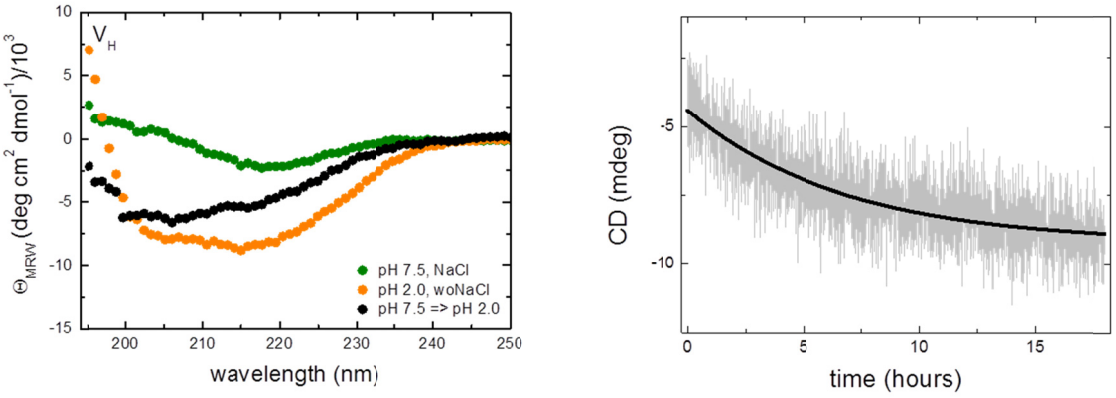

Figure 19: The pathway to the AFS for V_H under low salt conditions.
Left: The FUV-CD spectra of V_H directly after a single jump from physiological conditions [50 mM sodium phosphate (pH 7.5) and 100 mM NaCl] to low salt conditions [50 mM sodium phosphate (pH 2.0)] are shown (black circles). The jumps were repeated 12 times and averaged. For comparison, the FUV-CD spectra under physiological conditions (green circles) and low salt conditions (orange circles) are shown. Right: Changes in ellipticity observed at 220 nm after the transfer of the protein from physiological conditions to low salt conditions are shown. Measurements were performed in a 1-mm quartz cuvette at 25°C at a protein concentration of 10 µM.

3.1.1.5 Discussion

The structures adopted by antibodies at low pH are set apart from those of many other proteins [219-221] in that they are remarkably stable against unfolding and that they are oligomeric. In the original experiments performed with a complete IgG antibody [132], it was not clear whether all domains of the antibody contribute to this structural state. Subsequent studies on the Fab fragment revealed that this part of the antibody is capable of forming the AFS on its own [134]. Furthermore, the AFS could be induced also in the isolated light chain [140], a single chain fragment variable [222], and the C_H3 domain [133]. However, in these cases, slightly different conditions, such as pH values below pH 2 in the case of the light chain [140] or a certain ionic strength in the case of C_H3 [133], were necessary. Despite these variations in solvent conditions required to induce the AFS, the properties of the AFS seemed to be quite consistent for the different domains or fragments studied. However, the insights for individual domains were limited [133]. To expand our view of the AFS, we embarked on studying all four domains of the MAK33 IgG Fab fragment. The picture emerging from these

studies is that the domains show significantly different propensities to form the AFS. Three of the four domains adopt the AFS under low pH and intermediate ionic strength conditions, albeit with different characteristics. In contrast, the C_L domain unfolds at low pH and does not exhibit any detectable structure up to 175 mM NaCl. Only at higher ionic strength could C_L be poised to form secondary structure at pH 2. For C_H1, paradoxically, the isolated domain is completely unfolded at physiological pH [181], but forms a stable structure at pH 2. The variable domains both form an AFS, but with strikingly different characteristics. For V_H, which is unstable at pH 7.5, the most stable AFS of the Fab domains was observed, and spectroscopic evidence suggests that even defined tertiary structures exist in this case. Together with the data for the C_H1 domain, these results demonstrate that the stability of the native state is not directly correlated to the formation of an AFS. At present, it is not clear what sets C_L apart from the other domains in terms of its propensity to form an AFS. It should be noted, however, that C_L under acidic conditions and low ionic strength exposes hardly any hydrophobic surface (Figure 15). C_H1, which only forms smaller oligomers under AFS conditions (Figure 16), exposes only a small amount of hydrophobic surface (Figure 15), yet the variable domains, which form large oligomers under AFS conditions, expose significant hydrophobic surfaces (Figure 15). Formation and, in particular, stability of the AFS might thus partially depend on higher order assemblies induced by hydrophobic interactions. In this context, it should be noted that the AFS of the C_H1 domain is also the least stable against urea denaturation and that we observed a higher stability of the AFS against thermal denaturation with increasing protein concentrations. With higher concentrations the ratio between the different oligomeric species shifted but it did not change for three days (data not shown). This observation might explain the increase of the thermal stabilities with higher concentrations. Being diverse in sequence, the CDRs of the variable domains might play a role in AFS formation. A human V_L domain directed against HIV-1 [213] did not form an AFS under typical conditions (Figure 11), whereas we observed AFS formation for another murine anti-IL1β V_L domain [211] and a humanised anti-Her2 V_H domain [212]. The AFS formed of all variable domains investigated so far show very similar characteristics (Figure 10; Figure 11). Thus, the propensity to form an AFS and key structural features seem to be conserved for different variable domains, but might be tuned by the nature of the CDRs. It was speculated that the interactions in the oligomer(s) are required for the formation of a stably folded AFS [133, 134]. For the C_H1 domain, these interactions must be less pronounced than for the other domains. Thus, our results show that larger oligomers are not required *per se* for the formation and maintenance of an AFS, but domain interactions, which are still observed for the C_H1 domain, might nevertheless be a prerequisite for AFS formation.

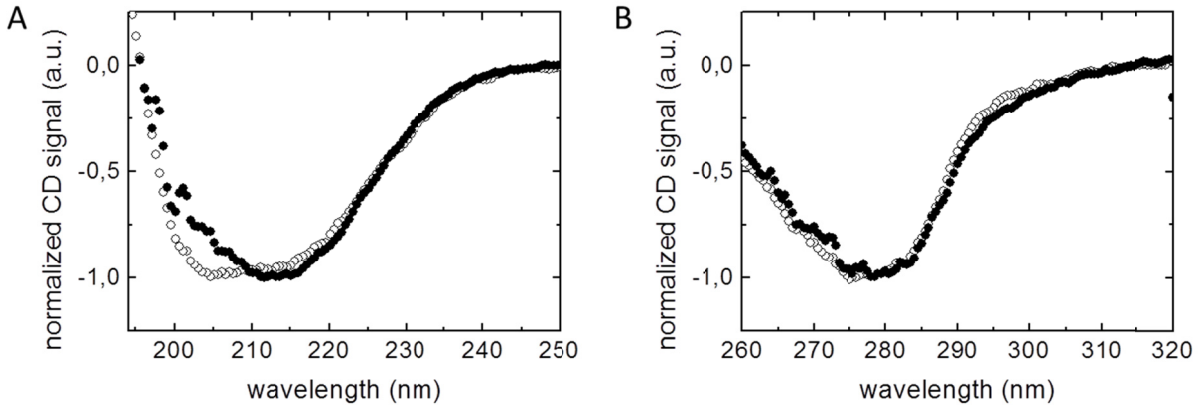

Figure 20: Comparison of the Fab fragment AFS to the sum of the single Fab domains.
To address the question whether the AFS adopted by the Fab fragment under acidic conditions can be described as the mere sum of the AFS of the single domains, FUV-CD (A) and NUV-CD (B) spectra were recorded. Closed circles represent the Fab; open circles the sum of the individual spectra of the single domains of the Fab fragment. For a better comparison the spectra were normalised to values between 0 and -1. The protein concentration was 50 µM for NUV and FUV-CD measurements. Spectra were recorded at 25°C

In this context, it is interesting to note that the NUV-CD spectrum of the AFS of the entire Fab fragment shows an identical shape to the sum of the spectra of the individual domains (Figure 20). It is thus dominated by the tertiary structure of V_H, the only domain found to form a defined tertiary structure under AFS conditions. Small differences, however, are observed for the FUV-CD spectrum. A slightly smaller random coil content is observed in the Fab fragment AFS spectrum in comparison to the sum of the individual domain spectra (Figure 20). This either might argue for rigidification of the AFS secondary structure in the context of the Fab fragment or the C_L domain, which is unfolded at low pH and a similar ionic strength that has been used to induce Fab or IgG AFS formation [132, 140], might gain some structure due to the induced proximity to other domains in the context of the Fab fragment.

The FUV-CD spectra of the AFS of all domains and fragments studied so far exhibit larger amplitudes than in the native state and a different shape of the spectra. This suggests that significant structural rearrangements take place upon its formation in comparison to the native state. A major cause for the observed alterations in the FUV-CD spectra are most likely secondary structure changes in the Fab domains under AFS conditions. Additionally, in particular for β-sheet proteins, a pronounced influence of aromatic residues on the FUV-CD spectra has been reported in several cases [131, 223, 224]. Changes in the molecular environment of the aromatic residues, as observed by altered NUV-CD spectra under AFS conditions, thus likely also contribute to the differences in the FUV-CD spectra. Our kinetic studies reveal that, generally, the protein has to pass through a partially or completely unfolded state to reach the AFS. This implies that the AFS may represent not just a rearranged native-like topology but a completely remodelled structure. Propensities to form an AFS and its biophysical features are therefore most likely encoded in the sequence and not in the structure of the individual domains. The theoretical pI of the individual domains ranges from 5.3 for V_H, via 5.9 for C_L and 6.7 for V_L, to 7.8 for C_H1 (Table 4). Furthermore, the

number and distribution of acidic and basic amino acids are very similar for the individual Fab domains (e.g., for V_H and C_L), which behave completely differently in terms of AFS formation. The number of hydrophobic amino acids, though, is slightly lower in C_L than in the other Fab domains, which might partially account for the differences in AFS formation. Conclusive evidence about the defining features of the AFS, however, cannot be deduced from the present spectroscopic analysis. NMR techniques, including solid-state NMR, could shed further light on this issue.

Table 4: Theoretical pIs of the Fab domains.

	V_L	C_L	V_H	C_H1
theoretical pI	6.7	5.9	5.3	7.8

Theoretical pIs were calculated from the amino acid sequences determined using ProtParam.

Taken together, our studies on the Fab domains expand our view on the AFS. The common theme is the remarkable stability of this low-pH structure possibly due to hydrophobic intermolecular interactions. Consistent with a complete rearrangement of the secondary structure elements, the antibody domains have to pass through a partially or completely unfolded state to reach the AFS. It will be interesting to identify which sequence elements support the formation of the AFS, whether it is restricted to antibodies or whether it can be found in other members of the immunoglobulin superfamily and maybe even beyond.

3.1.2 AFS formation of C_H2

C_H2 is the last domain of MAK33 to be analysed for the ability to adopt an AFS. As this domain contains a conserved glycosylation motif (Asn-Xaa-Ser/Thr) [153] the influence of the glycan should be included within this study. Four different C_H2 variants will be considered: refolded C_H2 wild type from *E. coli*, glycosylated C_H2 from *Sf9* expression, nonglycosylated C_H2 from *Sf9* and a C_H2 control construct (refolded from *E. coli* expressed inclusion bodies). The control construct contains the additional amino acids resulting from *Sf9* expression system (signal peptide for secretion into the medium and TEV restriction site to remove the His_6-tag; see Figure 21) and is refolded from inclusion bodies (*E. coli* expression).

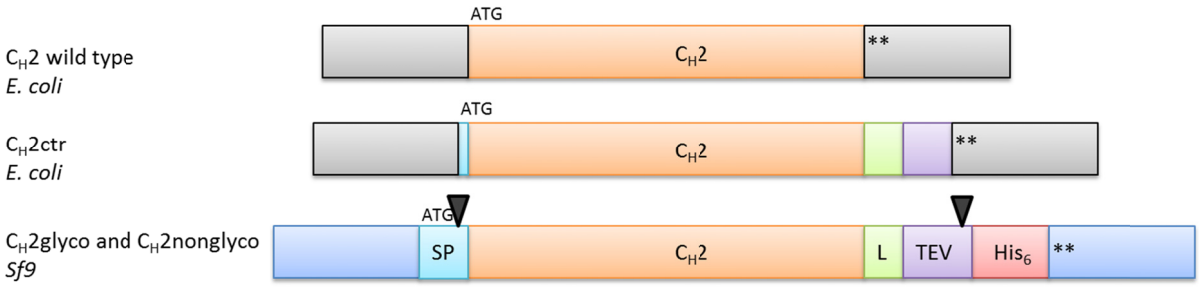

Figure 21: C_H2 constructs to analyse the ability of AFS formation.
MAK33 wild type C_H2 consists of 99 amino acids and was cloned in pET28-b (black) for expression in E. coli (top). For the expression in Sf9 cell lines C_H2 was cloned into pAcGP67-B (blue, bottom). The plasmid contains a gp67 signal sequence for secretion of the target protein (light blue) which is cleaved during secretion leaving ADLGSA at the N-terminus of the target. For purification a short SGSG-linker (light green), a TEV-restriction site (purple), and a His_6-tag (red) were introduced at the C-terminus. Asterisks indicate stop-codons. After TEV cleavage GSGSENLYFQ remains at the C- terminus. A control for expression in E. coli was designed containing the additional amino acids at the N- and C-terminus to be able to determine the effects of those additional residues (middle).

3.1.2.1 Glycan analysis of glycosylated C_H2

PNGase F can remove almost all types of Asn-linked (N-linked) glycosylation: high mannose, hybrid, bi-, tri- and tetra-antennary. The enzyme is suitable to remove all N-linked carbohydrates. Endo H is more specific and removes only high mannose and some hybrid types of N-linked carbohydrates. It is used to define more closely the type of N-linked glycosylation. To determine the amount of glycosylation within C_H2, *Sf9* expressed protein (before the Concanavalin A column) and *E. coli* expressed protein (wild type, refolded) were treated with Endo H and PNGase F according to the manufacturer`s protocol.

Sf9 expressed C_H2 shows three different bands which result from glycosylated and unglycosylated protein. Endo H treatment shows no influence on C_H2 wild type or *Sf9* expressed C_H2 (Figure 22, bottom). However, PNGase F cleaves the upper band from *Sf9* expressed C_H2, thus this band contains N-linked carbohydrates (Figure 22, top). Therefore with the expression of C_H2 in *Sf9* cells in one approach unglycosylated and glycosylated protein can be obtained. Glycosylated and nonglycosylated C_H2 could successfully be separated from each other by a Concanavalin A column (GE Healthcare; not shown).

A second approach to control glycosylation is a glycol-staining approach. Attempts to stain glycosylated C_H2 with the Pro-Q Emerald 488 Glycoprotein Stain Kit (Invitrogen) failed, as the dye did not only stain the glycosylated C_H2 band but also C_H2 wild type from *E. coli* expression used as a control. Posttranslational modifications, such as glycosylation, are not possible in *E. coli* BL21 thus binding to the control must be unspecific and prevents any conclusion.

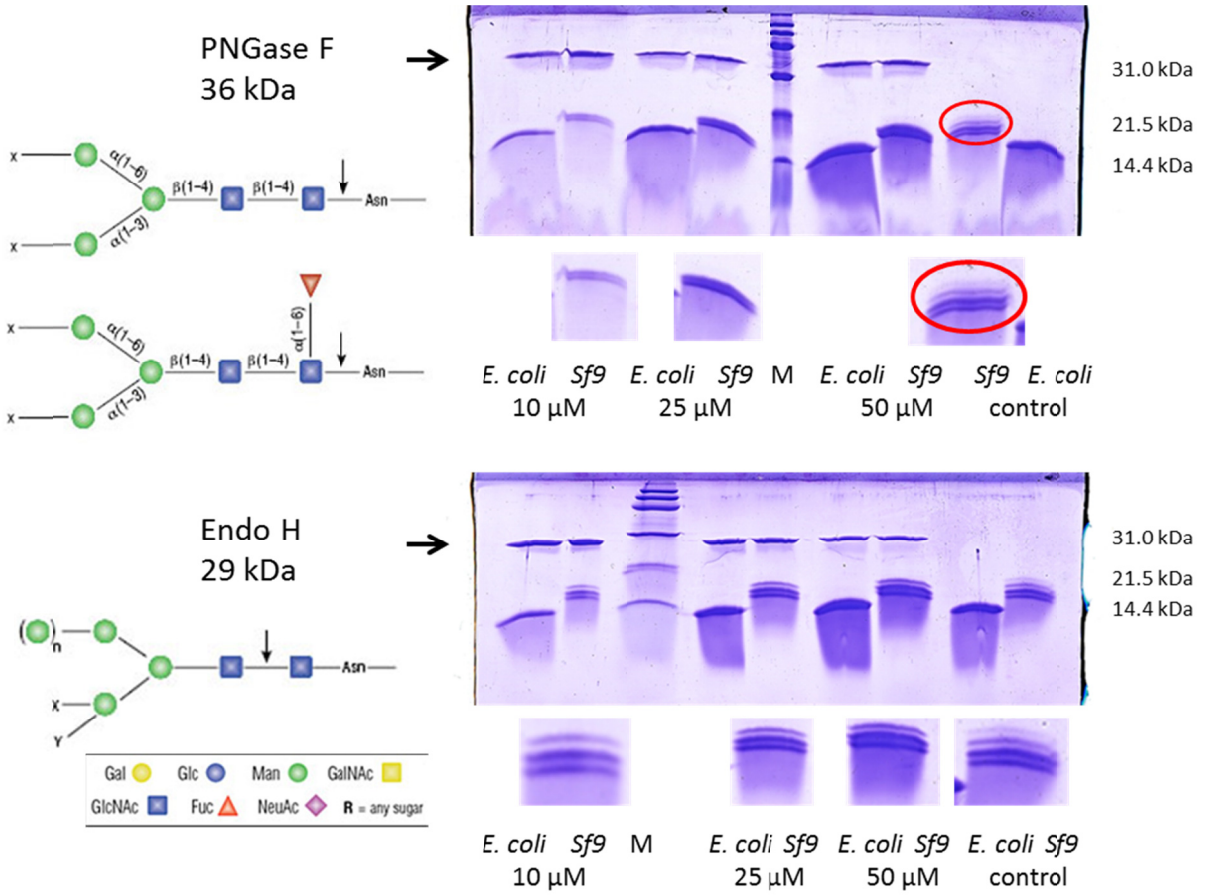

Figure 22: Test for glycosylation of C_H2 with endoglycosides.
Sf9 (before the Concanavalin A column) and E. coli expressed C_H2 was treated with PNGase F and Endo H to control the amount of glycosylated protein. Three different concentrations of protein (10 µM, 25 µM, 50 µM) were tested. The E. coli expressed C_H2 construct is not affected by any of the enzymes (top and bottom). For Sf9 expressed C_H2 treated with PNGase F, the upper band disappears which is present in the untreated control (red circle highlights the band which is not present after PNGase F treatment). For Endo H treatment no differences between control and treated samples could be observed (all 18% SDS-gels, Sf9 protein bands are enlarged at the bottom of each gel).

Mass spectrometry (MALDI-ToF) of the purified glycosylated C_H2 domain has shown that the mass difference between C_H2 glyco and C_H2 nonglyco is ~1040 Da indicating that the sugar part of C_H2 glyco consists of two N-acetylglycosamines and three mannoses.

3.1.2.2 Glycosylation does not influence the ability to adopt an AFS

In antibodies the Fc part is glycosylated (see 1.7.1) and numerous well-defined sugar-protein, protein-protein, and also sugar-sugar interactions determine the final complex structure. Under physiological conditions, C_H2 can adopt an immunoglobulin fold independent of glycosylation [154, 203]. Dimerization of the Fc part is mainly driven by C_H3, which also forms homo-dimers in isolation [133]. On the contrary, both C_H2 domains have no direct contact with each other and the oligosaccharides are positioned at the inner face of the dimer. The crystal structure of the unglycosylated Fc part of MAK33 has shown that a lack of oligosaccharides has no influence on the C_H3 conformation but some changes for C_H2

have been observed [154]. First, the distance between both C_H2 domains is decreased ("closed" conformation) and second, the C`E loop as well as the BC and FG loops show more mobility, although they are still well defined [154]. The thermodynamic stability of the unglycosylated Fc fragment was slightly decreased compared to the glycosylated counterpart and behaved less cooperative (in thermal- and GdmCl-induced transitions) [154].

Figure 23 A shows an alignment of the unglycosylated Fc fragment of MAK33 with a human glycosylated counterpart (PDB ID 1H3X). The overall structures of the two C_H2 and two C_H3 domains fit quite well. All four tested C_H2 variants (wild type, glycosylated, nonglycosylated and control) show defined FUV-CD spectra under physiological conditions [50 mM sodium phosphate (pH 7.5) and 100 mM NaCl] with two minima at 230 nm and 210 nm and two maxima at 222 nm and 200 nm which differ from the classical β-sheet FUV-CD spectrum with a minimum at 218 nm and a maximum at 200 nm. The observed FUV-CD spectra are typical for isolated C_H2 and indicate a well-defined secondary structure [225] (Figure 23 B-E; green). Under conditions that have been shown to induce AFS formation of antibody domains [133, 192] [50 mM sodium phosphate (pH 2.0) and 100 mM NaCl] all variants show defined secondary structure (Figure 23 B-EC; red). Wild type C_H2 shows a minimum at 212 nm characteristic for Fab domains (app. 215 nm minimum, see 3.1.1.1) [192]. The minima of C_H2 glycosylated, nonglycosylated and the control are further in the FUV range at ~208 nm (Figure 23 C-E, red). The control construct and the unglycosylated C_H2 behave similar and confirm that the refolding of inclusion bodies worked. The spectra differ clearly from chemically denatured proteins which show typical random coil characteristics with a minimum at 200 nm (data not shown).

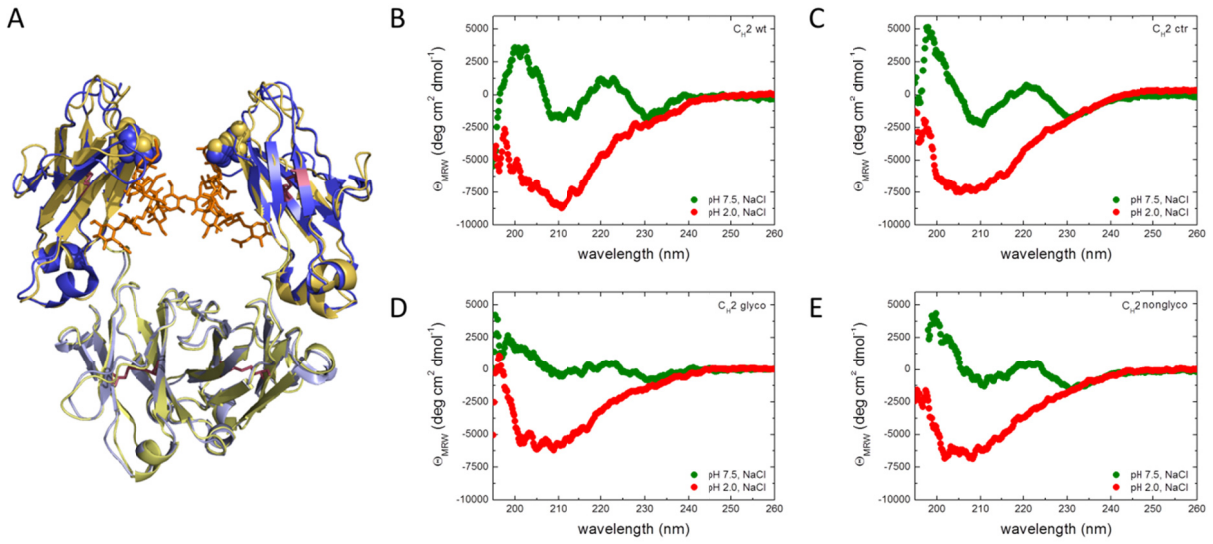

Figure 23: Structure of the different C_H2 domain constructs and secondary structure of the domains under physiological and AFS conditions.

In (A), an alignment of the structure of the unglycosylated Fc (PDB ID 3HKF, C_H2 in blue, C_H3 in pale blue) with its human glycosylated counterpart (PDB ID 1H3X, C_H2 in yellow, C_H3 in pale yellow) is shown. The conserved asparagine residues are highlighted as spheres, the internal disulphide bridges in raspberry. Oligosaccharides

are depicted as sticks (orange). (B-E) shows the FUV-CD spectra of the C_H2 variants - C_H2 wild type (B), C_H2 control (C) C_H2 glyco (D),and C_H2 nonglyco (E) - measured under physiological conditions [50 mM sodium phosphate (pH 7.5) and 100 mM NaCl] (green) and under AFS conditions [50 mM sodium phosphate (pH 2.0) and 100 mM NaCl] (red). All measurements were performed at a protein concentration of 20 µM at 25°C.

The ionic strength of the solvent and the internal disulphide bridge in immunoglobulin domains and fragments had been reported to influence the formation of the AFS [132-134, 140]. Additionally, for the Fab domains it could be shown that some are able to adopt the AFS under acidic low salt conditions (V_H; Figure 12 and [192]) whereas AFS formation of V_L, C_H1 and C_L was not possible in the presence of low ionic strength (Figure 12 and [192]). AFS formation of all C_H2 variants is dependent on the presence of salt, too, as all show random coil characteristics with a typical minimum at 200 nm in acidic buffer conditions without salt addition [50 mM sodium phosphate (pH 2.0)] (Figure 24).

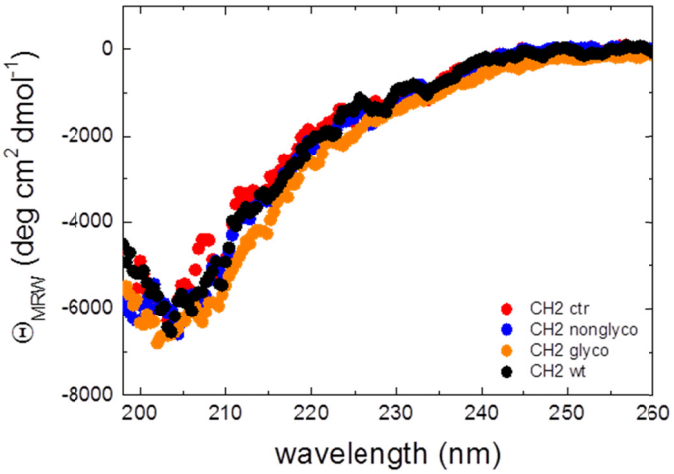

Figure 24: All C_H2 variants cannot adopt secondary structure at acidic pH in the absence of salt.
Under acidic conditions without the presence of salt [50 mM sodium phosphate (pH 2.0)] all four C_H2 variants show FUV-CD spectra typical for random coil structures. C_H2 wild type (black circles), C_H2 nonglyco (blue circles), C_H2 glyco (orange circles) and C_H2 control (red circles) are shown. Measurements were performed at a protein concentration of 20 µM at 25°C.

On the other hand, an increase of the salt concentration up to 200 mM had no remarkable effect on those Fab domains that were already able to adopt an AFS in the presence of 100 mM NaCl. However, C_L was able to adopt secondary structure in the presence of higher ionic strength what was not observed for lower salt concentrations (Figure 12 and [192]). An increase of the salt concentration up to 200 mM NaCl had only minor effects on the secondary structure formation of the four C_H2 variants under AFS conditions (Figure 25).

C_H2 wild type (black circles), C_H2 control (red circles), C_H2 glyco (orange circles) and C_H2 nonglyco (blue circles) show still the typical minimum at 210-208 nm observed for AFS conditions under acidic conditions with up to 200 mM salt (Figure 25 A and B). However, the signal intensity is increased compared to AFS conditions what might be a hind for a higher degree of secondary structure formation. As no chemical or thermal stability increase could be observed for C_H2 wild type in the presence of 200 mM NaCl (data not shown) it can be assumed that the changes are only small.

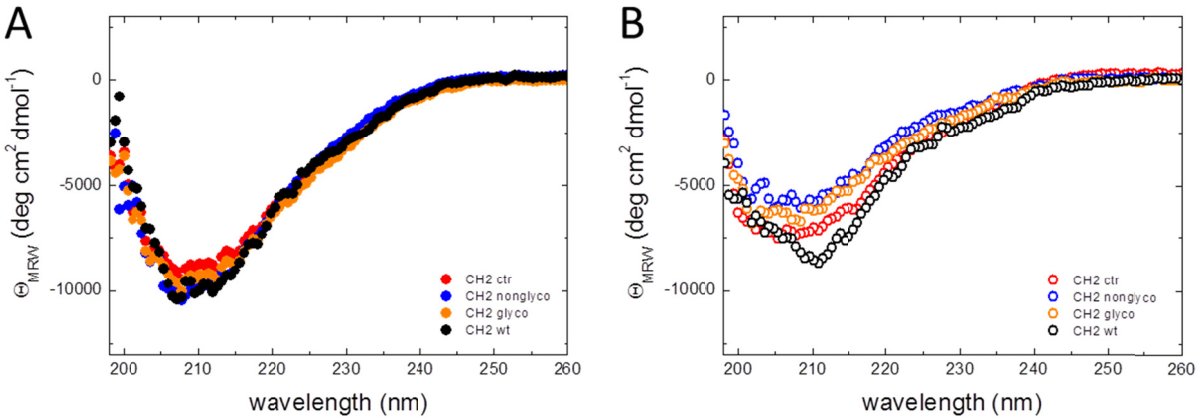

Figure 25: Higher ionic strength at acidic conditions has only slight effects on secondary structure formation.

To test the salt dependence, all four C_H2 variants were analysed under AFS conditions [50 mM sodium phosphate (pH 2.0) and 100 mM NaCl] or in the presence of 200 mM NaCl. In (A) the FUV-CD spectra for C_H2 wild type (black circles), C_H2 nonglyco (blue circles), C_H2 glyco (orange circles) and C_H2 control (red circles) under AFS conditions with 200 mM NaCl are shown. For a better comparison in (B) the same domains under our standard AFS conditions are depicted. Colour code is same as in (A). Protein concentrations used were 20 µM (25°C).

Disulphide bridges have been shown to influence AFS formation of the MAK33 Fab fragment and light chain [134, 140]. On single domains level (Fab domains) no influence of the presence of TCEP on either the native nor the AFS state could be observed (Figure 12 and [192]). To test the influence of the conserved internal disulphide bond within C_H2 FUV-CD spectra under physiological and AFS conditions for the four variants were recorded. Figure 26 shows the spectra under physiological (A) and AFS (B) conditions. For all variants no influence of the internal disulphide bridge neither on the native nor the AFS could be detected. Under AFS conditions (Figure 26 B) a slight increase of the signal intensity compared to the oxidised proteins for C_H2 glyco and C_H2 nonglyco is observable. However, the signal difference is very small and slight differences in the protein concentration determination could have influenced the result.

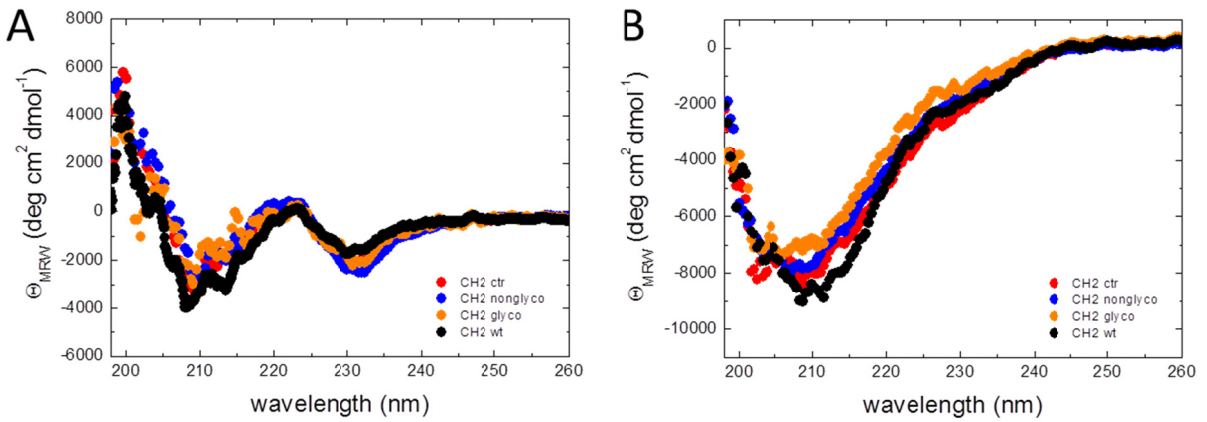

Figure 26: The presence of the internal disulphide bridge has no influence on secondary structure formation.

To analyse the effects of the internal disulphide bridge on the secondary structure formation of the four C_H2 variants, FUV-CD spectra under physiological conditions [50 mM sodium phosphate (pH 7.5) and 100 mM NaCl] (A) and AFS conditions [50 mM sodium phosphate (pH 2.0) and 100 mM NaCl] (B) in presence of 1mM TCEP were recorded. C_H2 wild type (black circles), C_H2 nonglyco (blue circles), C_H2 glyco (orange circles) and C_H2 control (red circles) are shown. Measurements were performed at a protein concentration of 20 µM at 25°C.

For most domains an important ability of the AFS is the reversibility of its formation when the buffer conditions change. For a low concentration (10 µM) all Fab domains adopted an AFS reversibly (V_H, V_L and C_H1; see Figure 13 and [192]). Indeed, the AFS formation was also reversible for all C_H2 variants within 10 µM protein concentration (Figure 27). C_H2 wild type was reversible even at higher concentrations, up to 100 µM without observable aggregation (data not shown).

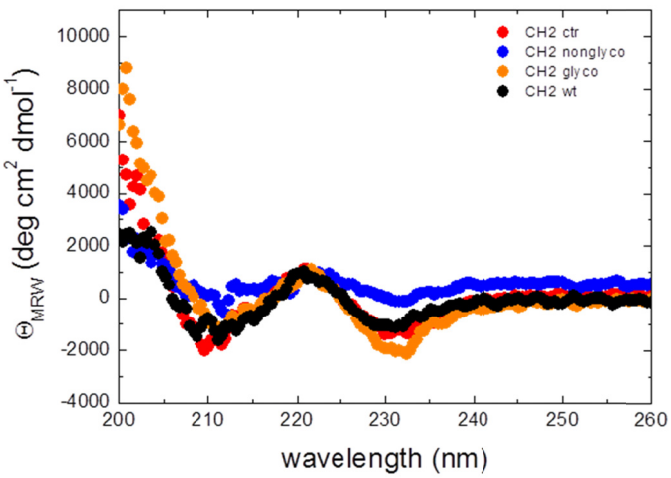

Figure 27: For all C_H2 variants the AFS is reversible.

To test reversibility, all C_H2 variants were dialysed overnight against 50 mM sodium phosphate (pH 2.0) and 100 mM NaCl and afterwards transferred into 50 mM sodium phosphate (pH 7.5) and 100 mM NaCl and dialysed overnight. The FUV-CD spectra for C_H2 wild type (black circles), C_H2 nonglyco (blue circles), C_H2 glyco (orange circles) and C_H2 control (red circles) are shown. Protein concentration was 20 µM (25°).

3.1.2.3 Structural characterisation of the AFS

To obtain a more detailed picture of structural changes that the C_H2 variants undergo upon formation of their AFS, NUV-CD spectroscopy and fluorescence spectroscopy were the methods of choice to monitor tertiary structural changes around the aromatic amino acids Phe, Tyr, and predominantly Trp.

Under physiological conditions, all domains, except for C_H2 control, show well-defined NUV-CD spectra indicative of a defined structure around the aromatic reporter groups with the typical minimum at 290 nm (Figure 28, top). The shape is similar to V_L and C_L NUV-CD spectra under physiological conditions (see Figure 14 and [192]) with the exception that C_H2 shows a small redshift from 280 nm to 290 nm within the minimum. The NUV-CD spectrum of C_H2 control is also indicating a well-defined tertiary structure but differs from the other C_H2 variants with a minimum at 290 nm but a signal increase in the range between 285 nm and 260 nm (Figure 28 B, top). This finding indicates that the environment around the aromatic amino acids is not exactly the same compared to C_H2 wild type, C_H2 glyco and C_H2 nonglyco. The difference between the control and wild type C_H2 can be explained through the additional amino acids which also include a phenylalanine (see Figure 21). However, this is no explanation for the differences compared to C_H2 glyco and C_H2 nonglyco as they have the same additional residues in common. The oligosaccharides bound to C_H2 glyco can have an influence on the conformation and thus influence the NUV-CD spectrum but C_H2 control and C_H2 nonglyco would be expected to behave the same. However, small changes in the environment of aromatic amino acids can have clear effects onto the NUV-CD fingerprint. This is probably the case here as there are no differences within the FUV-CD spectra visible (see Figure 23 B). The NUV-CD spectra at AFS conditions show a signal loss for C_H2 wild type and C_H2 control (Figure 28 A and B, top). Hence, there seems to be no defined tertiary structure present in these two variants. In contrast, for C_H2 glyco and C_H2 nonglyco a weak signal around 300 nm can be observed indicating that there might be a low amount of defined tertiary interactions (Figure 28 C and D, top). These structural arrangements are different compared to the native state. The fluorescence spectra of the different C_H2 variants demonstrate differences between the native and AFS state (Figure 28, bottom). C_H2 wild type shows a signal increase and a slight red-shift from 342 nm to 350 nm when the buffer conditions are changed from physiologic to AFS formation (Figure 28 A, bottom). The signal maximum at 342 nm indicates that the tryptophan residues are partially shielded from solvent. An increase in fluorescence was also observed for V_H and V_L when they adopt an AFS (Figure 14 A and B). C_H2 glyco and C_H2 nonglyco show similar behaviour with only a very small signal increase and a red-shift from 340 nm to 350 nm (Figure 28 C and D, bottom). For C_H2 control the same red-shift was observed as for C_H2 glyco and C_H2 nonglyco but the signal intensity decreases slightly (Figure 28 B, C, and D, bottom). The results indicate that the tryptophan residues are not completely solvent-exposed in the AFS state. C_H2 has, like the other antibody domains, a conserved tryptophan residue which is quenched in the native state by the internal disulphide bridge (Figure 23 A) within the hydrophobic core. The signal increase observed for C_H2 wild type and to a lesser extent for C_H2 glyco and C_H2 nonglyco (Figure 28 A, C, and D; bottom) argues for structural rearrangements moving the conserved tryptophan residue away from the internal disulphide bridge. The slight decrease in fluorescence intensity observed for C_H2 control might argue for other quenching reactions (e.g., due to solvent exposure) in the AFS (Figure 28 B; bottom).

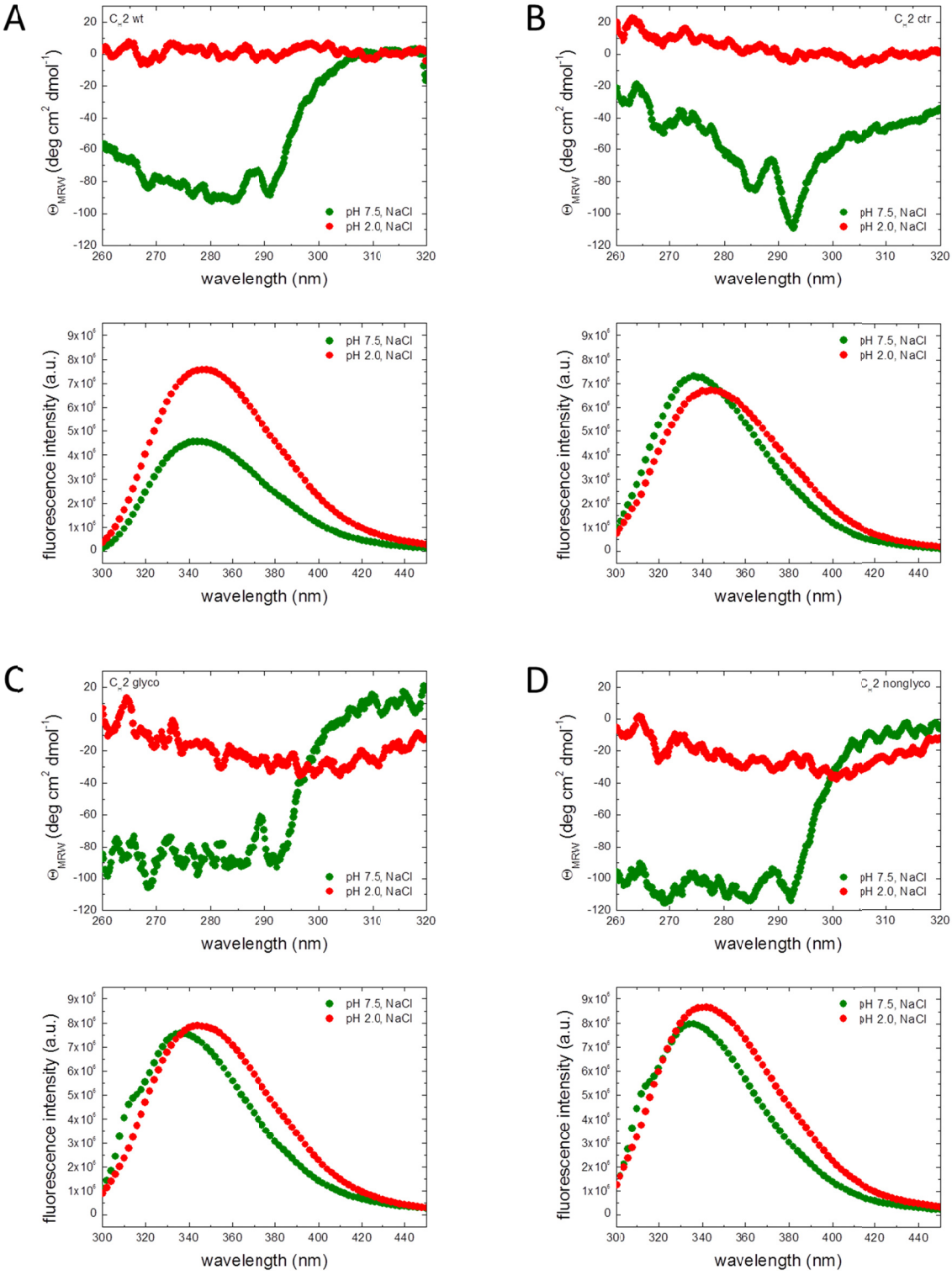

Figure 28: Tertiary structure of the different C_H2 variants under physiological and AFS conditions.
NUV-CD spectra (top) and fluorescence spectra (bottom) of C_H2 wild type (A), control (B), glyco (C), and nonglyco (D) were measured. Spectra were obtained under physiological conditions [50 mM sodium phosphate (pH 7.5) and 100 mM NaCl] (green), as well as under AFS conditions [50 mM sodium phosphate (pH 2.0) and 100 mM NaCl] (red). For NUV-CD measurements, a protein concentration of 50 µM was used; for fluorescence measurements, a protein concentration of 1 µM was applied. Protein fluorescence was excited at 280 nm. All measurements were performed at 25°C.

To gain some information about the hydrophobicity at the protein surface, ANS binding was determined. Under physiological conditions for none of the C_H2 variants an ANS signal could be observed, thus no hydrophobic patches were exposed (Figure 29 A-D). The fluorescence signal clearly increases for all four C_H2 variants under AFS conditions (Figure 29 A-D), which indicates a significant increase in the exposed hydrophobic surface. The amount of signal increase is comparable between C_H2 wild type, C_H2 nonglyco and C_H2 control (maximum at 450 nm is in the range of 3.7 x10^7; Figure 29 A, B, and D). The signal increase for C_H2 glyco is lower than for the other three variants with a maximum at 450 nm of 3.1 x10^7 (Figure 29 C). The magnitude order for all four variants is within the range of ANS binding observed for C_H1 (Figure 15 C).

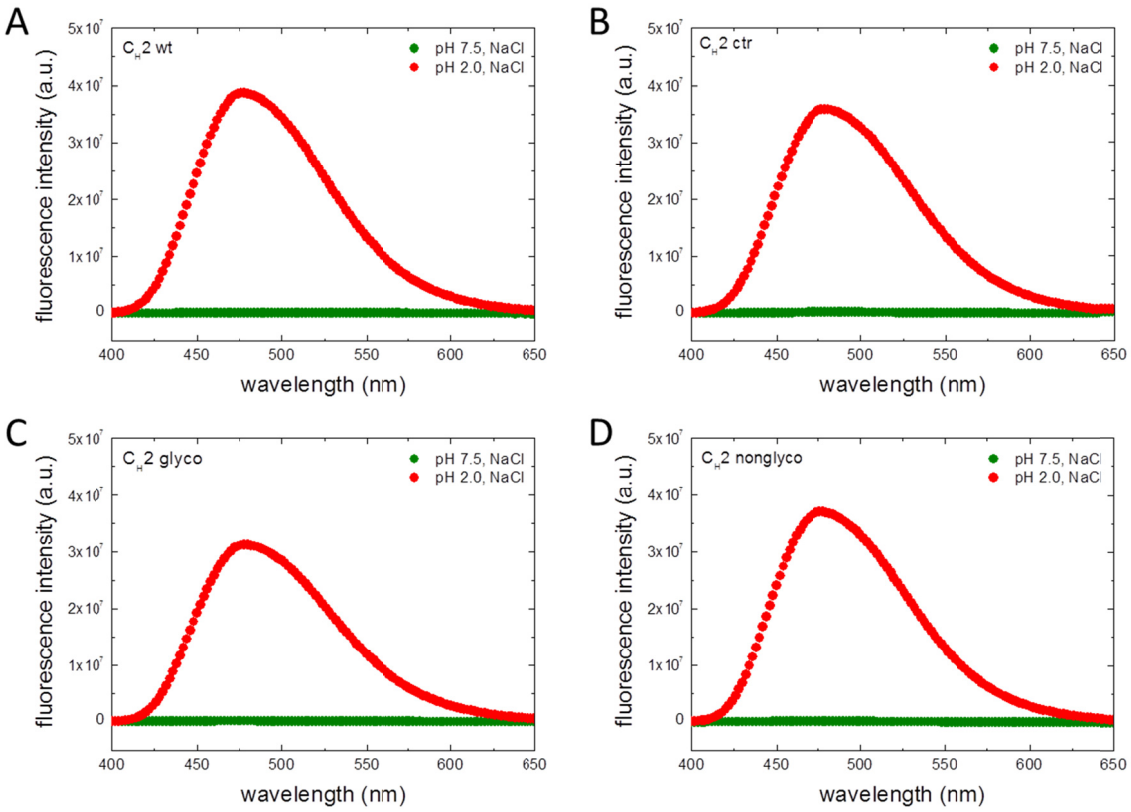

Figure 29: Exposure of hydrophobic surface upon AFS formation.

To analyse whether AFS formation leads to an increase of exposed hydrophobic surface, ANS binding assays were performed. ANS fluorescence spectra of the different C_H2 variants – wild type (A), control (B), glyco (C) and nonglyco (D) - under physiological conditions (50 mM sodium phosphate, pH 7.5, 100 mM NaCl) (green) and AFS conditions (50 mM sodium phosphate, pH 2.0, 100 mM NaCl) (red) in the presence of 100 µM ANS were measured. Protein concentrations were 10 µM for each domain. Excitation was set to 380 nm. Measurements were performed at 25°C.

Higher-order protein assemblies within the AFS were observed for C_H3 [133], V_H, V_L and C_H1 (Table 2 and [192]). Therefore sedimentation velocity analytical ultracentrifugation analysis of the different C_H2 variants was performed and revealed that all variants are monomeric under physiological conditions (Table 5). Under AFS conditions no higher-order oligomers were observed for C_H2 wild type and C_H2 nonglyco (Table 5). C_H2 glyco shows, besides ~40%

monomeric species, also ~60% species within a range of 25-60 kDa (Table 5). This size corresponds roughly to 2-4 monomers. C_H2 control is the second domain which forms higher oligomeric species. The main species at AFS conditions is the monomer (~53%) but dimers/tetramers are also formed (~35%). Furthermore, C_H2 control is the only domain which forms to a small amount even larger oligomers in a range of 60-80 kDa corresponding to 5-6 monomers (~12%) Interestingly, C_H2 glyco was the domain with the smallest signal increase upon ANS binding (Figure 29 C). Thus this interaction is probably not driven by hydrophobic interactions. This assumption is confirmed by the different behaviour between C_H2 nonglyco and C_H2 control which showed the same signal increase in the ANS binding assay but differ in their ability to adopt higher order oligomers. For the Fab domains a correlation between hydrophobicity at the surface and oligomerisation seemed to exist (see Figure 15 and Table 2). The only difference between C_H2 glyco and C_H2 nonglyco lies within the glycan, thus the interaction might be driven by oligosaccharides. On the contrary, C_H2 control is also able to form oligomers. In conclusion sugar moieties cannot be the only driving force for oligomerisation.

Table 5: Quaternary structure of the different C_H2 variants.

Domain	Buffer conditions	Monomer [%]	25-60 kDa [%]	60-80 kDa [%]
$C_H2_{wild\ type}$	pH 7.5, 100 mM NaCl	100	-	
	pH 2.0, 100 mM NaCl	100	-	
C_H2_{ctr}	pH 7.5, 100 mM NaCl	100	-	
	pH 2.0, 100 mM NaCl	~ 53	~ 35	~12
C_H2_{glyco}	pH 7.5, 100 mM NaCl	100	-	
	pH 2.0, 100 mM NaCl	~ 40	~ 60	
$C_H2_{nonglyco}$	pH 7.5, 100 mM NaCl	100	-	
	pH 2.0, 100 mM NaCl	100	-	

Sedimentation velocity analytical ultracentrifugation was used to determine the oligomeric state of the different C_H2 variants under physiological conditions [50 mM sodium phosphate (pH 7.5) and 100 mM NaCl] and AFS conditions [50 mM sodium phosphate (pH 2.0) and 100 mM NaCl]. The observed quaternary structure was grouped into three categories.

3.1.2.4 Stability of the AFS of the C_H2 variants

The main difference between the AFS and molten globule state is the high stability against temperature and denaturants and the often observed cooperative behaviour (see 1.6.3.2). To characterise the stabilities of the four C_H2 variants, thermal and urea-induced unfolding transitions were performed. Thermal-induced unfolding was monitored by FUV-CD spectroscopy; in contrast, urea-induced unfolding was monitored by fluorescence spectroscopy as the signal-to-noise ratio was too low for the domains in AFS conditions to be determined by CD measurements. Therefore for the urea-induced transitions no secondary structure changes but rearrangements within the environment of the aromatic amino acids are observable. Due to the low signal intensity, triplicates were measured and averaged to improve the signal quality. Reversibility of the transitions was not tested and thus thermodynamic stabilities could not be determined.

In Figure 30 A the thermal transitions of C_H2 wild type (in black), C_H2 control (in red), C_H2 glyco (in orange) and C_H2 nonglyco (in blue) are shown. C_H2 wild type is the least stable variant with a T_{melt}= 42.5 ± 0.2°C. C_H2 nonglyco and C_H2 control have comparable melting temperatures with T_{melt}= 64.5 ± 0.2°C and T_{melt}= 65.6 ± 0.2°C, respectively. For C_H2 glyco the melting temperature is even higher with a T_{melt}= 68.4 ± 0.3°C (Figure 30 A and Table 6). Glycosylation is known to influence protein stability [149], thus the stabilising effect of approximately 3.5°C is probably a result of the influence of the oligosaccharides. The strong thermal stability increase of ~20°C between C_H2 wild type and C_H2 nonglyco (Figure 30 A and Table 6) must be caused by the additional amino acids at the N- and C-terminus of C_H2 nonglyco and is not a refolding artefact as C_H2 control has the same stability compared to C_H2 nonglyco (Figure 30 A and Table 6).

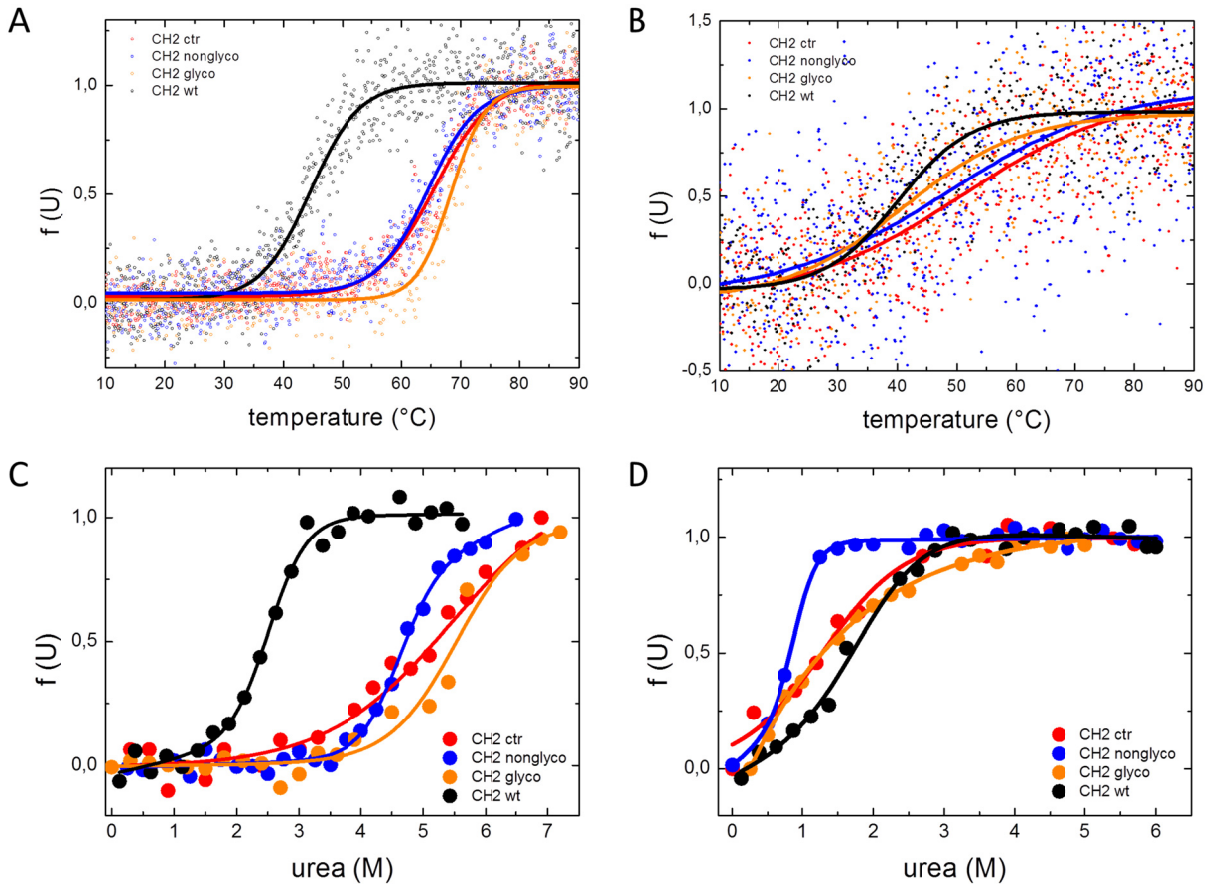

Figure 30: Stability of the four C_H2 variants in their native state and in the AFS.

To assess the stability of the C_H2 variants under physiological and AFS solvent conditions, temperature-induced (top) and urea-induced (bottom) unfolding experiments were performed. Data for thermal-induced (A and B) and urea-induced transitions (C and D) are shown. In (A) temperature-induced unfolding transitions for physiological conditions [50 mM sodium phosphate (pH 7.5) and 100 mM NaCl] for C_H2 wild type (black line and circles), C_H2 glyco (orange line and circles), C_H2 nonglyco (blue line and circles), and C_H2 control (red line and circles) are depicted and in (B) the same samples for AFS conditions [50 mM sodium phosphate (pH 2.0) and 100 mM NaCl]. For urea melts, the transitions for the C_H2 variants under physiological conditions (C) and for AFS conditions (D) are shown. Colour code is the same as for (A) and (B). Temperature-induced unfolding was monitored by FUV-CD spectroscopy at a fixed wavelength (210 nm at pH 7.5 and 205 nm at pH 2.0) with a heating rate of 20°C h^{-1}. Data were fitted to a Boltzmann function to obtain transition midpoints. Chemical denaturation of the proteins was induced by urea, and structural changes were monitored by fluorescence spectroscopy at a fixed wavelength. Three measurements were averaged to improve the signal quality. Data were evaluated according to a two-state unfolding model to obtain midpoints and cooperativity parameters of the transitions. Measurements were performed at 25°C at a protein concentration of 20 µM (temperature-induced transitions) or 1 µM (urea-induced transitions).

Table 6: Thermal and chemical stabilities of different C_H2 variants.

Domain	pH 7.5, 100 mM NaCl			pH 2.0, 100 mM NaCl		
	T_{melt} [°C]	$D_{1/2}$ [M]	m [kJ mol^{-1} M^{-1}]	T_{melt} [°C]	$D_{1/2}$ [M]	m [kJ mol^{-1} M^{-1}]
$C_H2_{wildtype}$	42.5 ± 0.2	2.3 ± 0.2	4.8 ± 0.8	37.9 ± 0.7	2.5 ± 1.2	6.1 ± 2.0
$C_H2_{control}$	65.6 ± 0.2	5.0 ± 1.3	2.2 ± 0.3	50.6 ± 1.8	1.3 ± 0.3	4.2 ± 0.7
C_H2_{glyco}	68.4 ± 0.3	5.5 ± 3.0	4.6 ± 1.8	41.2 ± 1.8	NA	NA
$C_H2_{nonglyco}$	64.5 ± 0.2	4.6 ± 1.9	7.6 ± 2.2	48.6 ± 2.8	0.9 ± 0.3	14.5 ± 3.2

Stabilities against the thermal and chemical denaturation (urea) of the C_H2 variants under physiological conditions [50 mM sodium phosphate (pH 7.5) and 100 mM NaCl] and AFS conditions [50 mM sodium phosphate (pH 2.0) and 100 mM NaCl] are shown. Midpoints of thermal transitions are shown as T_{melt}. Even though none of the urea-induced unfolding transitions was reversible, data were evaluated according to a two-state equilibrium unfolding model to derive the midpoint of transitions ($D_{1/2}$), as well as the cooperativity parameter (m), for a qualitative comparison of the data. The data for C_H2 glyco under AFS conditions were not evaluated (NA) as they could not be fitted with a two-state equilibrium unfolding model.

Considering the results for urea-induced transitions, under physiological conditions C_H2 wild type is still the least stable domain (Figure 30 C and Table 6). The stability distribution matches quite well with the observations from the thermal-induced transitions with the stability of C_H2 wild type < C_H2 nonglyco ≤ C_H2 control < C_H2 glyco for urea-induced transition (Figure 30 A and C; Table 6). The folding behaviour is cooperative, even though less pronounced for C_H2 control (Figure 30 C). Urea-induced unfolding transitions show also some differences in the stability within the variants: according to thermal-induced unfolding under AFS conditions the order of stabilities is C_H2 wild type < C_H2 glyco < C_H2 nonglyco ≤ C_H2 control (Figure 30 B; Table 6). According to urea-induced unfolding transitions under AFS conditions the order of stabilities is different with C_H2 nonglyco < C_H2 glyco ≤ C_H2 control < C_H2 wild type (Figure 30 D; Table 6).

Although the C_H2 domains are in their AFSs not as stable as under physiological conditions, no differences in the melting temperatures could be observed when the samples were incubated overnight under AFS conditions or within four days (kept at 8°C until sample analysis, data not shown). Within a concentration between 10 and 50 µM also no obvious differences in the melting temperatures could be observed (as far as tested) what fits to the AUC experiments were the domains were always monomeric (Table 5), except C_H2 glyco and C_H2 control under AFS conditions (data not shown).

3.1.2.5 The pathway to the AFS and reverse

From the three different putative mechanistic scenarios (direct reorganisation of the structure upon changes of solvent conditions, partially or completely unfolding before establishing an AFS, or population of other states *en route* to the AFS) one was determined for the Fab domains. V_L and V_H were first (partially) unfolding and then adopting the AFS, C_H1 which is intrinsically unfolded under physiological conditions started the formation with an unfolded state as well (3.1.1.4, [192]).

To determine the mechanism for the C_H2 variants, kinetic experiments were performed. As the formation of the AFS was very fast, especially C_H2 wild type was too fast for manual mixing approaches, stopped-flow measurements with fluorescence detection were employed. The proteins were dialysed against the physiological buffer and mixed with an acidic buffer to reach AFS conditions. For secondary structure information FUV-CD spectra were recorded directly after the pH jump. For C_H2 wild type the kinetic trace could be described best by a single exponential function. The kinetic could be represented by a single rate constant (k_{obs}= 5.34 ± 1.2x10^{-4} s^{-1}) (Figure 31 A, right). The reaction was concentration-independent, as measurements with 5 µM final protein concentration resulted in the same k_{obs} (data not shown). The AFS formation for C_H2 wild type is consequently much faster compared to the Fab domains which formed their AFS within minutes to hours (see Figure 18 and Figure 19, right). Yet, the C_H3 domain adopts its AFS still faster than C_H2 with a k_{obs} in the range of the dead time of the instrument (>2000 s^{-1}) [133]. FUV-CD spectra of C_H2 wild type recorded directly after pH change indicate that the AFS is indeed formed within the manual mixing time of a few seconds (Figure 31 A, left). However, it cannot be excluded, that there is an intermediate formed within the dead-time of the instrument.

For C_H2 glyco and C_H2 nonglyco two rate constants could be observed, $k_{obs}1$= 0.14 ± 5.6x10^{-4} s^{-1} and $k_{obs}2$= 0.025 ± 3.3x10^{-4} s^{-1}, and $k_{obs}1$= 0.14 ± 6.6x10^{-4} s^{-1} and $k_{obs}2$= 0.018 ± 2.5x10^{-4} s^{-1}, respectively (Figure 31 C and D; right), indicating the presence of at least one intermediate state. Hence, the AFS formation is besides more complex also slower compared to C_H2 wild type. The FUV-CD spectra after the single jumps indicate that this intermediate has random coil characteristics (Figure 31 C and D, left). As the observable rate constants are of the same magnitude, it seems that the oligisaccharide has no influence on the mechanism of AFS fomation. The FUV-CD spectra indicate the presence of a more random coil like structure directly after the pH conditions change (Figure 31 C and D; left) as the minimum is more shifted to 200 nm. This observation supports the assumption that the mechanism of AFS formation observed for Fab domains (Figure 18 and [192]) is also true for C_H2.

For C_H2 control two rate constants for the AFS formation could be observed ($k_{obs}1$= 0.31 ± 4.0x10^{-3} s^{-1} and $k_{obs}2$= 0.026 ± 1.4x10^{-3} s^{-1}) (Figure 31 B, right). Consequently, the AFS formation for C_H2 control is slightly faster compared to C_H2 glyco and C_H2 nonglyco but slower than C_H2 wild type (Figure 31). The mechanism seems to be similar to C_H2 glyco and C_H2 nonglyco with at least one intermediate (Figure 31 B, left). The additional amino acids have hence an influence on AFS fomation but not the oligosaccharide (Figure 31). The FUV-CD spectrum after the pH jump shares more similarities with C_H2 wild type than C_H2 nonglyco as the signal minimum is more in the direction of 208-210 nm typical for the AFS (Figure 31; left). As C_H2 control shows the fastest rate constants, the manual mixing approach is too slow to be abe to observe the intermediate (Figure 31; right).

As AFS formation of C_H2 is also in higher concentrations reversible, the mechanism from the AFS back to the native state was addressed, too. Here, the C_H2 variants were dialysed against AFS conditions and mixed with a buffer to end up in physiological conditions. Kinetics were

measured using a stopped-flow device with fluorescence detector and spectra after mixing were recorded in the FUV-CD. AFS formation is a reversible process but the mechanisms of AFS formation and refolding differ from each other. For C_H2 wild type for the reverse reaction three different rate constants were observed ($k_{obs}1$= 6.6 ± 1.2x10^{-1} s^{-1}, $k_{obs}2$= 1.5 ± 2.1x10^{-2} s^{-1}, and $k_{obs}3$= 0.05 ± 3.0x10^{-4} s^{-1}) indicating the presence of at least two intermediates (Figure 32 A, right). The formation of the native state starting from the AFS state is slower than AFS formation (Figure 31 A, right and Figure 32 A, right). The FUV-CD spectra after the pH jump cannot shed light on the structure of these intermediates, as the spectrum after mixing looks already similar to the native state. Considering the maxima and minima distribution one can assume that putative intermediates are probably native-like (Figure 32 A, left).

For C_H2 glyco and C_H2 nonglyco three rate constants were observed as well. For C_H2 glyco rate constants of $k_{obs}1$= 2.4 ± 1.7x10^{-2} s^{-1}, $k_{obs}2$= 0.099 ± 6.3x10^{-4} s^{-1}, and $k_{obs}3$= 3.17x10^{-3} ± 3.2x10^{-5} s^{-1} and for C_H2 nonglyco $k_{obs}1$= 0.17 ± 2.2x10^{-3} s^{-1}, $k_{obs}2$= 0.022 ± 1.7x10^{-4} s^{-1}, and $k_{obs}3$= 2.2x10^{-3} ± 1.9x10^{-5} s^{-1} were determined (Figure 32 C and D, right). In comparison to C_H2 wild type the reaction for both variants is slower. The FUV-CD spectra indicate that the intermediates might be native-like, especially for C_H2 nonglyco, as the spectra show aready some characteristics of the native state (Figure 32 C and D, left). For C_H2 glyco the native-like structure elements are less pronounced but the reaction is faster compared to C_H2 nonglyco which indicates that the oligosaccharide accelerates the formation of the native state (Figure 32 C and D).

C_H2 control refolds also with three observable rate constants ($k_{obs}1$= 1.33±4.7x10^{-2} s^{-1}, $k_{obs}2$= 0.05±8.8x10^{-4} s^{-1}, and $k_{obs}3$= 2.80x10^{3}±4.5x10^{-5} s^{-1}) and lies in between the rates for C_H2 glyco and C_H2 nonglyco (Figure 32 B, C and D, right). The FUV-CD spectra directly after a pH jump from AFS to physiological conditions show already all characteristic minima and maxima observable for the native structure although less intense compared to the native state (Figure 32 B, left).

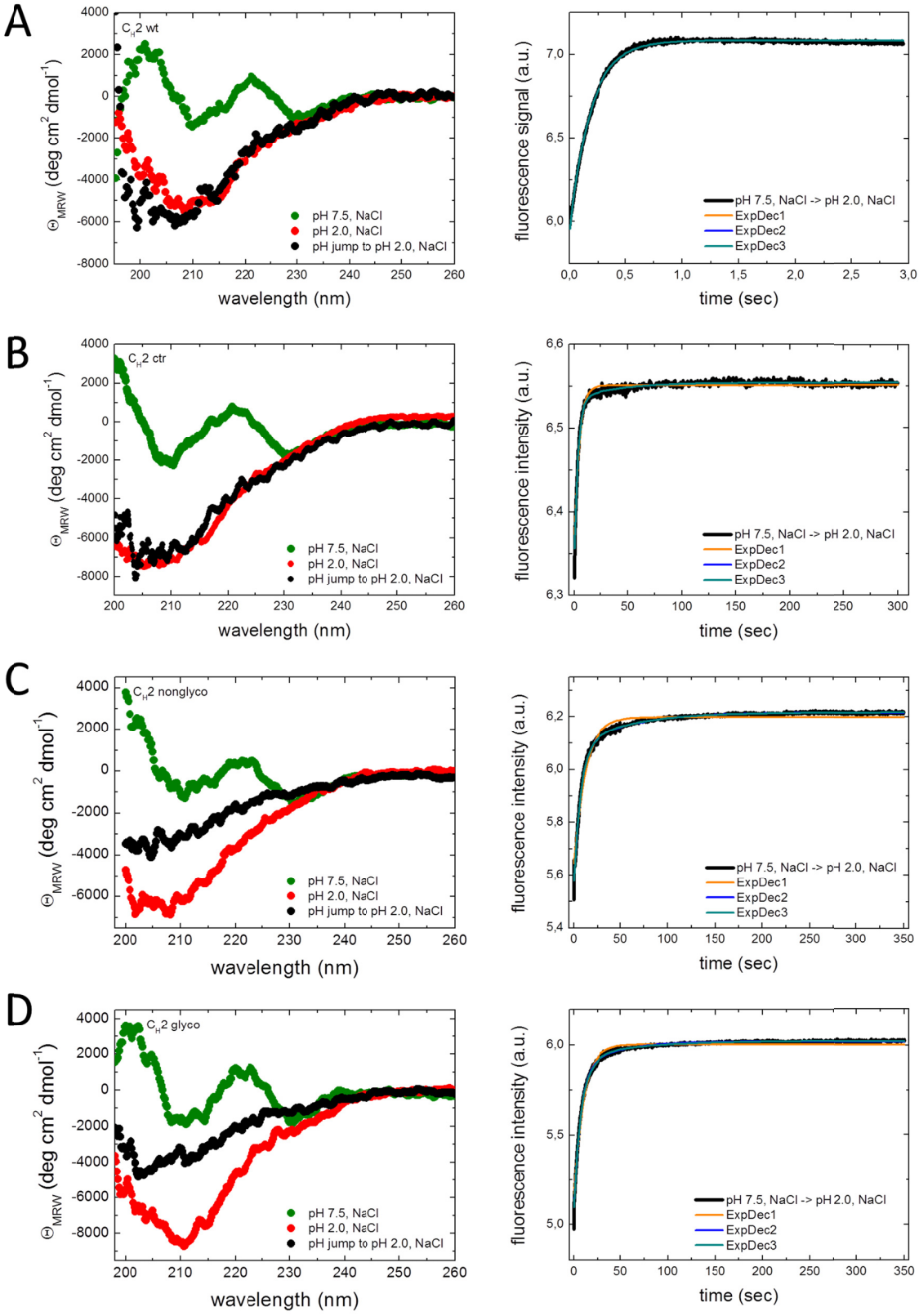

Figure 31: The pathway to the AFS the different C_H2 variants.

To obtain structural insights in the pathway leading to the AFS, kinetic stopped-flow fluorescence and FUV-CD experiments for all C_H2 variants were performed: C_H2 wild type (A), C_H2 control (B), C_H2 nonglyco (C), and C_H2 glyco (D). Left: The FUV-CD spectra of the domains directly after a single jump from physiological conditions [50 mM sodium phosphate (pH 7.5) and 100 mM NaCl] to AFS conditions [50 mM sodium phosphate (pH 2.0) and 100 mM NaCl] are shown (black circles). The jumps were repeated 12 times and averaged. As comparison, the

FUV-CD spectra under physiological (green circles) and AFS conditions (red circles) are shown. Right: Changes in fluorescence signal observed with a 305 nm cut-off filter and 280 nm excitation after the transfer of the protein from physiological conditions to AFS conditions are shown. The single exponential fit is marked in orange, double exponential in blue and triple exponential in cyan. All CD measurements were performed in a 1-mm quartz cuvette at 25°C at a protein concentration of 10 µM; stopped-flow experiments were performed with 1 µM final protein concentration at 25°C.

Spectroscopic techniques enable only limited insights into the structure of a protein. To gain a more detailed view for the AFS of C_H2, attempts to solve this state at an atomic level have been made in cooperation with the group of Professor Bernd Reif (TU München). The study focussed on C_H2 wild type as it can be easily and in high yield expressed in *E. coli* as ^{15}N- and ^{15}N-^{13}C-labelled protein.

Figure 33 shows the HSQC-spectra of ^{15}N-labelled C_H2 wild type under physiological [50 mM sodium phosphate (pH 7.5) and 100 mM NaCl] and AFS conditions [50 mM sodium phosphate (pH 2.0) and 100 mM NaCl] at 25°C. HSQC-spectra detect the NMR "fingerprint" of a protein, in which each peak represents an amino acid (i.e., the cross-peak between the nitrogen of the amide in the backbone and the amide proton). Under physiological conditions C_H2 wild type is well defined and most of the 99 amino acids can be distinguished (Figure 33 A). The well-defined pattern is lost when the domain is transferred to AFS conditions (Figure 33 B). The signals are in a range of flexible residues (8.0 ppm) indicating that the AFS is not one fixed state. This assumption is confirmed due the lost resolution of many residues that were visible in the native state (Figure 33 A and B).

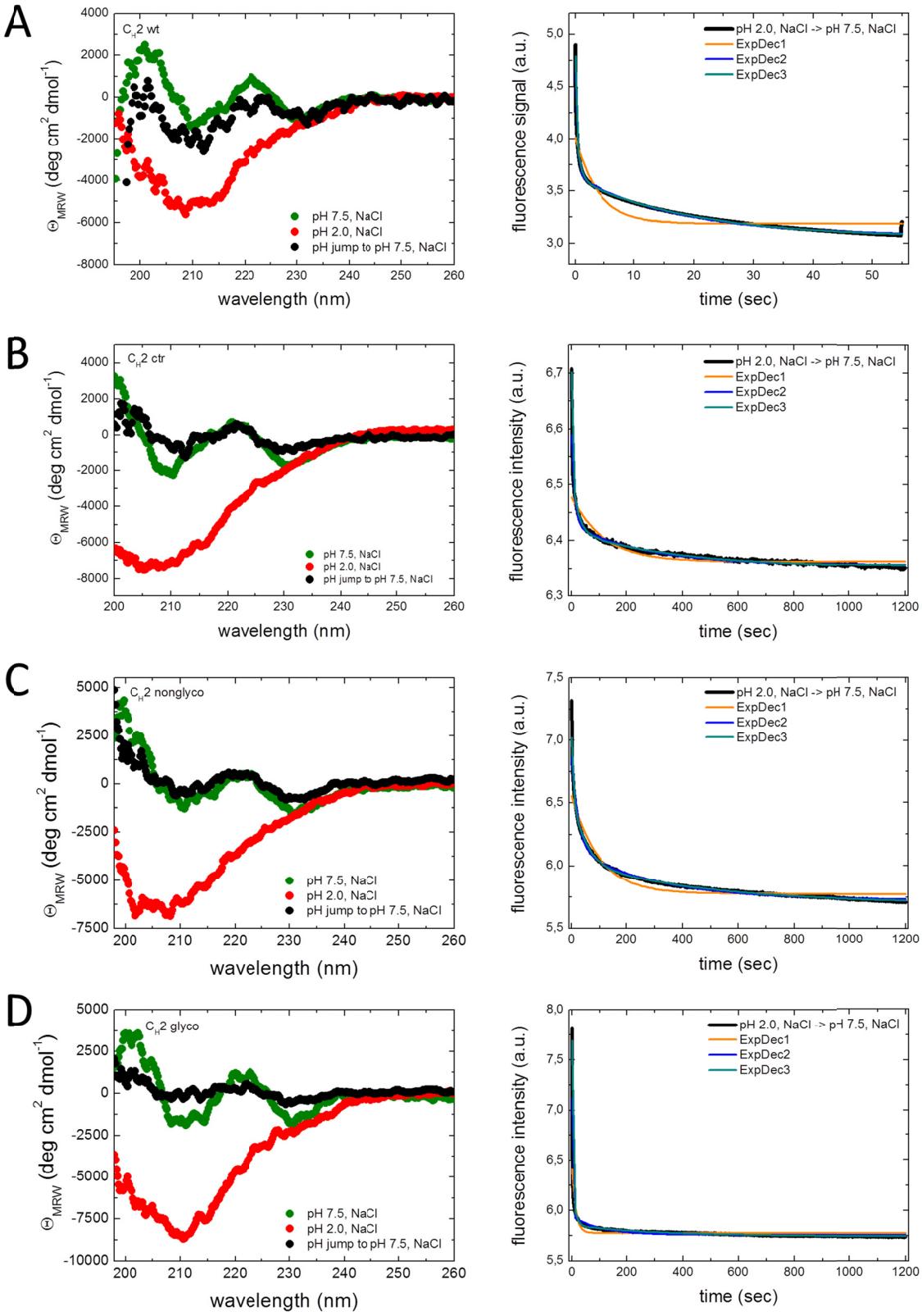

Figure 32: The pathway from the AFS to the native state of the different C_H2 variants.

To obtain structural insights into the reverse pathway leading back to the native state, kinetic stopped-flow fluorescence and FUV-CD experiments for all C_H2 variants were performed: C_H2 wild type (A), C_H2 control (B), C_H2 nonglyco (C), and C_H2 glyco (D). Left: The FUV-CD spectra of the domains directly after a single jump from AFS conditions [50 mM sodium phosphate (pH 2.0) and 100 mM NaCl] to physiological conditions [50 mM sodium phosphate (pH 7.5) and 100 mM NaCl] are shown (black circles). The jumps were repeated 12 times and averaged. As comparison, the FUV-CD spectra under physiological (green circles) and AFS conditions (red circles)

are shown. Right: Changes in fluorescence signal observed with a 305 nm cut-off filter and 280 nm excitation after the transfer of the protein from AFS conditions to physiological conditions are shown. The single exponential fit is marked in orange, double exponential in blue and triple exponential in cyan. All CD measurements were performed in a 1-mm quartz cuvette at 25°C at a protein concentration of 10 µM; stopped-flow experiments were performed with 1 µM final protein concentration at 25°C

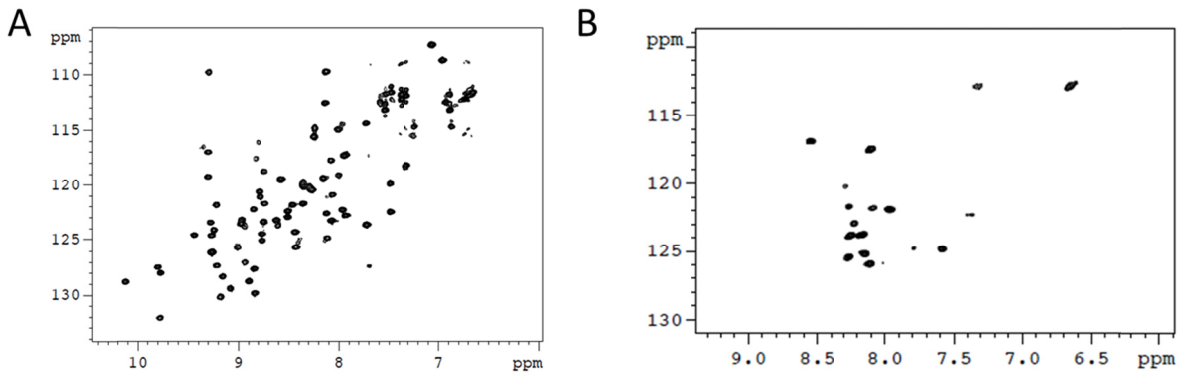

Figure 33: HSQC-spectra of ^{15}N-labelled C_H2 wild type at pH 7.5 and pH 2.0.
500 µM of ^{15}N-labelled C_H2 wild type in physiological [50 mM sodium phosphate (pH 7.5) and 100 mM NaCl] (A) and AFS conditions [50 mM sodium phosphate (pH 2.0) and 100 mM NaCl] (B) were prepared and HSQC-spectra recorded. Spectra were recorded at 25°C.

As double labelled ^{15}N-^{13}C-C_H2 wild type enabled an assignment of those residues that can still be detected in the AFS, we tried to identify them. To be able to assign residues, the information from nitrogen of the amides and nitrogen protons are not sufficient and a third dimension must be recorded to be able to detect the carbon atoms (HNCαCβ and HNCα – spectra). Ten residues could reliably be determined. They are positioned at the N- and C-terminus of the domain (Figure 34 A). Blotting these residues onto the native structure of C_H2 wild type (from PDB ID 3HKF) shows that they are located at β-strand A and strand G within the immunoglobulin fold (Figure 34 B and Figure 6 C). The signal range indicates flexibility of these residues and consequently these two β-strands are no longer present under AFS conditions. The remaining residues are too dynamic to be resolved by this NMR approach. Interestingly, several proline residues at the N- and C-terminus are present (Figure 34 B, orange sticks) and they might increase the flexibility.

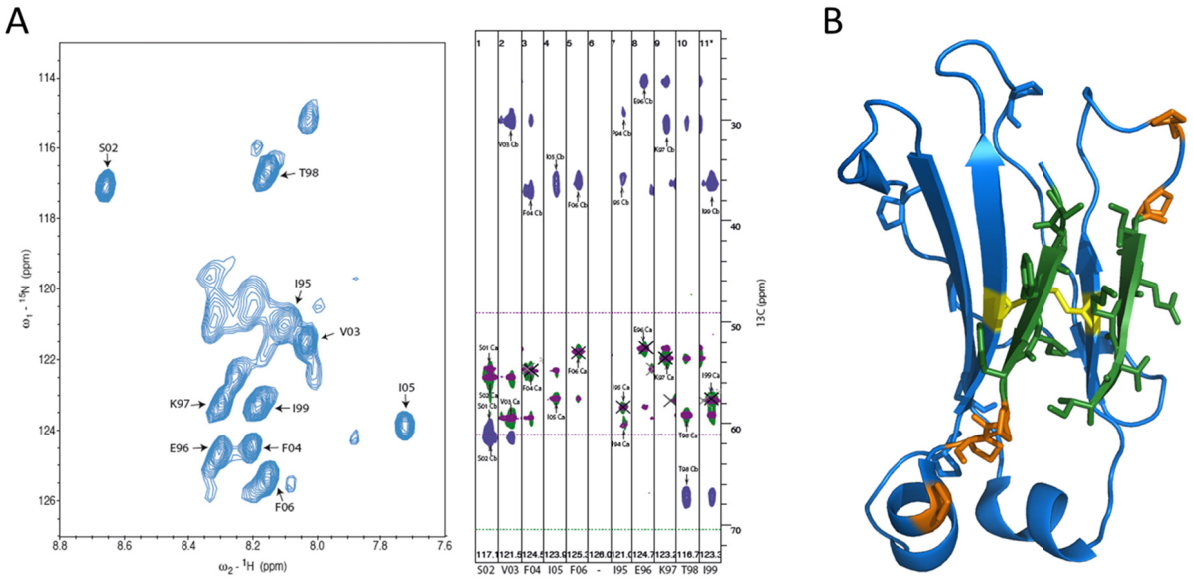

Figure 34: Assignment of residues in the AFS.

Ten residues from the ^{15}N-^{13}C-labelled C_H2 wild type in AFS conditions [50 mM sodium phosphate (pH 2.0) and 100 mM NaCl] could be reliably assigned. Five of the ten residues are located at the N-terminus, the other five at the C-terminus. The signal range indicates that these residues are flexible (A). HNCαCβ- and HNCα -spectra were recorded at 25°C. The combination of the information of HSQC- and HNCαCβ- and HNCα –spectra enables a reliable assignment of amino acids. In (B) the assigned residues are plotted onto the native structure of C_H2 wild type. Proline residues are labelled in orange, the internal disulphide bridge in yellow and the assigned residues in dark green (modification from PDB ID 3HKF).

The NMR results indicate that the AFS of C_H2 wild type is a mixture of at least two conformations which are in equilibrium. Hence, if one could shift the equilibrium to one of these states it should be possible to determine more residues. Different approaches are possible, for example the addition of higher salt concentrations and different kind of salts (shown for C_H3 to influence its AFS [133]), common stabilising reagents or pH variations [132, 133]. A buffer screen was carried out in which the influence of different types of salt and salt concentrations (NaCl, KCl, $KClO_4$; 25 mM-200 mM) as well as different additives (L-histidine, L-arginine, glycine, betaine) on the AFS of C_H2 wild type was examined. L-arginine and L-histidine are often used for refolding of proteins and have a stabilising effect [226]. Betaine is an osmolyte which is well known to influence protein stability and conformational exchanges [227]. In Table 7 all buffer compositions tested are summarised. To determine the most promising conditions for NMR measurements, FUV-CD spectra and thermal-induced unfolding transitions were recorded. Those conditions resulting in a FUV-CD spectrum indicating high secondary structure content and elevated thermal stability were selected as new NMR samples. In the end, three conditions were selected and NMR HSQC-spectra with ^{15}N-labelled C_H2 recorded. However, no new signals appeared (see Figure 35).

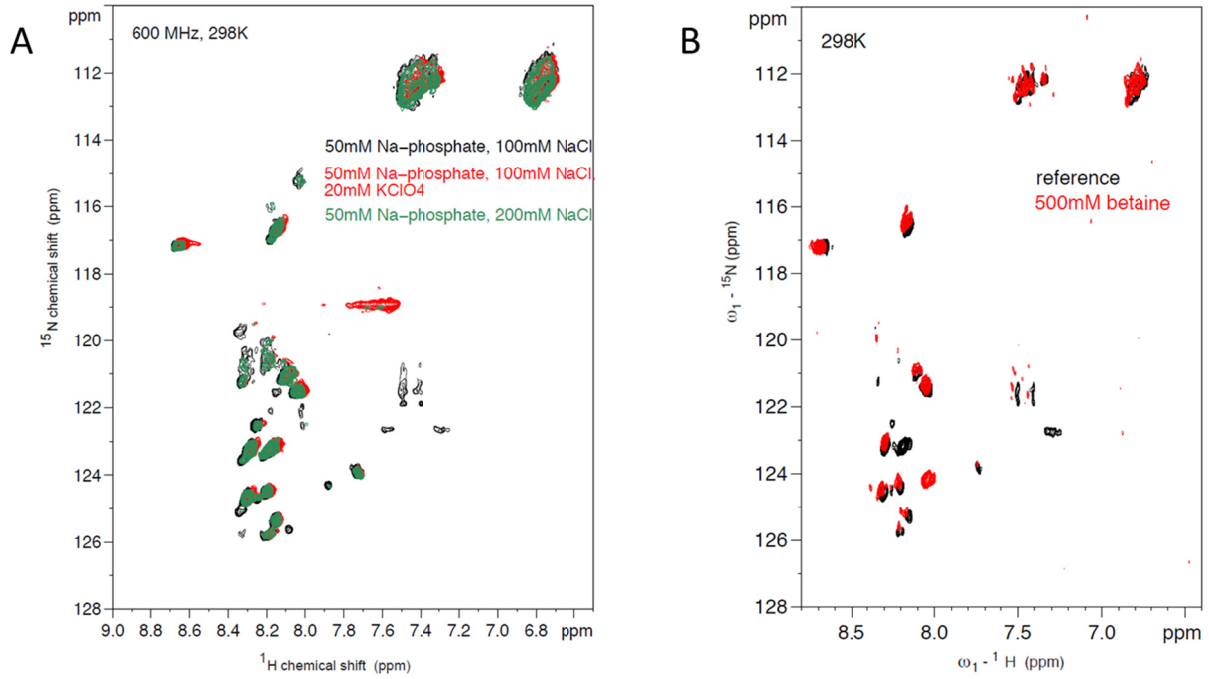

Figure 35: HSQC-spectra of ^{15}N-C_H2 wild type in different buffer conditions.

Three different buffer conditions were tested to be able to shift the equilibrium of the AFS conformations. In (A) the HSQC-spectra for 50 mM sodium phosphate (pH 2.0), 20 mM $KClO_4$ and 100 mM NaCl (red) and 50 mM sodium phosphate (pH2.0) and 200 mM NaCl (green) are shown. As a reference the black signals represent the AFS of C_H2 wild type under AFS conditions (50 mM sodium phosphate (pH 2.0) and 100 mM NaCl). In (B) the HSQC-spectra for the reference (black) and in presence of 500 mM betaine are shown (both in AFS buffer conditions). All spectra were recorded at 25°C.

Table 7: Screening buffer conditions for stabilisation of the AFS of C_H2 wild type.

	Secondary structure		
Buffer	**Minimum (nm)**	**Signal intensity**	**T_{melt} (°C)**
50 mM Na_2HPO_4, 100 mM NaCl pH 7.5	210 and 195	-	42.5 ± 0.2
50 mM NaH_2PO_4, 100 mM NaCl pH 2.0	210	-	37.9 ± 0.7
50 mM NaH_2PO_4, 200 mM NaCl pH 2.0	**210**	**↓**	**44.9 ± 1.1**
50 mM NaH_2PO_4, 25 mM $KClO_4$, pH 2.0	208	↓	NA
50 mM NaH_2PO_4, 100 mM NaCl, 100 mM KCl, pH 2.0	210	-	45.5 ± 0.4
50 mM NaH_2PO_4 ,100 mM NaCl, 20 mM $KClO_4$, pH 2.0	**217**	**↓**	**61.3 ± 1.4**
50 mM NaH_2PO_4 ,100 mM NaCl, 1 mM L-His, pH 2.0	210	-	40.8 ± 0.8
50 mM NaH_2PO_4 ,100 mM NaCl, 1 mM L-Arg, pH 2.0	210	↓	39.7 ± 0.6
50 mM NaH_2PO_4 ,100 mM NaCl, 100 mM Glycine, pH 2.0	210	↓↓	NA

50 mM NaH_2PO_4 ,100 mM NaCl, 1 mM L-His, 1 mM L-Arg, pH 2.0	210	↓	40.8 ± 0.5
50 mM NaH_2PO_4, 100 mM NaCl, 100 mM betaine pH 2.0	**NA**	**NA**	**NA**
50 mM NaH_2PO_4 ,200 mM NaCl, 1 mM L-His, pH 2.0	210	-	43.9 ± 0.6
50 mM NaH_2PO_4 ,200 mM NaCl, 1 mM L-Arg, pH 2.0	210	↓	44.2 ± 0.5

To test the influence of different buffer conditions, samples were dialysed overnight against the different buffers. To pre-screen the samples FUV-CD spectra were recorded and the influence on the thermal stability was determined. CD measurements were performed at 25°C and the thermal transitions were performed with a heating rate of 20°C h^{-1} at 205 nm. Those conditions in which secondary structure could be observed and the thermal stability was increased compared to the AFS conditions [50 mM sodium phosphate (pH 2.0) and 100 mM NaCl] were used for NMR measurements. Physiological conditions are highlighted in green, the AFS conditions in red. Bold are the variants that were tested by HSQC-NMR.

In a second approach the influence of the pH on the AFS of C_H2 wild type was tested. The buffer compositions were identical to the AFS buffer [50 mM sodium phosphate (pH 2.0) and 100 mM NaCl] but the pH range was set from pH 5 to pH 2.0 with an interval of 0.5 pH units. Due to the fact that C_H2 wild type aggregated in the pH 4.5 buffer, this sample was excluded from the study (data not shown). The calculated pI is at pH 5.0 (ProtParam) but according to this observation the actual pI is probably in the range of pH 4.5. For a read-out of the influence FUV-CD spectra were again recorded as well as thermal-induced unfolding transitions. C_H2 wild type at pH 5.0 still exhibited characteristics of the native state (Figure 36 green and light green traces). At pH 4.0 the secondary structure of the protein changes and the characteristic minimum at 210 nm appears but with a much lower intensity compared to the AFS (Figure 36 light yellow and red traces). Within the range between pH 4.0 and 2.5 the intensity increases just slightly but in a continuous manner (Figure 36). The thermal-induced unfolding transitions were not cooperative within a pH range of 4.0 and 2.5 (data not shown) indicating that these states are not as well organised as under AFS conditions and even more dynamic. On the contrary, for pH 5.0 a T_{melt}= 40.8 ± 0.5°C could be determined and the transition was cooperative (data not shown).

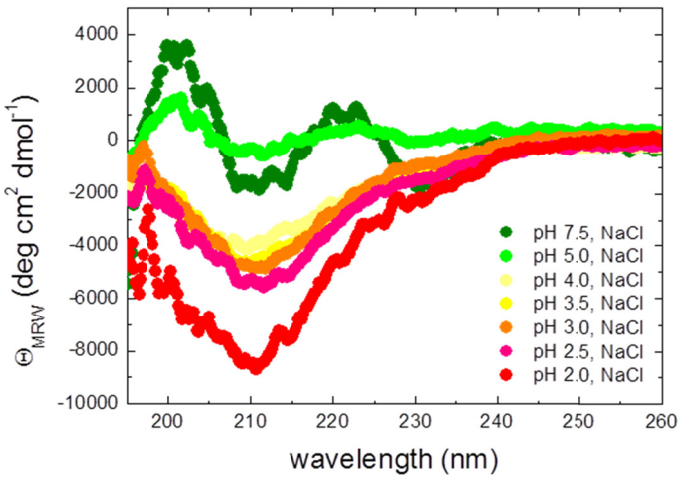

Figure 36: pH influence on the secondary structure of C_H2 wild type.
To determine at which pH an AFS can be adopted a pH titration was recorded. As a comparison the FUV-CD spectra of C_H2 wild type under AFS (red circles) and physiological conditions (green circles) are shown. Spectra were recorded in a 1-mm quartz cuvette with 10 µM protein concentration at 25°C. 16 accumulations were averaged and buffer-corrected.

In the end, additional HSQC-spectra with ^{15}N-labelled C_H2 wild type were recorded for pH 5.0, pH 4.0, and pH 3.0 at 25°C and 4°C to be able to cover the whole pH range from pH 7.5 to pH 2.0 (Figure 37 A). Additionally, the residues for C_H2 wild type under physiological conditions were assigned (Figure 37 B) to be able to determine those residues changing and to obtain information about the order of events.

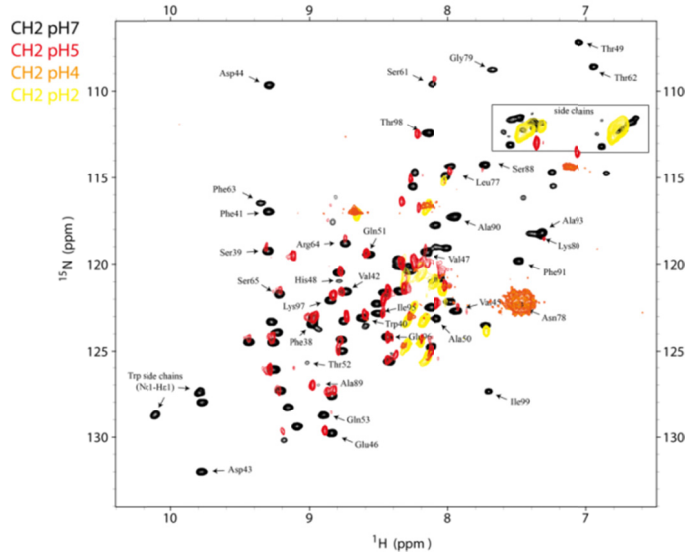

Figure 37: HSQC-spectra of ^{15}N-C_H2 wild type in different pH conditions with assigned residues at pH 7.5.
HSQC-spectra of C_H2 wild type were recorded at 25°C in 50 mM sodium phosphate buffer with 100 mM NaCl at different pH values. In black the reference for the native state under physiological conditions (pH 7.5) is shown; in red at pH 5.0, in orange at pH 4.0 and in yellow the AFS at pH 2.0. The residues which could be assigned at pH 7.5 are labelled.

Not all residues could be assigned at pH 7.5. Most of them were located in β-strands A, B and F as well as in the loop between A-B (Figure 38). When the pH is changed from pH 7.5 to pH 5.0 shifts mainly within the loop regions between strands C-D, D-E, E-F and F-G appear. At pH 4.0 only two residues could be determined, Val45 and Asn78. The signals of the amino acids in the β-strands C, D, E and G are lost. For pH 2.0 more signals are observed in the HSQC-spectrum (Figure 37). However, within the assigned regions is the state comparable to pH 4.0 (Figure 38). We knew already from previous experiments for the AFS at pH 2.0 that the β-strands A and G are unstructured (Figure 34). Thus, the combination of the analysis of the native as well as the AFS by NMR could expand the structural insights in the AFS.

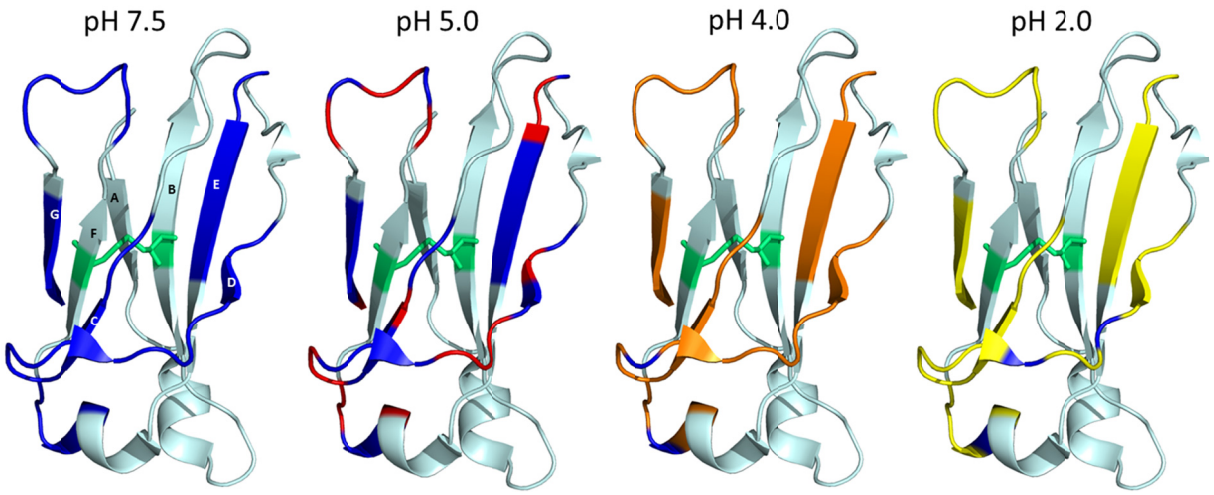

Figure 38: Assigned residues plotted on the native C_H2 structure dependent on pH
In order to gain insights in the structural changes upon different pH conditions, those residues which could be assigned successfully at pH 7.5 were plotted on the native structure of non-glycosylated C_H2 (modified from PDB ID 3HKF). The parts for which the residues could not be identified are labelled in pale cyan. The assigned residues which were present under certain pH conditions are dark blue and the ones which disappeared upon changing the pH are coloured in red (pH 5), orange (pH 4) and yellow (pH 2). The internal disulphide bridge is highlighted in green. For pH 7.5 the single β-strands of an Ig fold are labelled.

3.1.2.6 Discussion

The formation of an alternatively folded state is a particularity of antibodies and was first shown by Buchner and co-workers for MAK33 [132]. Not only intact antibodies, but also the Fab fragment [134], the light chain [140] and several isolated domains, such as C_H3 [133], V_L, C_H1, C_L and V_H [192] can adopt a stable secondary structure at acidic conditions with an appropriate ionic strength. C_H2 is the last domain which was not yet analysed for its ability to form an AFS. This domain has a conserved glycosylation site (Asn-297) and N-linked oligosaccharides in C_H2 are characteristic for intact antibodies [147]. It is already known that the glycan is not necessary for adopting the immunoglobulin fold [202] but some differences were observed when the crystal structure of the unglycosylated Fc of MAK33 was solved and compared to the glycosylated human counterpart [154]. The expression of glycosylated isolated C_H2 was challenging as our attempts to express them in mammalian cell culture (fibroblast from apes) failed due to an insufficient expression level. Therefore we changed the expression system to *Sf9* cells and transfected them with a Baculo virus system. On has to keep in mind that the glycosylation pattern of insect cells differs from mammalian cells and that the C_H2 glyco domain in this study has no oligosaccharides typical for mammalian antibodies (see Figure 39). According to the mass deviation between glycosylated and nonglycosylated C_H2 we consider the glycosylation pattern to consist of two N-acetylglycosamines and three mannoses. We assume that the glycosylation pattern is consistent with the first pattern in Figure 39 A. Mass spectrometry gave no hints for the presence of other types of glycosylation. Hence, we suppose that the sample is homogenous. In contrast, variations of glycosylation between IgG molecules as well as within different sites in the same molecule are not unusual in intact antibodies and result in

heterogenous mixtures of IgGs [147]. Consequently, the glycosylation pattern in our C_H2 glyco domain is very similar to the simplest pattern observed in human cells with one N-acetlyglucosamine difference (Figure 39 A and B).

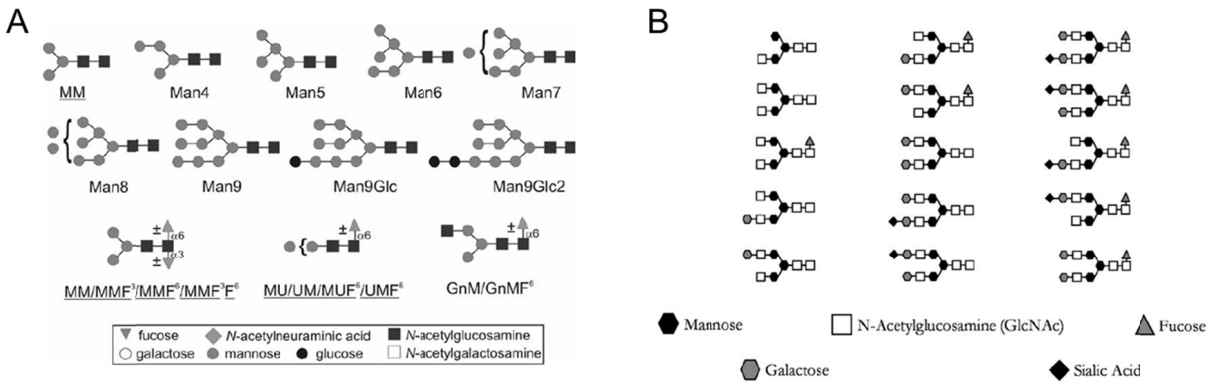

Figure 39: Glycosylation patterns differ between insect and mammalian cells.
The glycosylation patterns between the two systems are not identical. In (A) typical glycosylation patterns for insect cell lines are shown. A typical feature is the high content of mannose. Figure is taken from [228]. In (B) typical glycosylation patterns of human cells are shown. They also contain mannose but mostly to a different extend and contain at least one N-acetylglucosamine. Figure is taken from [155].

The predominant differences between the AFS and molten globule states are the high thermodynamic stabilities in the AFS [132, 192]. The thermal stability in the native state is influenced by the changes between the *E. coli* and *Sf9* constructs as C_H2 wild type has a T_{melt} of ~43°C and all other variants which include the additional amino acids have T_{melt}s in a range between ~65°C (C_H2 control/C_H2 nonglyco) to ~68.5°C (C_H2 glyco). Glycosylation is known to influence protein stability [229-231]. In our case, the actual stability increase due to the glycan is in the range of 3.5°C considering the errors (Table 6). This finding is consistent with the differences observed for other N-glycosylated proteins when their glycosylated and deglycosylated forms were compared [230]. Stability differences were in a range from equal up to 3.5°C higher stabilities for the glycosylated proteins [230]. In addition, the observation that the first transition midpoint of MAK33 which is considered to be the result of C_H2 unfolding shifts when the antibody is deglycosylated [232] suggests that the stability of C_H2 is influenced by the glycan. A much larger stabilising effect of ~22.5°C is caused by the additional amino acids at the N- and C-terminus of C_H2. 16 amino acids (ADLGSA at the N-terminus and GSGSENLYFQ at the C-terminus in a 99 amino acid long domain) are the consequence of the *Sf9* transfection vector pAcGP-67 B with its secretion signal and the linker with a His-Tag added for purification (see Figure 21). Attempts to design the construct in a way that only one additional amino acid will be left after purification failed because transfection was not possible so far. Similar effects were observed for the chemical stabilities of the four C_H2 variants. C_H2 wild type was clearly less stable ($D_{1/2}$= 2.3 ± 0.2 M urea) compared to the other constructs ($D_{1/2}$~ 5.0 M urea) (see Table 6). On the other hand, no clear improvement due to the glycan could be observed if the errors are taken into

account. As a consequence one should be careful with immunoglobulin domains in which amino acids were added to the termini.

Interestingly, all C_H2 variants tested had similar thermal stabilities in the AFS with T_{melt}s of ~38°C (C_H2 wild type) to ~51°C (C_H2 control) compared to their native states with T_{melt}s of ~43°C (C_H2 wild type) to ~68°C (C_H2 glyco) (see Figure 30 and Table 6). C_H2 wild type is the least stable domain in the native state as well as in the AFS. Thus, the additional amino acids have still some influence on the thermal stability in the AFS, although not as pronounced as for the native structure (differences of 13°C in AFS *vs.* 22.5°C in the native state without glycosylation). Interestingly, the additional stabilising effect of the glycan is lost at acidic conditions, as the thermal stability of C_H2 glyco is even slightly lower compared to C_H2 nonglyco. For those Fab domains which adopted an AFS the stability in the AFS was comparable or even higher than in the native state (see Table 3; [192]). This is not the case for any of the tested C_H2 variants. Yet, the variants showed no difference after overnight or 3-4 days incubation concerning FUV-CD spectra and thermal stability and in conclusion the AFS of C_H2 is a stable state. The urea-induced transitions affirm the observation that the stabilising effect of the glycan at physiological conditions is lost in the AFS and that there is still some influence of the additional amino acids (Figure 30).

The second characteristic feature for most antibody domains was the formation of oligomers in their AFS [132, 133] which was correlated to hydrophobicity of the surface. V_H and V_L formed the largest oligomers and showed the highest signal increase upon ANS binding whereas C_H1 which formed smaller oligomers could bind ANS less effective (see Table 2, Figure 15 and [192]). This correlation is not valid for C_H2. C_H2 glyo and C_H2 control were the only variants that formed some dimeric/tetrameric species besides the monomers (Table 5) although all variants showed similar ANS binding (Figure 29). The signal increase was comparable to C_H1 which can form oligomers of predominantly 5-20 subunits in the AFS (Table 2, Figure 15). C_H2 glyco binds slightly less ANS compared to the other variants indicating that sugar-sugar or sugar-protein interactions might enable the formation of dimers or tetramers. Glycans have several contacts in the Fc part of antibodies and influence C_H2 orientation [154]. On the other hand, C_H2 control could also adopt oligomeric states. This finding is surprising as C_H2 nonglyco is not able to self-associate although both variants share the same sequence (Table 5). C_H2 control was included to prove that some of the observed effects are really caused by the additional amino acids and not a consequence due to changes caused by expression in *Sf9* instead of refolding from inclusion bodies. Both variants showed the same behaviour in regard of secondary structure (Figure 23 C and E), thermal stability (Figure 30 A and B), and ANS binding (Figure 29 B and D). However, small differences were observed regarding tertiary structure. The NUV-CD spectra do not completely correspond to each other as well as the fluorescence behaviour at physiological and AFS conditions (Figure 28 B and D). We gather that C_H2 control and C_H2 nonglyco are not exactly the same from these observations and that they have small differences in their tertiary structure conformation.

The folding to the AFS of C_H3 was completed within the dead-time of the stopped-flow device but assumed to pass an unfolded state before the AFS is established [133]. For the Fab domains which adopt an AFS a general pathway *via* a (partially) unfolded state could be shown but the process was much slower compared to C_H3 (see Figure 18 and [192]). All four C_H2 variants seem to share this pathway. However, it is difficult to prove that an unfolded state precedes AFS formation as for the kinetic traces fluorescence detection (i.e., tertiary structure changes) and for the control spectra FUV-CD (i.e., secondary structure changes) were used. This was necessary as the FUV-CD signals at 205 nm were not sufficient to observe the structural changes properly. The time to adopt an AFS was clearly the fastest for C_H2 wild type (relaxation time ~1 s), followed by C_H2 control (~25 s), C_H2 glyco (~70 s) and C_H2 nonglyco (~100 s). The kinetic traces could be best described by a single or double exponential function for C_H2 wild type and the other domains, respectively, indicating the presence of one intermediate for the latter (Figure 31). The presence of an intermediate within the dead-time of the stopped-flow device cannot be excluded and the FUV-CD spectra after the pH jump of C_H2 wild type might indicate that a more unfolded intermediate precedes AFS formation as the minimum is slightly shifted to the FUV (Figure 31 A, left). As the only common difference between C_H2 wild type and the other species are the additional amino acids, these residues might influence the folding process (single *vs*. double exponential). Yet, one must be careful with these assumptions.

A striking feature of C_H2 is the reversibility in a high concentration range from 1 µM to 100 µM (Figure 27). The complete MAK33 (glycosylated) and C_H3 were also described to be able to get back the native state starting from the AFS [132, 133]. In contrast, the Fab domains could only adopt the native state in low concentrations (≤ 10 µM) otherwise aggregation occurred (Figure 13 and [192]). Hence, we were interested in the pathway from the AFS back to the native state. In general the process is slower compared to AFS formation, but C_H2 wild type is still the fastest domain (relaxation time ~50 s). C_H2 nonglyco and C_H2 control behave similar what is to be expected as they share the same sequence (~800 s). C_H2 glyco is interesting as it can adopt the native structure clearly faster compared to C_H2 nonglyco and C_H2 control (~200 s). As the glycan enhanced the stability in the native state, it can be assumed that it accelerates the native state formation due to its favourable effect. However, refolding of C_H2 glyco is slower compared to C_H2 wild type and consequently the additional amino acids are disadvantageous. The mechanism is more complex compared to AFS formation with at least two intermediates (traces were best described by a triple exponential function; see Figure 32). The FUV-CD spectra indicate that at least one intermediate is native-like.

One of the best characterised non-native states is the molten globule state. NMR is a powerful tool to determine structural details of this often dynamic state. For α-lactalbumin the molten globule state was analysed in great detail [233, 234]. In return, no structural details are known up to date about the AFS. Spectroscopic techniques, based on fluorescence or CD, have a very limited resolution. Attempts to resolve the AFS of C_H2 wild type at atomic resolution failed, as our data indicate the presence of at least two structural

states in the AFS which are in a dynamic equilibrium. We could just identify five residues at the C- and five residues at the N-terminus which are for sure not part of a defined structure (Figure 34). In these areas there are also several proline residues which might facilitate flexibility under AFS conditions. It was not possible to shift the equilibrium to one structure neither by buffer variations nor additives such as betaine, arginine or histidine which have been shown to influence protein stability [226, 227]. However, the pH has a strong influence on the structure. HSQC-spectra of ^{15}N-C_H2 wild type in different pH conditions show that the unfolding of the structure is a step-wise process (Figure 37). Together with the assigned residues at pH 7.5 the residues changing with differing pH values can be plotted on the structure. This will result in a map of the different regions changing during the pH gradient. It was not yet possible to reliably assign all residues at pH 7.5. However, we could expand our view on AFS formation. When the buffer conditions shift to acidic pH, rearrangements within some of the flexible loop regions appear first (pH 5; Figure 38). In a more acidic environment the outer β-strands (C, D, E and G) lose their contacts. The disulphide bridge between strands B and F might protect this area. However, the residues within these two strands could not yet be assigned at pH 7.5 but some additional ones in strand A in the AFS (Figure 34). The residues within β-strand A were also flexible. For the Fab and light chain of MAK33 Buchner and co-workers have shown, that the internal disulphide bridge impede the formation of an AFS [140]. This seems to be not the case for C_H2 as the domain is able to adopt an AFS in presence of TCEP (Figure 26). The same was observed for the Fab domains, which were able to form an AFS in the presence of a reducing agent (Figure 12; [192]).

Taken together, the C_H2 domain can adopt an AFS independent of its glycan. Not all characteristics of an AFS are fulfilled but the protein is stable for several days at pH 2. The AFS is dynamic and not one fixed conformation. Noteworthy, it is reversible even at high concentration and this property sets C_H2 apart from other domains able to adopt an AFS. The glycan had only a positive effect on the stability at physiological conditions and not in the AFS. It was not possible to resolve the structure of the AFS as it is dynamic. However, we could gain insights in the formation of an AFS.

3.2 Interaction mechanism between V_H and V_L

Antigen recognition is an important task of the humoral immune response mediated by immunoglobulins. The Fv fragment is responsible for specificity and antigen binding and consists of the variable domain of the light chain (V_L) and the variable domain of the heavy chain (V_H). These two domains share each three hypervariable regions (i.e., long and flexible loops, so called complementarity determining regions, CDRs) which represent the binding sites for the antigen and account for approximately 25% of the domains [160]. CDR-H3 (i.e., the third CDR of V_H) is the most flexible of these six regions concerning length and amino acid sequence [235]. The structure of the Fv fragment of MAK33 and the positions of the CDRs are shown in Figure 40. Due to the fact that CDR-H3 is highly flexible, this region could not be resolved by X-ray crystallography. Hence, CDR-H3 is indicated as a grey line.

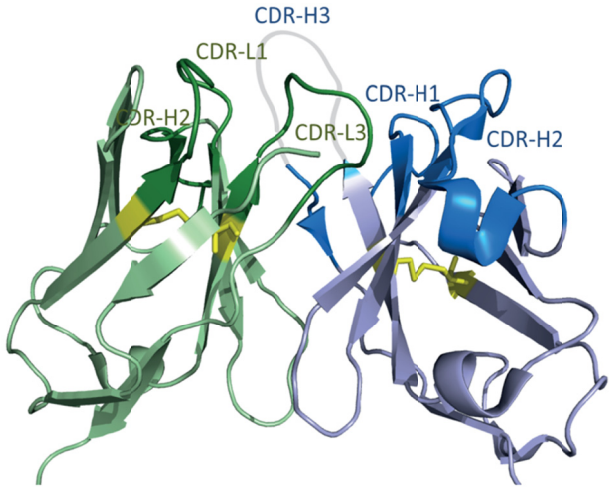

Figure 40: Structure of the Fv fragment of MAK33.

The Fv fragment of MAK33 consists of V_L (light green) and V_H (light blue). Both domains have three hypervariable regions (complementary determining regions, CDRs) each which are crucial for antigen recognition and binding. The CDRs of V_L are depicted in dark green, the ones from V_H in dark blue (according to http://www.bioinf.org.uk/). The conserved internal disulphide bridges are highlighted in yellow. Figure is modified from PDB ID 1FH5. As CDR-H3 is not resolved in this crystal structure, it is indicated as a grey loop.

Antibodies play an important role for therapeutic and diagnostic applications due to their high selectivity and specificity. Engineering of antibody sequences is often required to modify antigen recognition, stability, solubility, affinity, penetration abilities, and immunogenicity or effector functions. The Fv fragment is besides antigen binding properties also responsible for pharmacokinetics, pharmaceutical properties and immunogenicity (besides Fc) [236]. Unfortunately, an increase in affinity is often accompanied by a decrease in stability and *vice versa* especially when changes within the CDRs are made and consequences are difficult to predict [237]. Hence, for several years it has been tried to understand the relationship between structure, stability and binding affinity of V_H and V_L.

3.2.1 Conserved amino acids within the interface of V_H and V_L

One important approach to understand the mechanism of V_H and V_L interaction was covariation analysis to determine conserved amino acid clusters. The idea behind is that amino acids which are conserved within different species and germlines must play a functional role as they would otherwise not have an evolutionary pressure to be unchanged. One of the first to address conserved amino acids were Chothia and co-workers in 1985 [160]. They could already show that the residues at the positions 98, 44 and 36 (V_L) and 103, 47, 45 and 37 (V_H) might be conserved. Back then their data set was very small containing only three structures and thus they could only draw conclusions regarding the position of the amino acid and not of their type. In 1998 Chothia and co-worker published a new covariation analysis with a larger data set and their results fit quite well to previous selected positions [166]. The covariation analysis of Wang and co-workers in 2009 [164] was the most elaborate analysis of variable domain sequences that was performed up to date and to our knowledge. With this approach they aimed to identify naturally occurring amino acid networks that are generally important for antibody structure and function. Other studies to identify defining positions within different germlines were also performed. Altschuh and co-workers for example investigated the differences between murine and human germline V_H

and V_L domains [165, 171]. The majority of the most strongly conserved amino acids identified in the study of Wang were positioned at/or adjacent to the V_H-V_L interface [164]. For the V_H domain additionally some conserved residues at the V_H-C_H1 interface could be identified being probably important for framework orientation. Interestingly, the same network, yet not the same amino acids, for the interactions with V_L and with C_H1 are found in camelid V_{HH} domains. V_{HH} domains are functioning in the absence of a V_L domain [162, 163]. This finding indicates that these particular amino acids might influence the interaction between V_H and V_L.

V_L and V_H of the model antibody MAK33 fit the findings from covariation analysis with all amino acids on conserved positions. In Figure 41, the conserved amino acids within/adjacent to the interface of the variable domains are shown and the important residues highlighted. For V_L and V_H six and seven residues were considered, respectively. The only residue in MAK33 that does not match the covariation analysis from Wang and co-workers is the serine at position 43 in V_L. From the analysis an alanine would be expected and therefore the exchange to alanine might increase the interaction ability between V_H and V_L. Taken all data from covariation analysis together, it is justified to assume that these residues play a crucial role for the interaction between the variable domains.

The aim of this study will be to mutate those hits against alanine and to characterise the influence on structure, stability and interaction between both domains. As a model system the murine monoclonal antibody MAK33 (κ/IgG1 subclass) was used. Here, we focus on the analysis of isolated variable domains and not on Fab or scFv fragments as in previous studies [158, 170, 172]. These approaches draw conclusions from the binding ability of the antibodies or antibody fragments to their antigens onto the interaction abilities between both domains. In contrast, this study will consider the influence of isolated domains.

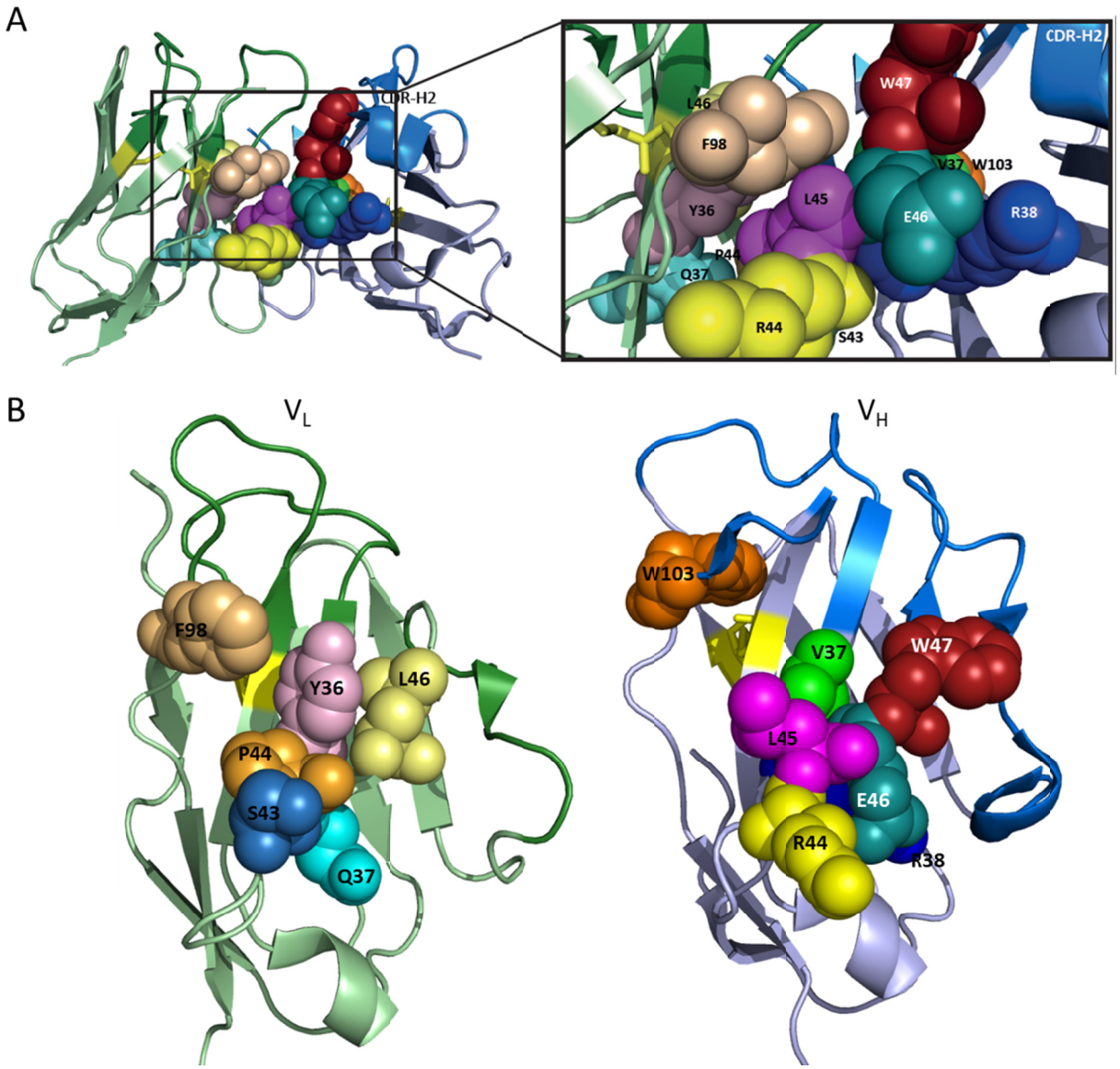

Figure 41: Conserved residues within the interface between V_L and V_H.
Covariation analysis of the variable domains revealed the presence of conserved amino acid residues in V_L and V_H. The residues with the highest scores are predominantly located in the interface between the variable domains. To determine their role for structure, function and interaction an alanine exchange study was set-up. In (A) the Fv fragment of MAK33 is shown. The V_L domain is depicted in light green with the three CDRs highlighted in dark green. V_H is shown in light blue and the CDRs are highlighted in dark blue. Conserved residues within the interface are illustrated as spheres and colour-coded. On the right the region within the rectangle is enlarged. For a better orientation CDR-H2 is indicated. The labelled residues were selected for an alanine-exchange mutational analysis. In (B) the top views of the interacting residues of the V_L (left) and V_H (right) domain are shown. Six residues were selected for V_L (Y36, Q37, S43, P44, L46, and F98) and seven for V_H (V37, R38, R44, L45, E46, W47, and W103). Structures are modified from PDB ID 1FH5.

3.2.1.1 The role of conserved residues for protein structure

The FUV-CD spectrum of V_L indicates the presence of β-sheets and random coil characteristics as two minima at 218 nm and 200 nm are visible (Figure 42 A; open circles). The result fits to the crystal structure, where besides the immunoglobulin fold, three long and flexible loop regions, the CDRs, are present (Figure 40). The characteristic shape is

observable for all tested alanine exchange mutants (F98A, Y36A, S43A, Q37A, P44A, and L45A, as well as the double mutants Y36AP44A and Y36AS43A; Figure 42 A). Small deviations in the intensity of the spectra are yet visible. For F98A (black circles) the minimum at 200 nm is more pronounced compared to the wild type and other mutants suggesting a higher flexibility due to the mutated area. For S43A (blue circles), the minimum at 218 nm is the least pronounced and more shifted to 212 nm. The FUV-CD indicates the presence of structural rearrangements but cannot give evidence of the nature of the structural changes. The same holds true for P44A (yellow circles), and Y36A (red circles). They show a signal increase of the maximum at 205 nm. Interestingly, the signal increase of both double mutants (Y36AP44A in pink circles and Y36AS43A in orange circles) is equal (Figure 42 A). One might have expected a higher increase for Y36AP44A as the single point mutants show each an increase but the effects do not seem additive. In general, the V_L domain is able to tolerate single point mutations of the conserved residues quite well and the β-sheet structure seems unaffected.

The FUV-CD spectrum of V_H wild type also shows characteristics of a β-sheet protein as it has a typical minimum at 218 nm (Figure 42 B; open circles). The signal intensity is lower compared to C_L (see Figure 10 B) but similar to V_L (Figure 42 A). As V_H also has three CDRs, their random coil characteristics diminish the β-sheet influence (see Figure 40). In general more severe effects for V_H point mutants were observable concerning the secondary structure formation than for V_L mutants (Figure 42 A and B). R38A (red circles) and W103A (pink circles) show a minimum shifted to ~205 nm (see Figure 42 B). Consequently, these domains are clearly different compared to the wild type and must have undergone structural rearrangements or might even be misfolded. Moreover, W47A (purple circles) shows a broad distributed minimum from 218 nm to 205 nm and indicates even higher random coil content like V_H wild type. For the mutants L45A (dark cyan circles) and E46A (yellow circles) the FUV-CD spectra are most similar to wild type V_H (Figure 42 B). A remarkable mutation is R44A (blue circles) which shows a more distinct minimum at 218 nm referring to a more pronounced β-sheet structure. V_H domains are in most cases less stable than their V_L counterparts [199] and hence it is not surprising that they are less tolerant against point mutations within the conserved residues.

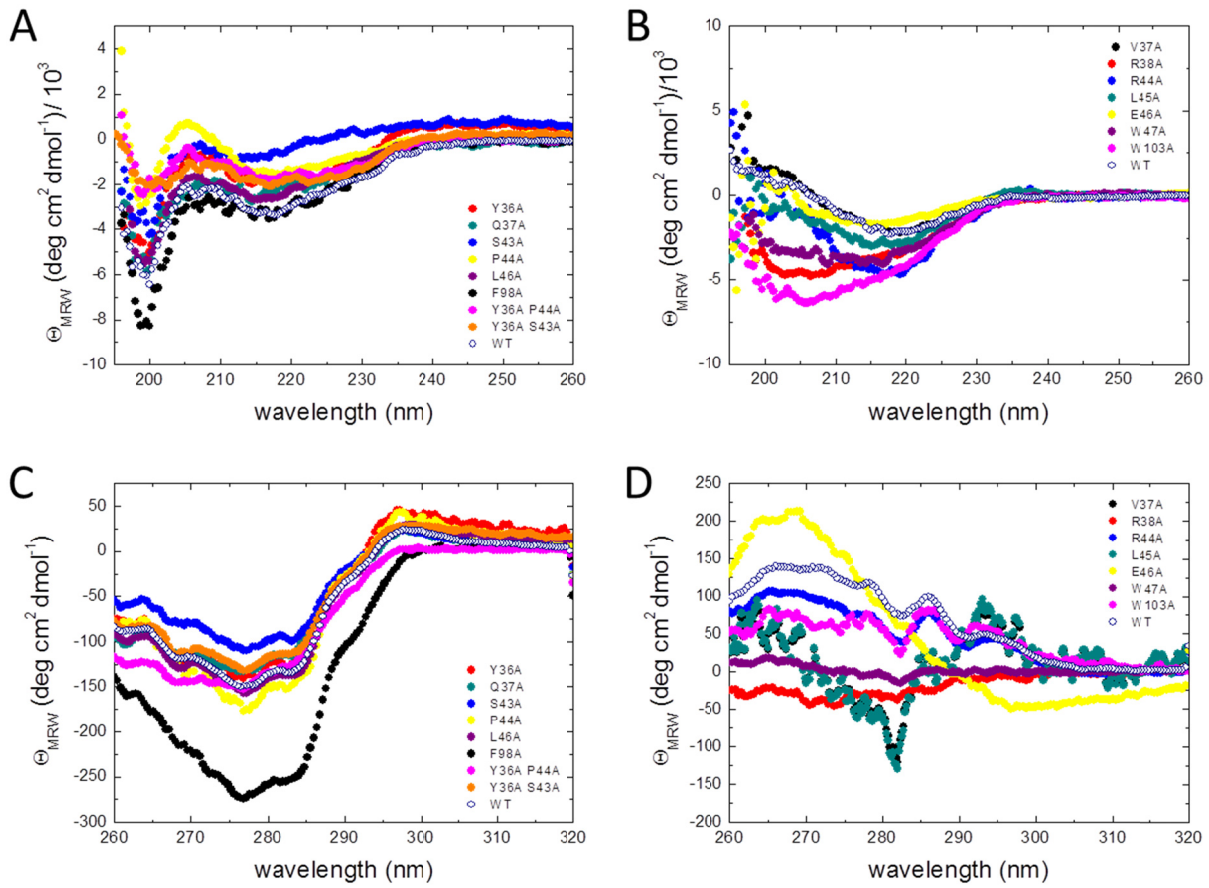

Figure 42: Secondary and tertiary structure of V_L and V_H alanine-exchange mutants.
In (A) and (B), FUV-CD spectra of V_L point mutants (A) and V_H point mutants (B) are shown. Colour code for V_L: Y36A is red, Q37A dark cyan, S43A blue, P44A yellow, L45A purple, F98A black, Y36AS43A orange, Y36AP44A pink (all closed circles/lines) and V_L wild type is royal blue (open circles/line). Colour code for V_H: V37A is black, R38A red, R44A blue, L45A dark cyan, E46A yellow, W47A purple, W103A pink (all closed circles/lines) and V_H wild type royal blue (open circles/line). In (C) and (D) the NUV-CD spectra of the V_L (C) and V_H (D) point mutants are shown. Colour code for (C) and (D) is analogue to (A) and (B), respectively. For the spectra 16 accumulations each were recorded and buffer-corrected (PBS). All measurements were performed at a protein concentration of 20 µM (FUV-CD) and 50 µM (NUV-CD) in 0.5-mm (FUV) or 5-mm (NUV) quartz cuvette at 20°C.

NUV-CD spectra can provide an indication of structural rearrangements in the environment of aromatic amino acids. V_L wild type (Figure 42 C; open circles) confirms a well-defined structure with a minimum at ~275 nm, typical for antibody domains [132, 134]. The most obvious difference between the NUV-CD spectra of wild type and mutants was observed for F98A (black circles) as it shows a ~2-fold signal increase in the range of 275 nm. As phenylalanine is, besides tryptophan and tyrosine, responsible for the NUV-CD signals, mutating one of the signal centres will influence the fingerprint. However, the shape is not affected. Thus, the overall structure seems uninfluenced (Figure 42 C). Besides F98A, slight differences were observed for S43A (blue circles) and Y36AP44A (pink circles) in the range between 270 and 260 nm (Figure 42 C). However, the overall shape is also not affected. Therefore NUV-CD spectra confirm the conclusions drawn from the FUV-CD that the V_L domain is quite tolerant against an exchange of conserved amino acids.

The V_H domain shows a well-defined NUV-CD spectrum, too. However, the NUV-CD signal is strikingly different from V_L wild type (Figure 42 C and D) and exhibits only a positive signal. The MAK33 V_H domain amino acid sequence includes three Trp, three Phe and 13 Tyr residues. In contrast, MAK33 V_L has only two Trp, five Phe and three Tyr residues. V_H is the domain with the highest content of aromatic amino acids in MAK33. The R44A exchange (blue circles) effects only slightly the tertiary structure, as the minima and maxima distribution is comparable to the wild type although the signal intensity between 290 nm and 260 nm is lower. This observation matches with the one from the FUV-CD analysis (Figure 42 B and D). Obviously different from the wild type fingerprint are the NUV-CD spectra of V37A (black circles) and L45A (dark cyan circles). They share a maximum at ~295 nm and a minimum at ~280 nm. As both mutants show the same differences the changes of the environments of the aromatic amino acids must be similar (Figure 42 D). As FUV-CD spectra of these mutations indicate β-sheet structure one can assume that rearrangements are not within the secondary structure level (Figure 42 B). The NUV-CD spectrum of E46A (yellow circles) is completely different as it shows a minimum at ~295 nm and a maximum at 270 nm (Figure 42 D). For this mutant also no differences in secondary structure were observed (Figure 42 B). The NUV-CD spectrum of R38A (red circles) is less pronounced compared to the other mutants. The signal is negative and of low intensity. Low signal intensity holds also true for W47A (purple circles) (Figure 42 D). Consequently, both mutants seem to be less defined on their tertiary structure level. The NUV-CD spectrum of W103A (magenta circles) is similar to the wild type domain with a less pronounced signal in the range between 280 nm and 260 nm. However, the domain seems to adopt defined tertiary interactions similar to the wild type, although the FUV-CD spectrum indicates differences on the secondary structure level (Figure 42 B and D). The NUV-CD spectra support the notion that V_H is less tolerant against changes within the conserved residues compared to V_L.

3.2.1.2 Conserved residues influence protein stability

Conserved residues can have different functions. Besides influencing the domain structure, stability might be affected by mutation. To address this point, thermal- and denaturant-induced (GdmCl) unfolding transitions were measured. In Figure 43, the results for V_L (A and C) and V_H (B and D) are shown. None of the thermal-induced transitions were reversible (data not shown); hence just information about the melting temperatures (T_{melt}) could be obtained. Most of the denaturant-induced transitions of V_H mutants were found to be irreversible, and the same holds true for some V_L mutants (data not shown). Nevertheless, all transitions were fitted to a two-state model to compare the results more easily.

The chemical stability of V_L is sensitive to the exchange of conserved residues within the interface to V_H as most mutations lead to a decrease in the concentration needed for half-maximum unfolding ($D_{1/2}$;Figure 43 A). The stability decline is in particular pronounced for Y36AP44A (pink circles), Y36A (red circles), Y36AS43A (orange circles) and L46A (purple circles). The findings fit well with the observations of the thermal-induced unfolding transitions (Figure 43 C). The T_{melt}s of Y36AP44A (pink circles), Y36A (red circles), Y36AS43A

(orange circles) and L46A (purple circles) are the lowest of the tested mutants. However, S43A (blue circles) is even more stable against GdmCl than the wild type and has a similar T_{melt} (Figure 43 A and C). As P44A (yellow circles) shows just a slight decrease in chemical and thermal stability and S43A (blue circles) is even more stable regarding denaturant-induced unfolding, the stability decrease must mainly be driven by the Y36A exchange (Figure 43 A and C). Therefore one can conclude that the residues Y36 and L46 are important for V_L stability. The cooperativity of GdmCl-induced unfolding transitions for all domains is in a comparable range when the errors are taken into account (Figure 43 A, Table 8). As the cooperativity parameter "m" correlates with the solvent-exposed hydrophobic surface upon unfolding and is comparable, one can assume that the structures of the mutants are still well folded. This assumption is supported by the FUV-CD and NUV-CD spectra (Figure 42 A and C). Regarding the thermal-induced unfolding transitions it is obvious that the cooperativity of Y36A (red circles/line) and P44A (yellow circles/line) is reduced compared to the wild type and the other mutants (Figure 42 C). A more detailed conclusion is not possible as the transitions were not reversible.

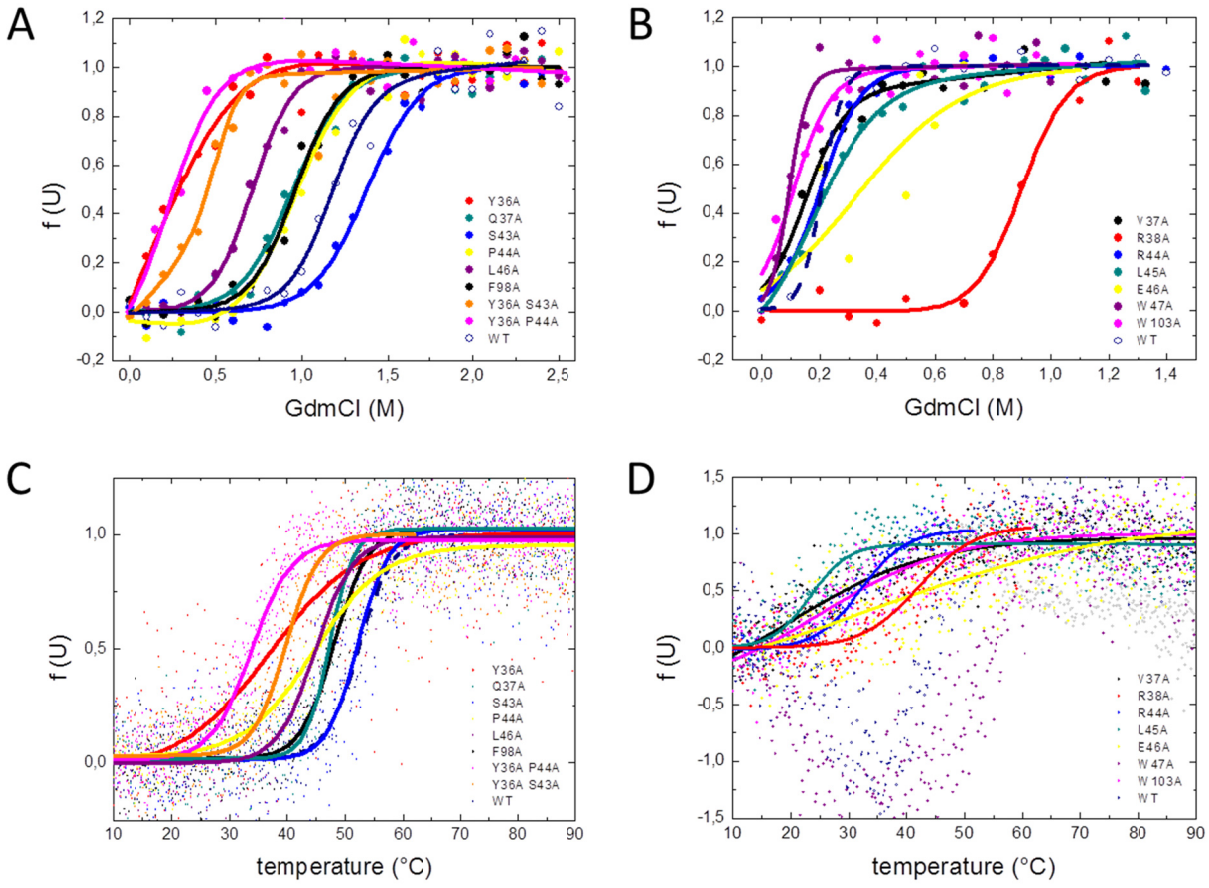

Figure 43: Influence of conserved residues onto protein stability

To assess the stability of the V_H and V_L alanine point mutants GdmCl-induced (top) and temperature-induced (bottom) unfolding experiments were performed. Data for denaturant-induced (A and B) and thermal-induced transitions (C and D) are shown. In (A) GdmCl-induced unfolding transitions for the V_L mutants are illustrated. Y36A is red, Q37A dark cyan, S43A blue, P44A yellow, L45A purple, F98A black, Y36AS43A orange, Y36AP44A pink (all closed circles/lines) and the wild type is royal blue (open circles/line). In (B) the denaturant-induced samples for the V_H mutants are shown. V37A is black, R38A red, R44A blue, L45A dark cyan, E46A yellow, W47A

purple, W103A pink (all closed circles/lines) and the wild type royal blue (open circles/broken line). In (C) and (D) the thermal-induced unfolding transitions for V_L mutants (C) and V_H mutants (D) are depicted. Colour code is the same for (C) and (A) and for (D) and (B). Temperature-induced unfolding was monitored by FUV-CD spectroscopy at 215 nm with a heating rate of 20°C h^{-1}. Data were fitted to a Boltzmann function to obtain transition midpoints. Chemical denaturation of the proteins was induced by GdmCl, and structural changes were monitored by fluorescence spectroscopy at a fixed wavelength (mainly 355 nm emission; 280 nm excitation). Data were evaluated according to a two-state unfolding model to obtain midpoints and cooperativity parameters of the transitions. Measurements were performed at 20°C at a protein concentration of 20 µM (temperature-induced) or 1 µM (GdmCl-induced).

V_H wild type (open circles) is much less stable than V_L independent of the transition type (Figure 42). It is conspicuous that all mutants besides R38A (red circles) start directly to unfold in the presence of GdmCl indicating a very low stability against denaturants (Figure 42 B). The cooperativity m is higher compared to the V_L mutants but must be considered carefully as the missing plateau of the native protein distort the two-state folding fit. The thermal stability is in general also decreased compared to V_L. V_H wild type and W47A show the same uncooperative behaviour and unfold in the range of 30°C (Figure 42 D). Both proteins aggregate at temperatures > 30°C. The low cooperativity of V37A (black circles/line), W103A (pink circles/line), and E46A (yellow circles/line) is noticeable (Figure 42 D). Stability analysis confirms the observations made in the structural analysis that V_H is very sensitive to manipulations within the conserved residues. Reliable beneficial effects regarding stability could only be made for R44A, which is reversible in GdmCl, and R38A, which shows the most explicit stability increase for thermal- and chemical-induced unfolding transitions (Figure 42 B and D). The observation for R38A is in contrast to the low signal intensity in the NUV-CD spectrum (Figure 42 D).

Table 8: Thermal and chemical stability of V_L and V_H alanine exchange mutants.

Domain	Mutation	T_{melt} [°C]	$D_{1/2}$ [M]	m [kJ mol^{-1} M^{-1}]
V_L	wild type	52.3 ± 0.2	1.17 ± 0.68	20.1 ± 8.2
	Y36A	37.1 ± 1.7	0.30 ± 0.06	16.1 ± 2.3
	Q37A	47.3 ± 0.3	0.94 ± 0.46	16.0 ± 5.2
	S43A	52.1 ± 0.4	1.38 ± 1.04	15.8 ± 8.6
	P44A	44.5 ± 0.5	0.95 ± 19.02	17.8 ± 25.0
	L46A	44.8 ± 0.4	0.71 ± 0.16	21.2 ± 3.5
	F98A	48.1 ± 0.3	0.96 ± 0.20	19.0 ± 2.8
	Y36A P44A	33.8 ± 0.2	0.52 ± 0.43	33.7 ± 18.6
	Y36A S43A	39.1 ± 0.3	0.23 ± 0.17	17.1 ± 6.0
V_H	wild type	~30	0.21 ± 0.10	68.6 ± 23.3
	V37A	14.2 ± 6.8	0.15 ± 0.04	35.3 ± 6.7
	R38A	42.2 ± 0.8	0.90 ± 0.69	30.8 ±16.7
	R44A	32.6 ± 0.3	0.20 ± 0.03	34.3 ± 3.7
	L45A	23.2 ± 0.7	0.17 ± 0.14	21.5 ± 8.5
	E46A	34.8 ± 11.1	0.32 ± 0.79	75.2 ± 15.4
	W47A	~30	0.09 ± 0.03	75.2 ±19.4
	W103A	25.2 ± 3.7	0.11 ± 0.03	37.1 ± 6.3

Stabilities against the thermal and chemical denaturation (GdmCl) of the V_L and V_H mutants in PBS are shown. Midpoints of thermal transitions are shown as T_{melt}. Even though most of the GdmCl-induced unfolding transitions were not reversible, data were evaluated according to a two-state equilibrium unfolding model to derive the midpoint of transitions ($D_{1/2}$), as well as the cooperativity parameter (m), for a qualitative comparison of the data.

3.2.1.3 Interaction between V_H and V_L is influenced by conserved residues

In contrast to previous studies on the Fv and Fab level, in this work isolated domains are considered. To be able to characterise the interaction between V_H and V_L sedimentation equilibrium AUC and SPR experiments were performed. In the AUC set-up both interaction partner are in solution and have the maximum degree of freedom. For SPR binding a ligand is immobilised and the analyte is added with continuous flow.

For sedimentation equilibrium experiments three samples with different concentrations (in general 7.5 µM, 5.0 µM and 2.5 µM; for low K_Ds down to 500 nM) in a 1:1 ratio were prepared. The samples were loaded in six channel epon centrepieces. Three different velocities were observed until the samples reached their equilibrium. In Figure 44 an example for the interaction of V_L Y36AS43A *vs.* V_H wild type is shown. In the best case more than six traces could be included in the fit model, in the worst case just three. The fit errors were in a range between 6% and 25%.

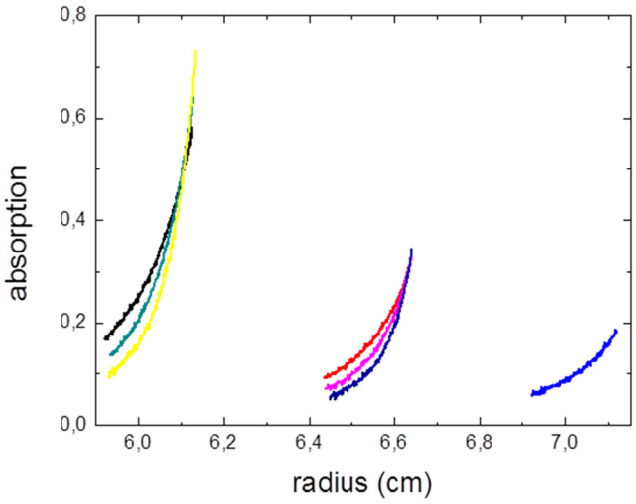

Figure 44: Example of an AUC sedimentation equilibrium experiment.
K_D values were determined by AUC sedimentation equilibrium experiments. As an example the experiment for V_H wild type vs. V_L Y36AS43A is shown. Three different concentrations (from left to right: high to low concentrations) and velocities (from top to bottom: slow to fast velocities) in a 1:1 ratio were measured and fitted to a model for heterogeneous interaction between two components. Each colour represents one concentration at one certain velocity. Temperature was set to 20°C and 15°C for V_L and V_H mutants, respectively.

Analysis of the interaction behaviour of V_H mutants shows a decrease of the binding affinity. According to the K_D values determined by AUC, V37A, R38A, L45A and W103A are those domains with the worst binding behaviour. W47A, R44A and E46A also have a larger K_D compared to the wild type but to a lesser extent. An extremely poor binding behaviour was observed for L45A and indicates a loss of the interaction with V_L. Yet one has to keep in mind that the equilibration equilibrium experiments take several days and some of the V_H mutants are only marginally stable. Consequently, some might not be stable enough for the long measuring time. For E46A, for example, in some experiments degradation was observed. The K_D values are summarised in Table 9.

Table 9: K_D determination by AUC SE experiments and SPR

	AUC	SPR		AUC	SPR
V_L	**K_D [µM]**	**K_D [µM]**	**V_H**	**K_D [µM]**	**K_D [µM]**
wild type	0.2	1.3	wild type	0.2	1.3
Y36A	0.4	ND	V37A	33.0*	4.4
Q37A	0.4	ND	R38A	17.3*	2.3
S43A	0.5	ND	R44A	1.8	2.1
P44A	15.2	ND	L45A	292.0*	50.0
L46A	0.6	ND	E46A	2.0*	1.3
F98A	0.6	ND	W47A	1.2*	5.7
Y36AS43A	0.4	ND	W103A	26.4*	6.9
Y36AP44A	6.8	ND			

K_D values were determined by AUC sedimentation equilibrium experiments. Three concentrations in the range between 0.5 µM and 7.5 µM with V_H and V_L in a 1:1 ratio were measured at three velocities and globally fitted to a model heterogeneous interaction between two components. All mutants were tested with the wild type V_H or V_L domain. Experiments were performed at 20°C; asterisks indicate that samples were measured at 15°C. Errors of the fits for AUC data were in a range between 6%-25%. SPR experiments were performed with a Biacore X100. Ligand was immobilised via an introduced cysteine at the N-terminus. ND indicates that the samples could not be analysed.

The Biacore data confirm the poor binding behaviour of L45A. V37A, W47A and W103A share also higher K_D values. However, their increase is not as pronounced as for L45A. In general, the same effects were observed as in the AUC experiments although not the same K_D values were obtained. The only exception is E46A, which shows poor binding abilities in the AUC experiments, whereas it binds V_L wild type with the same affinity when analysed by SPR.

Due to problems with the stability of V_H with an N-terminal cysteine immobilised to a CM5 chip the K_D values of the V_L mutants could not be determined. V_L wild type was the only sample which could be measured before the performance of the chip decreased. The K_D of the wild type V_H domain with the V_H N-terminal cysteine domain and V_L was the same as for V_H wild type and V_L with an N-terminal cysteine (K_D= 1.3 µM) (see Table 9).

In Figure 45 the results for the AUC and SPR experiments are summarised. For V_L, conserved residues are labelled according to the AUC data as no SPR measurements were possible. For V_H, both methods could be used. P44A (highlighted in red) was the only mutation which strongly decreased the interaction between V_L and V_H. The other residues had only minor effects (orange; Figure 45 B). In the case of V_H, different results were observed for E46A and R38A when analysed by AUC or SPR (Figure 45 bottom). W103A, V37A and L45A showed the strongest effects on association.

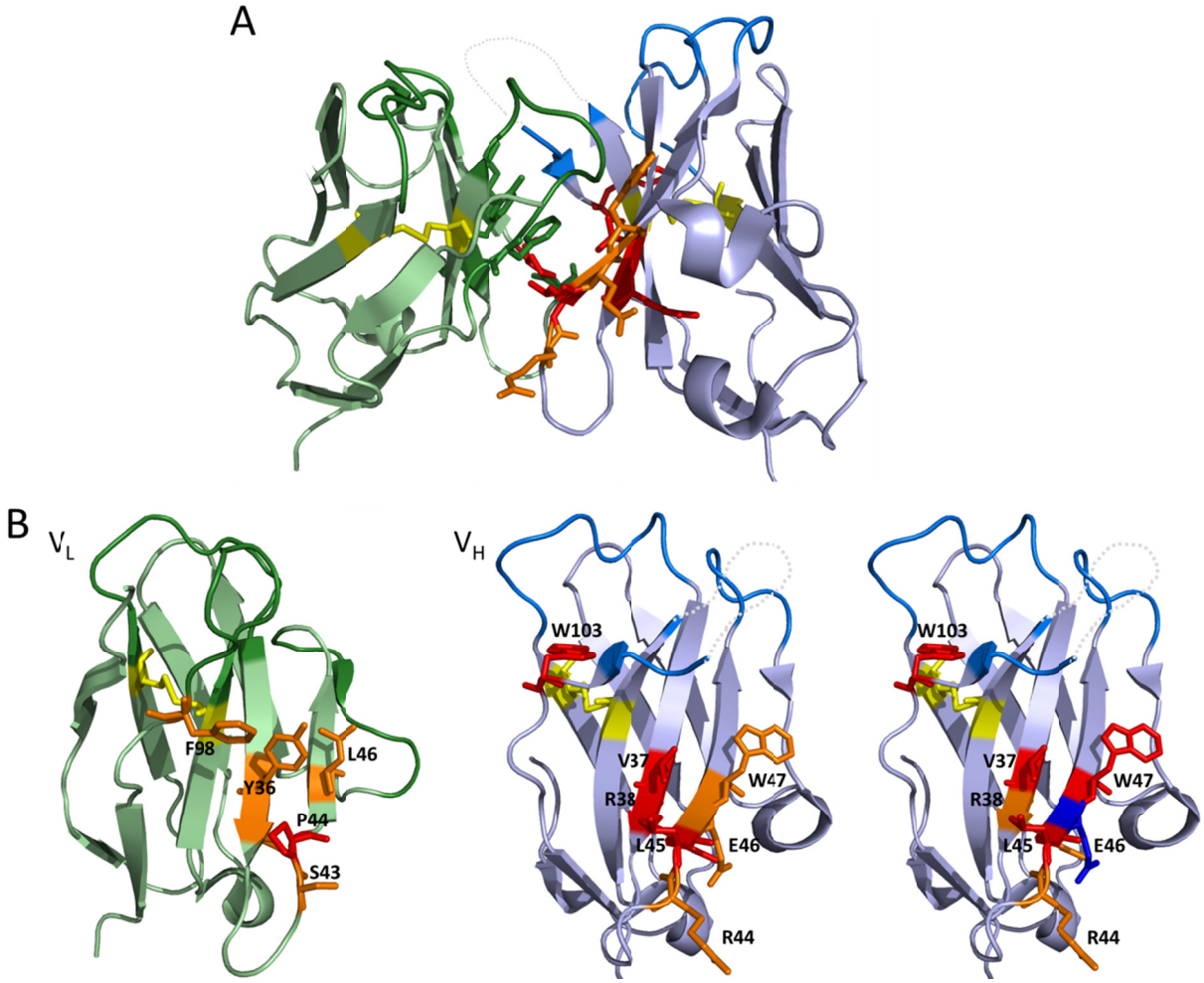

Figure 45: Critical residues for V_L-V_H interaction.

In (A) the heterodimer of V_L and V_H is shown. The conserved residues from the interaction study are shown as sticks. Yellow sticks represent the internal disulphide bridge. Residues highlighted in orange had only minor effects on the K_D of V_L and V_H when mutated to alanine, whereas in red labelled residues showed a strong effect on the interaction. Residues without a clear influence are dark green (V_L) or dark blue (V_H). In (B) the top-views for V_L and V_H are shown. For some residues AUC and SPR experiments resulted in different observations. For a comparison both results for the V_H domain are shown. The grey dotted line indicates CDR-H3 which is not resolved in PDB ID 1FH5.

3.2.1.4 Functionality of V_H and V_L is influenced by conserved residues

Besides structure, stability and association conserved amino acids might also influence antigen recognition. Consequently, the functionality of the Fv fragment can be disturbed. To address this question an ELISA with isolated variable domains was established. MAK33 is directed against human creatine kinase. The biotinylated antigen was immobilised to a streptavidine coated 96-microwell plate. For detection the wild type domain was labelled with a FLAG-tag at the C-terminus. The tag does not influence the interaction between the wild type variable domains (determined by AUC, data not shown). An anti-FLAG antibody coupled to a horseradish peroxidase allows detection of the bound wild type domains. For the case that the mutants interact with the wild type domains and are functional, the read-out should fit to the wild type domains. As this assay is complex due to several K_D values

(tagged domain to antigen, mutant to antigen, mutant to tagged domain) only qualitative and no quantitative conclusions are possible. Figure 46 shows representatives for inactive, mediate-active and wild-type like mutants.

The binding ability of V_L-Flag is strongly decreased compared to wild type V_H when W47 is exchanged to an alanine as no binding could be observed at all. L45A and R38A show a weaker interaction, however, stronger compared to V37A and W103A (see Table 10). The low activity of W47A and L45A fit well with the high K_D determined by AUC and SPR (see Table 9). Interestingly, R38A cannot bind to the antigen as efficient as the wild type indicating that the K_D from the AUC experiment might not be abnormal although the K_D determined by SPR clearly differs. E46A, which had a similar K_D in SPR and a higher K_D in AUC as the wild type (see Table 9), is able to bind to human creatine kinase as efficient as the wild type (Table 10). The same holds true for R44A. To sum up the observations, functionality is not always correlated with association.

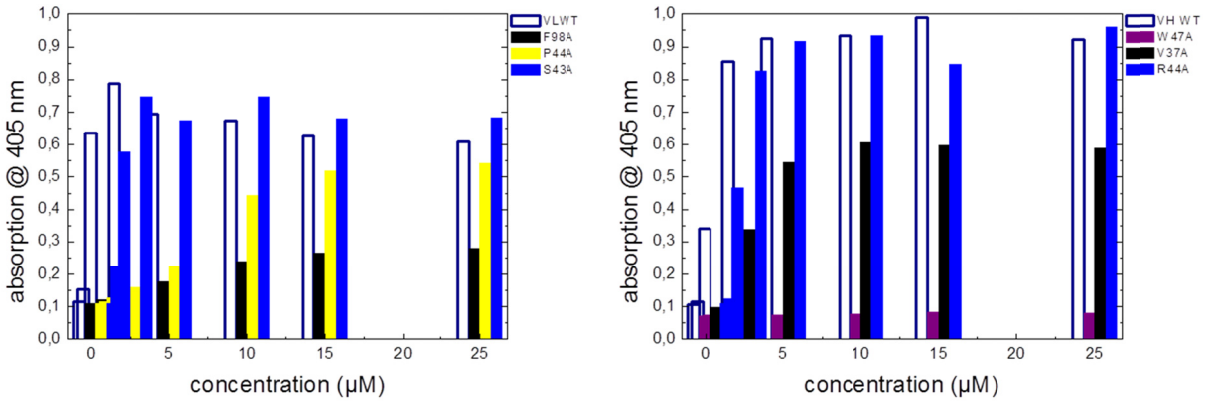

Figure 46: ELISA with isolated variable domains to determine the influence of conserved amino acids on antigen binding.

To determine the functionality of the single mutants an ELISA with isolated domains was performed. Biotinylated human creatine kinase was coupled to a streptavidine-coated microwell plate. Wild type V_H and V_L were labelled with a FLAG-tag at the C-terminus and could be detected with an anti-FLAG antibody coupled to horseradish peroxidase. For V_L (left side) and V_H (right side) the wild type, a representative of an inactive, a low active and a mutant with similar activity as the wild type is shown, respectively. The absorption at 405 nm at the signal maximum (after 15 minutes) was corrected for the signal of the variable labelled domains alone. Experiments were performed at 20°C.

The V_H-FLAG domain is less efficient in interacting with the antigen when V_L with a L46A or F98A mutation is the binding partner. Y36A, P44A and the double mutants Y36AS43A, as well as Y36AP44A, also have a negative influence on the functionality but to a lesser extent. On the contrary, Q37A and S43A do not influence antigen binding. The only domain with a clearly reduced K_D (AUC, Table 9) was P44A. However, this domain shows only a slight decrease in its ability to bind the antigen. On the contrary, L46A and F98A could bind the V_H domain with a K_D similar to the wild type but they are not able to interact with the antigen any more (Table 9 and Table 9). Consequently, the analysis of V_L mutants enhances the

assumption that interaction between V_H and V_L is not necessarily correlated with the ability to bind the antigen.

Table 10: Functionality is influenced by mutations of conserved residues.

	V_H mutant	relative signal		V_L mutant	relative signal
V_L-FLAG	wild type	wt	V_H-FLAG	wild type	wt
	V37A	-		Y36A	-
	R38A	--		Q37A	wt
	R44A	wt		S43A	wt
	L45A	--		P44A	-
	E46A	wt		L46A	---
	W47A	---		F98A	---
	W103A	-		Y36AS43A	-
				Y36AP44A	-

To determine the functionality of the mutants a qualitative analysis of ELISA with isolated variable domains was performed. Detection of V_H and V_L wild type domains was possible via an introduced FLAG-tag at the C-termini. As a reference the signal of the wild type domains vs. FLAG-tagged domains after 15 minutes of incubation were chosen. Samples were corrected with the signal from the single FLAG-tagged variable domains. One minus up to three was chosen to group the different active mutants (from slightly lower [-] binding to no binding [---]). Experiments were performed at 20°C.

3.2.2 The influence of the framework and CDRs

In terms of structure, amino acid sequence diversity and germline usage seven V_H and seven V_L (four V_κ and three V_λ) germline families can be distinguished in the human antibody repertoire. They cover more than 95% of the human antibody diversity [238]. A systematic evaluation of these human variable domain germline families as isolated domains was performed by Ewert and co-workers in 2003 [199]. They found differences between the representatives of V_H and V_L germline families regarding stability, oligomeric state and expression behaviour. The domain with the highest thermodynamic stability and yield of soluble protein was the V_H3 family (modelled structure PDB ID 1DHU) and $V_\kappa3$ (modelled structure PDB ID 1DH5) [199].

As the stability of isolated MAK33 V_L and especially V_H is lower compared to these two consensus sequences a sequence alignment (framework and CDRs) was performed to determine differences on the amino acid sequence level. Interestingly, MAK V_L and 1DH5 share 65% identity and 83% homology. For MAK33 V_H and 1DHU even 74% sequence identity and 82% positive hits were identified (Figure 47). The results of these alignments are shown in Figure 47 A and B. Interestingly, CDR-L1, CDR-H1 and CDR-H2 are similar between MAK33 and the consensus sequences.

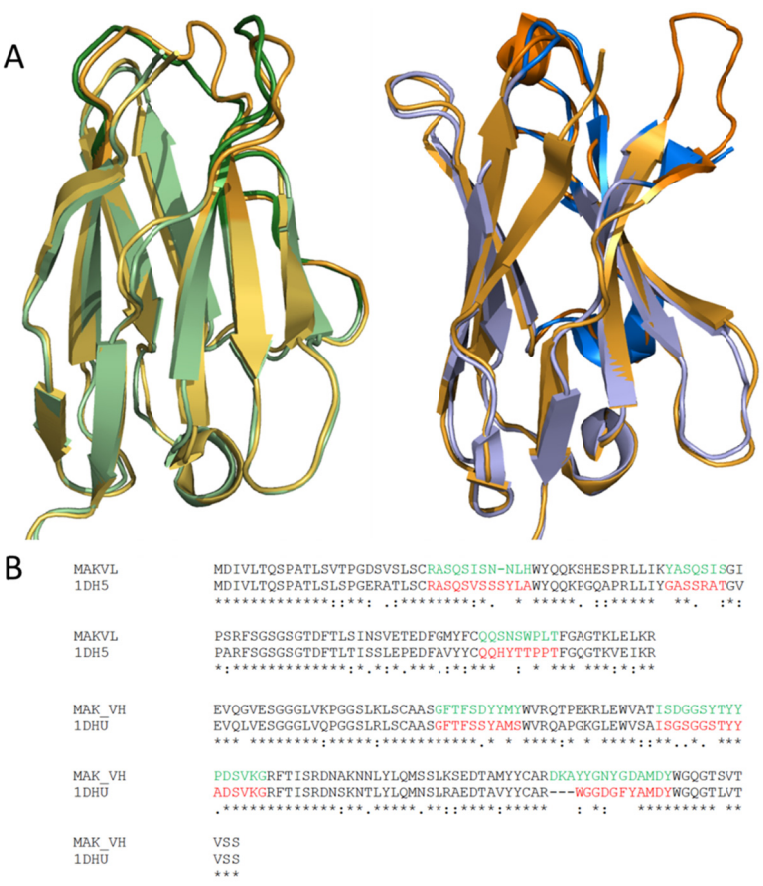

Figure 47: Structural and sequence alignment between MAK33 variable domains and human consensus sequences.

In (A) the structural alignment between MAK33 V_L and 1DH5 (left) and MAK33 V_H and 1DHU (right) is shown. MAK33 V_L is depicted in light green with the three CDRs highlighted in green. The most stable human V_L consensus sequence is yellow-orange with orange CDRs. MAK33 V_H is illustrated in light blue with blue CDRs and the human V_H consensus sequence in yellow-orange with orange CDRs. Structures are based on PDB IDs 1FH5, 1DHU and 1DH5. (B) depicts the sequence alignment between the variable domains of the light chain (top) and of the heavy chain (bottom) shown. CDRs are highlighted in green and red (assignment according to http://www.bioinf.org.uk/abs/).

For an analysis of the influence of framework and CDR regions, grafting mutants between MAK33 V_L and 1DH5 as well as MAK33 V_H and 1DHU were designed. The constructs used for this study are shown in Figure 48.

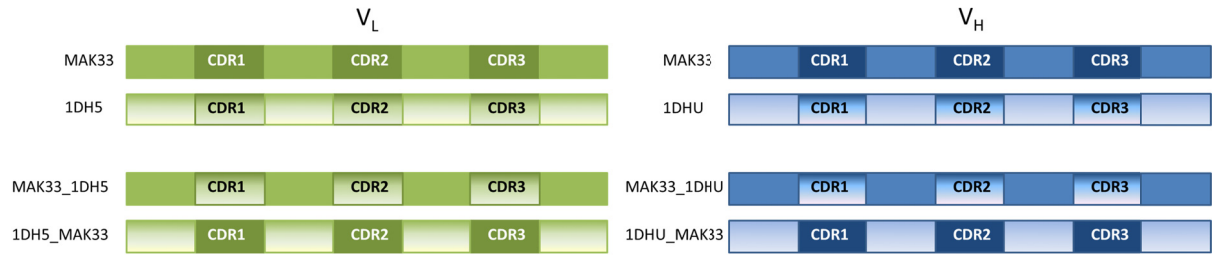

Figure 48: CDR grafting mutants.

To determine the influence of the framework (FW) residues and CDR regions on the structure and function of the variable domains, CDR grafting mutants were designed. The candidate for V_L grafting mutants (left side) was 1DH5, which is the most stable human consensus sequence for V_L domains (from HuCAL [238]). Consequently, four constructs will be compared: MAK33 wild type, 1DH5 wild type, MAK33_1DH5 and 1DH5_MAK33. For the MAK33 V_H domain (right side) 1DHU, which is the most stable human consensus sequence for V_H domains, was selected. Four V_H constructs will be analysed: MAK33 wild type, 1DHU wild type, MAK33_1DHU and 1DHU_MAK33. (Nomenclature: FW_CDR).

3.2.2.1 Structure of grafting and CDR mutants

In Figure 49 A the FUV-CD spectra for MAK33 V_L and 1DH5 V_L as well as the grafting mutants are shown. MAK33 V_L (black, open circles) is a β-sheet with random coil characteristics (see 3.2.1.1). The FUV-CD spectrum of 1DH5 V_L (green, closed circles) shows a minimum at 218 nm typical for β-sheet proteins. The signal intensity at 200 nm is positive, another typical feature for β-sheet proteins. Hence, one can assume that the influence of the flexible CDRs is less pronounced in 1DH5 compared to MAK33 V_L (Figure 49 A). The grafting mutant based on the 1DH5 framework (1DH5_MAK V_L; blue circles) has a similar spectrum as 1DH5 V_L, encouraging the assumption that the framework is less influenced by the CDRs (Figure 49 A). However, the grafting mutant based on the MAK33 V_L framework shows a clearly different FUV-CD spectrum (Figure 49 A, red circles). Besides the minimum at 218 nm and the maximum at 200 nm there is a second minimum at ~230 nm. In conclusion, the framework of MAK33 V_L is more sensitive to CDR changes than 1DH5 V_L.

The FUV-CD spectrum of MAK33 V_H indicates a protein with β-sheet structure (Figure 49 B, black, open circles; see also 3.2.1.1). The 1DHU V_H domain is characterised by a minimum at 218 nm and a positive signal at 200 nm typical for antiparallel β-sheet proteins (Figure 49 B, green circles). The minimum is less pronounced compared to MAK33 V_H or the grafted mutants which might be an influence of the consensus CDRs. Both grafting mutants, MAK_1DHU (red circles) and 1DHU_MAK V_H (blue circles) show a similar FUV-CD spectrum with a minimum at 218 nm and a maximum at 200 nm (Figure 49 B). The maximum intensity at 200 nm is higher for 1DHU_MAK compared to MAK_1DHU V_H. One might conclude that the influence of the CDRs from MAK33 is smaller than the one of the CDRs from 1DHU and consequently the FUV-CD spectrum shows less random coil characteristics. In contrast, the observation of the wild type domains cannot support this assumption (Figure 49 B).

The NUV-CD spectrum of MAK33 V_L indicates a well-defined structure with the typical minimum at ~275 nm (Figure 49 C, black, open circles; see also 3.2.1.1). The spectrum of 1DH5 (green circles) corresponds to a defined structure as well. Small differences seem to exist within the environment of the aromatic amino acids as the minimum is shifted to ~280 nm. Sequence alignment of MAK33 and 1DH5 V_L demonstrates differences in amino acids including aromatic ones (see Figure 47). This might explain the differences. The grafting mutants are well defined and the 1DH5_MAK V_L (blue circles) is more similar to 1DH5; in contrast, MAK_1DH5 (red circles) is more like MAK V_L (see Figure 49 C). The framework residues have a stronger influence on the NUV-CD spectrum compared to the CDRs, which account for ~25% of the variable domain sequence.

The NUV-CD spectra of the V_H domains of the heavy chain gave further insights in the structure of the variants. The NUV-CD of MAK33 V_H (see Figure 49 D, black, open circles; see also 3.2.1.1) and 1DHU V_H (green circles) are clearly different (Figure 49 D). The spectrum of MAK_1DHU V_H (red circles) is quite similar to MAK33, whereas the 1DHU_MAK V_H (blue circles) differs strongly from the spectrum of 1DHU (Figure 49 D). The role of the CDRs seems to be more crucial for the tertiary structure of V_H than for V_L (Figure 49 C and D).

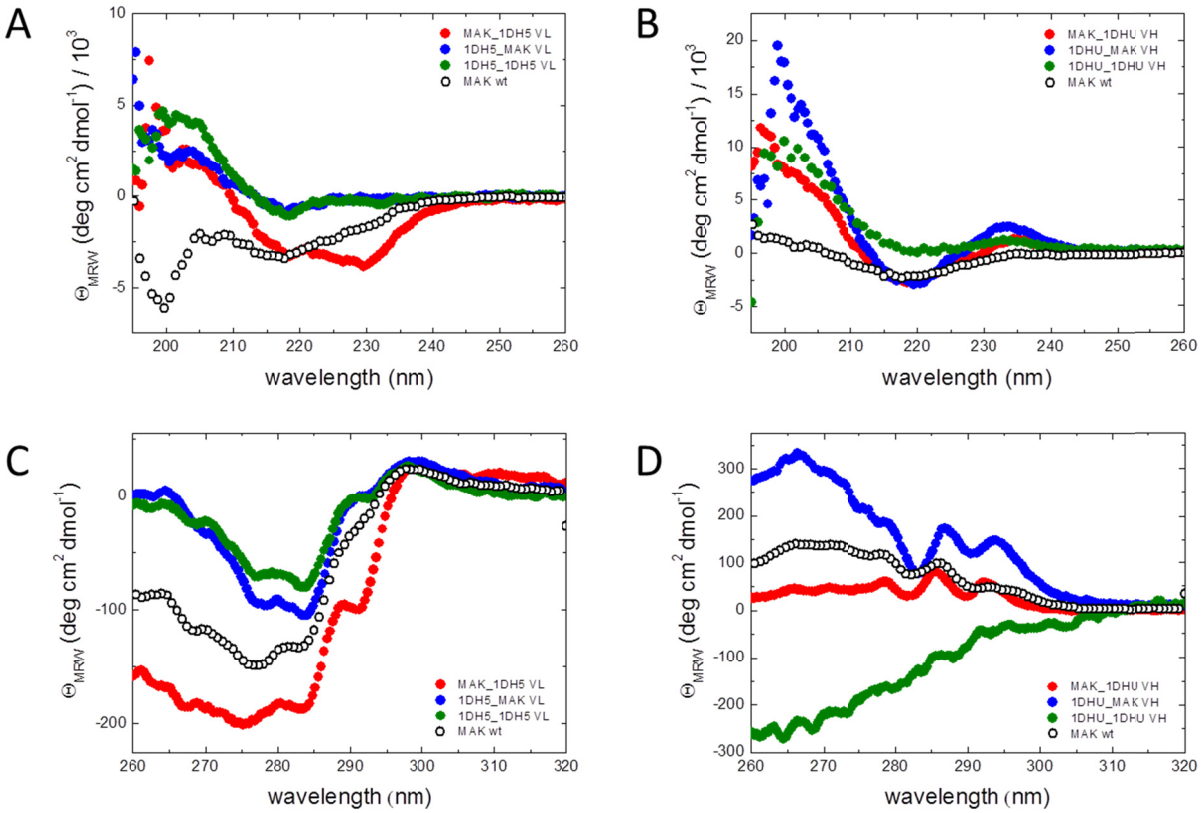

Figure 49: The influence of CDR grafting on the secondary and tertiary structure of V_L and V_H.
In (A) and (B), FUV-CD spectra of V_L (A) and V_H (B) CDR grafting mutants are shown. Colour code for V_L (A) is: MAK_1DH5 in red, 1DH5_MAK in blue (closed circles); wild type 1DH5 is shown in green (closed circles) and MAK33in black (open circles). Colour code for V_H (B) is: MAK_1DHU in red, 1DHU_MAK in blue (closed circles) and the wild type 1DHU in green (closed circles) and MAK33 in black (open circles). In (C) and (D) NUV-CD spectra of the V_L (C) and V_H (D) point mutants are shown. Colour code for (C) and (D) is analogous to (A) and (B), respectively. For all spectra were 16 accumulations each recorded and buffer-corrected (PBS). All measurements were performed at a protein concentration of 20 µM (FUV-CD) and 50 µM (NUV-CD) in 0.5-mm (FUV) or 5-mm (NUV) quartz cuvette at 20°C.

3.2.2.2 Protein stability can be influenced by framework sequences and CDRs

An interesting observation is that the human consensus variable domains are quite stable [199] against chemical and thermal denaturation whereas the MAK33 V_L and V_H are less stable (especially V_H) [192], although, the sequence identity between MAK33 and the appropriate consensus sequences is quite high (see Figure 47 B). To determine the influence of the framework and CDR sequences on protein stability, GdmCl-induced and thermal induced unfolding transitions were performed. Thermal transitions were observed by FUV-CD at 215 nm; GdmCl-induced transitions were followed by fluorescence changes of the internal tryptophan residues at 355 nm emission (excitation at 280 nm).

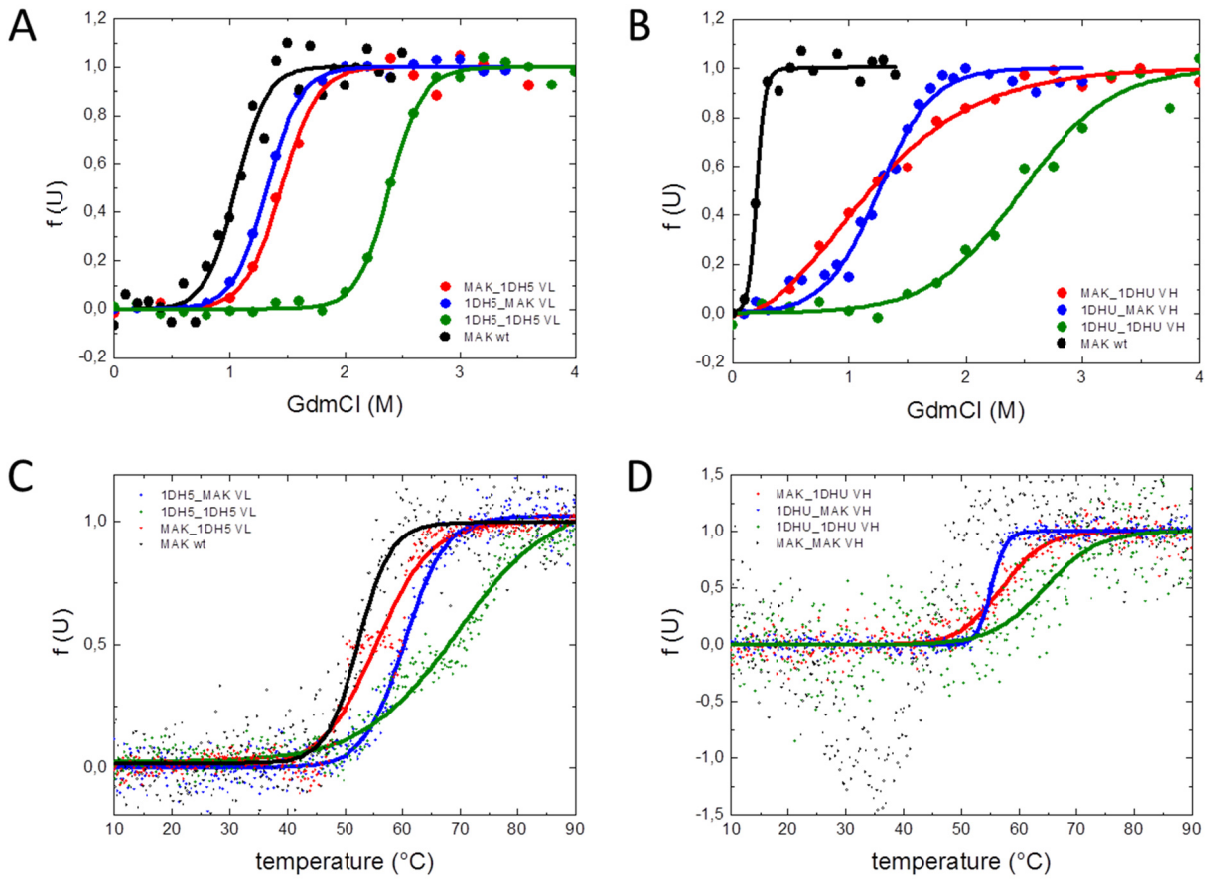

Figure 50: CDRs and framework influence protein stability

To assess the stability of the V_H and V_L grafting mutants GdmCl-induced (top) and temperature-induced (bottom) unfolding experiments were performed. Data for denaturant-induced (A and B) and thermal-induced transitions (C and D) are shown. In (A) GdmCl-induced unfolding transitions for the V_L mutants are illustrated. MAK33 wild type is shown in black, 1DH5 wild type in green, MAK_1DH5 in red and 1DH5_MAK in blue (in each case circles and line). In (B) the denaturant-induced samples for the V_H mutants are shown. MAK33 wild type is coloured in black, 1DHU wild type in green, MAK_1DHU in red and 1DHU_MAK in blue. In (C) and (D) the thermal-induced unfolding transitions for V_L mutants (C) and V_H mutants (D) are depicted. Colour code is for (C) the same as for (A) and for (D) the same as for (B). Temperature-induced unfolding was monitored by FUV-CD spectroscopy at 215 nm with a heating rate of 20°C h^{-1}. Data were fitted to a Boltzmann function to obtain transition midpoints. Chemical denaturation of the proteins was induced by GdmCl, and structural changes were monitored by fluorescence spectroscopy at a fixed wavelength (mainly 355 nm emission; 280 nm excitation). Data were evaluated according to a two-state unfolding model to obtain midpoints and cooperativity parameters of the transitions. Measurements were performed at 20°C at a protein concentration of 20 µM (temperature-induced transitions) or 1 µM (GdmCl-induced transitions).

GdmCl-induced transitions show that MAK33 V_L is the least stable domain against chemical denaturation (Figure 50 A, black circles/line; Table 11). The most stable domain is 1DH5 V_L (green circles/line) with a $D_{1/2}$ of ~2.4 M (Figure 50 A, Table 11). The grafted mutants have stabilities in between MAK33 V_L and 1DH5 (Figure 50 A). Consequently, the CDRs of MAK33 decrease the stability of the 1DH5 framework as 1DH5_MAK (blue circles) is roughly as half as stable as 1DH5 (green circles/line) (Figure 50 A). In contrast, the stability of the MAK33 V_L framework is increased when the CDRs are exchanged against the consensus CDRs of 1DH5. MAK_1DH5 (red circles/line) is about 20% more stable than MAK33 V_L (black circles/line)

(Figure 50 A, Table 11). The m-values of all domains are similar if one takes the errors of the fit into account (Table 11).

MAK33 V_H is obviously the least stable domain in the presence of GdmCl (black circles/line, Figure 50 B and Table 11). The most stable domain is the consensus sequence of V_H, 1DHU (green circles/line). The grafting mutants are more stable than MAK33 V_H but less stable compared to 1DHU (Figure 50 B and Table 11). When the CDRs of MAK33 V_H are exchanged against the CDRs of 1DHU, the thermal stability increases on the one hand but on the other hand the transition is clearly less cooperative compared to MAK33 V_H (Figure 50 B, Table 11). The exchange of the CDRs of 1DHU against MAK33 CDRs decreases the stability compared to 1DHU, yet the cooperativity change is less pronounced compared to MAK33 (Figure 50 B and Table 11). The analysis of the chemical stability allows the same conclusion for V_L and V_H: the CDRs influence the stability of the variable domains but the effects can be positive as well as negative.

MAK33 V_L is a stable domain against thermal-induced denaturation with a T_{melt}= 52.3 ± 0.2°C (black line; Figure 50 C and Table 11). However, 1DH5 is clearly more stable with a T_{melt}= 69.0 ± 0.6°C (green line; Figure 50 C and Table 11). The grafting mutants were in between those two domains. The thermal transition of MAK_1DH5 was determined in the NUV-CD as the transition in the FUV-CD was too noisy to be analysed. It has to be kept in mind, that for this domain tertiary structure changes and not changes on the secondary structure level were observed. The cooperativity is lower for 1DHU compared to MAK33 V_L. Interestingly, 1DH5_MAK V_L seems more cooperative than MAK_1DH5 (Figure 50 C). These observations lead to the conclusion that the CDRs can have an effect on the stability and also on the cooperativity of unfolding.

In Figure 50 D, the thermal-induced transitions for the V_H domains are shown. MAK33 V_H is the least stable domain and unfolds non-cooperative (black circles). At temperatures > 30°C the protein aggregates. However, 1DHU is a very stable domain with a T_{melt}= 64.4 ± 0.8°C. An exchange of the 1DHU CDRs against the MAK33 CDRs leads to a decrease in stability, on the other hand is the stability for MAK33 increased if the CDRs of 1DHU are grafted onto the framework (Figure 50 D and Table 11). In principle the observations for V_H match the finding for the V_L domains.

Table 11: Thermal and chemical stability of V_H and V_L grafting mutants

Domain	Mutation	T_{melt} [°C]	$D_{1/2}$ [M]	m [kJ mol^{-1} M^{-1}]
V_L	wild type	52.3 ± 0.2	1.17 ± 0.68	20.1 ± 8.2
	MAK_1DH5	55.5. ± 0.2*	1.44 ± 0.66	15.6 ± 4.8
	1DH5_MAK	60.4 ± 0.1	1.32 ± 0.21	16.5 ± 1.8
	1DH5_1DH5	69.0 ± 0.6	2.39 ± 1.21	17.1 ± 6.1
V_H	wild type	~30	0.21 ± 0.07	67.3 ± 17.5
	MAK_1DHU	57.0 ± 0.3	0.52 ± 0.56	4.5 ± 0.4
	1DHU_MAK	55.2 ± 0.1	1.26 ± 0.16	10.7 ± 0.9
	1DHU_1DHU	64.4 ± 0.8	2.48 ± 1.91	6.3 ± 3.6

Stabilities against the thermal and chemical denaturation (GdmCl) of the V_L and V_H mutants in PBS are shown. Midpoints of thermal transitions are given as T_{melt}. Even though most of the GdmCl-induced unfolding transitions were not reversible, data were evaluated according to a two-state equilibrium unfolding model to derive the midpoint of transitions ($D_{1/2}$), as well as the cooperativity parameter (m), for a qualitative comparison of the data. Asterisks indicate that the thermal transition was determined by NUV-CD.

3.2.2.3 Interaction between V_H and V_L

Although 1DHU and 1DH5 are consensus sequences and no naturally occurring variable domains, the K_Ds for their association were determined using AUC sedimentation equilibrium approach and SPR. For AUC experiments all domains were tested against MAK33 V_L or V_H, for SPR measurements MAK33 V_L or V_H with an N-terminal cysteine immobilised and tested against the different mutants.

CDR grafting has an influence on the binding between V_L and V_H. AUC experiments had shown that MAK33 V_H binds MAK33 V_L with a K_D of 0.2 µM (Table 9). A similar K_D could be determined for 1DHU_MAK and MAK33 V_L with 0.4 µM, whereas MAK_1DHU binds less efficient to MAK33 V_L with a K_D of 1.0 µM (see Table 12). This observation lead to the conclusion that the CDRs have a strong influence on the interaction between MAK33 V_L and V_H as MAK33 V_H and 1DHU_MAK can bind with a similar efficiency whereas MAK_1DHU is a worse binding partner. The consensus sequence 1DHU is an insufficient binding partner for MAK33 V_L, yet its abilities for interaction can be improved by CDR grafts (Table 12).

Binding experiments between V_L and V_H mutants with SPR cannot confirm all observations made with the AUC experimental set-up. The K_D between MAK V_L and V_H is slightly higher compared to the AUC experiments (1.5 µM SPR; 0.2 µM AUC). While AUC experiments can be performed with both interaction partners in solution, for SPR one partner needs to be immobilised and consequently the degrees of freedom are restricted. Additionally, the immobilisation can be disadvantageous for the accessibility of the biding site. Nevertheless, the K_Ds for MAK33 V_H and 1DHU_MAK are comparable and binding for MAK_1DHU is less efficient (Table 12). While these findings match well with the observations of the AUC experiments, the binding efficiency of 1DHU V_H according to the SPR experiments is better than observed by AUC.

Table 12: K_D determination by AUC SE experiments and SPR.

Domain	Mutation	AUC K_D [µM]	SPR K_D [µM]	Domain	Mutation	AUC K_D [µM]	SPR K_D [µM]
	wild type	0.2	1.5		wild type	0.2	1.5
	MAK_1DH5	0.3	ND		MAK_1DHU	1.0	2.8
V_L	1DH5_MAK	3.5	ND	V_H	1DHU_MAK	0.4	1.4
	1DH5_1DH5	3.8	ND		1DHU_1DHU	8.8	0.5

K_D values were determined by AUC sedimentation equilibrium experiments. Three concentrations in the range between 0.5 µM and 7.5 µM with V_H and V_L in a 1:1 ratio were measured at three velocities and globally fitted to a model for heterogeneous interaction between two components. Experiments were performed at 20°C. SPR experiments were performed with a Biacore X100. The ligand was immobilised via an introduced cysteine at the N-terminus. Errors of the fits for AUC data were in a range between 17%-25%. ND indicates that the samples could not be analysed.

As mentioned in 3.2.1.3, the stability of V_H coupled to a CM5 chip was unsatisfactory. Therefore no analysis of the V_L grafting mutants by SPR was possible.

3.2.2.4 Functionality of grafted variable domains

1DHU and 1DH5 are consensus sequences from a large set of variable domains and published in the HuCAL (human combinatorial library, see [238]). Consequently, their CDRs are not directed against a specific antigen. Nevertheless, the influence of the framework and CDR grafts between MAK33 and the consensus sequences on antigen binding was tested. None of the V_H grafting mutants (1DHU, MAK_1DHU and 1DHU_MAK) could recognise human creatine kinase, even in presence of MAK33 CDRs (data not shown). On the other hand, 1DH5 with MAK33 CDRs could recognise and bind the antigen, however, to a lower extend compared to MAK33 wild type V_L. 1DH5 and MAK33_1DH5 showed no binding at all (data not shown). This observations lead to the conclusion that the framework orientation between 1DH5 and MAK33 V_L must be similar and thus the grafted CDRs can adopt a conformation still able to recognise and bind human creatine kinase. On the V_H side, the framework conformations are probably different as no binding for 1DHU even with MAK33 CDRs could be observed.

3.2.3 The influence of the CDR-H3

CDR-H3 is known to be the most variable in sequence length and structure of all CDRs [235]. The study of Ewert and co-workers on the biophysical properties of human antibody domains comes to the conclusion that CDR-H3 might also be important for protein solubility as a long CDR is able to cover the hydrophobic interface of V_H which is the task of the V_L domain in the heterodimer [199]. The authors tested the seven V_H and seven V_L consensus

sequences from the HuCAL additionally with an 11 residue long CDR-H3 of 1FVC V_H (PDB ID; [239]) and a 17 amino acid long CDR-H3 from one of the selected V_H3 domains which had CDR-H3s with more than 15 amino acids (in the following termed as CDR-H3 of V_H3). They observed that CDR-H3s with more than 15 amino acids were favourable for V_H stabilities [199]. Additionally, a multiple sequence alignments of the CDR-H3 loops had shown, that in the consensus sequences many charged amino acids are present. In the sequence of 1DHU with the long CDR-H3 (from selected V_H3; YNHEADMLIRNWLYSDV) the same can be observed. This is one starting-point to modify the sequence of the MAK33 CDR-H3 (DKAYYGNYGDAMDY). To get a general idea if additional charged residues can influence the V_L-V_H interaction, single (A103E/D, G106R) and double (A103DG106R) point mutants in the CDR-H3 were generated. Additionally, double mutants between R44A (conserved residue of the interface) and charged CDR-H3 residues were combined. R44A had favourable effects on V_H stability but decreased the association of the variable domains (see section 3.2.1). In this approach should be tested if the effects of conserved interface and CDR-H3 residues will be additive. The effect of an extended CDR-H3 on the V_L-V_H complex formation was addressed with two constructs with a 17 amino acid long CDR-H3 in MAK33 V_H and the grafted domain MAK_DHU V_H.

3.2.3.1 Role of CDR-H3 for structure

In a first step structural changes due to point mutations and CDR-H3 exchanges were addressed (Figure 51). The FUV-CD spectrum of MAK V_H wild type indicates secondary structure of a β-sheet with random coil characteristics due to the low minimum at 218 nm and the low maximum at 200 nm (Figure 51 A and B; open circles in black and 3.2.1.1). All CDR-H3 point mutants show a minimum at 218 nm, except the MAK33 double mutants R44AA103E (purple circles) where the minimum is shifted into the FUV range to ~215 nm (Figure 51 A). All CDR-H3 point mutants share a positive signal at 200 nm which is most intense for MAK_1DHU A103E (blue circles). The signal intensities at 200 nm of MAK_1DHU G106R (light blue circles), MAK_MAK A103E (red circles), MAK_MAK G106R (yellow circles), and MAK_MAK A103DG106R (orange circles) are in between the MAK33 V_H and MAK_DHU A103E signal (Figure 51 A). The two double mutants of conserved residues and CDR-H3 have similar signal intensities at 200 nm as the MAK33 V_H domain. In general, all point mutants are well-defined β-sheet structured proteins.

An exchange of CDR-H3 of MAK33 (red circles, Figure 51 B) had no influence on the FUV-CD spectrum of MAK33 V_H. However, changes can be observed when the long CDR-H3 of V_H3 is introduced. The FUV-CD spectra between MAK_MAK CDR3 V_H3 (blue circles) and MAK_1DHU CDR V_H3 (green circles) are comparable with a clear minimum at 218 nm and a pronounced positive signal at 200 nm. Summing up, none of the CDR-H3 manipulations had any effects onto β-sheet formation of V_H (Figure 51 A and B). However, there might be an influence on the random coil characteristics and consequently on CDR flexibility.

NUV-CD spectra were recorded to determine a fingerprint of the three-dimensional structure of the single mutants. In Figure 51 C the results for the point mutants in the CDR-H3 are shown. The smallest effects on the environment of aromatic amino acids could be observed for MAK33 A103E (red circles), MAK33 R44AA103E (purple circles), and MAK33 R44AG106R (green circles) indicating a tertiary structure quite similar to MAK V_H wild type (black, open circles). MAK_1DHU point mutants (A103E in blue circles; G106R in light blue circles) show two stronger pronounced maxima at ~293 nm and ~285 nm which were characteristic for MAK_1DHU (Figure 51 C and Figure 49 D). The NUV-CD spectrum of MAK G106R (yellow circles) is interesting as it shows a minimum at 300 nm and a clear maximum at ~282 nm (Figure 51 C). However, for all point mutants "fingerprints" could be obtained indicating the presence of defined tertiary structure.

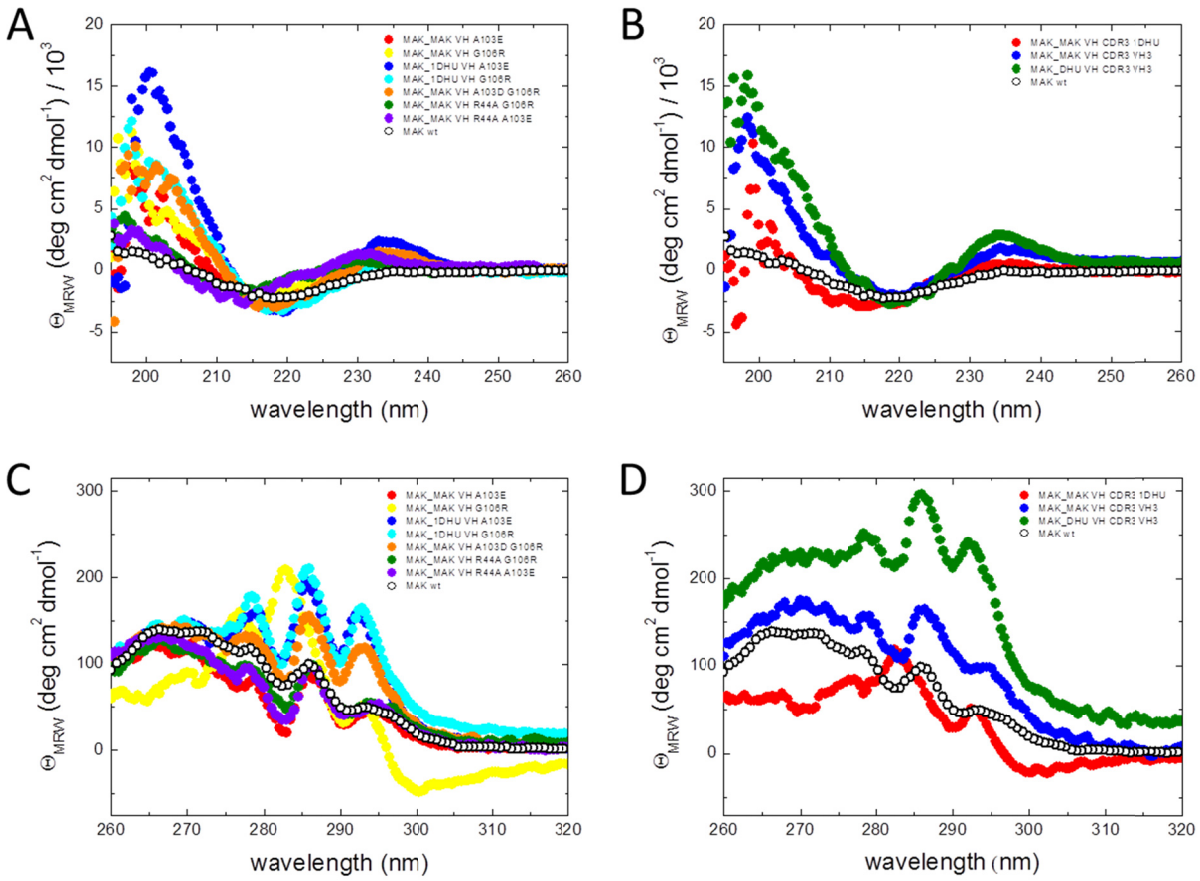

Figure 51: Mutations of CDR-H3 have only marginal effects on V_H secondary and tertiary structure.
The FUV-CD spectra of V_L (A) and V_H (B) CDR grafting mutants are shown. For V_L (A) MAK_MAK A103E (red circles), MAK_MAK G106R (yellow circles), MAK_1DHU A103E (blue circles), MAK_1DHU G106R (pale cyan circles), MAK_MAK A103D G106R (orange circles), MAK_MAK R44A G106R (green circles) and MAK_MAK R44A A103E (purple circles) were analysed. Wild type V_L is shown in black (open circles). The colour code for V_H CDR-H3 exchange mutants (B) is: MAK_MAK CDR3 from 1DHU in red circles, MAK_MAK CDR3 from V_H3 in blue circles, MAK_1DHU with CDR3 from V_H3 in green circles and MAK wild type in black (open circles). In (C) and (D) the NUV-CD spectra of the V_L (C) and V_H (D) CDR-H3 mutants are shown. Colour code for (C) and (D) is analogue to (A) and (B), respectively. For the spectra each 16 accumulations were recorded and buffer-corrected (PBS). All measurements were performed at a protein concentration of 20 µM (FUV-CD) and 50 µM (NUV-CD) in 0.5-mm (FUV) or 5-mm (NUV) quartz cuvette at 20°C.

NUV-CD spectra of V_H domains with exchanged CDR-H3 also show a well-defined tertiary structure for all constructs (Figure 51 D). MAK_MAK CDR3 V_H3 (blue circles) is most similar to MAK33 V_H (black; open circles) although the sequences between the CDR-H3s differ in the amount of aromatic amino acids (see 3.2.3). The introduction of CDR-H3 of 1DHU into MAK33 has more effect on the NUV-CD behaviour (red circles). MAK_1DHU CDR-H3 V_H3 (green circles) shows a strong signal increase due to the CDR-H3 modification (Figure 51 D and Figure 49 D).

3.2.3.2 Protein stability can be influenced by CDR-H3 modifications

Chemical stability was tested by GdmCl-induced unfolding transitions. Although not all transitions were reversible, data were fitted to a two-state model to be able to compare the stabilities more easily (data not shown). In Figure 52 A the transitions for the CDR-H3 point mutants are shown. MAK33 V_H is a very unstable domain with a $D_{1/2}$≈ 0.2 ± 0.1 M and a highly cooperative transition behaviour (Figure 52 A, black circles/line and Table 13). The introduction of charges residues within CDR-H3 had only a marginal effect on the stability of MAK33 V_H (MAK_MAK A103E in red circles/line and MAK_MAK G106E in light green circles/line, Figure 52 A). However, the cooperativity is decreased compared to MAK33 V_H (Table 13). MAK_MAK G106E cannot be described properly with the two-state transition model and the fit results in a negative $D_{1/2}$. This result is obviously not a realistic value. The double mutation of charged residues in CDR-H3 resulted in a yet higher stability with a $D_{1/2}$≈ 0.75 ± 0.39 M and an even lower cooperativity (cyan circles/line; Figure 52 A and Table 13).

A graft of the three CDRs of 1DHU on the MAK33 V_H framework increased the stability, but decreased the cooperativity (m≈ 4.5 ± 0.4 kJ mol^{-1} M^{-1}; see Table 11 and Figure 50 B). In this case, the introduction of a single charged residue in CDR-H3 leads to a further stabilisation and an increase in the cooperative behaviour (MAK_1DHU A103E in blue circles/line and MAK_1DHU G106R in pale cyan circles/line; Figure 52 A and Table 13). Hence, the influence of charged residues on stability and cooperative unfolding behaviour can differ depending on the overall CDR composition.

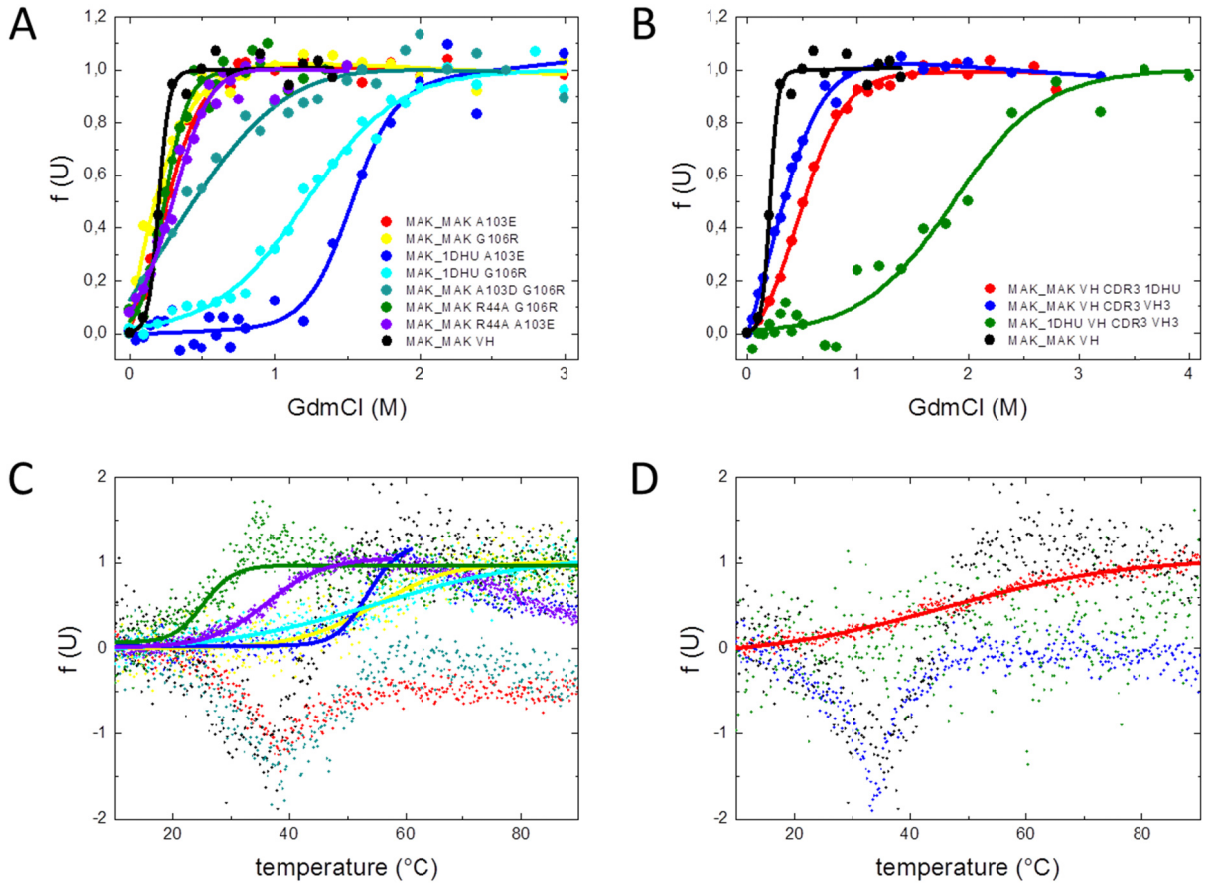

Figure 52: Changes in CDR-H3 influence protein stability

To assess the stability of the V_H CDR-H3 point mutants as well as CDR-H3 exchange mutants, GdmCl-induced (top) and temperature-induced (bottom) unfolding experiments were performed. Data for denaturant-induced (A and B) and thermal-induced transitions (C and D) are shown. In (A) GdmCl-induced unfolding transitions for the V_H CDR-H3 point mutants are illustrated. MAK_MAK A103E (red circles), MAK_MAK G106R (yellow circles), MAK_1DHU A103E (blue circles), MAK_1DHU G106R (pale cyan circles), MAK_MAK A103D G106R (orange circles), MAK_MAK R44A G106R (green circles) and MAK_MAK R44A A103E (purple circles) were analysed. In (B) the denaturant-induced samples for the V_H CDR-H3 exchange mutants are shown. Samples are labelled the following: MAK_MAK CDR3 from 1DHU in red circles, MAK_MAK CDR1 from V_H3 in blue circles, MAK_1DHU with CDR3 from V_H3 in green circles and MAK wild type in black (open circles). In (C) and (D) the thermal-induced unfolding transitions for V_H CDR-H3 point mutants (C) and V_H CDR-H3 exchange mutants (D) are depicted. Colour code for (C) is the same as for (A) and for (D) the same as for (B). Temperature-induced unfolding was monitored by FUV-CD spectroscopy at 215 nm with a heating rate of 20°C h^{-1}. Data were fitted to a Boltzmann function to obtain transition midpoints. Chemical denaturation of the proteins was induced by GdmCl, and structural changes were monitored by fluorescence spectroscopy at a fixed wavelength (mainly 355 nm emission; 280 nm excitation). Data were evaluated according to a two-state unfolding model to obtain midpoints and cooperativity parameters of the transitions. Measurements were performed at 20°C at a protein concentration of 20 µM (temperature-induced transitions) or 1 µM (GdmCl-induced transitions).

The combination of single charged residues in CDR-H3 and the R44A point mutation (see 3.2.1.2) was tested to analyse if synergistic effects might be observed. However, no further stabilisation could be observed when the double mutants are compared to the single charged point mutants. This hold true either for MAK_MAK R44AA103E (purple circles/line) or for MAK_MAK R44AG106R (green circles/line) as shown in Figure 52 A and Table 13.

In Figure 52 B the effects of CDR-H3 exchanges on protein stability are shown. When CDR-H3 of MAK33 V_H is exchanged against CDR-H3 of 1DHU (red circles/line) or against the long CDR-H3 of V_H3 (blue circles/line), the chemical stability is marginally increased and unfolding is less cooperative compared to MAK33 V_H wild type (Figure 52 B and Table 13). Yet, the chemical stability of MAK33 with CDR-H1 and CDR-H2 of 1DHU in combination with CDR-H3 of V_H3 V_H is increased. In contrast, the cooperativity is only little affected (see Figure 52 B, Table 13 and Table 11, Figure 50 B). In general, the CDR-H3 has an influence on stability and unfolding behaviour of V_H domains.

Table 13: Thermal and chemical stability of V_H CDR-H3 mutants

Domain	Mutation	T_{melt} [°C]	$D_{1/2}$ [M]	m [kJ mol^{-1} M^{-1}]
	wild type	~30	0.21 ± 0.10	68.6 ± 23.3
	MAK_MAK A103E	~30	0.23 ± 0.07	17.2 ± 2.4
	MAK_MAK G106R	54.7 ± 0.8	-0.07 ± 0.21	12.1 ± 3.2
	MAK_MAK A103DG106R	~25	0.75 ± 0.39	6.3 ± 3.2
	MAK_MAK R44AA103E	36.1 ± 0.1	0.30 ± 0.14	22.4 ± 6.1
V_H	MAK_MAK R44AG106R	25.2 ± 0.4	0.25 ± 0.08	30.4 ± 6.8
	MAK_1DHU A103E	53.4 ± 0.4	1.52 ± 0.85	16.1 ± 6.4
	MAK_1DHU G106R	55.3 ± 1.5	1.21 ± 0.09	7.7 ± 0.4
	MAK_MAK_CDR3 1DHU	46.2 ± 0.8	0.46 ± 0.09	11.7 ± 1.4
	MAK_MAK_CDR3 V_H3	~25	0.22 ± 0.08	10.8 ± 1.3
	MAK_1DHU_CDR3 V_H3	ND	1.87 ± 0.35	5.9 ± 0.8

Stabilities against the thermal and chemical denaturation (GdmCl) of the V_L and V_H mutants in PBS are shown. Midpoints of thermal transitions are shown as T_{melt}. Even though most of the GdmCl-induced unfolding transitions were not reversible, data were evaluated according to a two-state equilibrium unfolding model to derive the midpoint of transitions ($D_{1/2}$), as well as the cooperativity parameter (m), for a qualitative comparison of the data. ND indicates that the data quality was not sufficient for T_{melt} determination.

3.2.3.3 Interaction between V_H and V_L

CDR-H3 is the most flexible region regarding length and sequence [235]. To test the influence of charged residues and combinations between conserved amino acids and CDR-H3 residues single and double mutants were analysed for their interaction abilities. K_Ds were determined using AUC sedimentation equilibrium approach and SPR. For AUC experiments all domains were tested against MAK33 V_L or V_H, for SPR measurements MAK33 V_L with an N-terminal cysteine immobilised was tested against the different mutants.

The results from the AUC show a clear increase of the K_D for MAK_MAK V_H R44AA103A and MAK_MAK V_H R44AG106R compared to the wild type (Table 14). However, MAK_MAK V_H A103E and MAK_MAK V_H A103DG106R show a binding behaviour similar to the wild type. Consequently, the unfavourable influence of R44A is not compensated. MAK33 with CDRs of 1DHU binds V_L wild type slightly less efficient compared to MAK_MAK V_H A103E but very

similar compared to MAK_DHU (see Table 14 and Table 12). This finding enhances the assumption that A103E has no crucial effect on the binding of V_H and V_L.

The SPR results do not always corresponded to the AUC experiments. MAK_MAK V_H A103E behaves like a poor interaction partner in contrast to the AUC experiments. It is possible that the different buffer conditions (AUC: PBS and SPR: HBS-P+) might lead to a different behaviour.

Table 14: K_D determination by AUC SE experiments and SPR.

Domain	Mutation	AUC K_D [µM]	SPR K_D [µM]
	wild type	0.2	1.3
	MAK_MAK V_H A103E	0.4	4.3
	MAK_MAK V_H G106R	ND	1.3
	MAK_MAK V_H R44AA103E	3.5	4.1
	MAK_MAK V_H R44AG106R	7.1	3.4
V_H	MAK_MAK V_H A103DG106R	ND	2.3
	MAK_1DHU V_H A103E	0.9	2.9
	MAK_1DHU V_H G106R	ND	2.0
	MAK_MAK V_H CDR3 1DHU	ND	1.6
	MAK_MAK V_H CDR3 V_H3	ND	NB
	MAK_1DHU V_H CDR3 V_H3	ND	NB

K_D values were determined by AUC sedimentation equilibrium experiments. Three concentrations in the range between 0.5 µM and 7.5 µM with V_H and V_L in a 1:1 ratio were measured at three velocities and globally fitted to a model for heterogeneous interaction between two components. The errors ranged from 9%-20% Experiments were performed at 20°C. SPR experiments were performed with a Biacore X100. Ligand was immobilised via an introduced cysteine at the N-terminus. NB indicates that no binding could be observed. ND indicates that the sample was excluded due to homo-dimerization.

3.2.3.4 CDR-H3 is important for functionality

To determine if single point mutations within CDR-H3 influence the recognition and binding of the antigen, an ELISA with isolated domains was performed.

A single point mutation in MAK33 CDR-H3 at residue A103 to glutamic acid has no effect on antigen binding (see Figure 53 and Table 15). The same could be observed in combination with the mutation R44A, which does not influence antigen binding (Table 15 and Table 10).

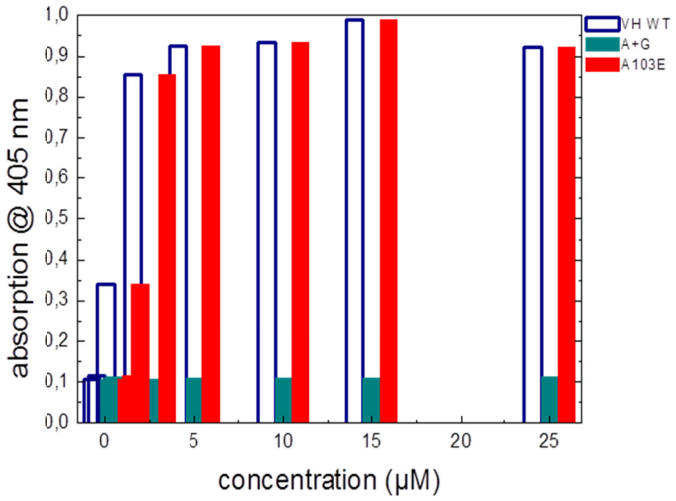

Figure 53: ELISA with isolated variable domains to determine the influence of CDR-H3 on antigen binding.

To determine the functionality of the single mutants an ELISA with isolated domains was performed. Biotinylated human creatine kinase was coupled to a streptavidine-coated microwell plate. Wild type V_L was labelled with a FLAG-tag at the C-terminus and could be detected with an Anti-FLAG-tag antibody coupled to a horseradish peroxidase. A representative of an inactive and a mutant with similar activity as the wild type are shown. The absorption at 405 nm at the signal maximum (after 15 minutes) was corrected for the signal of the variable labelled domains alone and compared. Experiments were performed at 20°C.

The exchange of G106 to an arginine is tolerated as the binding was comparable to V_H wild type. However, this mutation leads to a loss of antigen binding when combined with R44A or A103D (Table 15). That is an interesting observation, as all mutations alone do not abolish the interaction with the antigen (see Table 15, Figure 46 and Table 10). The CDR-H3 exchanges of MAK_MAK and MAK_1DHU resulted in a complete loss of the binding to human creatine kinase what leads to the assumption that CDR-H3 is a very important region for antigen binding and changes in CDR-H3 can be enough to prevent antigen recognition (Table 15).

The results confirm the observation which was made for the point mutants of conserved amino acid residues. Antigen recognition and affinity between the variable domains are not necessarily correlated.

Table 15: CDR-H3 is important for antigen binding.

	V_H mutant	relative signal		V_H mutant	relative signal
V_L-FLAG	wild type	wt	V_L-FLAG	wild type	wt
	MAK_MAK V_H A103E	wt		MAK_MAK V_H CDR3 1DHU	---
	MAK_MAK V_H G106R	wt		MAK_MAK V_H CDR3 V_H3	---
	MAK_MAK V_H R44AA103E	wt		MAK_1DHU V_H CDR3 V_H3	ND
	MAK_MAK V_H R44AG106R	---			
	MAK_MAK V_H A103DG106R	---			
	MAK_1DHU V_H A103E	---			
	MAK_1DHU V_H G106R	---			

To determine the functionality of the mutants a qualitative analysis of ELISA with isolated variable domains was performed. Detection of V_H and V_L wild type domains was possible via an introduced FLAG-tag at the C-terminus. As a reference the signal of the wild type domains vs. FLAG-tagged domains after 15 minutes of incubation were chosen. Samples were corrected with the signal from the single FLAG-tagged variable domains. One minus up to three was chosen to group the different active mutants (from slightly lower [-] binding to no binding [---]). ND indicates that the sample was not included in the study. Experiments were performed at 20°C.

3.2.4 Discussion

Antibodies and antibody derivatives are important for therapeutically and diagnostically applications. Their high target specificity qualifies them for the treatment of cancer, autoimmune and inflammatory diseases. However, engineering of stability, affinity or humanisation are often adverse to other properties. During CDR grafting, for example, frequently a conflict occurs between stability and antigen specificity [169]. Besides an exchange of CDRs or just the specificity determining regions (SDRs), single point mutations in the hypervariable regions, at the interfaces between the variable and constant domains or in the framework can also influence the stability in a positive as well as a negative manner [197, 240-243]. Changes in the residues H6, H7 and H9 of the framework region have experimentally been shown to influence affinity, stability and folding yield [244]. Covariation analysis has demonstrated that not only amino acid clusters but also particular residues are conserved within the variable domains [160, 164-166]. This is predominantly the case at/adjacent to the interface between V_H and V_L [164]. Different reasons for the presence of conserved residues are possible. I) these residues can be crucial for the immunoglobulin fold 2) they might be important for stability or 3) they are necessary for the interaction between both domains and 4) influence the functionality. However, covariation analysis cannot prove the importance of the single residues nor depict the functions. Hence, we performed an

alanine exchange mutational study of the conserved residues with the highest scores in the covariation study. The influence of conserved residues on the structure of the isolated domains was analysed by CD spectroscopic methods. Interestingly, the exchange of any of the conserved residues in the V_L domain to alanine had no considerable influence on the secondary or tertiary structure (Figure 42 A and C). Results of *in silico* analysis suppose that only very few residues (< 10) are sufficient to adopt an immunoglobulin fold [245]. In the case of V_L, none of the interface conserved residues seemed to be crucial for folding as the single alanine mutations had no clear effects on the structure. On the other hand, V_H is very sensitive to exchanges of the conserved residues in the interface as clear differences in the FUV-CD and NUV-CD spectra were observed (Figure 42 B and D). The exchange of the conserved tryptophans had in particular a negative influence on the secondary and tertiary structure formation. R38A was also not well tolerated. One residue with a positive influence on the structure of V_H could also be determined: the exchange of R44 to alanine. The biophysical properties of isolated V_H domains are in general compromised compared to V_L [199]. As none of the mutations lead to the formation of a random coil structure, one can assume that the tryptophans and R38 are important for a proper Ig fold but not essential for folding. For a detailed correlation between these residues and the Ig fold more sophisticated methods such as NMR or X-ray crystallography are necessary to determine the influence on the atomic level.

A critical parameter for proteins used for biomedical or biotechnological applications is their stability. Simulations of scFvs have shown that the stability of the interface between the two variable domains plays a critical role for the overall stability of an antibody (or fragment) as dissociation precedes unfolding [246]. Hence, the influence of the conserved residues on domain stability should be addressed. In terms of stability, mutations of conserved residues for V_L were predominantly rather non-favourable except of S43A which behaved similar to the wild type (Figure 43 A and C, Table 8). In contrast, for V_H, which is in isolation a very unstable domain, not only further destabilising but also a few slightly stabilising mutations (T_{melt}s of R38A, R44A and E46A; Figure 43 B and D, Table 8) could be identified. In the case of R38A this finding is surprising due to the variation within the changes in the FUV-CD and NUV-CD (Figure 42 B and D). The mutations of the conserved tryptophans had a negative effect on stability. W47 was already previously mutated for improved solubility and stability but none of the mutations were favourable (W47L, W47R [247]). In summary, the effects on the stability were in most cases for V_L as well as for V_H disadvantageous and lead to the assumption that the conserved residues indeed influence protein stability. However, a clearly visible influence of the mutations on isolated domains does not necessarily mean that it will lead to the same results in the context of Fab fragments or even complete antibodies. Here, more favourable interactions are present which might be able to compensate these effects. In the study of Ewert and co-workers was shown that the stability of the human consensus sequences was higher in the context of scFvs compared to the isolated domains [199].

Of special interest was the influence of the alanine exchange on the interaction between the variable domains as it is expected to have consequences for antigen recognition, too. However, the results could not prove a direct correlation between interaction of the variable domains and antigen binding. In the case of V_L only the mutation P44A resulted in a decreased V_H binding affinity. The result fits to the study of Chailyan and co-worker [248]. They analysed the mode of interaction between V_H and V_L in scFvs and could identify two mechanisms which are correlated to the presence of particular residues. In one of the two clusters, P44 plays a crucial role for the interaction due to its influence on a 7:7-type hairpin between L38 and L44 [248]. On the other hand, this mutation had only a slightly non-favourable effect on antigen binding indicating that the domain had not lost all functionality (Table 9 and Table 10). *Vice versa*, L46A and F98A exhibited K_D values similar to the wild type domain but were clearly disturbed in antigen binding. The same holds true for V_H. No clear correlation between interaction and antigen binding could be observed. SPR and AUC data fit in most cases well, although the absolute K_D values differ (Table 9). For example, E46A binds less efficient to V_L according to AUC but with a similar efficiency according to SPR. The latter would be expected to be more reliable as antigen binding is comparable to V_H wild type. Yet, the correlation is not supported by the other mutant and hence it is not possible to determine which method is more precise. It is to mention that E46A tends to aggregate during longer incubation times (in some samples degradation was observed as well) and it cannot be excluded that some protein is lost during the long measuring time in the AUC sedimentation equilibrium experiments. In addition, the buffer composition between AUC and SPR differs (phosphate- and HEPES-based buffer, respectively) and this may influence interaction as well. These results lead to the assumption that a strong interaction between V_H and V_L might disturb antigen recognition as flexibility between the two domains might be necessary to bind the antigen. Further analysis will be necessary to prove this idea. However, these results challenge studies on the influence of mutations on the interaction between V_H and V_L where the conclusions that were drawn based on an indirect approach using Fabs or scFvs and antigen recognition as a read-out [168-172]. Conserved residues in V_H are very important for the interaction with V_L but not automatically also for antigen recognition. In V_L, most mutations were tolerated concerning domain interaction. In contrast, strong effects were observed on antigen binding.

It is difficult to make reliable assumptions for the reasons why some residues disturb the interaction quite strongly from the known X-ray structure of the variable domains. There is only one resolved X-ray structure of a MAK33 Fab available (PDB ID 1FH5) which has not the optimal resolution. In addition, drawbacks are the unresolved CDR-H3 and some disagreement on the sequence level (see Figure 54 E). However, for a few selected mutants of V_H possible effects can be discussed. To prove these assumptions a high-resolved NMR or X-ray structure of each mutant will be crucial. Attempts to crystallise V_H variants in the presence of V_L wild type (with and without human creatine kinase) failed so far. As the CDRs are quite flexible, NMR might be the more promising approach as crystallisation can be

inhibited by too dynamic structures. It was for example possible to solve the structure of molten globule state of α-lactalbumin by NMR which is a very dynamic state [233, 234].

R38 and E46 are close enough to interact with each other either due to the formation of a salt bridge or hydrogen bond formation (this depends on the charge in this environment) according to the crystal structure of MAK33 (Figure 54 A). If this interaction is real and important for a stable β-sheet formation (R38 is in strand C and E46 in strand F) one would expect the same consequences when either of these two residues is changed. An alanine exchange at positions R38 and E46 resulted in a clearly higher K_D for R38A (17.3 µM) compared to the wild type (0.2 µM) and a less pronounced increase for E46A (2.0 µM). Moreover, the binding affinities were the same for E46A and wild type in the SPR experiments (see Table 9). In addition, the functionalities differed as E46A was able to bind the antigen with a similar efficiency as the wild type. In conclusion, the theory of a salt bridge or hydrogen bonds between the two residues cannot be confirmed. In case of the two hydrophobic residues V37A and L45A, an obvious decrease of the interaction between V_H and V_L as well as a decrease in the functionality could be determined (Table 9 and Table 10). No peculiar hydrophobic clusters were observed around V37 and L45 (Figure 54 B). It is discussed that the balance between hydrophobic and polar residues is crucial for a stable interface formation [246]. Yet, an exchange to alanine should not shift the ratio between polar and hydrophobic residues. The only difference is the size of the residues; they might play a role for the orientation of V_L to V_H. For the conserved tryptophans W103 (Figure 54 C) and W47 (Figure 54 D) also no specific interactions are noticeable. Tryptophan is the bulkiest of the 20 naturally occurring amino acids and for W103 the size might be important for the conformation of CDR-H3 which is adjacent. Rotations of this residue are found in V_{HH} domains and thus play a role for association [249, 250]. V_{HH} domains cannot form dimers with V_L [251]. For W47 the environment is more complex. It is near enough to influence a proline residue which probably influences the conformation of CDR-H2 as it can introduce a kink. Jespers and co-workers discuss that the orientation of W47 is important [252]. They observed that in conventional V_H-V_L antibody pairs W47 lies flat on the V_H surface such that its hydrophobic phenyl ring engages with the V_L domain. However, this hypothesis needs to be further tested.

CDR grafting is frequently performed to decrease the immunogenicity or to improve the stability of the scFvs or Fabs but this step often impairs other favourable characteristics. Ewert and co-workers characterised the propensities of the identified human consensus sequences (from HuCAL [238] and identified V_H3 (PDB ID 1DHU) and V_K3 (PDB ID 1DH5) as the most stable domains [199]. Indeed, it was possible to improve the biophysical stabilities of MAK33 V_H and V_L by CDR grafting with the stable consensus sequences as well as their folding behaviour. On the other hand, the stability of the consensus sequences was decreased when MAK33 CDRs were grafted (see Figure 50 and Table 11). The border sequences between framework and CDR regions (according to *http://www.bioinf.org.uk/abs/)* are highly similar between MAK33 and the consensus sequences (Figure 47 B); therefore no fine tuning was performed. However, one has to keep

in mind that some of the framework residues can be necessary for antigen binding [253, 254].

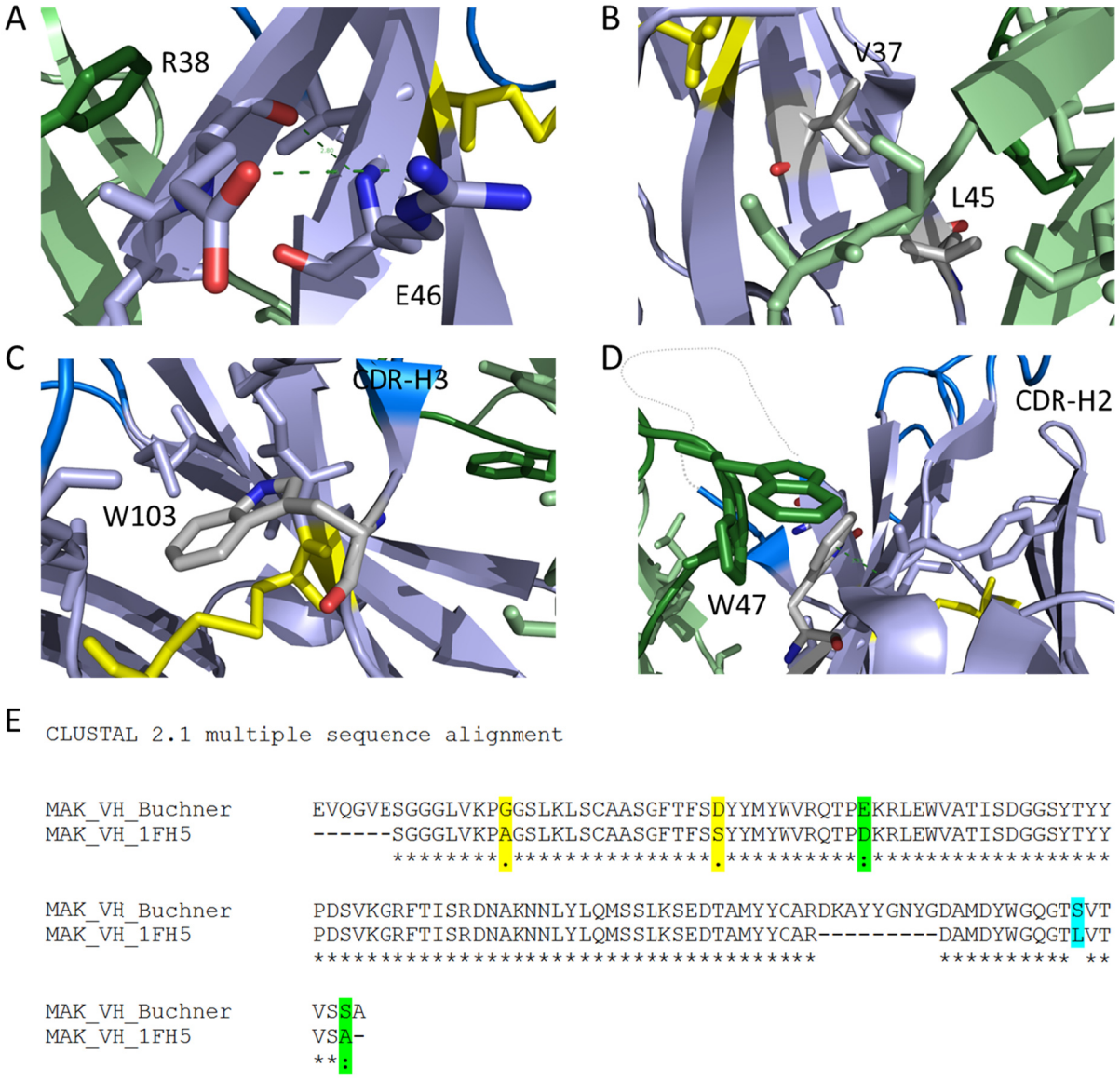

Figure 54: The environment of conserved residues of V_H and sequence alignment of V_H.
Based on the crystal structure of MAK33 (PDB ID 1FH5) the environment of important conserved residues of V_H was analysed. In A) R38 of β-strand C and E46 of strand F are shown. They are close enough to interact with each other. In B) the two hydrophobic residues V37 and L45 are highlighted. No important interaction partner could be identified but mutations to alanine had severe effects on stability and interaction. W103 C) might influence the orientation of CDR-H3 because the bulky residue is adjacent to the loop. D) W47 would be close enough to influence the proline residue nearby CDR-H2 and might influence its conformation. V_H is depicted in light blue and the CDRs in blue. In D) CDR-H3 is indicated as a grey dashed line. For V_L the domain is coloured in light green and the CDRs in green. Disulphide bonds are indicated as yellow sticks and distances < 3 Å as a dark green dashed line. The residues of interest are shown as grey sticks with nitrogen atoms labelled in red and oxygen atoms in blue. In E) a sequence alignment of MAK33 V_H from the crystal structure (PDB ID 1FH5) and MAK33 V_H used in the Buchner`s lab is shown. Alignment was performed with CLUSTALW (version 1.83). Residues are highlighted according to CLUSTAW output.

It is not surprising that the interaction between the MAK33 wild type domains and their consensus sequence partners was not as efficient as for the MAK33 V_H and V_L domain. The consensus sequences result from a large alignment and do not occur naturally [238]. However, two grafting constructs were almost as efficient in binding as the MAK33 wild type domains: 1DHU_MAK33 V_H and MAK33_1DH5 V_L. CDR-L1, L2 and to some part L3 are highly similar between V_L and 1DH5 (Figure 47 B). This indicates that for V_L the framework and CDR orientation in the MAK33_1DH5 graft is not much different from MAK33 V_L. Yet, both grafting mutants were different in the FUV-CD and NUV-CD. In conclusion, the structure must somehow differ.

The same holds true for the V_H mutants which behaved more similar in the FUV-CD but were clearly distinct on the tertiary structure level. Antigen binding was not detected for the CDR grafting mutants between MAK33 and the consensus sequences, 1DH5 or 1DHU. This is not surprising as consensus CDRs are not naturally occurring sequences directed against a specific antigen. Additionally, no fine-tuning of the MAK33 CDR grafted mutants was performed. Interestingly, 1DH5_MAK33 V_L showed some antigen binding although to a lower extent than MAK33 V_L (3.2.2.4). This supports the conclusion from the mutational analysis of the conserved interface residues: there is no correlation between domain interaction and antigen binding. A stable framework had a positive influence on the marginally stable V_H domain and also for the quite stable V_L domain clearly stabilising effects were observed. The CDRs have an influence on stability that should not be underestimated. Hence, it is a good approach to perform SDR instead of CDR grafting [254]. It was not foreseen that any of the grafting mutants will be able to bind the human creatine kinase as no fine-tuning was performed. However, 1DH5_MAK, although it is a poor interaction partner for MAK33 V_H, can improve antigen binding compared to V_H alone. From this experiment it cannot be predicted if the framework or CDRs are more important for the interaction between the variable domains as on the one hand the framework of the actual domain (MAK33_1DH5 V_L) and on the other hand the consensus framework (1DHU_MAK33 V_H) had been the better interaction partner.

In a third step, the role of CDR-H3 was further analysed. It is known that this region is the most flexible CDR concerning sequence and length [235]. The sequence length can vary between three and 25 amino acids. Additionally, CDR-H3 can influence V_H stability as long CDR-H3s are assumed to be able to cover the hydrophobic surface which would interact with V_L if present [199]. Charged residues within CDR-H3 are also considered to be favourable in terms of stability [199] but they might have severe effects on the interaction between V_H and V_L or antigen binding. Indeed, all domains showed better biophysical stabilities (especially in terms of chemical denaturation, see Figure 52 and Table 13). Interestingly, several mutants gained the ability to form homo-dimers at low concentrations (≤ 10 µM). This is disadvantageous for the AUC analysis as homo- and heterodimers have a comparable size and cannot be distinguished. MAK33_MAK33 A103E and G106R, as well as MAK_MAK with CDR-H3 of 1DHU or V_H3 and MAK_DHU with CDR3 of V_H3 were composed of mixtures of monomeric and dimeric protein. On the contrary, all single point mutants of the

conserved residues in the interface were still monomeric in low protein concentrations like MAK33 V_H and V_L wild type (data not shown). Introduction of a charged residue at position A103 was favourable in terms of stability (Table 13 and Figure 52 A and C) as well as for affinity to V_L in the case of MAK_MAK and MAK_DHU V_H (Table 14). Antigen binding was also not diminished for MAK_MAK by this mutation, whereas MAK_DHU cannot bind (Table 15). This was already observed for MAK_DHU without the mutation (3.2.2.4). Introduction of a charged residue at position G106 also had a positive influence on stability and did not diminish the binding to V_L although NUV-CD indicates some structural rearrangements (Figure 51 B). Two types of double mutants were tested. On the one hand the combination between charged residues and on the other hand R44A with charged residues. The double mutants of MAK_MAK showed all a less efficient binding to V_L compared to MAK33 wild type (Table 14) and a diminished antigen recognition except for R44AA103E (Table 15). A103E cannot compensate the negative binding abilities of R44A. Interestingly, the combination of A103E and G106R is also non-favourable for the interaction between the domains and the antigen although the single mutations did not disturb binding. An exchange of the complete CDR-H3 improved the stability against denaturants (Table 13 and Figure 52 B) but abolished the interaction with V_L for MAK_MAK and MAK_1DHU with CDR-H3 of V_H3. The CDR of V_H3 (17 amino acids) is even 3 amino acids longer than CDR-H3 of MAK33 and might partially shield the interface. For none of the CDR-H3 exchange mutants a sufficient antigen binding was observed (Table 15). Due to the fact that none of the MAK33 grafted mutants was able to recognise human creatine kinase anymore (3.2.2.4), it is clear that mutations of this CDR are sufficient to abolish binding. In the crystal structure of the MAK33 Fab domain (PDB ID 1FH5) the CDR-H3 is not resolved. That indicates a high flexibility of this region. However, the insufficient resolution does not allow making any predictions of the precise structural influences of the particular singe point mutants within the CDR-H3.

In conclusion, stability of the variable domains depends on the presence of conserved amino acids within the interface of the variable domains. This is especially pronounced for V_H. On the other hand, these residues are not crucial for folding but they can influence the structure of V_H. Several of these residues were crucial for the interaction between V_H and V_L. For V_L, residue P44 had a strong influence. Its stability was lower than the wild type and the recognition of the antigen decreased. On the V_H side V37, R38, L45, and W104 were the worst interaction partner. V37A, L45A and W103A had a decreased stability and could recognise the antigen less efficient. In contrast, R38A was more stable than the wild type and disturbed antigen binding. In addition, R44A had a stabilising effect but interacted less efficient than the wild type with V_L whereas the antigen recognition was not disturbed. AUC and SPR data are in most cases concordant although the K_Ds differ in their absolute values. Exceptions were in particular E46A, MAK_MAK V_H A103E and 1DHU_1DHU. In summary, stability, interaction and functionality showed no clear correlation. However, a tendency can be observed that less stable mutants interact less efficiently although there are exceptions. The limited correlation between interaction of the variable domains and antigen binding should be emphasised because many earlier studies are based on the assumption that they

are correlated. In terms of CDR grafting CDR-H3 should be highlighted. Introduction of single charged residues can have a positive influence on all properties addressed in this study but combinations can be disadvantageous. In total small changes can have already severe effects concerning, structure, stability, specificity and interaction. Yet, one must prove first that the observations for MAK33 are valid in general and whether these effects are compensated in antibodies, scFvs or Fabs by other interactions.

3.3 The folding behaviour of V_H domains

Antibodies are ideal model systems for studying protein folding, dynamics and association due to the presence of homo- and heterodimeric interactions, conserved proline residues, disulphide bonds and population of intermediate states. These features increase the complexity of folding of antibody domains compared to the classical two-state folding mechanism observed for many small proteins [255]. All antibody domains share a common structural motif, the immunoglobulin fold. This fold consists of two β-sheets which form a β-sandwich. Variable and constant domains belong to different subclasses and differ in their β-strand composition. Variable domains depict a 4+5 orientation (ABED/CC`C``FG) whereas constant domains are characterised by 4+3 strands (ABED/CFG) (see Figure 6 and [157]). Another important difference between variable and constant domains are the complementarity determining regions in V_H and V_L which form the antigen binding site. This regions are flexible loops which differ in length and sequence. Isolated V_L and V_H domains differ in their biophysical properties whith the latter being less stable and more prone to aggregate [194, 195].

The folding behaviour of C_H3 [180], C_H2 [202] and C_L [191], as well as V_L [119] of MAK33 could already be studied in great detail. Differences between the folding of V_L and V_H is of special interest as some V_L domains are associated with misfolding diseases [196] whereas the latter are less often pathogenic. The comparison of the two pathways will help to determine differences between V_L and V_H domains and to correlate sequence and structural features with biophysical properties. Besides insight into basic aspects of protein structure, these results may also be useful for diagnostic and therapeutic applications of antibodies [197] and allow insights into the mechanisms leading to amyloid formation when compared to the folding pathway of amyloidogenic V_L domains. Another aspect to focus on is to extend the experiments from the murine MAK33 V_H domain to human V_H domains. Murine and human V_H domains differ in their CDRs but also in their framework sequences [169] and can be clustered according to sequence homology analysis in different germline families [198]. Not only sequences but also the properties between germline families can differ [199]. Therefore it will be interesting to see if and how the folding behaviour will vary.

Five V_H domains were selected and are shown in Figure 55. V_H of the murine monoclonal antibody MAK33 and a stabilising mutant R44A which was developed for the interaction study between V_H and V_L were selected. 1OPG is also a murine antibody which is highly

similar to MAK33 on the structural level. It recognises the gpIIb/IIIa receptor [256]. 1FVC is a humanised IgG1 κ antibody directed against p185HER2 [212]. The last candidate is a human IgM antibody (1HEZ) directed against *Peptostreptococcus magnus* protein L [257]. They all have three tryptophan residues in common; two are exposed and one is near to the conserved internal disulphide bridge which can be used to probe folding by tryptophan fluorescence. 1FVC V_H has an additional exposed tryptophan.

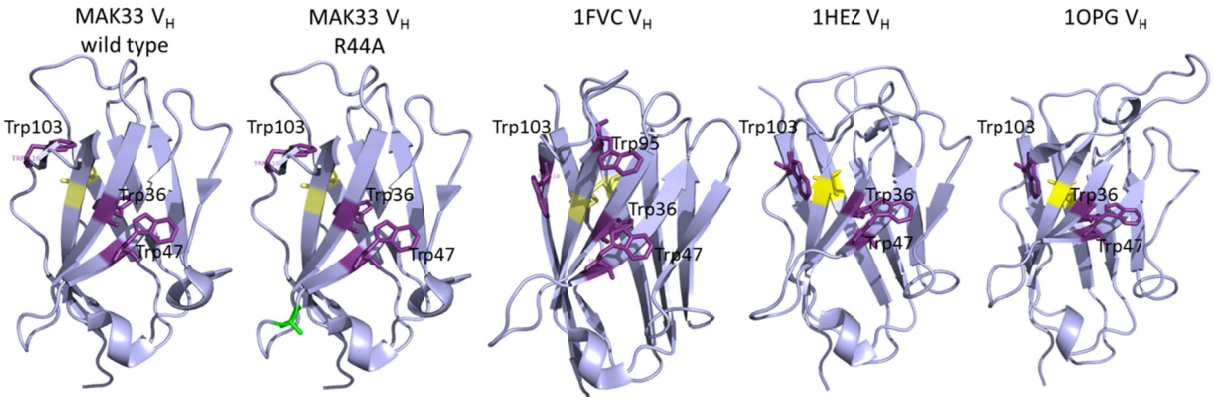

Figure 55: An overview of the different variable domains analysed for the folding study of V_H.
Different V_H domains were analysed for their abilities to unfold and refold. MAK33 V_H and its stabilising mutant R44A (light green stick) are murine monoclonal antibodies with three Trp residues (purple sticks). Trp36 is located near a conserved internal disulphide bridge (indicated in yellow) and can be used as a probe for folding and unfolding. The structure was modified from PDB ID 1FH5. 1FVC V_H (PDB ID) is the variable domain from a humanised IgG1 κ antibody and has one additional Trp residues. The antigen of 1FVC is p185HER2 [212]. 1HEZ (PDB ID) V_H is from a human IgM antibody directed against Peptostreptococcus magnus protein L [257]. The last domain considered, is 1OPG V_H. It is from a murine monoclonal antibody which binds to the gpIIb/IIIa receptor [256] and has a similar sequence compared to MAK33 V_H.

3.3.1 Structural characterisation of V_H domains

Crystal structures of the different antibodies have shown the characteristic β-sheet structure for all V_H domains. In isolation, the five selected V_H domains share a minimum at ~218 nm in FUV-CD and a different pronounced maximum at 200 nm (Figure 56, top, closed circles) which are typical for β-sheet proteins. The minimum at 218 nm is not as pronounced as for example for C_L [192], which is probably caused by the CDRs which form a more random coil structure. In the presence of GdmCl the signal changes to a minimum at ~200 nm indicating the formation of a random coil structure (Figure 56, top, open circles). This effect is the least pronounced for 1FVC V_H. Near-UV CD spectra of the different V_H domains support the observations from the FUV-CD measurements. All domains show a characteristic fingerprint, which differs from the spectra typically observed for antibody domains [132, 134], but are well defined. MAK33 wild type and R44A, as well as 1FVC V_H, are similar with an exclusively positive signal and three maxima peaks at ~295 nm, 285 nm and 265 nm (Figure 56, middle, closed circles). 1HEZ and 1OPG V_H differ and share the presence of a negative minimum. In summary, the spectra were all well-defined and indicate the presence of defined tertiary

structure around the aromatic amino acids. All domains have in common that they lose their tertiary structure in the presence of GdmCl (Figure 56, middle, open circles).

As all antibody domains they share the presence of a conserved tryptophan near the internal disulphide bridge. Hence, intrinsic tryptophan fluorescence spectra were recorded in the absence (closed circles) and presence of GdmCl (open circles) (Figure 56, bottom). The fluorescence signal in the native state is higher compared to V_L [119, 192] due to the presence of the additional tryptophans. W103 and W47 are always solvent-exposed, whereas W36 is influenced by the internal disulphide bridge (see Figure 55; numbering according to Kabat [258]). 1FVC V_H has an additional exposed tryptophan at position 95 and shows consequently the highest fluorescence signal in the native state. The presence of GdmCl increases the fluorescence signal for all domains and is accompanied by a small red shift (Figure 56, bottom, open circles). The effect was the least pronounced for 1FVC V_H which shows only a slight signal increase. The small signal increase makes the domain difficult to study by fluorescence.

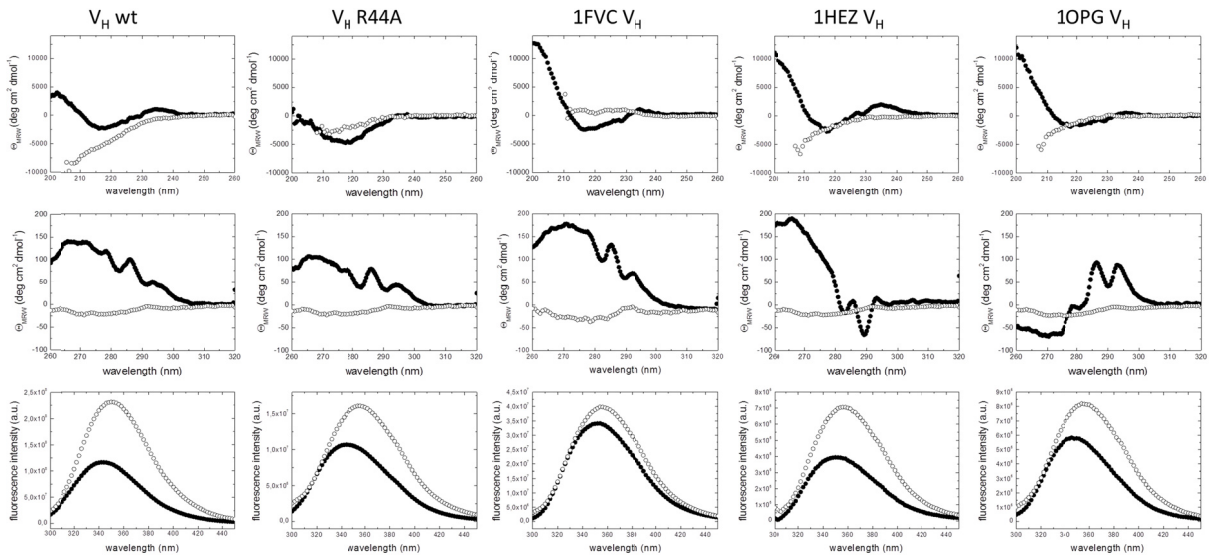

Figure 56: Spectroscopic characterisation of different V_H domains.
FUV-CD spectra (top), NUV-CD spectra (middle), and intrinsic tryptophan fluorescence spectra (bottom) of native (closed circles) and denatured (opened circles; MAK V_H in 1.2 M GdmCl, R44A in 1.5 M GdmCl, 1FVC in 4 M GdmCl, 1HEZ in 2.5 M GdmCl, and 1OPG in 2.5 M GdmCl) V_H domains in PBS buffer at 20°C. V_H wild type and V_H R44A are from MAK33. Proteins were denatured in GdmCl over night to ensure that the samples were in equilibrium.

3.3.2 Thermodynamic stability of V_H domains

The thermal stability of the V_H domains included in this study varies strongly. MAK33 V_H wild type unfolds in an uncooperative manner and aggregates at temperatures above ~30°C (Figure 57, top and Table 16). The mutation of the arginine at position 44 to alanine improves the stability, with a more cooperative folding behaviour in a temperature range between 10-50°C. At higher temperatures the domain aggregates like the wild type. 1OPG

V_H, which has the highest sequence similarity to MAK33 V_H, unfolds less cooperatively compared to the humanised 1FVC or the 1HEZ V_H domain but more like MAK33 wild type and has a T_{melt}= 30.4 ± 1.3°C (Figure 57, top and Table 16). Consequently, the stabilities between MAK33 and 1OPG V_H are comparable but 1OPG V_H is less prone to aggregation. 1HEZ V_H has a T_{melt}= 51.6 ± 0.3°C and its stability is comparable to V_L or C_L of MAK33 [192]. 1FVC V_H, the humanised domain is even more stable with a T_{melt}= 60.0 ± 0.2°C (Figure 57, top and Table 16). At high temperatures all domains showed aggregation and consequently none of the transitions were reversible (data not shown).

Equilibrium unfolding and refolding experiments were carried out to determine the chemical stability of the different V_H domains. The transitions were observed by intrinsic tryptophan fluorescence at the wavelength where the signal difference between the native and unfolded protein has its maximum (355 nm and 360 nm for MAK33 and the others, respectively). These experiments allow the observation of changes on the tertiary structure level. Additionally, the transitions were observed in FUV-CD at 215 nm to determine changes on the secondary structure level. For all domains a concentration range between 1 µM and 10 µM (for some 7.5 µM) was tested. No differences could be observed, indicating that second-order or higher order processes do not influence the equilibrium unfolding process (data not shown). The only exception is 1OPG V_H which tends to aggregate during refolding at concentrations > 1 µM (data not shown) and MAK33 wild type. The results of the transitions at 20°C are shown in Figure 57 (middle).

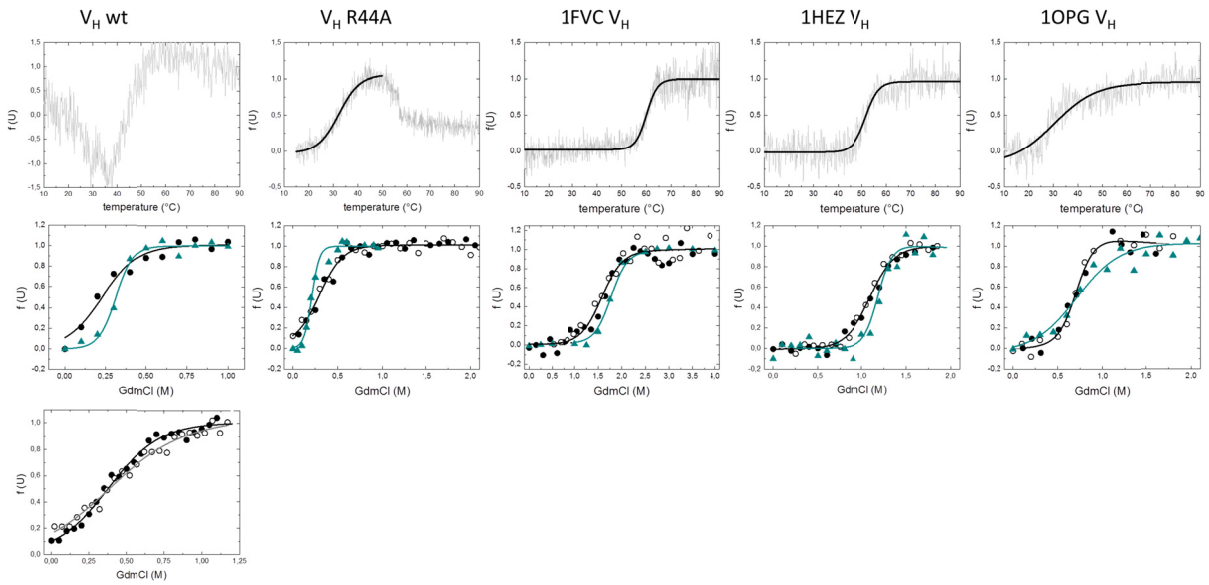

Figure 57: Biophysical stability of different V_H domains varies.

The top panels show the thermal-induced unfolding transitions for the different V_H domains. Temperature-induced unfolding was monitored by FUV-CD spectroscopy at 212 nm with a heating rate of 20°C h^{-1}. Data were fitted to a Boltzmann function to obtain transition midpoints. None of the transitions were reversible due to aggregation (not shown). The middle panels show chemical denaturation of the proteins induced by GdmCl. Triangles represents CD-data (measured at 215 nm). Circles represent fluorescence data. Closed circles indicate unfolding of the protein and opened circles refolding. Structural changes were monitored by fluorescence at a fixed wavelength (355 nm emission for V_H wt and R44A, 360 nm for 1FVC, 1HEZ, and 1OPG; excitation

wavelength was 280 nm). Data were evaluated according to a two-state unfolding model to obtain midpoints and cooperativity parameters of the transitions (fits are shown as continuous lines). Measurements were performed at 20°C at a protein concentration of 10 µM (for some 7.5 µM; temperature-induced transitions) or 10 (7.5) µM/1 µM (GdmCl-induced transitions). 1OPG is only reversible at low concentrations (≤ 1 µM). The transition observed with fluorescence is shown here for 1 µM concentration. The panel at the bottom shows the GdmCl-induced transition of V_H wild type at 1 µM concentration and 8°C.

Assuming a two-state transition, the data were fitted using linear extrapolation [96], resulting in different stabilities for the tested domains. The results of the transitions observed by fluorescence (circles) and FUV-CD (triangles) are shown in Figure 57 (middle panels). MAK33 V_H wild type and R44A are the least stable domains with a $D_{1/2}$ of ~0.25 M GdmCl (considering the error; see Table 16). The wild type is not reversible at 20°C (data not shown), but at 1 µM and 8°C at least a fraction can refold and the $D_{1/2}$ is slightly higher with ~0.37 ± 1.0 M GdmCl (Figure 57, bottom). The domain is not suited for further folding analysis as non-reversible proteins cannot be analysed quantitatively. The stability was not higher when urea was used as a denaturant (data not shown). The similar 1OPG V_H is, considering the errors, slightly more stable compared to MAK33, not only in terms of its thermal stability, with a $D_{1/2}$= 0.71 ± 0.30 M GdmCl. 1HEZ V_H is quite stable with a $D_{1/2}$= 1.09 ± 0.18 M GdmCl and has a well-defined plateau for the native state in low GdmCl concentrations (Figure 57, middle and Table 16). The most stable domain is 1FVC V_H, with a $D_{1/2}$= 1.57 ± 0.45 M GdmCl. However, due to the poor signal difference between the native and unfolded state, the transition could only be observed in NUV-CD (see Figure 56, bottom). When the $D_{1/2}$- values from the FUV-CD measurements are considered, most V_H domains seem to be slightly more stable (MAK33 V_H wild type, 1FVC V_H, and 1HEZ V_H; see Table 16). When the errors are taken into account the differences annihilate. The m-values ranged from ~11.0 kJ mol^{-1} M^{-1} for 1FVC V_H to ~24.4 kJ mol^{-1} M^{-1} for 1OPG V_H. V_L, for example, had an m-value of 16.0 ± 1.9 kJ $mol^{-1}M^{-1}$ [119] and lies in between the stable and unstable V_H domains. In general, the m-values were higher for the less stable domains and *vice versa*. The m-values from the FUV-CD measurements are larger compared to the fluorescence results for MAK33 V_H wild type and R44A, for 1OPG V_H in contrast the cooperativity decreases when the transition is observed on the secondary structure level (Table 16). For the other domains the values are comparable considering the errors. ΔG_{N-U} could not be determined for MAK33 V_H wild type as unfolding is not reversible. However, the other domains are able to refold under certain conditions. MAK V_H R44A is a very unstable domain with a ΔG_{N-U} of 5.5 ± 1.4 kJ mol^{-1}. V_L and C_L of MAK33 are clearly more stable with ΔG_{N-U} values of 17.4 kJ $mol^{-1}$and 14.6 kJ mol^{-1}, respectively [119]. 1FVC and 1HEZ V_H have ΔG_{N-U} values in a comparable order to V_L (17.3 ± 3.6 kJ mol^{-1} and 19.1 ± 2.1 kJ mol^{-1}, respectively).

Table 16: Biophysical stabilities of different V_H domains.

Domain	Method	T_{melt} [°C]	$D_{1/2}$ [M]	ΔG_{N-U} [kJ mol^{-1}]	m [kJ mol^{-1} M^{-1}]
MAK33 V_H wt	FL		0.23 ± 0.06	(5.1 ± 1.0)	22.2 ± 4.0
	FL (10°C)		0.37 ± 0.10	(3.8 ± 0.8)	10.2 ± 1.5
	CD	~30	0.31 ± 0.08	(14.3 ± 2.4)	45.7 ± 7.8
MAK33 V_H R44A	FL		0.30 ± 0.09	5.5 ± 1.4	18.7 ± 3.2
	CD	32.3 ± 0.2	0.21 ± 0.05	11.7 ± 1.9	55.7 ± 9.1
1FVC V_H	NUV-CD		1.57 ± 0.45	17.3 ± 3.6	11.0 ± 2.4
	CD	60.0 ± 0.2	1.76 ± 0.46	24.7 ± 4.5	14.0 ± 2.6
1HEZ V_H	FL		1.09 ± 0.18	19.1 ± 2.1	17.5 ± 2.2
	CD	51.6 ± 0.3	1.18 ± 0.83	36.2 ± 18.0	30.6 ± 15.4
1OPG V_H	FL		0.71 ± 0.30	17.3 ± 5.4	24.4 ± 7.0
	CD	30.4 ± 1.3	0.71 ± 0.33	7.2 ± 2.6	10.2 ± 2.9

Stabilities against thermal- and chemical-induced (GdmCl) denaturation of the different V_H domains in PBS are shown. Midpoints of thermal transitions are shown as T_{melt}. The GdmCl-induced unfolding transitions of MAK V_H were not reversible. Nevertheless; the data were evaluated according to a two-state equilibrium unfolding model to derive the midpoint of transitions ($D_{1/2}$), as well as the cooperativity parameter (m). The other V_H domains were reversible at least in a low concentration range (≤ 1 µM). ΔG_{N-U} values for MAK33 wild type are given in brackets to indicate that the values were determined from a protein which unfolds irreversibly.

The FUV-CD results differ from the analysis by intrinsic tryptophan fluorescence (Figure 57, middle, triangles). For most domains tested, the $D_{1/2}$-values seem higher when the samples were analysed by FUV-CD, except for MAK33 V_H R44A and 1OPG V_H (Table 16). However, the difference observed for 1FVC, 1HEZ and MAK33 V_H wild type are only marginal when the errors are taken into account and the transition behaviour is quite noisy for the domains in the FUV-CD due to the already high absorption of GdmCl at 215 nm. Hence, the results should not be overrated. Yet, the transition behaviour of these three domains suggests that an equilibrium molten globule state [81, 259] might be formed. A molten globule is characterised by high secondary structure content and only few defined tertiary interactions [123, 124]. Thus, the transition midpoint appears earlier when the samples are analysed by fluorescence (tertiary structure information) compared to FUV-CD (secondary structure information).

The reversibility of the transitions was confirmed by measuring spectra of the native, denatured, refolded and the native protein in the remaining GdmCl concentration. This is necessary as 1OPG and MAK33 V_H do not reach a well-defined plateau for the native state in the presence of GdmCl due to their low stability. Intrinsic tryptophan fluorescence and FUV-CD spectra show the same observations and confirm the reversible transition behaviour observed in the equilibrium unfolding and refolding experiments (Figure 58). MAK V_H wild type is not completely reversible even at low concentrations such as 1 µM and 8°C, but to a higher extend compared to higher concentrations (Figure 58, top and bottom). 1OPG V_H is also only reversible at concentrations ≤ 1 µM. The fluorescence spectra at 1 µM

concentration indicate reversibility, whereas the FUV-CD spectra cannot confirm this result (7.5 µM). Interestingly, aggregation was not only observed for the refolded protein at 7.5 µM concentration but also for the control (native protein in the low remaining GdmCl concentration; Figure 58). On the contrary, MAK33 V_H R44A and 1HEZ V_H are reversible in a range between 1 µM and 10 µM, although R44A cannot reach the native state at 0 M GdmCl due to the missing plateau for the native state (Figure 58 and Figure 57). In the case of 1FVC V_H reversibility was just tested by FUV-CD as the fluorescence signal difference is very low (Figure 56, bottom). The FUV-CD spectra prove the reversibility observed in equilibrium unfolding and refolding experiments (Figure 58 and Figure 57).

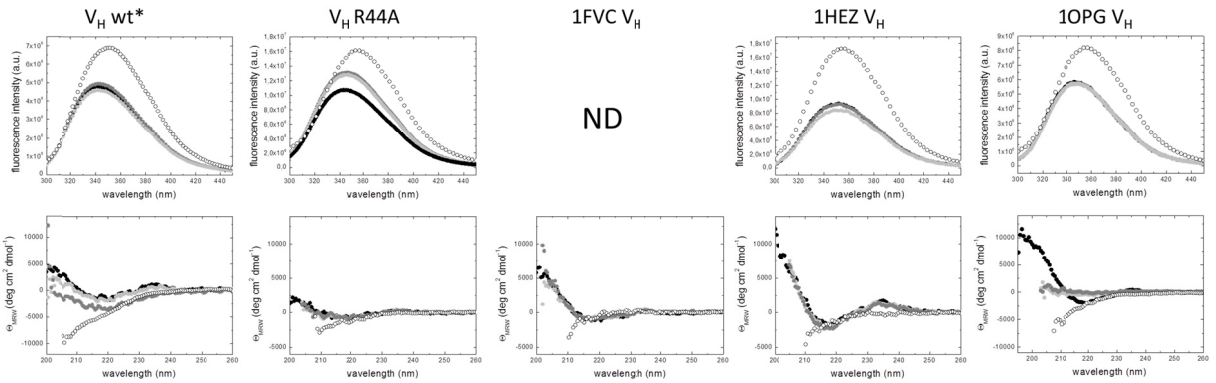

Figure 58: For most V_H domains unfolding is reversible at low concentrations.
To prove that the unfolding process is reversible, intrinsic tryptophan fluorescence (top panels) and FUV-CD (bottom panels) spectra were recorded. The samples were denatured overnight in 1.2-4 M GdmCl and refolded by dilution in PBS (grey circles). As some of the domains show no plateau for the native state (MAK33 V_H and MAK33 V_H R44A) native protein was incubated in the GdmCl which is remaining after dilution (light grey circles). For a better comparison the spectra of the native (black circles) and unfolded (black open circles) proteins are included. FUV-CD experiments were performed with 7.5 µM – 10 µM protein, fluorescence experiments with 1 µM. Temperature was 20°C. Asterisk indicates that the samples were treated at 8°C. ND indicates that the spectra were not determined.

We could observe for some of our V_H domains that they have the ability to form homo-dimers under certain conditions. Interestingly, 1HEZ V_H forms exclusively homo-dimers at high (10 µM) and low concentrations (1 µM) in AUC sedimentation velocity experiments. Dimerization increases the complexity of the folding analysis as association and dissociation reactions will appear in parallel to folding and unfolding events. Hence, the folding pathway will probably be more complex compared to V_H domains which cannot self-associate in this concentration range. As a consequence this domain is not suited for this folding analysis as the main domain of interest is MAK33 V_H. 1FVC V_H is monomeric at low concentrations (1 µM) but can also self-associate at higher concentrations. At 10 µM protein concentration already ~10% of the protein is existent as a dimer (Table 17).

Table 17: Some V_H domains can form homo-dimers.

Domain	Monomer [%]	Dimer [%]
MAK33 V_H	100	-
MAK33 V_H R44A	100	-
1FVC V_H*	100	-
1FVC V_H	90	10
1HEZ V_H	-	100
1OPG V_H	100	-

Sedimentation velocity analytical ultracentrifugation was used to determine the ability of the different V_H domains to form homo-dimers. The protein concentration tested was 10 µM to 1 µM. Experiments were performed in PBS at 20°C. For all cases where no differences between 10 µM and 1 µM could be observed, the results for 10 µM are shown. 1FVC V_H was the only domain differing in the concentration range tested. The asterisk indicates the species formed at 1 µM concentration.

To account for the dimeric state of 1HEZ V_H, the equilibrium unfolding and refolding transitions were fitted with a model including dimerization (adopted from [154]). As the errors of the fit for the FUV-CD were too large, just the fluorescence data were considered. According to the model the stability is more than twice as determined by the monomer model ($\Delta G_{N\text{-}U,\ dimer}$ = 54.5 ± 6.0 kJ mol^{-1} compared to $\Delta G_{N\text{-}U,\ monomer}$ = 19.1 ± 6.0 kJ mol^{-1}). In case of dimeric C_H3 the calculated intrinsic stability was 66.5 ± 1.5 kJ mol^{-1} [180]. Both domains are in a range of 42 kJ mol^{-1} to 113 kJ mol^{-1} described for dimeric proteins [260]. In contrast, the cooperativity values are comparable (m_{dimer}= 24.0 ± 5.3 kJ mol^{-1}M^{-1} and $m_{monomer}$ = 17.5 ± 2.2 kJ mol^{-1}M^{-1} for monomer model; see Table 16).

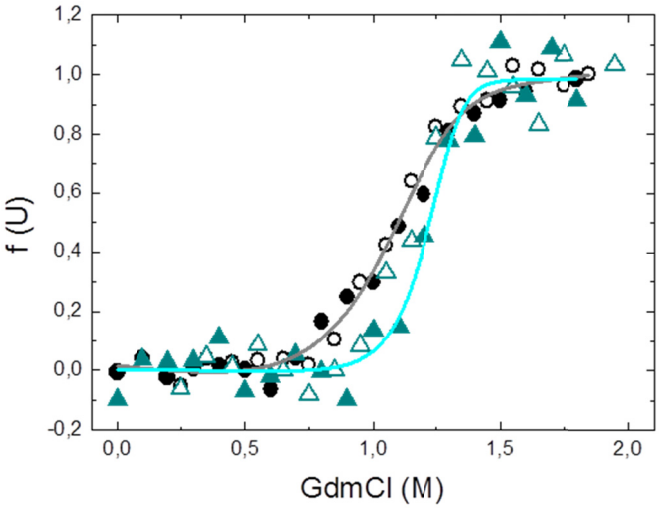

Figure 59: Dimer model for transition of 1HEZ V_H.
Due to the observation that 1HEZ V_H forms dimers also at concentrations of 1 µM, the GdmCl-induced transitions for unfolding and refolding were fitted to a model including dimer formation [154]. Triangles represent CD-data (measured at 215 nm). Circles represent fluorescence data (measured at 360 nm). Closed symbols indicate unfolding of the protein and opened symbols refolding. Straight lines indicate the fit (dark grey for fluorescence and cyan for CD).

3.3.3 Refolding and unfolding kinetics

In order to describe the folding mechanism of a protein, it is necessary to look at the folding kinetics [261]. Protein folding is a first order, unimolecular reaction, as long as no aggregation/oligomerisation (second order process) occurs. Hence, the kinetics of unfolding and refolding of the different V_H domains at different GdmCl concentrations was determined at different protein concentrations. The proteins were denatured in a sufficient amount of GdmCl to ensure that the proteins were fully denatured (1.2 M-4 M GdmCl; see Figure 57). To guarantee that the equilibrium was reached, the samples were incubated overnight. Afterwards, the samples were diluted manually or by a stopped-flow device into PBS to induce refolding and the changes in tryptophan fluorescence were monitored. For 1FVC V_H the reaction was followed by FUV-CD because the fluorescence signal difference is too small for a good signal-to-noise ratio.

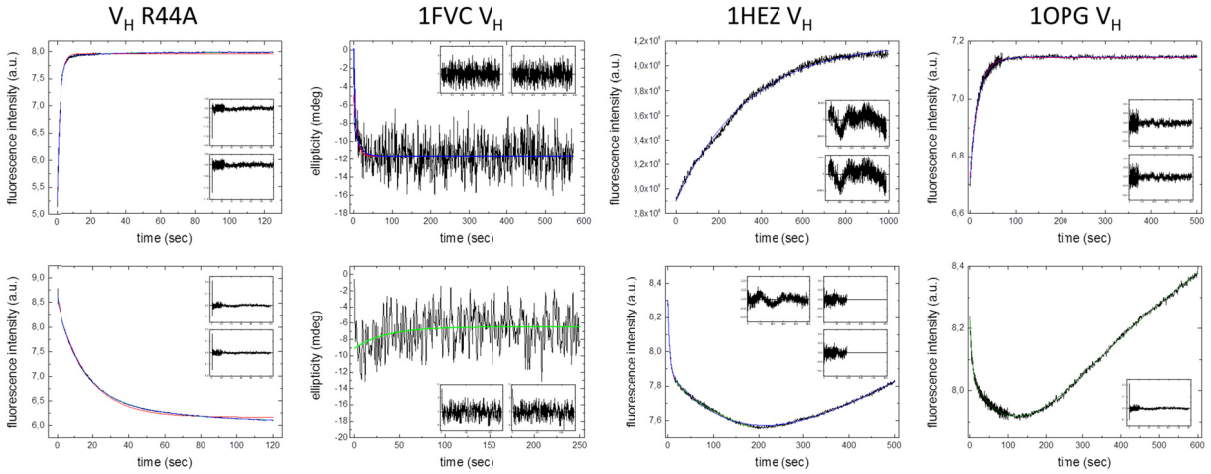

Figure 60: Representative kinetic traces for unfolding and refolding single jump experiments.
For all domains which unfold reversibly, single jump experiments to observe unfolding and refolding behaviour were performed. As most of the domains unfold (top panels) and refold (bottom panels) including a fast phase, stopped-flow measurements were preferred. All samples were measured using intrinsic tryptophan fluorescence as a probe except 1FVC V_H which was measured by FUV-CD. Wavelengths were the same as for the equilibrium unfolding and refolding experiments. Residual for the single, double or triple exponential fits are shown as insets to demonstrate the fit quality. R44A: unfolding at 0.7 M GdmCl; refolding at 0.06 M GdmCl. 1FVC: unfolding at 2.5 M GdmCl; refolding at 0.5 M GdmCl. 1HEZ: unfolding at 2.4 M GdmCl; refolding at 0.1 M GdmCl. 1OPG: unfolding at 0.8 M GdmCl; refolding at 0.04 M GdmCl

Examples for kinetic traces of unfolding (upper panels) and refolding (lower panels) which could be observed for MAK33 V_H R44A, 1FVC V_H, 1HEZ V_H and 1OPG V_H are shown in Figure 60. These examples denote already that MAK33 V_H R44A unfolds and refolds the fasted. 1HEZ V_H and 1OPG V_H show a complex refolding behaviour with a signal decrease which is followed by an increase (Figure 60). As 1HEZ V_H forms homo-dimers also at low concentrations (Table 17), association might occur after refolding and explain some signal change. On the contrary, 1OPG V_H is a monomer and behaves reversible at the applied concentration. However, it refolds in a complex manner similar to 1HEZ V_H. The FUV-CD signal of 1FVC V_H is very noisy although the traces are averages of at least five repeats. One

challenge is the absorption of GdmCl at 215 nm. In conclusion, the fits have high errors and are not reliable. Residual of the fits are shown as inlets to prove the qualities of the fits (Figure 60).

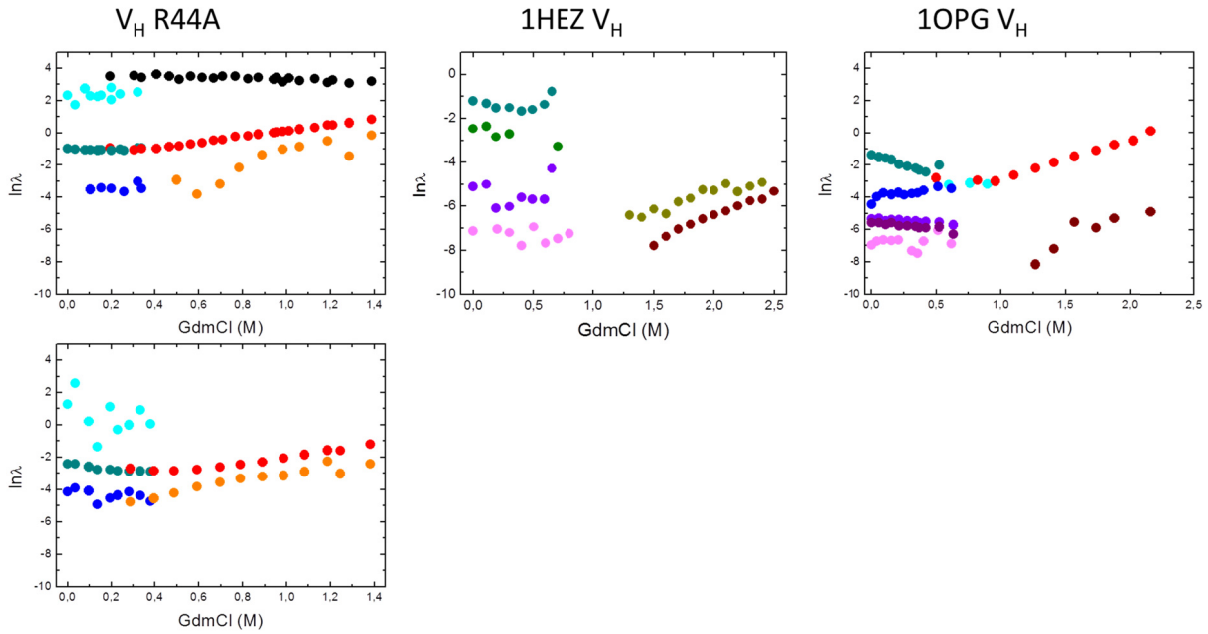

Figure 61: Chevron plots for the different V_H domains from single jump experiments.
Chevron plots showing the variation in the observed rate constants obtained from exponential fits of refolding and unfolding kinetics with GdmCl concentration for MAK33 R44A V_H, 1HEZ V_H and 1OPG V_H. The V_H domains in 0 or 1.5 M or 2.5 M GdmCl were diluted into an unfolding or refolding buffer (PBS) with varying GdmCl concentrations, respectively, and the change in fluorescence intensity was monitored. All apparent rate constants (λ) were determined by stopped-flow and manual mixing measurements. Cyan (λ1), dark cyan (λ2), blue (λ3), purple (λ4), violet (λ5) and light magenta (λ6) circles represent the apparent rate constants of refolding experiments. Black (λ1), red (λ2), orange (λ3), dark yellow (λ4) and brown (λ5) circles represent the apparent rate constants of unfolding experiments. For V_H R44A all apparent rate constants could be determined by stopped-flow experiments. In the case of 1OPG V_H, λ3 (blue) and λ6 (light magenta) for refolding were only observed by manual mixing experiments. λ2 (red) for unfolding was observed in all measurements (stopped-flow and manual mixing). For 1HEZ V_H λ4 (unfolding) and λ6 (refolding) are a result of manual mixing and the other λs were from stopped-flow measurements.

Due to the poor quality of the kinetic traces for 1FVC V_H no reliable statement about its unfolding and refolding behaviour is possible (see Figure 60). Two unfolding rates were observed in the case of 1HEZ V_H which were slow. λ4 (dark yellow circles) was observed in manual mixing and λ5 (brown circles) stopped-flow experiments. For the unfolding of 1OPG V_H a similar rate constant could be observed (λ5; brown circles) and one faster one (λ2; red circles) which was also observed for MAK33 V_H R44A (λ2; red circles) (Figure 61, top panels). Interestingly, one very fast phase (λ1; black circles) which behaves GdmCl independent at 20°C could be observed. When the kinetic traces were measured at 8°C this phase was not visible (Figure 61, top panel and bottom panel). However, the source of this rate could not yet be identified. For R44A, 1OPG and 1HEZ V_H a concentration range from 0.5 µM to 5 µM was tested and no changes within this concentration range were detected for R44A and

1HEZ V_H (data not shown). 1OPG V_H differs in the apparent rate constants due to aggregation observed at concentrations > 1 µM.

Refolding of three domains (MAK33 V_H R44A, 1HEZ V_H and 1OPG V_H) is a complex process with at least three observable rate constants (λ; Figure 61, top panels). 1FVC V_H cannot be included due to the low signal quality. R44A V_H is the fastest refolding domain with one very fast refolding phase which was not observed for the other domains (λ1; cyan circles). The phase λ2 (dark cyan circles) is detectable for all three domains, λ3 (blue circles) besides for R44A just for 1OPG V_H. 1HEZ and 1OPG V_H share two apparent rate constants (λ4; purple circles and λ6; light magenta circles; Figure 61, top panels, right). When the complete kinetic traces for 1OPG V_H were analysed refolding was even more complex with one more observable rate constants (λ5; violet circles; Figure 61, top panel and Figure 60 bottom panel).

3.3.4 Contribution of proline isomerisation to the refolding process

All V_H domains show multiple phases in both refolding and unfolding kinetics suggesting that the folding pathway of all V_H domains is more complicated than a simple two-state process. 1OPG V_H aggregates during refolding at concentrations > 1 µM and 1HEZ V_H forms homo-dimers during unfolding and refolding even at low concentrations (1 µM). On the contrary, for MAK33 V_H R44A a contribution of aggregation/oligomerisation reactions during refolding could be excluded by performing the experiments over a 10-fold concentration range (final protein concentration, 0.5 µM-5 µM). Within this range no differences in the observed rate constant were detected (data not shown). In the following just MAK33 V_H R44A will be addressed.

For several antibody domains it was observed that one source of slow refolding is *cis-trans* isomerisation of peptidyl-prolyl peptide bonds [35, 119, 191, 202]. Peptide bonds have a partial double bond character which enables a *cis* or *trans* conformation. Most peptide bonds are more stable in *trans* conformation independent if the protein is in its native or unfolded state [262]. When a peptide bond precedes a proline residue both conformations exist as the two states are in this case almost isoenergetic [263]. Yet, the *trans* conformation is more prevalent in native proteins [264] and the activation barrier to interconversion is quite high. Hence, populations of isomerisation-limited slow folding molecules can appear [263, 264]. The X-ray structure of MAK33 Fab (PDB ID 1FH5) was inspected for the presence of Xaa-Pro in *cis* conformation. None of the three proline residues in this structure give any hints for the presence of a *cis* conformation. However, to confirm this observation refolding in the presence of a peptidyl prolyl *cis-trans* isomerase (PPIase) Cpr6 was observed. Cpr6 is a cyclophilin from yeast [265] and could successfully accelerate the *cis-trans* isomerisation reaction in V_L [119]. Cpr6 had no effect on any of the three observed refolding phases which corresponded to λ1, λ2 and λ3 (of refolding) as no dependency on the presence of PPIase could be observed (Figure 62 A). It was surprising that not even λ3 (blue circles) was affected as it seemed independent of the GdmCl concentration (i.e., a marker for isomerisation

reaction) and was the slowest phase observed. However, it is possible that Cpr6 is not specific for V_H although it can recognise V_L. To confirm this observation other PPIases should be tested.

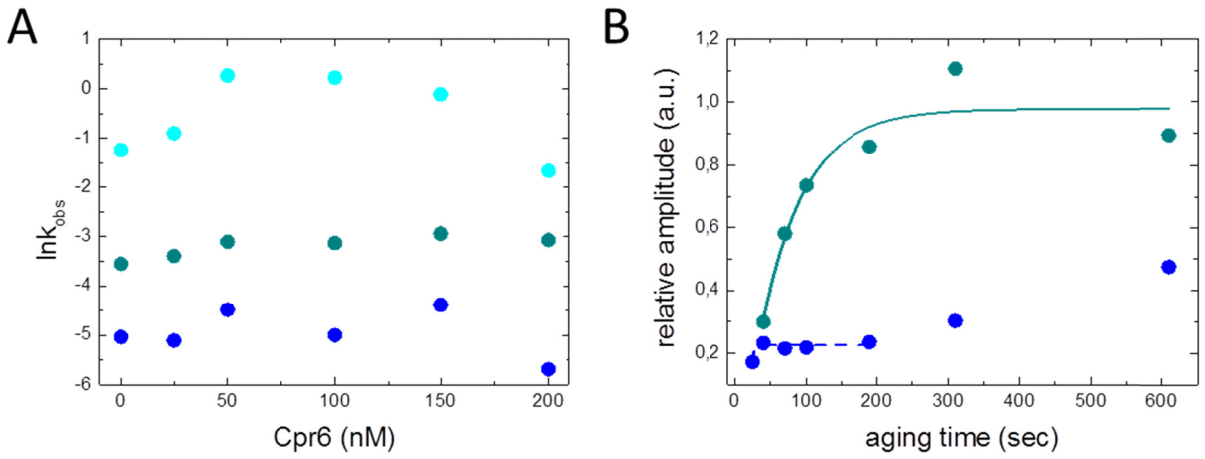

Figure 62: Proline isomerisation does not influence the refolding process.
(A) Effects on refolding rate constants of MAK33 R44A V_H upon addition of Cpr6. Cyan, dark cyan and blue circles denote the refolding rates of λ1, 2 and λ3, respectively, in 0.05 M GdmCl in the presence of increasing amounts of Cpr6. (B) Double-jump experiment of MAK33 R44A V_H. The native protein was unfolded in 1.5 M GdmCl for various aging times before being refolded in 0.05 M GdmCl. Two refolding phases were observed at all aging times. The increase in amplitude of each of the refolding phases corresponding to λ2 (dark cyan) and λ3 (blue circles)

Other slow equilibration processes can be detected by double-jump experiments [263]. For this experiment MAK33 V_H R44A was unfolded in 1.5 M GdmCl for varying aging times. A high denaturant concentration ensures that unfolding is faster than putative isomerisation phases. Afterward, the protein was transferred to renaturing conditions (0.05 M GdmCl). The refolding process was monitored by intrinsic tryptophan fluorescence. Long aging times enable the unfolded state to equilibrate and to observe all processes. On the other hand, at short aging times only directly unfolded proteins will be present to refold. As a consequence, phases occurring due to refolding and due to slow equilibration processes in the unfolded state can be distinguished. This experiment was performed to rule out other slow equilibration processes besides proline isomerisation which were observed for some proteins [266]. Although non-prolyl peptide bonds appear in their *cis* conformation in populations $< 0.15\%$ in unfolded proteins, the large number of peptide bonds in a protein can generate a significant fraction of slow-folding molecules and consequently heterogeneous populations in an unfolded protein can induce complex folding kinetics on various time scales [266]. For MAK33 V_H R44A two kinetic phases were observed at all aging times. The amplitudes correspond to the two slower apparent rate constants of refolding (λ2 and λ3; Figure 62 B). The amplitude for the fastest phase could not be detected at any aging time. This reaction is probably too fast to be determined by a manual mixing approach and might be observed if the experiment is repeated with a stopped-flow device. The result indicates that the slower phases might not be a consequence of slow equilibration processes

in the unfolded state. Otherwise the appearance of an isomerisation phase would be expected. One of the observed amplitudes increases in an exponential manner (λ2) and corresponds to the refolding phase in the Chevron plot λ2 (refolding; see Figure 62 B and Figure 61). The second amplitude was more difficult to describe but corresponds to λ3 (refolding) in the Chevron plot. In the case of 1OPG V_H just three different amplitudes could be observed at all aging times. Only one of the observed amplitudes increased with longer aging times. However, except for this one none of them could be clearly assigned to the refolding phases observed for 1OPG V_H in the Chevron plot (data not shown). They referred probably to λ2 and λ4/λ5. PPIase and double-jump experiments give only hints for the absence of *cis-trans* isomerisation of proline or non-proline residues. These observations are in accordance with the information of the X-ray structure. However, it cannot be excluded yet that a *cis-trans* isomerisation takes place. To confirm the observations interrupted refolding experiments [267] are crucial.

3.3.5 Discussion

Since the first folding studies on immunoglobulins, fragments and single domains [174-176, 268-270], insight in the folding mechanisms could be improved for the folding of C_H3 [180], C_H2 [202], C_L [191], and V_L [119] of MAK33. C_H1, which is intrinsically disordered in isolation, cannot be analysed as folding depends on the presence of C_L [181]. All domains studied so far show complex folding behaviour with multiple folding phases [35, 119]. V_H is the only domain without a known folding mechanism. The domain is highly flexible especially in the CDR-H3 [235, 271] but also in the framework region within different germline representatives [198]. Besides differences on the sequence level also variations in the properties between germline families were observed [199]. As a consequence different behaviour between various V_H domains cannot be excluded. Two murine IgG1 domains (MAK33 and 1OPG) which share ~80% sequence identity and one stabilising point mutant (MAK33 R44A) which was identified in the interaction analysis of V_L and V_H (see 3.2.1) were selected for analysis of the folding process. In addition, a humanised IgG1 domain (1FVC) and one IgM domain (1HEZ) were included in the folding analysis.

On the structural level, the five domains showed similar secondary structures in the FUV-CD but 1HEZ and 1OPG V_H differ in some aspects of their NUV-CD spectra. However, all domains depict a characteristic fingerprint which is influenced by the different contents of aromatic amino acids (Phe, Tyr, and Trp) in the different domains. In the presence of high concentrations of GdmCl the domains unfold and adopt random coil characteristic (Figure 56, top and middle). In tryptophan fluorescence spectra all share a signal increase in the presence of GdmCl due to the exposure of two conserved tryptophans near the internal disulphide bond (Trp36; Figure 55). This was already observed for the V_L domain [119]. However, the signal difference is lower for 1FVC V_H due to a second exposed tryptophan, Trp95, which is not present in the other domains and increases the background signal in the native state. The biophysical stabilities vary clearly between the different domains. MAK33

V_H wild type is the least stable domain in terms of temperature- and GdmCl-induced equilibrium unfolding (T_{melt}≈ 30°C, $D_{1/2}$≈ 0.2 M GdmCl), whereas 1OPG V_H which shares ~80% sequence identity has a slightly better stability (T_{melt}≈ 30.4°C; $D_{1/2}$≈ 0.7 M GdmCl; Table 16 and Figure 57). C_H2, which was already considered as a relatively instable antibody domain compared to the previous described (T_{melt}≈ 41°C; [202]), is more stable than 1OPG V_H. 1FVC and 1HEZ V_H are the most stable domains in this study. The thermal stability of the optimised 1FVC V_H (T_{melt}≈ 60°C) domain is comparable to the most stable V_H consensus sequence 1DHU (T_{melt}≈ 64.4°C; Table 16 and [199]). 1HEZ V_H is less stable (T_{melt}≈ 51.6°C) compared to 1FVC V_H and behaves more like V_L (T_{melt}≈ 50.6°C, [119]) or C_L (T_{melt}≈ 51°C, [191]). A special feature of 1FVC and 1HEZ V_H is their ability to self-associate even at low concentrations. This was not observed for MAK33 or 1OPG V_H (Table 17). This homo-dimerization may stabilise the exposed hydrophobic interface covered by V_L when present [199]. For V_H domains it was assumed that long CDR-H3s might shield to some extent the hydrophobic surface and therefore stabilise the domain [199]. Their assumption does not fit with the results for the four V_H domains selected here. 1OPG and MAK33 V_H are the least stable domains but their CDR-H3s are longer compared to 1HEZ and 1FVC V_H. The domains differ also in their ability to refold reversibly. MAK33 V_H cannot completely refold even at low concentrations and low temperature (Figure 58) and consequently no quantitative evaluation was possible. Interestingly, the mutation of R44 to alanine increases the refolding ability. Due to the single point mutation, the domain refolds even at 20°C and up to concentrations of 10 µM (higher concentrations were not tested), whereas the similar 1OPG V_H domain is only reversible at low concentrations ≤ 1 µM at 20°C (Figure 58). This leads to the conclusion that already small changes within the domain can have severe effects on the biophysical properties and on the folding behaviour hence, it will be unlikely to find a common folding mechanism valid for all V_H domains. The two stable domains, 1FVC and 1HEZ V_H, had both the ability to completely refold. Interestingly, the transitions for MAK33 wild type, 1FVC and 1HEZ V_H indicate the presence of a molten globule intermediate. The transition midpoints of FUV-CD transition appear at higher GdmCl concentrations compared to the fluorescence traces. Consequently, an intermediate might be formed which has still some secondary structure whereas the tertiary contacts are already lost – a typical feature of molten globules [81, 123]. Yet, one must be careful with these assumptions as the transitions are the mean values of three experiments and the data quality in the FUV-CD is noisy < 215 nm due to GdmCl absorption. Triplicates were necessary to improve the signal-to-noise ratio.

Besides equilibrium unfolding and refolding transitions, kinetic experiments were performed to address the question of folding intermediates. Unfolding for 1HEZ V_H could be best described by a single exponential function. However, for stopped-flow and manual mixing experiments different apparent rate constants were observed (λ4 and λ5; unfolding; Figure 61). The difference might at least to some extend be explained by the very slow unfolding reaction. The traces in the stopped-flow experiments were not monitored as long as for the manual mixing experiments. Thus, the fit could not describe the traces well because the plateau of the reaction was not reached. As 1HEZ V_H forms homo-dimers in the native state,

the denaturation process is expected to comprise dissociation and unfolding reactions. In a concentration range between 1 µM and 5 µM no differences between the kinetic traces could be observed. Therefore, at least in the tested range, folding is not concentration-dependent. However, this is not surprising as in this concentration range still only dimers are present. C_H3 showed also no concentration-dependence and can also still form dimers at very low concentrations [202] (as low as 10^{-10} M [177]). The refolding traces for 1HEZ and 1OPG V_H were complex with a signal decrease followed by an increase (Figure 60) and could be described only with complex exponential functions. For 1HEZ V_H it would be expected that association of the refolded domains leads to a signal change as it forms dimers even at low concentrations. In addition, two tryptophans are located near the interface to V_L (Trp47 and Trp103) which is assumed to be the interaction site for self-association, too. Due to several hydrophobic residues in the interface dimerization is probably driven by hydrophobic interactions. However, one would expect a signal decrease if one of these residues or even both would be covered due to association. The intrinsic tryptophan fluorescence spectra indicate no such effect as their intensity is comparable to monomeric MAK33 V_H at the same concentration (Figure 56, bottom). Another aspect against this assumption is the observation that also for refolding no differences at various concentrations appeared (data not shown). If dimerization would be a rate-limiting step during refolding and cause a change in the signal, an influence on the rate constants should be observable. In the case of dimeric C_H3, also no effects of concentration on refolding were detected but the domain has no tryptophan in the interaction site and hence association might not influence the signal intensity [180]. These observations raise the question whether homo-dimerization does not appear in the interface area or if dimerization is not a rate-limiting step. For 1OPG V_H the effects are even more challenging to describe. As aggregation was observed during refolding at concentration > 1 µM it might be possible that the complex kinetic is a result of the formation of some unspecific assemblies before the native structure is adopted (at low protein concentrations such as 1 µM). Another reason for the differences between V_H domains in general might be the influence of the CDRs which differ between the different domains. There has not yet been a systematic study of the CDR influence on stability but V_L domains have been shown to differ in stability in isolation [272]. For V_H predominantly the effects of CDR-H3 were addressed in terms of the influence of the CDRs on antigen binding but less the consequences for stability (especially in isolated domains) [271, 273]. The preferred model systems for CDR exchanges have been scFvs and Fabs [168-172]. The loop regions are considered to influence the initial stages of protein folding as they can bring different regions of the protein into close proximity [274] and might hence influence the folding mechanism.

The Chevron plots obtained from single jump experiments are complex and exclude a simple two-state folding process as a possible mechanism. At the current state, no clear assumption of the characteristics of the intermediates or a reliable number of them can be made. With the current data set no global analysis was possible so far because a proper model could not yet be identified (global fitting with Dynafit; data not shown). However, some peculiarities of

the plots are worth to mention. Most obvious for MAK33 R44A V_H, at 20°C a fast and GdmCl-independent apparent rate constant for unfolding is observable which is not detected at 8°C (λ1; black circles in Figure 61, top and bottom). Different reasons are possible why this apparent rate constant disappears at lower temperatures. One explanation might be a higher energy barrier for this pathway compared to the other two observed, which cannot be overcome at low temperatures. The consequence would be that there is a clear change within the energy landscape. It might also be caused by a similar conformation between the sensors thus no signal of the denatured fast phase at 8°C could be observed. However, this would presume that a temperature-dependent pronounced conformational change takes place. Without further experiments it will not be possible to differentiate between these possibilities. The Chevron plot of a two-state folding protein will have only two limbs, one for refolding and one for unfolding. The intercept of the fit of both arms will result in the denaturation midpoint (C_m). This value should be concordant with $D_{1/2}$ determined in the equilibrium unfolding experiment [72]. In the case of MAK R44A and 1OPG V_H the values obtained for C_m and $D_{1/2}$ are consistent within the error range when λ2 (cyan circles) and λ2 (red circles) are analysed accordingly. However, for 1HEZ V_H C_m and $D_{1/2}$ are concordant when λ4 (purple circles; refolding) and λ4 (dark yellow circles; unfolding) or λ6 (light magenta circles, refolding) and λ5 (brown circles; unfolding) are combined (see Figure 61). In both cases apparent rate constants of a stopped-flow and manual mixing experiment were combined. This might be a hint that indeed all observed apparent rate constants are real and not an artefact. In conclusion, MAK33 and 1OPG V_H probably share some common process during unfolding and refolding and these two processes are dominant in the equilibrium unfolding and refolding mechanism. As long as aggregation could be excluded a rollover in the refolding arm of the Chevron plot is a good hint for the presence of an intermediate [275]. For MAK R44A, aggregation was not observed between 0.5-5 µM (data not shown), whereas 1OPG V_H aggregated at concentrations > 1 µM. Within the limited range where refolding could be observed for MAK33 R44A V_H (0- ~0.25 M GdmCl), there is no hint for the presence of a rollover in any of the three phases (Figure 61). The refolding arms of the Chevron plot of 1OPG V_H at 1 µM might indicate the presence of a rollover in λ3 (blue circles) and λ6 (light magenta circles) though some of the values near the transition point scatter. Due to the complex refolding kinetics and the limited GdmCl concentration range in which refolding can be observed one must be careful with the interpretation at the current state. Some intermediates form too rapidly to be directly detected in stopped-flow experiments [129, 276]. Besides rollover at low GdmCl concentrations the presence of a burst phase in the amplitude data can be a hint for an intermediate [276]. R44A V_H is challenging as no plateau for the native state could be gained. On the one hand, the limited range in which refolding can be observed is probably not sufficient to elucidate all events. On the other hand, a normalisation of the amplitudes is more challenging. In addition, the Chevron plot should be measured as close to 0 M GdmCl as possible to show that there is no deviation from the kinetics in pure aqueous solution [72]. This is difficult for many V_H domains as they tend to aggregate in high concentrations and this restricts the dilution range.

For all antibody domains of MAK33 analysed so far, proline isomerisation was an important step within the different folding pathways [119, 180, 191, 202]. This slow process can either be detected directly through the addition of a PPIase which catalyses the *cis-trans* isomerisation of proline bonds or, more indirectly, by double jump experiments [263]. According to the crystal structure of MAK33 all three proline residues of V_H are in a *trans* conformation in the native state (PDB ID 1FH5). This does not exclude the presence of one or more *cis* conformations in the unfolded state. However, none of the three observed refolding phases (λ1, λ2 and λ3) of R44A V_H showed acceleration in the presence of Cpr6. In the case of V_L, one observed phase demonstrated a linear dependency on the Cyclophilin B (CypB) concentration as well as on Cpr6 [119]. For a control, a second PPIase, for example also CypB [179] should be tested to exclude that the enzyme simply cannot recognise V_H as a substrate. It is known that different PPIases can have different substrate specificities [277, 278]. Double jump experiments can be used to determine the presence of different slow isomerisation processes which can appear besides peptidyl-prolyl isomerisation [266]. For MAK33 R44A V_H two refolding phases were observed at all aging times during the double jump experiment corresponding to λ2 and λ3 of the refolding limp. As the experiments were so far only performed by manual mixing, λ1 might not be observed because it is too fast for the ~8 seconds dead-time of mixing. For this refolding phase, repeating the experiment with a stopped-flow device (dead-time within a few ms) will be useful. The current results suggest that the two observed phases are not a result of slow isomerisation processes although they show no strong dependence on the concentration of GdmCl (Figure 61). If slow equilibration processes would take place a delay in the appearance of the affected phase would be observable [119]. In addition, the rate of refolding would be independent of the denaturation time [119]. In contrast, the amplitudes of the two observed rate constants changed in an exponential manner with an increasing aging time as observed for V_L [119]. So far the results suggest that λ2 and λ3 are indeed folding phases although λ3 shows no obvious GdmCl dependence and that the refolding process is not related to *cis-trans* isomerisation. According to Feige and co-workers an increase of amplitudes with longer aging times (especially observed λ2) can be a hint for isomerisation [202]. To affirm this assumption, it would be important to control if the amplitude of the fastest phase (λ1) decreases in a similar manner as the amplitude of the slow phase increases. This effect was observed for C_H2 and lead to the assumption that isomerisation plays a role for the folding of C_H2 [202].

The properties of the intermediates could not yet be determined. This will require more experiments, for example an interrupted refolding experiment (N-test) to determine the nature of the intermediates [267]. This experiment can also shed light on isomerisation processes. Not till then, reliable assumptions on the differences between the folding mechanism of V_H and V_L and their different propensity to form fibrillar structure can be made.

Conclusion

Within this thesis several important topics in the field of antibody research were addressed. New insights in the folding and association behaviour of antibody domains could be gained, on the one hand for V_L and V_H association and V_H folding and on the other hand for several antibody domains in extreme solvent conditions. In addition, the role of the conserved amino acids in the interface of the variable domains could for the first time be determined on the single domain level.

The analysis of the single domains of the Fab fragment of MAK33 and C_H2 in acidic conditions had demonstrated surprising differences between the single domains. Despite the common folding motif in the native state, differences in their requirements for an AFS formation as well as in their stabilities were observed. The mechanism of AFS formation seems to be common for all observed domains so far. For a complete rearrangement on the secondary structure level, the antibody domains pass a partially or completely unfolded state. NMR analysis of C_H2 had enabled deeper insight on the AFS structure. Surprising, the AFS is a highly dynamic system contrary to its high thermodynamic stability. It will be interesting to see if these findings can be transferred to other immunoglobulin superfamily members. The folding behaviour of V_H differs from the other domains characterised do far. For MAK33 V_H there was no clear hint for the presence of slow isomerisation processes. This finding was surprising, as *cis-trans* proline isomerisation was important for the refolding of all other domains of MAK33. However, the comparison of different V_H domains has shown that they differ in their abilities to refold and their folding behaviour. In conclusion, also their rate-limiting steps may differ. V_H is also more sensitive to changes in the conserved amino acids than V_L. Structure and stability were strongly influenced as well as the interaction ability with V_L. For V_L just one proline was identified to be important for stability and interaction. However, antigen binding was not correlated with the interaction between the variable domains, although for some residues the binding affinities were influenced in the same way. Several mutations could not confirm this assumption.

The variable domains often behaved different compared to the constant domains. In the AFS they exposed the most hydrophobic surface and formed the largest oligomers. The folding pathway of V_L was demonstrated to differ from the constant domains [119]. V_H differed already between different V_H domains. Besides differences in the β-strand organisation of the immunoglobulin fold [157] between variable and constant domains, the variety within the CDRs is most obvious. They could also explain the variations observed between the V_H domains analysed in this study. Single point mutations in CDR-H3 could be shown to influence structure and stability without disturbing interaction or functionality. It would be interesting to see if they have also a beneficial influence on the folding of V_H as the R44A mutation in MAK33 which granted reversibility. Thus, for engineering of basic properties the CDRs should also be considered. However, it is important to consider that antigen

recognition and binding is not impeded. These new findings of a well-studied class of proteins will provide novel guidelines for future antibody engineering.

Abbreviations

ADCC	antibody-dependent cellular cytotoxicity
AFS	alternatively folded state
AL	light chain amyloidosis
AH	heavy chain amyloidosis
ER	endoplasmic reticulum
Ig	immunoglobulin
Fc	fragment, crystallisable
FcγR	Fcγ receptor
Fd	fragment V_H+C_H1
Fab	fragment, antigen binding
Fv	fragment of variable domains
scFv	single chain variable fragment
V_H	variable domain of heavy chain
V_{HH}	variable domain of heavy chain of heavy chain antibodies
V_L	variable domain of light chain
C_H1	first constant domain of heavy chain
C_H2	second constant domain of heavy chain
C_H3	third constant domain of heavy chain
CDR	complementarity determining region
CDR-H3	CDR-3 of the heavy chain variable domain
SDR	specificity determining region
FW	framework
HuCAL	human combinatorial library
wt	wild type
ELISA	enzyme linked immunosorbent assay
SEC	size exclusion chromatography
EM	electron microscope
TEM	transmission electron microscope
NMR	nuclear magnetic resonance
SAXS	small-angle X-ray scattering
CD	circular dichroism
Θ_{MRW}	mean residue ellipticity
AUC	analytical ultracentrifugation
SE	sedimentation equilibrium
SV	sedimentation velocity
PPIase	peptidyl-prolyl-*cis*-*trans* isomerase
Cpr6	Cyclosporin-sensitive Proline Rotamase 6
CypB	Cyclophilin B
PDI	protein disulphide isomerase

Abbreviations

HRP	horse reddish peroxidase
PDB ID	protein data bank identification
RT	room temperature
GdmCl	Guanidinium chloride
ANS	1-anilinonaphtalene-8-sulfonate
APS	ammonium persulfate
TCEP	tris(2-carboxyethyl) phosphine hydrochloride
TEMED	*N,N,N′,N′*-Tetramethylethylendiamin
Tris	tris(hydroxymethyl)-aminomethan
DTT	di-thiothreitol
CM5	carboxymethylated dextran
EDC	N,N'-1-ethyl-3-(3-dimethylaminopropyl)-carbodiimid
EMCH	3,3´-N-(ε-maleimidocaproic acid) hydrazide, trifluoroacetic acid salt
GSH	L-Glutathione reduced
GSSG	L-Glutathione oxidized
IPTG	isopropyl-β-D-thiogalactopyranoside
NHS	*N*-Hydroxysuccinimid
PBS	phosphate buffered saline
PCR	polymerase chain reaction
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecylsulfate polyacrylamide gel electrophoresis
EDTA	ethylenedinitrilotetraacetic acid
LB	Lysogeny Broth
SOB	Super Optimal Broth

References

1. Ellis, R.J. and F.U. Hartl, *Principles of protein folding in the cellular environment.* Current Opinion in Structural Biology, 1999. **9**(1): p. 102-110.
2. Ellis, R.J. and A.P. Minton, *Cell biology: join the crowd.* Nature, 2003. **425**(6953): p. 27-8.
3. Hartl, F.U., *Molecular chaperones in cellular protein folding.* Nature, 1996. **381**(6583): p. 571-9.
4. Walter, S. and J. Buchner, *Molecular Chaperones—Cellular Machines for Protein Folding.* Angewandte Chemie International Edition, 2002. **41**(7): p. 1098-1113.
5. Gething, M.J. and J. Sambrook, *Protein folding in the cell.* Nature, 1992. **355**(6355): p. 33-45.
6. Dobson, C.M., *The structural basis of protein folding and its links with human disease.* Philos Trans R Soc Lond B Biol Sci, 2001. **356**(1406): p. 133-45.
7. Wang, S. and R.J. Kaufman, *The impact of the unfolded protein response on human disease.* J Cell Biol, 2012. **197**(7): p. 857-67.
8. Clore, G.M. and A.M. Gronenborn, *Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy.* Crit Rev Biochem Mol Biol, 1989. **24**(5): p. 479-564.
9. Eisenberg, D. and C.P. Hill, *Protein crystallography: more surprises ahead.* Trends in Biochemical Sciences, 1989. **14**(7): p. 260-264.
10. Smyth, M.S. and J.H. Martin, *x ray crystallography.* Mol Pathol, 2000. **53**(1): p. 8-14.
11. Tzakos, A.G., et al., *NMR TECHNIQUES FOR VERY LARGE PROTEINS AND RNAS IN SOLUTION.* Annual Review of Biophysics and Biomolecular Structure, 2006. **35**(1): p. 319-342.
12. Chothia, C., *Proteins. One thousand families for the molecular biologist.* Nature, 1992. **357**(6379): p. 543-4.
13. Addou, S., et al., *Domain-Based and Family-Specific Sequence Identity Thresholds Increase the Levels of Reliable Protein Function Transfer.* Journal of Molecular Biology, 2009. **387**(2): p. 416-430.
14. Sjolander, K., et al., *Ortholog identification in the presence of domain architecture rearrangement.* Brief Bioinform, 2011. **12**(5): p. 413-22.
15. Dobson, C.M., *Protein-misfolding diseases: Getting out of shape.* Nature, 2002. **418**(6899): p. 729-730.
16. E., C.T., *The Protein folding problem*, in *Mechanism of protein folding*, P.R. H., Editor. 1994: IRL press at Oxford University Press.
17. Denny, R.A., L.K. Gavrin, and E. Saiah, *Recent developments in targeting protein misfolding diseases.* Bioorganic & Medicinal Chemistry Letters, (0).
18. Daggett, V. and A.R. Fersht, *Is there a unifying mechanism for protein folding?* Trends in Biochemical Sciences, 2003. **28**(1): p. 18-25.
19. Ivarsson, Y., et al., *Mechanisms of protein folding.* Eur Biophys J, 2008. **37**(6): p. 721-8.
20. Orengo, C.A., A.E. Todd, and J.M. Thornton, *From protein structure to function.* Current Opinion in Structural Biology, 1999. **9**(3): p. 374-382.
21. Thornton, J.M., et al., *Protein folds, functions and evolution.* Journal of Molecular Biology, 1999. **293**(2): p. 333-342.
22. Tramontano, A., *Of men and machines.* Nat Struct Mol Biol, 2003. **10**(2): p. 87-90.

References

23. Daggett, V. and A. Fersht, *The present view of the mechanism of protein folding.* Nat Rev Mol Cell Biol, 2003. **4**(6): p. 497-502.

24. Baldwin, R.L. and G.D. Rose, *Is protein folding hierarchic? II. Folding intermediates and transition states.* Trends Biochem Sci, 1999. **24**(2): p. 77-83.

25. Baldwin, R.L. and G.D. Rose, *Is protein folding hierarchic? I. Local structure and peptide folding.* Trends Biochem Sci, 1999. **24**(1): p. 26-33.

26. Daggett, V. and A.R. Fersht, *Is there a unifying mechanism for protein folding?* Trends Biochem Sci, 2003. **28**(1): p. 18-25.

27. Dill, K.A. and J.L. MacCallum, *The protein-folding problem, 50 years on.* Science, 2012. **338**(6110): p. 1042-6.

28. Anfinsen, C.B., et al., *THE KINETICS OF FORMATION OF NATIVE RIBONUCLEASE DURING OXIDATION OF THE REDUCED POLYPEPTIDE CHAIN.* Proceedings of the National Academy of Sciences, 1961. **47**(9): p. 1309-1314.

29. Anfinsen, C.B., *Principles that govern the folding of protein chains.* Science, 1973. **181**(4096): p. 223-30.

30. Morris, E.R. and M.S. Searle, *Overview of protein folding mechanisms: experimental and theoretical approaches to probing energy landscapes.* Curr Protoc Protein Sci, 2012. **Chapter 28**: p. Unit 28 2 1-22.

31. Creighton, T.E., *Protein folding*, T.E. Creighton, Freeman, Editor. 1992: New York.

32. Creighton, T.E., *Toward a better understanding of protein folding pathways.* Proc Natl Acad Sci U S A, 1988. **85**(14): p. 5082-6.

33. Levinthal, C., *Are there pathways for protein folding?* J. Chim. Phys. Phys.-Chim. Biol, 1968. **65**: p. 44-45.

34. Gianni, S., et al., *Unifying features in protein-folding mechanisms.* Proceedings of the National Academy of Sciences, 2003. **100**(23): p. 13286-13291.

35. Feige, M.J., L.M. Hendershot, and J. Buchner, *How antibodies fold.* Trends Biochem Sci, 2010. **35**(4): p. 189-98.

36. Koga, N., et al., *Principles for designing ideal protein structures.* Nature, 2012. **491**(7423): p. 222-227.

37. Invernizzi, G., et al., *Protein aggregation: Mechanisms and functional consequences.* The International Journal of Biochemistry & Cell Biology, 2012. **44**(9): p. 1541-1554.

38. Soto, C., L. Estrada, and J. Castilla, *Amyloids, prions and the inherent infectious nature of misfolded protein aggregates.* Trends in Biochemical Sciences, 2006. **31**(3): p. 150-155.

39. Ross, C.A. and M.A. Poirier, *Protein aggregation and neurodegenerative disease.* Nat Med, 2004. **10 Suppl**: p. S10-7.

40. Blancas-Mejía, L.M. and M. Ramirez-Alvarado, *Systemic Amyloidoses.* Annual Review of Biochemistry, 2013. **82**(1): p. null.

41. Sipe, J.D., et al., *Amyloid fibril protein nomenclature: 2010 recommendations from the nomenclature committee of the International Society of Amyloidosis.* Amyloid, 2010. **17**(3-4): p. 101-4.

42. Dobson, C.M., *Principles of protein folding, misfolding and aggregation.* Seminars in Cell & Developmental Biology, 2004. **15**(1): p. 3-16.

43. Wilson, M.R., J.J. Yerbury, and S. Poon, *Potential roles of abundant extracellular chaperones in the control of amyloid formation and toxicity.* Mol Biosyst, 2008. **4**(1): p. 42-52.

44. Soto, C., *Protein misfolding and disease; protein refolding and therapy.* FEBS Lett, 2001. **498**(2-3): p. 204-7.

45. Sunde, M. and C. Blake, *The structure of amyloid fibrils by electron microscopy and X-ray diffraction.* Adv Protein Chem, 1997. **50**: p. 123-59.

46. Rochet, J.C. and P.T. Lansbury, Jr., *Amyloid fibrillogenesis: themes and variations.* Curr Opin Struct Biol, 2000. **10**(1): p. 60-8.

47. Jahn, T.R. and S.E. Radford, *The Yin and Yang of protein folding.* FEBS Journal, 2005. **272**(23): p. 5962-5970.

48. Bedrood, S., et al., *Fibril structure of human islet amyloid polypeptide.* J Biol Chem, 2012. **287**(8): p. 5235-41.

49. Prusiner, S.B., *Scrapie prions.* Annu Rev Microbiol, 1989. **43**: p. 345-74.

50. Haass, C., A.Y. Hung, and D.J. Selkoe, *Processing of beta-amyloid precursor protein in microglia and astrocytes favors an internal localization over constitutive secretion.* J Neurosci, 1991. **11**(12): p. 3783-93.

51. Obici, L., et al., *Clinical aspects of systemic amyloid diseases.* Biochim Biophys Acta, 2005. **1753**(1): p. 11-22.

52. Fernandez-Escamilla, A.-M., et al., *Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins.* Nat Biotech, 2004. **22**(10): p. 1302-1306.

53. Pastor, M.T., A. Esteras-Chopo, and M.L. de la Paz, *Design of model systems for amyloid formation: lessons for prediction and inhibition.* Current Opinion in Structural Biology, 2005. **15**(1): p. 57-63.

54. López de la Paz, M., et al., *De novo designed peptide-based amyloid fibrils.* Proceedings of the National Academy of Sciences, 2002. **99**(25): p. 16052-16057.

55. Ventura, S., et al., *Short amino acid stretches can mediate amyloid formation in globular proteins: The Src homology 3 (SH3) case.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(19): p. 7258-7263.

56. Feige, M.J., et al., *The structure of a folding intermediate provides insight into differences in immunoglobulin amyloidogenicity.* Proc Natl Acad Sci U S A, 2008. **105**(36): p. 13373-8.

57. Klabunde, T., et al., *Rational design of potent human transthyretin amyloid disease inhibitors.* Nat Struct Biol, 2000. **7**(4): p. 312-21.

58. Martins, P.M., *True and apparent inhibition of amyloid fibril formation.* Prion, 2013. **7**(2): p. 136-9.

59. Tanford, C., *Protein Denaturation: Part C. Theoretical Models for The Mechanism of Denaturation*, in *Advances in Protein Chemistry*, J.T.E. C.B. Anfinsen and M.R. Frederic, Editors. 1970, Academic Press. p. 1-95.

60. Baldwin, R.L., *Energetics of protein folding.* J Mol Biol, 2007. **371**(2): p. 283-301.

61. Dill, K.A., *Dominant forces in protein folding.* Biochemistry, 1990. **29**(31): p. 7133-7155.

62. Baldwin, R.L., *Weak interactions in Protein Folding: Hydrophobic free energy, van der Waals interactions, Peptide Hydrogen Bonds, and Peptide Solvation*, in *Protein Folding Handbook*, J. Buchner, Kiefhaber, T., Editor. 2005: Wiley-VCH Verlag GmbH & Co. KGaA.

63. Kauzmann, W., *Some Factors in the Interpretation of Protein Denaturation*, in *Advances in Protein Chemistry*, M.L.A.K.B. C.B. Anfinsen and T.E. John, Editors. 1959, Academic Press. p. 1-63.

64. Baldwin, R.L., *Properties of hydrophobic free energy found by gas–liquid transfer.* Proceedings of the National Academy of Sciences, 2013. **110**(5): p. 1670-1673.

65. Jaenicke, R., *Folding and association versus misfolding and aggregation of proteins.* Philos Trans R Soc Lond B Biol Sci, 1995. **348**(1323): p. 97-105.

66. Pace, C.N., *Conformational stability of globular proteins.* Trends in Biochemical Sciences, 1990. **15**(1): p. 14-17.

67. Garcia-Mira, M.M. and F.X. Schmid, *Key Role of Coulombic Interactions for the Folding Transition State of the Cold Shock Protein.* Journal of Molecular Biology, 2006. **364**(3): p. 458-468.

68. Sanchez-Ruiz, J.M. and G.I. Makhatadze, *To charge or not to charge?* Trends in Biotechnology, 2001. **19**(4): p. 132-135.

69. Wunderlich, M., A. Martin, and F.X. Schmid, *Stabilization of the Cold Shock Protein CspB from Bacillus subtilis by Evolutionary Optimization of Coulombic Interactions.* Journal of Molecular Biology, 2005. **347**(5): p. 1063-1076.

70. lt, et al., *Surface electrostatic interactions contribute little to stability of barnase.* Journal of Molecular Biology, 1991. **220**(3): p. 779-788.

71. Sun, D.P., et al., *Contributions of engineered surface salt bridges to the stability of T4 lysozyme determined by directed mutagenesis.* Biochemistry, 1991. **30**(29): p. 7142-7153.

72. Fersht, A., *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding* 1999: W.H. Freeman and Company.

73. Dyer, R.B., *Ultrafast and downhill protein folding.* Current Opinion in Structural Biology, 2007. **17**(1): p. 38-47.

74. Liu, Z. and H.S. Chan, *Desolvation is a Likely Origin of Robust Enthalpic Barriers to Protein Folding.* Journal of Molecular Biology, 2005. **349**(4): p. 872-889.

75. Plaxco, K.W., K.T. Simons, and D. Baker, *Contact order, transition state placement and the refolding rates of single domain proteins.* Journal of Molecular Biology, 1998. **277**(4): p. 985-994.

76. Wetlaufer, D.B., *Nucleation, Rapid Folding, and Globular Intrachain Regions in Proteins.* Proceedings of the National Academy of Sciences, 1973. **70**(3): p. 697-701.

77. Kim, P.S. and R.L. Baldwin, *Specific Intermediates in the Folding Reactions of Small Proteins and the Mechanism of Protein Folding.* Annual Review of Biochemistry, 1982. **51**(1): p. 459-489.

78. Karplus, M. and D.L. Weaver, *Protein-folding dynamics.* Nature, 1976. **260**(5550): p. 404-6.

79. Karplus, M. and D.L. Weaver, *Protein folding dynamics: the diffusion-collision model and experimental data.* Protein Sci, 1994. **3**(4): p. 650-68.

80. Tanford, C., *Contribution of Hydrophobic Interactions to the Stability of the Globular Conformation of Proteins.* Journal of the American Chemical Society, 1962. **84**(22): p. 4240-4247.

81. Ptitsyn, O.B., *How the molten globule became.* Trends Biochem Sci, 1995. **20**(9): p. 376-9.

82. Creighton, T.E., *How important is the molten globule for correct protein folding?* Trends in Biochemical Sciences, 1997. **22**(1): p. 6-10.

83. Bieri, O. and T. Kiefhaber, *Elementary steps in protein folding.* Biol Chem, 1999. **380**(7-8): p. 923-9.

84. Fersht, A.R., *Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications.* Proc Natl Acad Sci U S A, 1995. **92**(24): p. 10869-73.

85. Otzen, D.E., et al., *Structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding.* Proceedings of the National Academy of Sciences, 1994. **91**(22): p. 10422-10425.

86. Fersht, A.R., *Nucleation mechanisms in protein folding.* Current Opinion in Structural Biology, 1997. **7**(1): p. 3-9.

87. Bryngelson, J.D., et al., *Funnels, pathways, and the energy landscape of protein folding: a synthesis.* Proteins, 1995. **21**(3): p. 167-95.

88. Dill, K.A. and H.S. Chan, *From Levinthal to pathways to funnels.* Nat Struct Biol, 1997. **4**(1): p. 10-9.

89. Onuchic, J.N. and P.G. Wolynes, *Theory of protein folding.* Curr Opin Struct Biol, 2004. **14**(1): p. 70-5.

90. Friel, C.T., G.S. Beddard, and S.E. Radford, *Switching Two-state to Three-state Kinetics in the Helical Protein Im9 via the Optimisation of Stabilising Non-native Interactions by Design.* Journal of Molecular Biology, 2004. **342**(1): p. 261-273.

91. Mogensen, J.E., et al., *Elimination of a Misfolded Folding Intermediate by a Single Point Mutation.* Biochemistry, 2004. **43**(12): p. 3357-3367.

92. Conrad, J.C. and P.J. Flory, *Moments and Distribution Functions for Polypeptide Chains. Poly-L-alanine.* Macromolecules, 1976. **9**(1): p. 41-47.

93. Dobson, C.M., *Unfolded proteins, compact states and molten globules: Current Opinion in Structural Biology 1992, 2:6–12.* Current Opinion in Structural Biology, 1992. **2**(1): p. 6-12.

94. Shortle, D., *The denatured state (the other half of the folding equation) and its role in protein stability.* The FASEB Journal, 1996. **10**(1): p. 27-34.

95. Wong, K.-B., et al., *Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding.* Journal of Molecular Biology, 2000. **296**(5): p. 1257-1282.

96. Pace, C.N., *Determination and analysis of urea and guanidine hydrochloride denaturation curves.* Methods Enzymol, 1986. **131**: p. 266-80.

97. Matthews, C.R., *Pathways of protein folding.* Annu Rev Biochem, 1993. **62**: p. 653-83.

98. Brockwell, D.J., D.A. Smith, and S.E. Radford, *Protein folding mechanisms: new methods and emerging ideas.* Current Opinion in Structural Biology, 2000. **10**(1): p. 16-25.

99. Frauenfelder, H., G.A. Petsko, and D. Tsernoglou, *Temperature-dependent X-ray diffraction as a probe of protein structural dynamics.* Nature, 1979. **280**(5723): p. 558-63.

100. Austin, R.H., et al., *Dynamics of carbon monoxide binding by heme proteins.* Science, 1973. **181**(4099): p. 541-3.

101. Eisenmesser, E.Z., et al., *Intrinsic dynamics of an enzyme underlies catalysis.* Nature, 2005. **438**(7064): p. 117-21.

102. Henzler-Wildman, K.A., et al., *Intrinsic motions along an enzymatic reaction trajectory.* Nature, 2007. **450**(7171): p. 838-44.

103. Dunker, A.K., et al., *Function and structure of inherently disordered proteins.* Curr Opin Struct Biol, 2008. **18**(6): p. 756-64.

104. Uversky, V.N., C.J. Oldfield, and A.K. Dunker, *Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling.* J Mol Recognit, 2005. **18**(5): p. 343-84.

105. Fink, A.L., *Natively unfolded proteins.* Curr Opin Struct Biol, 2005. **15**(1): p. 35-41.

106. Dill, K.A. and D. Shortle, *Denatured states of proteins.* Annu Rev Biochem, 1991. **60**: p. 795-825.

107. Creighton, T.E., *The protein folding problem*, in *Mechanisms of protein folding*, H.R. Pain, Editor. 1994.

108. Bowler, B.E., *Residual structure in unfolded proteins.* Curr Opin Struct Biol, 2012. **22**(1): p. 4-13.

109. Creighton, T.E., N.J. Darby, and J. Kemmink, *The roles of partly folded intermediates in protein folding.* FASEB J, 1996. **10**(1): p. 110-8.

110. Lee, J.C. and S.N. Timasheff, *Partial specific volumes and interactions with solvent components of proteins in guanidine hydrochloride.* Biochemistry, 1974. **13**(2): p. 257-65.

111. Tanford, C., *Isothermal Unfolding of Globular Proteins in Aqueous Urea Solutions.* Journal of the American Chemical Society, 1964. **86**(10): p. 2050-2059.

112. Watlaufer, D.B., et al., *Nonpolar Group Participation in the Denaturation of Proteins by Urea and Guanidinium Salts. Model Compound Studies.* Journal of the American Chemical Society, 1964. **86**(3): p. 508-514.

113. Schellman, J.A., *Solvent denaturation.* Biopolymers, 1978. **17**(5): p. 1305-1322.

114. Schellman, J.A., *Fifty years of solvent denaturation.* Biophysical Chemistry, 2002. **96**(2–3): p. 91-101.

115. Moglich, A., F. Krieger, and T. Kiefhaber, *Molecular basis for the effect of urea and guanidinium chloride on the dynamics of unfolded polypeptide chains.* J Mol Biol, 2005. **345**(1): p. 153-62.

116. Matouschek, A., et al., *Extrapolation to water of kinetic and equilibrium data for the unfolding of barnase in urea solutions.* Protein Eng, 1994. **7**(9): p. 1089-95.

117. Segawa, S. and M. Sugihara, *Characterization of the transition state of lysozyme unfolding. I. Effect of protein-solvent interactions on the transition state.* Biopolymers, 1984. **23**(11 Pt 2): p. 2473-88.

118. Jahn, T.R., et al., *Amyloid formation under physiological conditions proceeds via a native-like folding intermediate.* Nat Struct Mol Biol, 2006. **13**(3): p. 195-201.

119. Simpson, E.R., E.M. Herold, and J. Buchner, *The folding pathway of the antibody V(L) domain.* J Mol Biol, 2009. **392**(5): p. 1326-38.

120. Kameda, A., et al., *Nuclear magnetic resonance characterization of the refolding intermediate of beta2-microglobulin trapped by non-native prolyl peptide bond.* J Mol Biol, 2005. **348**(2): p. 383-97.

121. Ikeguchi, M., et al., *Evidence for identity between the equilibrium unfolding intermediate and a transient folding intermediate: a comparative study of the folding reactions of alpha-lactalbumin and lysozyme.* Biochemistry, 1986. **25**(22): p. 6965-72.

122. Kuwajima, K., et al., *Transient intermediates in the folding of dihydrofolate reductase as detected by far-ultraviolet circular dichroism spectroscopy.* Biochemistry, 1991. **30**(31): p. 7693-703.

123. Arai, M. and K. Kuwajima, *Role of the molten globule state in protein folding.* Adv Protein Chem, 2000. **53**: p. 209-82.

124. Ptitsyn, O.B., *Molten globule and protein folding.* Adv Protein Chem, 1995. **47**: p. 83-229.

125. Uversky, V.N., *Natively unfolded proteins: a point where biology waits for physics.* Protein Sci, 2002. **11**(4): p. 739-56.

126. Fink, A.L., *Compact intermediate states in protein folding.* Annu Rev Biophys Biomol Struct, 1995. **24**: p. 495-522.

127. Fujiwara, K., et al., *Folding-unfolding equilibrium and kinetics of equine beta-lactoglobulin: equivalence between the equilibrium molten globule state and a burst-phase folding intermediate.* Biochemistry, 1999. **38**(14): p. 4455-63.

128. Arai, M. and K. Kuwajima, *Rapid formation of a molten globule intermediate in refolding of alpha-lactalbumin.* Fold Des, 1996. **1**(4): p. 275-87.

129. Baldwin, R.L., *On-pathway versus off-pathway folding intermediates.* Fold Des, 1996. **1**(1): p. R1-8.

130. Bhattacharyya, S. and R. Varadarajan, *Packing in molten globules and native states.* Curr Opin Struct Biol, 2013. **23**(1): p. 11-21.

131. Uversky, V.N. and O.B. Ptitsyn, *Further evidence on the equilibrium "pre-molten globule state": four-state guanidinium chloride-induced unfolding of carbonic anhydrase B at low temperature.* J Mol Biol, 1996. **255**(1): p. 215-28.

132. Buchner, J., et al., *Alternatively folded states of an immunoglobulin.* Biochemistry, 1991. **30**(28): p. 6922-9.

133. Thies, M.J., et al., *The alternatively folded state of the antibody C(H)3 domain.* J Mol Biol, 2001. **309**(5): p. 1077-85.

134. Lilie, H. and J. Buchner, *Domain interactions stabilize the alternatively folded state of an antibody Fab fragment.* FEBS Lett, 1995. **362**(1): p. 43-6.

135. Uversky, V.N. and Y. Goto, *Acid denaturation and anion-induced folding of globular proteins: multitude of equilibium partially folded intermediates.* Curr Protein Pept Sci, 2009. **10**(5): p. 447-55.

136. Goto, Y., N. Takahashi, and A.L. Fink, *Mechanism of acid-induced folding of proteins.* Biochemistry, 1990. **29**(14): p. 3480-8.

137. Goto, Y., L.J. Calciano, and A.L. Fink, *Acid-induced folding of proteins.* Proc Natl Acad Sci U S A, 1990. **87**(2): p. 573-7.

138. Uversky, V.N., et al., *Anion-induced folding of Staphylococcal nuclease: characterization of multiple equilibrium partially folded intermediates.* J Mol Biol, 1998. **278**(4): p. 879-94.

139. Fink, A.L., K.A. Oberg, and S. Seshadri, *Discrete intermediates versus molten globule models for protein folding: characterization of partially folded intermediates of apomyoglobin.* Fold Des, 1998. **3**(1): p. 19-25.

140. Buchner, J., R. Rudolph, and H. Lilie, *Intradomain disulfide bonds impede formation of the alternatively folded state of antibody chains.* J Mol Biol, 2002. **318**(3): p. 829-36.

141. Welfle, K., et al., *Conformation, pH-induced conformational changes, and thermal unfolding of anti-p24 (HIV-1) monoclonal antibody CB4-1 and its Fab and Fc fragments.* Biochim Biophys Acta, 1999. **1431**(1): p. 120-31.

142. Vlasov, A.P., Z.I. Kravchuk, and S.P. Martsev, *[Non-native conformational states of immunoglobulins: thermodynamic and functional analysis of rabbit IgG].* Biokhimiia, 1996. **61**(2): p. 212-35.

143. Kanmert, D., et al., *Thermal induction of an alternatively folded state in human IgG-Fc.* Biochemistry, 2011. **50**(6): p. 981-8.

144. Williams, A.F. and A.N. Barclay, *The immunoglobulin superfamily--domains for cell surface recognition.* Annu Rev Immunol, 1988. **6**: p. 381-405.

145. Harpaz, Y. and C. Chothia, *Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains.* J Mol Biol, 1994. **238**(4): p. 528-39.

146. Xu, Z., et al., *Immunoglobulin class-switch DNA recombination: induction, targeting and beyond.* Nat Rev Immunol, 2012. **12**(7): p. 517-531.

147. Schroeder, H.W., Jr. and L. Cavacini, *Structure and function of immunoglobulins.* J Allergy Clin Immunol, 2010. **125**(2 Suppl 2): p. S41-52.

148. Hendershot, L.M.a.S., R., *Immunoglobulin assembly and secretion*, in *Molecular Biology of B cells*, F.a.H. Alt, T., Editor. 2004: Academic Pr Inc.

149. Arnold, J.N., et al., *The impact of glycosylation on the biological function and structure of human immunoglobulins.* Annu Rev Immunol, 2007. **25**: p. 21-50.

150. Radaev, S. and P. Sun, *Recognition of immunoglobulins by Fcγ receptors.* Molecular Immunology, 2002. **38**(14): p. 1073-1083.

151. Kanda, Y., et al., *Comparison of biological activity among nonfucosylated therapeutic IgG1 antibodies with three different N-linked Fc oligosaccharides: the high-mannose, hybrid, and complex types.* Glycobiology, 2007. **17**(1): p. 104-18.

152. Natsume, A., et al., *Fucose removal from complex-type oligosaccharide enhances the antibody-dependent cellular cytotoxicity of single-gene-encoded antibody comprising a single-chain antibody linked the antibody constant region.* J Immunol Methods, 2005. **306**(1-2): p. 93-103.

153. Dwek, R.A., A.C. Lellouch, and M.R. Wormald, *Glycobiology: 'the function of sugar in the IgG molecule'.* J Anat, 1995. **187 (Pt 2)**: p. 279-92.

154. Feige, M.J., et al., *Structure of the murine unglycosylated IgG1 Fc fragment.* J Mol Biol, 2009. **391**(3): p. 599-608.

155. del Val, I.J., C. Kontoravdi, and J.M. Nagy, *Towards the implementation of quality by design to the production of therapeutic monoclonal antibodies with desired glycosylation patterns.* Biotechnology Progress, 2010. **26**(6): p. 1505-1527.

156. Arnold, J.N., et al., *Mannan binding lectin and its interaction with immunoglobulins in health and in disease.* Immunol Lett, 2006. **106**(2): p. 103-10.

157. Bork, P., L. Holm, and C. Sander, *The immunoglobulin fold. Structural classification, sequence patterns and common core.* J Mol Biol, 1994. **242**(4): p. 309-20.

158. Chatellier, J., et al., *Functional mapping of conserved residues located at the VL and VH domain interface of a Fab.* J Mol Biol, 1996. **264**(1): p. 1-6.

159. Poljak, R.J., et al., *Three-dimensional structure of the Fab' fragment of a human immunoglobulin at 2,8-A resolution.* Proc Natl Acad Sci U S A, 1973. **70**(12): p. 3305-10.

160. Chothia, C., et al., *Domain association in immunoglobulin molecules. The packing of variable domains.* J Mol Biol, 1985. **186**(3): p. 651-63.

161. Novotny, J. and E. Haber, *Structural invariants of antigen binding: comparison of immunoglobulin VL-VH and VL-VL domain dimers.* Proc Natl Acad Sci U S A, 1985. **82**(14): p. 4592-6.

162. Hamers-Casterman, C., et al., *Naturally occurring antibodies devoid of light chains.* Nature, 1993. **363**(6428): p. 446-8.

163. Ewert, S., et al., *Biophysical properties of camelid V(HH) domains compared to those of human V(H)3 domains.* Biochemistry, 2002. **41**(11): p. 3628-36.

164. Wang, N., et al., *Conserved amino acid networks involved in antibody variable domain interactions.* Proteins, 2009. **76**(1): p. 99-114.

165. Choulier, L., et al., *Covariance analysis of protein families: the case of the variable domains of antibodies.* Proteins, 2000. **41**(4): p. 475-84.

166. Chothia, C., I. Gelfand, and A. Kister, *Structural determinants in the sequences of immunoglobulin variable domain.* J Mol Biol, 1998. **278**(2): p. 457-79.

167. Kabat, E.A. and T.T. Wu, *Identical V region amino acid sequences and segments of sequences in antibodies of different specificities. Relative contributions of VH and VL genes, minigenes, and complementarity-determining regions to binding of antibody-combining sites.* J Immunol, 1991. **147**(5): p. 1709-19.

168. Honegger, A. and A. Pluckthun, *The influence of the buried glutamine or glutamate residue in position 6 on the structure of immunoglobulin variable domains.* J Mol Biol, 2001. **309**(3): p. 687-99.

169. Honegger, A., et al., *The influence of the framework core residues on the biophysical properties of immunoglobulin heavy chain variable domains.* Protein Eng Des Sel, 2009. **22**(3): p. 121-34.

170. Rothlisberger, D., A. Honegger, and A. Pluckthun, *Domain interactions in the Fab fragment: a comparative evaluation of the single-chain Fv and Fab format engineered with variable domains of different stability.* J Mol Biol, 2005. **347**(4): p. 773-89.

171. Hugo, N., et al., *Functional aspects of co-variant surface charges in an antibody fragment.* Protein Sci, 2002. **11**(11): p. 2697-705.

172. Tan, P.H., B.M. Sandmaier, and P.S. Stayton, *Contributions of a highly conserved VH/VL hydrogen bonding interaction to scFv folding stability and refolding efficiency.* Biophys J, 1998. **75**(3): p. 1473-82.

173. Halaby, D.M., A. Poupon, and J. Mornon, *The immunoglobulin fold family: sequence analysis and 3D structure comparisons.* Protein Eng, 1999. **12**(7): p. 563-71.

174. Goto, Y., T. Azuma, and K. Hamaguchi, *Refolding of the immunoglobulin light chain.* J Biochem, 1979. **85**(6): p. 1427-38.

175. Goto, Y. and K. Hamaguchi, *Unfolding and refolding of the reduced constant fragment of the immunoglobulin light chain. Kinetic role of the intrachain disulfide bond.* J Mol Biol, 1982. **156**(4): p. 911-26.

176. Goto, Y. and K. Hamaguchi, *Unfolding and refolding of the constant fragment of the immunoglobulin light chain.* J Mol Biol, 1982. **156**(4): p. 891-910.

177. Isenman, D.E., D. Lancet, and I. Pecht, *Folding pathways of immunoglobulin domains. The folding kinetics of the Cgamma3 domain of human IgG1.* Biochemistry, 1979. **18**(15): p. 3327-36.

178. Lang, K., F.X. Schmid, and G. Fischer, *Catalysis of protein folding by prolyl isomerase.* Nature, 1987. **329**(6136): p. 268-70.

179. Lilie, H., et al., *Prolyl isomerases catalyze antibody folding in vitro.* Protein Sci, 1993. **2**(9): p. 1490-6.

180. Thies, M.J., et al., *Folding and association of the antibody domain CH3: prolyl isomerization preceeds dimerization.* J Mol Biol, 1999. **293**(1): p. 67-79.

181. Feige, M.J., et al., *An unfolded CH1 domain controls the assembly and secretion of IgG antibodies.* Mol Cell, 2009. **34**(5): p. 569-79.

182. Buxbaum, J.N., et al., *Monoclonal immunoglobulin deposition disease: light chain and light and heavy chain deposition diseases and their relation to light chain amyloidosis. Clinical features, immunopathology, and molecular analysis.* Ann Intern Med, 1990. **112**(6): p. 455-64.

183. Buxbaum, J. and G. Gallo, *Nonamyloidotic monoclonal immunoglobulin deposition disease. Light-chain, heavy-chain, and light- and heavy-chain deposition diseases.* Hematol Oncol Clin North Am, 1999. **13**(6): p. 1235-48.

184. Hurle, M.R., et al., *A role for destabilizing amino acid replacements in light-chain amyloidosis.* Proc Natl Acad Sci U S A, 1994. **91**(12): p. 5446-50.

185. Kelly, J.W., *The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways.* Curr Opin Struct Biol, 1998. **8**(1): p. 101-6.

186. Uversky, V.N. and A.L. Fink, *Conformational constraints for amyloid fibrillation: the importance of being unfolded.* Biochim Biophys Acta, 2004. **1698**(2): p. 131-53.

187. Baden, E.M., et al., *Structural insights into the role of mutations in amyloidogenesis.* J Biol Chem, 2008. **283**(45): p. 30950-6.

188. Connors, L.H., et al., *Heterogeneity in primary structure, post-translational modifications, and germline gene usage of nine full-length amyloidogenic kappa1 immunoglobulin light chains.* Biochemistry, 2007. **46**(49): p. 14259-71.

189. Poshusta, T.L., et al., *Mutations in specific structural regions of immunoglobulin light chains are associated with free light chain levels in patients with AL amyloidosis.* PLoS One, 2009. **4**(4): p. e5169.

190. Stevens, P.W., et al., *Recombinant immunoglobulin variable domains generated from synthetic genes provide a system for in vitro characterization of light-chain amyloid proteins.* Protein Sci, 1995. **4**(3): p. 421-32.

191. Feige, M.J., et al., *Influence of the internal disulfide bridge on the folding pathway of the CL antibody domain.* J Mol Biol, 2007. **365**(4): p. 1232-44.

192. Feige, M.J., et al., *Dissecting the alternatively folded state of the antibody Fab fragment.* J Mol Biol, 2010. **399**(5): p. 719-30.

193. Beck, A., et al., *Strategies and challenges for the next generation of therapeutic antibodies.* Nat Rev Immunol, 2010. **10**(5): p. 345-52.

194. Jager, M. and A. Pluckthun, *Folding and assembly of an antibody Fv fragment, a heterodimer stabilized by antigen.* J Mol Biol, 1999. **285**(5): p. 2005-19.

195. Ohage, E. and B. Steipe, *Intrabody construction and expression. I. The critical role of VL domain stability.* J Mol Biol, 1999. **291**(5): p. 1119-28.

196. Stevens, F.J., *Four structural risk factors identify most fibril-forming kappa light chains.* Amyloid, 2000. **7**(3): p. 200-11.

197. Ewert, S., A. Honegger, and A. Pluckthun, *Stability improvement of antibodies for extracellular and intracellular applications: CDR grafting to stable frameworks and structure-based framework engineering.* Methods, 2004. **34**(2): p. 184-99.

198. Lefranc, M.P., et al., *IMGT, the international ImMunoGeneTics information system.* Nucleic Acids Res, 2005. **33**(Database issue): p. D593-7.

199. Ewert, S., et al., *Biophysical properties of human antibody variable domains.* J Mol Biol, 2003. **325**(3): p. 531-53.

200. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4.* Nature, 1970. **227**(5259): p. 680-5.

201. Fairbanks, G., T.L. Steck, and D.F. Wallach, *Electrophoretic analysis of the major polypeptides of the human erythrocyte membrane.* Biochemistry, 1971. **10**(13): p. 2606-17.

202. Feige, M.J., S. Walter, and J. Buchner, *Folding mechanism of the CH2 antibody domain.* J Mol Biol, 2004. **344**(1): p. 107-18.

203. Thies, M.J. and F. Pirkl, *Chromatographic purification of the C(H)2 domain of the monoclonal antibody MAK33.* J Chromatogr B Biomed Sci Appl, 2000. **737**(1-2): p. 63-9.

204. Kelly, S.M., T.J. Jess, and N.C. Price, *How to study proteins by circular dichroism.* Biochim Biophys Acta, 2005. **1751**(2): p. 119-39.

205. Schmid, F.X., *Spectroscopic techniques to study protein folding and stability*, in *The Protein Folding Handbook*, J. Buchner, Kiefhaber, T., Editor. 2005: Wiley-VCH Verlag GmbH & Co. KGaA.

206. Bolen, D.W. and M.M. Santoro, *Unfolding free energy changes determined by the linear extrapolation method. 2. Incorporation of delta G degrees N-U values in a thermodynamic cycle.* Biochemistry, 1988. **27**(21): p. 8069-74.

207. Cole, J.L., et al., *Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium.* Methods Cell Biol, 2008. **84**: p. 143-79.

208. Schanda, P., H. Van Melckebeke, and B. Brutscher, *Speeding up three-dimensional protein NMR experiments to a few minutes.* J Am Chem Soc, 2006. **128**(28): p. 9042-3.

209. Schanda, P. and B. Brutscher, *Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds.* J Am Chem Soc, 2005. **127**(22): p. 8014-5.

210. Buckel, P., et al., *Cloning and nucleotide sequence of heavy- and light-chain cDNAs from a creatine-kinase-specific monoclonal antibody.* Gene, 1987. **51**(1): p. 13-9.

211. Wilkinson, I.C., et al., *High resolution NMR-based model for the structure of a scFv-IL-1beta complex: potential for NMR as a key tool in therapeutic antibody design and development.* J Biol Chem, 2009. **284**(46): p. 31928-35.

212. Eigenbrot, C., et al., *X-ray structures of the antigen-binding domains from three variants of humanized anti-p185HER2 antibody 4D5 and comparison with molecular modeling.* J Mol Biol, 1993. **229**(4): p. 969-95.

213. Cardoso, R.M., et al., *Structural basis of enhanced binding of extended and helically constrained peptide epitopes of the broadly neutralizing HIV-1 antibody 4E10.* J Mol Biol, 2007. **365**(5): p. 1533-44.

214. Semisotnov, G.V., et al., *Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe.* Biopolymers, 1991. **31**(1): p. 119-28.

215. Stryer, L., *The interaction of a naphthalene dye with apomyoglobin and apohemoglobin. A fluorescent probe of non-polar binding sites.* J Mol Biol, 1965. **13**(2): p. 482-95.

216. Khurana, R., et al., *Partially folded intermediates as critical precursors of light chain amyloid fibrils and amorphous aggregates.* Biochemistry, 2001. **40**(12): p. 3525-35.

217. Martsev, S.P., et al., *Amyloid fibril formation of the mouse V(L) domain at acidic pH.* Biochemistry, 2002. **41**(10): p. 3389-95.

218. Vassilenko, K.S. and V.N. Uversky, *Native-like secondary structure of molten globules.* Biochim Biophys Acta, 2002. **1594**(1): p. 168-77.

219. Kuznetsova, I.M., et al., *Effect of self-association on the structural organization of partially folded proteins: inactivated actin.* Biophys J, 1999. **77**(5): p. 2788-800.

220. Santra, M.K. and D. Panda, *Acid-induced loss of functional properties of bacterial cell division protein FtsZ: evidence for an alternative conformation at acidic pH.* Proteins, 2007. **67**(1): p. 177-88.

221. Nath, D. and M. Rao, *Acid-induced partly folded conformation resembling a molten globule state of xylanase from an alkalothermophilic Bacillus sp.* Biochem Biophys Res Commun, 2001. **288**(5): p. 1218-22.

222. Martsev, S.P., et al., *Antiferritin single-chain Fv fragment is a functional protein with properties of a partially structured state: comparison with the completely folded V(L) domain.* Biochemistry, 2000. **39**(27): p. 8047-57.

223. Abramov, V.M., et al., *Structural and functional similarity between Yersinia pestis capsular protein Caf1 and human interleukin-1 beta.* Biochemistry, 2001. **40**(20): p. 6076-84.

224. Perczel, A., et al., *Convex constraint analysis: a natural deconvolution of circular dichroism curves of proteins.* Protein Eng, 1991. **4**(6): p. 669-79.

225. Thies, M.J., *Struktur, Stabilität und Faltung von Antikörperdomänen*, in *Biotechnologie, Department Chemie*2002, Technische Universität München.

226. Tian, F., et al., *Spectroscopic evaluation of the stabilization of humanized monoclonal antibodies in amino acid formulations.* International Journal of Pharmaceutics, 2007. **335**(1–2): p. 20-31.

227. Street, T.O., D.W. Bolen, and G.D. Rose, *A molecular mechanism for osmolyte-induced protein stability.* Proc Natl Acad Sci U S A, 2006. **103**(38): p. 13997-4002.

228. Rendic, D.W., I. and Paschinger, K., *The Glycosylation Capacity of Insect Cells.* Croatic Chemica Acta, 2008. **81**(1): p. 7-21.

229. Wormald, M.R. and R.A. Dwek, *Glycoproteins: glycan presentation and protein-fold stability.* Structure, 1999. **7**(7): p. R155-60.

230. Wang, C., et al., *Influence of the carbohydrate moiety on the stability of glycoproteins.* Biochemistry, 1996. **35**(23): p. 7299-307.

231. Shental-Bechor, D. and Y. Levy, *Folding of glycoproteins: toward understanding the biophysics of the glycosylation code.* Curr Opin Struct Biol, 2009. **19**(5): p. 524-33.

232. Ionescu, R.M., et al., *Contribution of variable domains to the stability of humanized IgG1 monoclonal antibodies.* J Pharm Sci, 2008. **97**(4): p. 1414-26.

233. Alexandrescu, A.T., et al., *Structure and dynamics of the acid-denatured molten globule state of alpha-lactalbumin: a two-dimensional NMR study.* Biochemistry, 1993. **32**(7): p. 1707-18.

234. Improta, S., et al., *Probing protein structure by solvent perturbation of NMR spectra. Photochemically induced dynamic nuclear polarization and paramagnetic perturbation techniques applied to the study of the molten globule state of alpha-lactalbumin.* Eur J Biochem, 1995. **227**(1-2): p. 87-96.

235. Morea, V., et al., *Conformations of the third hypervariable region in the VH domain of immunoglobulins.* J Mol Biol, 1998. **275**(2): p. 269-94.

236. Igawa, T., et al., *Engineering the variable region of therapeutic IgG antibodies.* MAbs, 2011. **3**(3): p. 243-52.

237. Worn, A. and A. Pluckthun, *Stability engineering of antibody single-chain Fv fragments.* J Mol Biol, 2001. **305**(5): p. 989-1010.

238. Knappik, A., et al., *Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides.* J Mol Biol, 2000. **296**(1): p. 57-86.

239. Carter, P., et al., *Humanization of an anti-p185HER2 antibody for human cancer therapy.* Proc Natl Acad Sci U S A, 1992. **89**(10): p. 4285-9.

240. Nieba, L., et al., *Disrupting the hydrophobic patches at the antibody variable/constant domain interface: improved in vivo folding and physical characterization of an engineered scFv fragment.* Protein Eng, 1997. **10**(4): p. 435-44.

241. Kugler, M., et al., *Stabilization and humanization of a single-chain Fv antibody fragment specific for human lymphocyte antigen CD19 by designed point mutations and CDR-grafting onto a human framework.* Protein Eng Des Sel, 2009. **22**(3): p. 135-47.

242. Honegger, A., *Engineering antibodies for stability and efficient folding.* Handb Exp Pharmacol, 2008(181): p. 47-68.

243. Weidenhaupt, M., et al., *Functional mapping of conserved, surface-exposed charges of antibody variable domains.* J Mol Recognit, 2002. **15**(2): p. 94-103.

244. Jung, S., et al., *The importance of framework residues H6, H7 and H10 in antibody heavy chains: experimental evidence for a new structural subclassification of antibody V(H) domains.* J Mol Biol, 2001. **309**(3): p. 701-16.

245. Galitsky, B., *Revealing the set of mutually correlated positions for the protein families of immunoglobulin fold.* In Silico Biol, 2003. **3**(3): p. 241-64.

246. Wang, T. and Y. Duan, *Probing the stability-limiting regions of an antibody single-chain variable fragment: a molecular dynamics simulation study.* Protein Eng Des Sel, 2011. **24**(9): p. 649-57.

247. Perchiacca, J.M., M. Bhattacharya, and P.M. Tessier, *Mutational analysis of domain antibodies reveals aggregation hotspots within and near the complementarity determining regions.* Proteins, 2011. **79**(9): p. 2637-47.

248. Chailyan, A., P. Marcatili, and A. Tramontano, *The association of heavy and light chain variable domains in antibodies: implications for antigen specificity.* FEBS J, 2011. **278**(16): p. 2858-66.

249. Desmyter, A., et al., *Antigen specificity and high affinity binding provided by one single loop of a camel single-domain antibody.* J Biol Chem, 2001. **276**(28): p. 26285-90.

250. Desmyter, A., et al., *Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme.* Nat Struct Biol, 1996. **3**(9): p. 803-11.

251. Vu, K.B., et al., *Comparison of llama VH sequences from conventional and heavy chain antibodies.* Mol Immunol, 1997. **34**(16-17): p. 1121-31.

252. Jespers, L., et al., *Crystal structure of HEL4, a soluble, refoldable human V(H) single domain with a germ-line scaffold.* J Mol Biol, 2004. **337**(4): p. 893-903.

253. Nishibori, N., et al., *Humanization of chicken monoclonal antibody using phage-display system.* Mol Immunol, 2006. **43**(6): p. 634-42.

254. Kashmiri, S.V., et al., *SDR grafting--a new approach to antibody humanization.* Methods, 2005. **36**(1): p. 25-34.

255. Jackson, S.E., *How do small single-domain proteins fold?* Fold Des, 1998. **3**(4): p. R81-91.

256. Kodandapani, R., et al., *Crystal structure of the OPG2 Fab. An antireceptor antibody that mimics an RGD cell adhesion site.* J Biol Chem, 1995. **270**(5): p. 2268-73.

257. Graille, M., et al., *Complex between Peptostreptococcus magnus protein L and a human antibody reveals structural convergence in the interaction modes of Fab binding proteins.* Structure, 2001. **9**(8): p. 679-87.

258. Kabat, E.A., Wu, T. T., Perry, H. M., Gottesman, K. S. and Foeller, C., *Sequences of Proteins of Immunological Interest.* National Institutes of Health, Bethesda, MD., 1991.

259. Mizuguchi, M., et al., *Equilibrium and kinetics of the folding of equine lysozyme studied by circular dichroism spectroscopy.* J Mol Biol, 1998. **283**(1): p. 265-77.

260. Neet, K.E. and D.E. Timm, *Conformational stability of dimeric proteins: quantitative studies by equilibrium denaturation.* Protein Sci, 1994. **3**(12): p. 2167-74.

References

261. Myers, J.K. and T.G. Oas, *Mechanism of fast protein folding.* Annu Rev Biochem, 2002. **71**: p. 783-815.

262. Odefey, C., L.M. Mayr, and F.X. Schmid, *Non-prolyl cis-trans peptide bond isomerization as a rate-determining step in protein unfolding and refolding.* J Mol Biol, 1995. **245**(1): p. 69-78.

263. Brandts, J.F., H.R. Halvorson, and M. Brennan, *Consideration of the Possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues.* Biochemistry, 1975. **14**(22): p. 4953-63.

264. Stewart, D.E., A. Sarkar, and J.E. Wampler, *Occurrence and role of cis peptide bonds in protein structures.* J Mol Biol, 1990. **214**(1): p. 253-60.

265. Mayr, C., et al., *Cpr6 and Cpr7, two closely related Hsp90-associated immunophilins from Saccharomyces cerevisiae, differ in their functional properties.* J Biol Chem, 2000. **275**(44): p. 34140-6.

266. Pappenberger, G., et al., *Nonprolyl cis peptide bonds in unfolded proteins cause complex folding kinetics.* Nat Struct Biol, 2001. **8**(5): p. 452-8.

267. Schmid, F.X., *Mechanism of folding of ribonuclease A. Slow refolding is a sequential reaction via structural intermediates.* Biochemistry, 1983. **22**(20): p. 4690-6.

268. Isenman, D.E., D. Lancet, and I. Pecht, *Folding pathways of immunoglobulin domains. The folding kinetics of the C.gamma.3 domain of human IgG1.* Biochemistry, 1979. **18**(15): p. 3327-3336.

269. Goto, Y. and K. Hamaguchi, *The role of the intrachain disulfide bond in the conformation and stability of the constant fragment of the immunoglobulin light chain.* J Biochem, 1979. **86**(5): p. 1433-41.

270. Buckley, C.E., 3rd, P.L. Whitney, and C. Tanford, *The Unfolding and Renaturation of a Specific Univalent Antibody Fragment.* Proc Natl Acad Sci U S A, 1963. **50**: p. 827-34.

271. Kuroda, D., et al., *Structural classification of CDR-H3 revisited: a lesson in antibody modeling.* Proteins, 2008. **73**(3): p. 608-20.

272. Helms, L.R. and R. Wetzel, *Destabilizing loop swaps in the CDRs of an immunoglobulin VL domain.* Protein Sci, 1995. **4**(10): p. 2073-81.

273. Davies, J. and L. Riechmann, *Affinity improvement of single antibody VH domains: residues in all three hypervariable regions affect antigen binding.* Immunotechnology, 1996. **2**(3): p. 169-79.

274. Kurchan, E., H. Roder, and B.E. Bowler, *Kinetics of loop formation and breakage in the denatured state of iso-1-cytochrome c.* J Mol Biol, 2005. **353**(3): p. 730-43.

275. Silow, M. and M. Oliveberg, *Transient aggregates in protein folding are easily mistaken for folding intermediates.* Proc Natl Acad Sci U S A, 1997. **94**(12): p. 6084-6.

276. Roder, H. and W. Colon, *Kinetic role of early intermediates in protein folding.* Curr Opin Struct Biol, 1997. **7**(1): p. 15-28.

277. Harrison, R.K. and R.L. Stein, *Substrate specificities of the peptidyl prolyl cis-trans isomerase activities of cyclophilin and FK-506 binding protein: evidence for the existence of a family of distinct enzymes.* Biochemistry, 1990. **29**(16): p. 3813-6.

278. Ranganathan, R., et al., *Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent.* Cell, 1997. **89**(6): p. 875-86.

Acknowledgement

First of all I would like to thank my supervisor Johannes Buchner for a stimulating PhD project, for all the ideas and advice he has given me over the past years and for the freedom he granted me in my fields of interest.

I thank also the members of the antibody group, past and present, in particular Matthias Feige, who I worked with on the alternatively folded state of antibodies and who introduced me to the world of protein characterisation, folding and kinetics; Roger Müller and Moritz Marcinowski for their practical advice and fruitful discussions, Mathias Rosam and Natalia Sarmiento who also gave me good advice and company and Christine John who will succeed me in the work of antibody association. It was a pleasure to work with you in this group! I am also very grateful to Emma Simpson, who I got to know after quite some Email communication on the folding of V_L. Thanks for never getting tired to answer my questions.

I need to thank Katrin Back, Jirka Peschek, Alexander Bepperling and Christine John for their help with the AUC experiments and data analysis; Bettina Zierer for her support with the stopped-flow device and Edgar Boczek and Alexandra Rehn for their support with the *Sf9* expression. I gave my best to avoid contaminations...

I want to thank Elke Prade and Bernd Reif for their cooperation on the AFS of C_H2. What started as an easy project took its time and turned out to be tricky. Good things come to those who wait and never give up!

It was a pleasure to use the Biacore X100 of the Dynamic Biosensors lab. I want to thank Ralf Strasser and Paul Hampel for their kind introduction to the system and support.

For their hard work I also want to thank my army of internship students. I want to emphasise Jonathan Eras for his excellent work (Master thesis), as well as Maria Trumpfheller, Sabrina Deubler and Marten Plambeck (Bachelor thesis). I am sorry that I cannot list you all.

Thanks to all the admins who were always present when computer problems appeared or programmes were needed etc. Here, I want to mention Mathias Rosam again for his tireless effort!

A special thanks to the secretaries Mrs. Hilber and Mrs. Rubinstein who accompanied me through the “jungle of bureaucracy”. It was for sure not always easy with me and my disorder.

I would also like to thank our technicians Bettina Richter, Ruby Khan, Anja Liebscher and Bianca Ludwig and last but not least Helmut Krause for they support and Martin Haslbeck for keeping things ticking over.

For friendship I thank all members of the Buchner lab; especially Natalia, Betty, Katrin, Christine, Alexandra, Eva S, Jasmin, Yan, Priyanka, Patricia, Eva K and Mathias.

Acknowledgement

I would like to thank my parents and my grandparents for their never-ending belief in me and their support throughout my university career. Finally I would like to thank my boyfriend Dominik Maslak, for his patience, love and for always being there.

List of publications

Simpson ER, Herold EM, Buchner J.

The folding pathway of the antibody V_L domain.

J Mol Biol. 2009 Oct 9; 392 (5):1326-38.

Feige MJ*, Simpson ER*, Herold EM*, Bepperling A, Heger K, Buchner J.

Dissecting the alternatively folded state of the antibody Fab fragment.

J Mol Biol. 2010 Jun 25; 399 (5):719-30.

*these authors contributed equally to this work

Feige MJ, Gräwert MA, Marcinowski M, Hennig M, Behnke J, Ausländer, D, Herold EM, Peschek J, Hendershot M, Sattler M, Groll M, Buchner J

Shark antibody domains reveal the structural and biophysical evolution of the immunoglobulin fold

Manuscript submitted

Declaration

I, Eva-Maria Herold, hereby declare that this thesis was prepared by me independently. I was using only the references and resources stated here. The work has so far not been submitted to any audit commission. Parts of this work have been published in scientific journals.

Hiermit erkläre ich, Eva-Maria Herold, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Die Arbeit wurde bisher keiner Prüfungskommission vorgelegt. Ein Teil dieser Arbeit wurde bereits in wissenschaftlichen Journalen veröffentlicht.

_____

Eva-Maria Herold

München,