TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl II für Organische Chemie - Institute for Advanced Study

About the Interplay of NMR Spectroscopy and Molecular Modeling within Bioorganic and Medicinal Chemistry

Andreas Oliver Frank

TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl II für Organische Chemie - Institute for Advanced Study

About the Interplay of NMR Spectroscopy and Molecular Modeling within Bioorganic and Medicinal Chemistry

Andreas Oliver Frank

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Christian F. W. Becker

Prüfer:
1. Univ.-Prof. Dr. Horst Kessler
2. Univ.-Prof. Dr. Luis Moroder
3. Univ.-Prof. Dr. Michael Groll

Die Dissertation wurde am 06.05.2009 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 23.06.2009 angenommen.

...für meine Eltern und meinen Bruder

"There he goes. One of God's own prototypes. A high-powered mutant of some kind never even considered for mass production. Too weird to live, and too rare to die."

(Raoul Duke)

Die vorliegende Arbeit wurde in der Zeit von Mai 2004 bis April 2009 am Department Chemie der Technischen Universität München unter Anleitung von Herrn Prof. Dr. Horst Kessler angefertigt.

Meinem Doktorvater, Herrn Prof. Dr. Horst Kessler, danke ich für die Bereitstellung einer einzigartigen wissenschaftlichen Infrastruktur, für seine vielseitigen und wertvollen Rat- und Vorschläge, für die nicht-selbstverständlichen Freiheiten, ohne die sich Wissenschaft nicht entfalten kann, und vor allem für seine Menschlichkeit.

Von ganzem Herzen danke ich...

… meinem alten Kumpi Sese und damit natürlich auch Blaui & Gelbi: „Meine Forschungen haben bewiesen, daß Obstsalat letztendlich keine gelben Zähne verursacht!“

… meinem Avukat Akku, meinem mexikanischen Zornwolf und ihrer zauberhaften Clara Victoria: „Die Wanderer müssen sterben, es sei denn, Vasio-Schweine greifen uns an!“

… meinem besten Swidi für die unvergessliche Zeit in Bayreuth und die bisher wunderbaren Monate in München: „Franconia Fantastica...hab' ich geschrieen?!“

… meinem weltweit besten Lbb Orest und seiner Irene: „Danke für die großartige Unterstützung und Euren Glauben an mich!“

… meiner lieben Rike für das „immer für mich da sein“.

… meiner lieben Anne für die „guten alten“ Bayreuther Tage.

… meiner lieben Cathinca für das „durch Dick und Dünn Gehen“ in all den Jahren.

... meinem alten Weggefährten aus Bayreuther und Stockholmer Tagen, Franz „The Zäpf-Man“ Hagn.

… meinen Cimbern für die in letzter Zeit viel zu kurz gekommenen Momente der Zerstreuung und Freundschaft („ich hoffe auf eine faire Dr.cimb.-Prüfung“).

… meinen alten Bandkollegen Pat, Wunk, Rob und Mel, meinen ehemaligen „urKultlern" Ultra und Isa, sowie meinen Club-Freunden Frank Scherzer, Dirk Bienoszek und Stephan „Pavel“ Schurig für die geile Zeit.

… Frau Mensch für „Sie wissen schon wieso".

… Rainer Häßner für seine unzähligen Hilfestellungen, Tipps&Tricks und die äußerst guten email-Gespräche.

… meinen Kooperationspartnern Burkhardt Laufer, Burkhard Luy, Grit Kummerlöwe, Florian Manzenrieder, Timo Huber, Stephan Lagleder, Oliver Demmer, Thomas Reiner, Alice Schlichtiger, Bernhard Wahl, Behrooz Yousefi und Gjermund Henriksen für die freundschaftliche und fruchtbare Zusammenarbeit.

… meinen hervorragenden Knechten Feline Benisch, Maximilian Kern, Alexander Eichner, Adrian Drazic und (vor allem) Matthias Friedrich, der es vom rechtelosen studentischen Mitarbeiter über einen sehr fleißigen Hiwi bis hin zum Top-Bacheloranden der TU geschafft hat (und wieder hinunter und bald wieder hinauf ...auf eine gute Masterarbeit!).

… allen ehemaligen und jetzigen Kollegen der Arbeitskreise Kessler, Luy, Spiteller, Glaser und Eppinger ("pauschal", damit ich niemanden vergesse, was für mich typisch wäre) für das feine Arbeitsklima und den Spaß, den wir zusammen hatten und noch haben werden.

… Dr. Ulrich Wendt, Dr. Stefan Bartoschek, Dr. Herbert Kogler, Dr. Michael Kurz, und Prof. Dr. Peter Hofmann für Ihren Tipp, in der Arbeitsgruppe Kessler zu promovieren.

… meinen ehemaligen Kollegen aus der Bayreuther Arbeitsgruppe Rösch, vor allem Stephan Schwarzinger, und Prof. Dr. Mathias Sprinzl stellvertretend für das tolle Biochemie-Studium an der Universität Bayreuth.

… Martha Fill und Gerd Gemmecker für ihre große Hilfsbereitschaft bei vielen wissenschaftlichen und nicht-wissenschaftlichen Problemen.

… Frau Bruckmaier und Frau Machule für Ihre Hilfe in allen täglichen Belangen.

… dem Club (mal mehr, mal weniger).

… allen, die ich wegen der Hektik der letzten Tage vor der Abgabe aus Versehen verdrängt habe...vergessen habe ich Euch nicht.

… und schließlich und am allermeisten meinen Eltern und meinem Bruder Thomas, ohne die ich es niemals bis hierher geschafft hätte. Danke für alles!

Teile der vorliegenden Arbeit wurden bereits wie folgt publiziert:

1) F. Manzenrieder, A.O. Frank, T. Huber, C. Dorner-Ciossek, H. Kessler. Synthesis and biological evaluation of phosphino dipeptide isostere inhibitor of human beta-secretase (BACE1). *Bioorg. Med. Chem.* **2007**, *15*, 4136-4143.

2) B. Luy, A.O. Frank, H. Kessler. Conformational analysis of drugs by NMR, in *Drug Properties: Measurement and Computation*, R. Mannhold (Ed.) **2008**, p. 207-254.

3) F. Manzenrieder*, A.O. Frank*, H. Kessler. Phosphorus NMR spectroscopy as versatile tool for compound library screening. *Angew. Chem. Int. Ed.* **2008**, *47*, 2608-2611. (* equally contributing authors)

4) A.O. Frank & H. Kessler. Biophysics: The sweetest candy for the virus. *Nature* **2008**, *452*, 822-823.

5) G. Kummerlöwe, S. Knör, A.O. Frank, T. Paululat, H. Kessler, B. Luy. Deuterated Polymer Gels for Measuring Anisotropic NMR Parameters with Strongly Reduced Artefacts. *Chem. Commun.* **2008**, *44*, 5722-5724.

6) B. Laufer, J. Chatterjee, A.O. Frank, H. Kessler. Can *N*-Methylated Amino Acids Serve as Substitute for Prolines in Conformational Design? *J. Pept. Sci.* **2009**, *15*, 141-146.

7) O. Demmer, A.O. Frank, H. Kessler. Design of cyclic peptides, in *Design of Peptides and Proteins: Applications to Therapeutic Agents and Biomedical Research,* K. Jensen (Ed.) **2009**, *in press*.

Table of Contents

The Idea and its Architecture

A Short Synopsis

In order to give a general but short overview, the architecture and content of my Ph. D. thesis are illustrated at this point.

As the title of this book states, the golden thread of the work is the interplay of nuclear magnetic resonance spectroscopy and computational chemistry. By means of a tutorial that describes the procedure of peptide structure determination and nine different projects from the research fields of physical, organic, biological, and medicinal chemistry, the fruitful meshing of both methods is demonstrated. In this context, it seemed - in my opinion - to be more important to focus on the ways experimental and computational techniques contribute to solve scientific problems than to extensively review the chemical and / or biological backgrounds of the treated topics. The architecture and the content of my thesis mirrors this philosophy. However, all nine projects are accompanied by brief background informations. The project presentations are subdivided into an introduction, a “Scientific Question” section, the preparation, performance, and analysis of experiments and calculations, a conclusion, a materials and methods section, and a personal declaration. The distinct chapters are structured as listed in the right column.

NMR Spectroscopy and Computational Chemistry: All theoretical and practical aspects which are of importance for this work are presented in three introductory chapters. First, two short overviews of essential NMR spectroscopic and molecular modeling approaches are given. Finally, a tutorial is provided that explains how peptide structures can be characterized.

Determination of Molecular Arrangements: Three topics are described which all deal with the determination of molecular constitutions, configurations, and conformations.

Simulation of Molecular Dynamics: Three projects are summarized in which flexibilities and motions of small molecules are studied via molecular dynamics simulations.

Elucidation of New Drug Candidates: Three chapters are presented which have the search for new compounds that display affinity for a given receptor molecule in common.

Conclusions and Perspectives: The obtained results are explicitly and shortly summarized. In addition, some perspectives in the interplay of nuclear magnetic resonance spectroscopy and computational chemistry are discussed.

1 Both Sides of the Medal

NMR Spectroscopy and Computational Chemistry

Nuclear magnetic resonance spectroscopy, usually called NMR, is the only experimental method that provides information about the structural arrangement, the dynamical behavior, and intermolecular interactions of almost all kinds of molecules at atomistic resolution. This insight in complex chemical states and mechanisms is of great importance, in particular with respect to medicinal chemistry. However, all data collected via NMR spectroscopy only represent averages over time and molecular ensembles. Since there is almost no experimental methodology that gives insight into the thermodynamic or kinetic properties of a single molecule over a certain period of time, computational approaches based on classical or quantum physics and empirical observations have been developed in order to explore time series and ensemble distributions. Moreover, some complex processes like intermolecular recognition can be scaled down by introduction of mathematical algorithms that have, in parts, the power to model molecular interactions or to predict the structural requirements of molecules for being well-binding receptor ligands. Regarding bioorganic and medicinal chemistry, NMR spectroscopy and computational chemistry complement each other in a very fruitful way. For example, data obtained via NMR spectroscopy can be explained by sophisticated molecular dynamics simulations; in addition, results gained by virtual library screening, docking or QSAR studies can guide NMR based screening experiments. Thus, the potential of NMR spectroscopy and computational chemistry and their interplay are the main topics of the following introductory chapter.

1.1 The Versatile Application Fields of Solution-State NMR Spectroscopy

NMR spectroscopy has emerged to one of the most valuable experimental tools in chemistry and physics during the last decades.[1] Its power results from the detection of almost every atom type in a molecule in its specific environment. Each of the magnetically active isotopes is characterized by its specific resonance frequency at a given static magnetic field, determined by its gyromagnetic constant γ. In addition, chemically different nuclei of the same isotope have different resonance frequencies (chemical shift; CS). The mutual interactions of these nuclei which occur through chemical bonds or through space can e.g. be used to elucidate a molecular structure. Furthermore, NMR spectroscopy does not only yield information about the connectivity of all atoms within a molecule (the constitution) or the spatial arrangements of atoms or groups (the configuration and conformation); it also provides precious aspects about dynamics within a molecule and interactions to other molecules, such as to surrounding solvent molecules or to specific binding partners. The most prominent application areas of NMR spectroscopy are listed below:

1) *Structure Determination*
2) *Investigation of Molecular Dynamics*
3) *Detection of Intermolecular Interactions*
4) *In vivo / vitro Analysis*
5) *Magnetic Resonance Imaging*
6) *Facility for Quantum Physics*

Since the latter three points are not of relevance for this work, only the first three topics described in more detail in the following subsections. It should be noted that all given aspects refer to solution-state NMR spectroscopy and to methods that have been developed to investigate small and medium-sized molecules and their interactions with biomacromolecules.

1.1.1 Structure Determination of Small Molecules in Solution

NMR based structural studies always begin with the elucidation of the constitution of a molecule. This is almost identical with the assignment of NMR signals to atoms in the molecule of interest. Whereas assignment of as many

as possible signals is required for the elucidation of a three-dimensional (3D) structure, a full assignment of all NMR signals is not necessary for establishing the constitution. However, when spatial distances between atoms are used for conformational analysis, it is important to identify all specific constitutional positions. The main source for establishing the constitution of a molecule are scalar couplings which are mediated between two nuclei by electrons via covalent bonds. Their strength strongly depend on the kind, the number, and the orientation of bonds between the magnetically active nuclei, e.g. ^{1}H, ^{13}C, ^{15}N, ^{19}F, ^{31}P. J-couplings directly provide informations about the connectivity of atoms which could be used for the determination of the constitution of molecules. Elucidation of the stereostructure - the configuration and the conformation - is the next step in structural analysis. Three main parameters are used to elucidate the stereochemistry: Scalar couplings - as addressed above - are in turn of great value since some J-couplings bear information about the 3D arrangement of atoms and atom groups that are connected by covalent bonds. The main source for differing scalar couplings is thus determined by different bond angles and dihedrals. Another way to obtain this information is the use of cross-correlated relaxation (CCR), but this is rarely used for drug or drug-like molecules.

The second and probably most important parameter for 3D structure elucidation is the nuclear Overhauser effect (NOE) which serves as basis to derive spatial distances between hydrogen atoms or, in rare cases, hetero-nuclei. The NOE emerges via relaxation processes between magnetic dipoles that are in a non-equilibrium state. In principle, relaxation processes can best be understood by looking at a simple model: if a hot piece of metal is placed in porridge, first the porridge molecules very close to the metal will be heated; subsequently, after a certain time, the heat of the metal will further dissipate into the highly viscous “porridge bath” until an equilibrium is reached in which the temperature of the metal and the porridge is identical. In other words, the temperature of a molecule depends on its distance to the metal in the beginning of the experiment. As the build-up of the NOE is inversely proportional to the sixth power of the distances between the nuclei, it can be used to determine intramolecular distances by following the kinetics of the build-up. This means that the intensity of an NOE cross-peak could be converted into a pairwise distance restraint. An example for this procedure is given in figure 1.1.

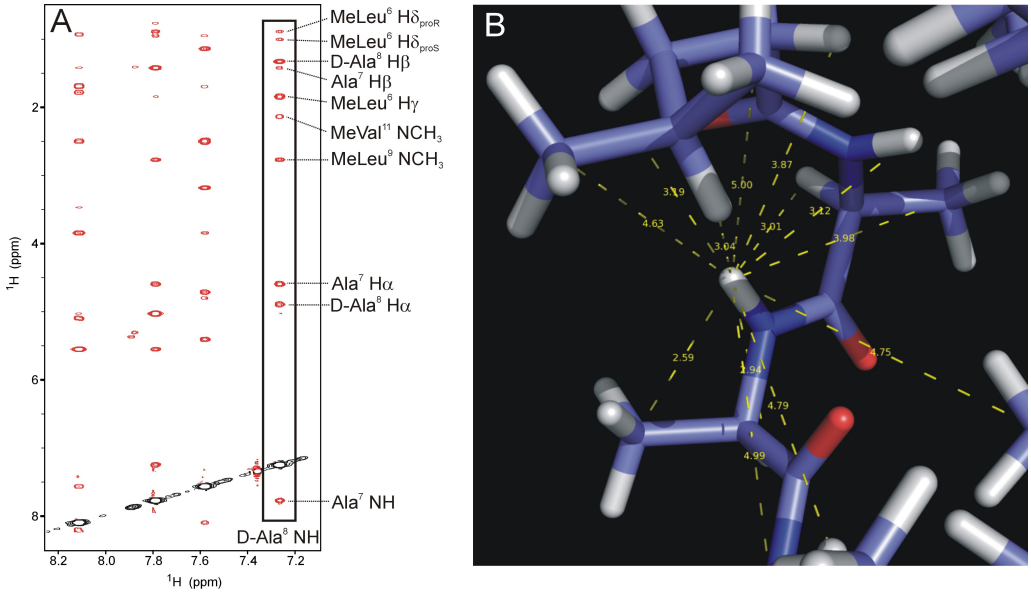


Figure 1.1: Fixation of the Ala^8-H^N proton in Cyclosporin A via NOE-derived distance restraints as an illustrative example for NMR based 3D structure determination.

A) The amide region of the 2D-NOESY spectrum of Cyclosporin A in $CDCl_3$ at 243 K with the cross peaks to Ala^8-H^N indicated by a box.

B) Distances in the final structure shown by dashed yellow lines as a result of the NOE-derived distance restraints. The distance to the methyl carbon is given as a representative for Me group protons. *(taken from reference [1])*

More recently, a third experimental source for NMR based conformational analysis has been established: residual dipolar couplings, shortly called RDCs.[2] To measure anisotropic parameters like RDCs, a solute must be partially aligned; this means, that the solute can not freely tumble in solution but prefers some orientations. The ongoing development of suitable alignment media makes the approach applicable to a wide range of drug and drug-like molecules.[3,4] Since anisotropic parameters are proportional to the averaged orientation of a molecule relative to the static magnetic field (which serves as an external reference), long-range correlations within a molecule can be elucidated which allow the determination of an averaged structure with a very high precision. In a simplistic picture, spins can be looked at as magnets with an inherent rotation at the Larmor frequency. Although spins are not oriented directly along the static magnetic field B_0, the integration over time of the fast rotating magnets yields a resulting magnetic moment parallel or antiparallel to B_0 (see Fig. 1.2). The magnetic moment of a spin results in the same magnetic field as a classical magnet with the typical r^{-3} dependence of the magnetic field with respect to the distance to the magnet and the $(3\cdot\cos^2\theta - 1)$ dependence with respect to the angle θ relative to the axis of the magnetic moment. Since the magnetic moment of the spin is oriented along the static magnetic field B_0, the angle θ is identical to the angle Θ with respect to B_0 (see Fig. 1.2). A second spin close in space feels the magnetic moment of the first spin and therefore resonates at a slightly different frequency as compared to an isolated spin. Since spins parallel and antiparallel to B_0 are about equally populated, the field contributions of neighboring spins lead to a signal split (which is

usually called the residual dipolar coupling) with the distance of the doublet components given in Hertz.[5] In addition, there are several other anisotropic parameters like residual quadrupolar couplings, residual chemical shift anisotropy, or pseudo-contact shifts that could fruitfully contribute to the structure determination process. Since the mentioned parameters are not routinely used for the elucidation of small molecule structures, they are not further described.

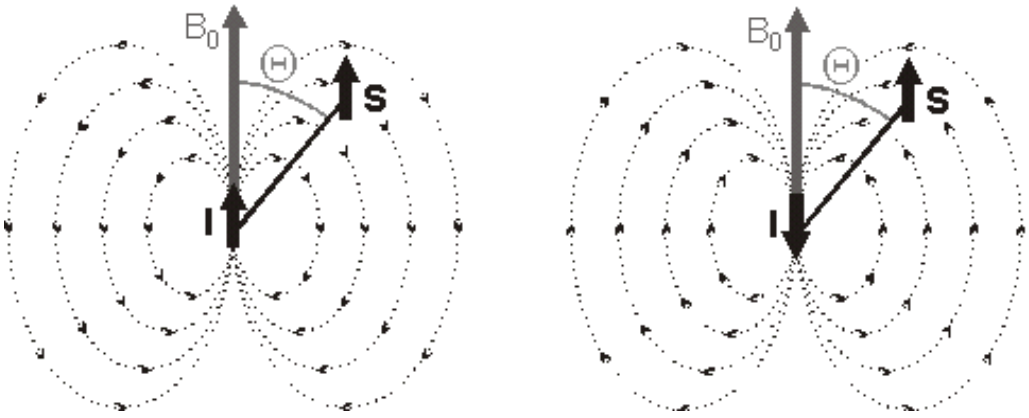


Figure 1.2: Illustration of the dipolar interaction. The magnetic field induced by spin *I* adds up to a static magnetic field B_0 and leads to a shift of the resonance frequency of the close-by spin *S*. Spins oriented parallel and antiparallel to the B_0 field are equally distributed, resulting in signal splitting with a certain dipolar coupling given in *Hz*. *(taken from reference [1])*

From a medicinal chemist's point of view, one is not only interested in the conformation of a ligand molecule freely tumbling in solution. When a compound has the potential to interact with a target molecule, e.g. a protein or nucleic acids, its “active” conformation - the three-dimensional structure in the bound state - is of even more importance. In this scope, a lot of methods have been developed that mostly take advantage of so-called exchange-transfer processes.[6] To name but a few, techniques based on exchange-transferred NOEs, exchange-transferred RDCs, exchange-transferred pseudo-contact shifts, or exchange-transferred cross-correlated relaxation are versatile tools to elucidate the conformation of a ligand bound to a receptor.[1] Although a lot of important informations can be obtained via the mentioned methods, the majority of structures of bound ligands is still determined by X-ray crystallography since the precision of crystal structures of molecular complexes is usually higher than conformations of receptor ligand complexes solved by NMR spectroscopy.[7]

A deeper and also more physical insight in the detection and application of J-couplings, NOEs, and RDCs, respectively, is spared at this point as a detailed explanation of the NMR driven structure elucidation procedure is given in chapter 1.3.

1.1.2 Tracing Molecular Dynamics

In chapter 1.1.1, methods for the elucidation of three-dimensional structures of small and medium-sized molecules were shortly described. For a full understanding of the biological function of molecules and the way interactions with binding partners are realized, dynamical processes must be taken into account.[8] In general, the connection between structure, dynamics, and function is of great importance. For example, if dynamics has consequences for function, insight in molecular flexibility is crucial for both protein and ligand design. A small but dynamic molecule, perfectly fitting into an active site of a protein when only its lowest energy conformation is considered, consequently loses affinity when it adopts another, higher energy structure. The same phenomenon holds even more true in case of protein dynamics. In contrast, the overwhelming view on molecules emerging from structural biology is static, and relevant dynamic information is still limited.[9] In order to animate the plethora of "molecular snapshots" present today, NMR based approaches have been developed for the elucidation of the dynamical behavior of atom groups, small compounds, and biomacromolecules (see Fig. 1.3).[8]

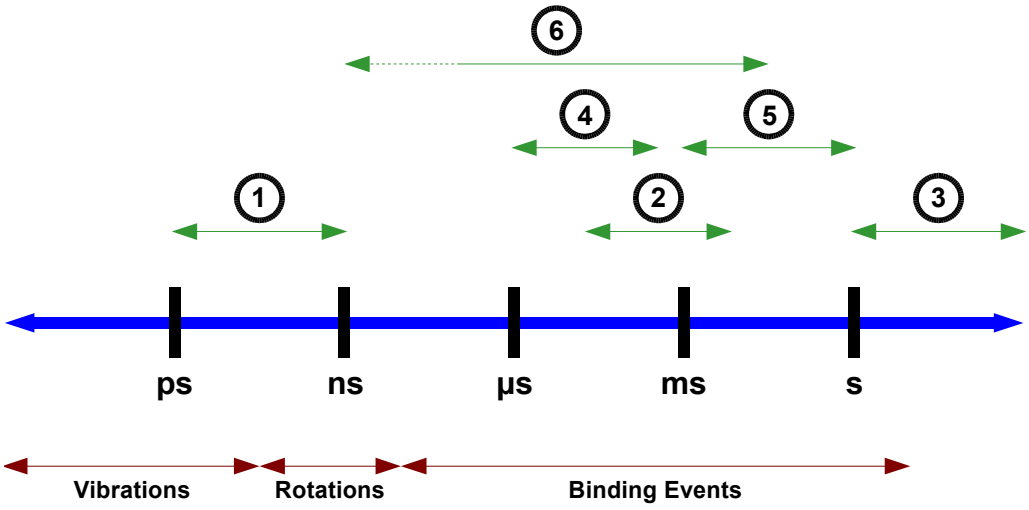


Figure 1.3: Illustration of the time range between femtoseconds and minutes. The time regime between femto- and picoseconds comprises in particular bond vibrations. Molecular rotations take place between pico- and nanoseconds while binding events occur between nanoseconds and sometimes even seconds. A lot of NMR techniques has been developed to cover the full range of molecular dynamics: (1) Both R_1 or R_2 relaxation measurements and NOE-type experiments; (2) CPMG filter measurements; (3) proton-deuterium (H-D) exchange experiments; (4) $R_{1\rho}$ relaxation dispersion experiments; (5) *ZZ* exchange measurements; (6) RDC based experiments. The nano- to microsecond time regime can only partially covered via RDCs (dashed green arrow).

By help of the techniques which are presented in figure 1.3, it is possible to cover almost all time ranges that are of interest in peptide and protein science.[10-12] One exception is the interval between nano- and microsecond dynamics that can only partially investigated via residual dipolar couplings.[13]

Motions in the ultra-fast limit, that is the sub-picosecond range, are usually not of special interest as their contribution to the more important overall dynamical

behavior of a molecule is usually negligible. Dynamics taking place in this regime are mainly of vibrational character.[14] Since motions of this kind are fully averaged, they are not extractable via NMR spectroscopic techniques.

In contrast to the sub-picosecond time regime, insight into the fast dynamics between pico- and nanoseconds is of more important weight. In this time range, backbone and side-chain fluctuations occur which are essential for receptor-ligand affinity. Motions in the fast regime can be studied via NMR spectroscopy by measuring three relaxation rates: the longitudinal relaxation rate R_1, the transversal relaxation rate R_2, and the steady-state heteronuclear NOE (hetNOE).[15] NMR measurements exploit the fact that molecular reorientations lead to fluctuating magnetic fields which cause transitions between nuclear spin states, finally resulting in dephased coherences. To express the amplitude of molecular motions, the so-called Lipari-Szabo S^2 order parameter is by far the most common motional descriptor.[16-19] An S^2 value of 1.0 indicates complete restriction of internal motion whereas a value of 0 refers to unrestricted isotropic flexibility. HetNOE experiments as usually applied in protein dynamics studies are of limited use in peptide chemistry since the nuclei of interest, ^{13}C and ^{15}N, are in most cases not enriched in these molecules which leads to hardly analyzable spectra.

The time regime between nano- and microsecond where - compared to the very fast dynamics at the sub-nanosecond time scale - the slower side-chain or backbone motions like peptide bond flips around adjacent φ and ψ torsional angles occur, can not be fully covered by NMR spectroscopy.[20-22] Those motions are averaged, therefore resulting in only one signal representing two or more conformational states. Due to the fast exchange, no line broadening of NMR resonances can be detected as it is possible when slow exchange is present. A recent approach used residual dipolar couplings to model such conformational transitions.[13] In principle, RDCs are sensitive reporters of dynamics in the nano- to microsecond range because structural fluctuations lead to variations of the RDC size in the order of ca. 10 Hz. Hence, dynamics equal or slower than the rotational correlation time τ_c and faster than about 10 ms can be elucidated by projection of residual dipolar couplings onto conformational models. The presented approach is very elaborate with respect to experiments and modeling procedures;[13] thus, the near future will show if the new technique becomes a routine application to cover motions in the ns-μs time regime.

The next time range, that is micro- to milliseconds, can be nicely investigated via NMR spectroscopy. Beside the utilization of RDCs, several other techniques based on relaxation exchange rates (R_{ex}) have been developed for this purpose.[13] The most prominent methods are the measurement of $R_{1\rho}$ (a spin-locking sequence is used in the pulse programs) and of R_2 relaxation dispersion (a so-called CPMG filter is applied).[23-26] As the most common spin probe for exchange relaxation measurements is ^{15}N, the mentioned methods are only rarely used for studying dynamics of unlabeled peptides. Many biological processes occur at the medium regime, including protein folding, substrate or drug binding, allosteric regulation, or catalysis.[10,11,27] Regarding peptides, slow conformational exchanges like transitions of peptide bonds between trans and cis conformations take place in this or at slower time regimes, depending on the height of energetic barriers between the two states.

Finally, NMR spectroscopy can also be used to elucidate molecular flexibility in the millisecond to second time range and even dynamics slower than seconds. Especially slow binding events occur in this time regime, but also slow protein folding or chemical exchange are present.[8] Induced fit motions can be traced by methods based on so-called ZZ exchange techniques. Chemical exchange events like the substitution of polar hydrogens by solvent hydrogens or deuterium (H-D exchange) can easily be observed by the detection of exchange peaks in NOE-type spectra.[8,12] Whereas the first mentioned technique is almost exclusively applied for protein NMR studies, the latter experiments are also very popular in peptide chemistry.

Most of the presented approaches have been developed to determine protein dynamics. For the investigation of peptide fluctuations, special procedures are routinely used which, on one side, do not depend on heteronuclear labeling, and, on the other side, are of practical use and mostly simple.[28]

Linear peptides intrinsically have many degrees of freedom and even small species can adopt thousands of different conformations which are in fast exchange. Thus, those structures can not be observed via NMR spectroscopy. Only in some cases (e.g. in *β*-peptides) stable secondary structural elements are formed that are in slow exchange.[29,30] Compared to linear peptides, cyclic templates (especially when the chemical bridging is formed via head-to-tail cyclization) not only differ in structural characteristics but also display a strongly modified dynamical behavior.[28] In order to investigate the residual

internal motions of a cyclic peptide, the simple ^{1}H-1D experiment is mainly chosen. If dynamics are slower than about 10 ms, a doubled signal set arises in the spectrum.[13] This means that two structural arrangements are separated by a high energy barrier, making the exchange slow or even impossible. In addition, 2D-NOESY or 2D-ROESY experiments can be recorded; here, exchange peaks - cross peaks with different sign for small molecules - become visible in spectra if conformations slowly interchange with each other.

Detection of only one signal set in a ^{1}H-1D does not exclude the presence of fast dynamics. Whereas fluctuations in the micro- to millisecond time regime result in a detectable line broadening, motions faster than ca. 50 µs do not contribute to the line width of NMR signals and are, therefore, only indirectly observable.[13] This can be achieved e.g. by tracing errors of the pairwise distance restraints in the structure determination process. If several restraints can not be fulfilled when only one structure is calculated, several conformations which are in fast exchange are often present.[1] The same is true for deviations between experimental and calculated scalar or residual dipolar couplings. Another way for qualitatively tracing medium to fast dynamics in peptides is the measurement of temperature gradients. By incrementally heating an NMR sample from room temperature up to 320-330 K, both chemical exchange processes are triggered and the chemical shifts of polar protons are perturbed. Resonances of protons which are solvent accessible are usually strongly affected whereas the chemical shift of solvent shielded protons does not change significantly; those atoms are often part of a hydrogen bond and thus, reveal restricted dynamics. However, it should be stated that solvent shielded protons do not need to be part of a hydrogen bridge but can also be buried by spatially demanding groups.[28]

In cyclic structures, not only the backbone flexibility is strongly reduced but also the dynamics of amino acid side chains. This is particularly true for the common three rotamers around the Cα-Cβ bond. Often, preferred conformations of the χ_1 dihedral angle are observed which can be easily identified and assigned by NMR spectroscopy (see section 1.3.3.1). For example, measurement of J-couplings reveals if side chain rotations are still present (an averaged value is yielded) or if one orientation is preferred, thus indicating that fast dynamical processes have more or less vanished. In addition, the occurrence of dynamically restricted side chains indicates a high probability of a preferred backbone conformation as well.

1.1.3 Elucidation of Receptor-Ligand Interactions

The biological function of a biomacromolecule (e.g. a protein) usually depends on the interaction with ligand molecules or other macromolecules. It is obvious that the investigation of intermolecular interactions plays a key role in order to understand biological processes which are based on the functions of proteins. Hence, a lot of methods has been developed to investigate the so-called “interactome” of proteins and their ligands.

In this scope, NMR spectroscopic techniques seem to be the most powerful tools since unknown interactions can directly be observed at atomistic resolution.[6] In principle, two general NMR spectroscopy methodologies are available to prove molecular recognition processes:

a) *Ligand Based Techniques*
b) *Target Based Methods*

Most of the ligand based screening techniques rely on the so-called transfer NOE (trNOE) effect; the most prominent methods are saturation transfer difference (STD) NMR spectroscopy, WaterLOGSY, cross-saturation, and transient transferred NOE experiments.[6] Furthermore, there are alternative approaches which make use of translational diffusion times, of longitudinal or transversal relaxation rates or of paramagnetic agents and spin labels. Some of the ligand based techniques have been extended to more robust methods like fluorine, competition, fragment based, or receptor-immobilized ligand screening. In contrast to ligand based approaches, almost all target based techniques exclusively rely on the detection of chemical shift perturbations. Whereas 1D NMR experiments are mostly used for ligand based screening, target based methods usually take advantage of two- or three-dimensional measurements.[6,31-34]

The basic concept in NMR aided screening is the discriminability of ligand molecules and their receptors with respect to NMR detectable properties (see Fig. 1.4), e.g. differing molecular weights of both species. Different masses have dramatic impact on rotational and translational correlation times, resulting in e.g. dissimilar relaxation rates or diffusion times. Since a ligand adopts most of the physical properties of the usually much bigger receptor molecule when being bound, a discrimination of binders and non-binders is possible.[35-40] Another major approach for distinguishing ligand and target

molecules depends on the different chemical shift dispersion. Target molecules have intrinsically a wider shift dispersion and can, therefore, exclusively be excited.[6,41] The magnetization is further transferred to bound ligands and can subsequently be detected when the ligands are released from their receptor. Another screening technique is the detection of chemical shift perturbations.[42-45] This strategy relies on changes in the chemical environment upon a binding event. Both the ligand and the target molecules are affected regarding their chemical shifts when they interact with each other. Especially when using isotopically labeled receptors, target based NMR screening is the most powerful technique since it not only provides informations about binding ligands but also about the (protein) binding site.

To present the most prominent screening methods in a more conceptual framework, the single stages of drug exploration and their correlation to NMR based approaches are shortly explained below (for a short tabular summary, see Fig. 1.5 on page 16).

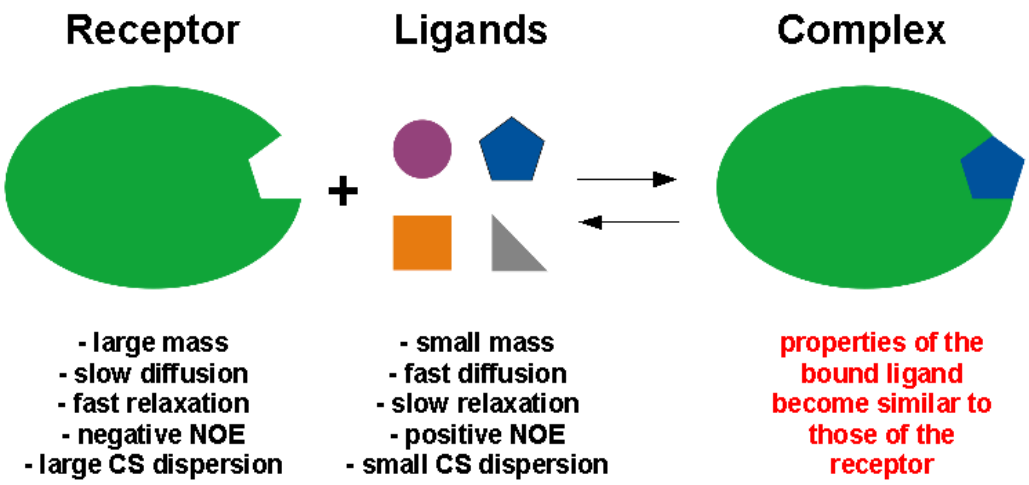


Figure 1.4: Cartoon explaining intermolecular recognition, e.g. between a protein and a much smaller molecule. Only those compounds that fit into the binding site of a receptor are bound. Since small ligands adopt the physical properties of a huge receptor, binders can be easily discriminated from non-binding molecules via NMR spectroscopy.

In the early stage of drug development it is essential to find so-called first hits which usually have a low to medium affinity to the receptor of interest (hit finding).[46] For this purpose, large libraries of small compounds are screened for target affinity. Since library screening is very time-consuming, one-dimensional NMR techniques are dominating this stage of drug research with the STD and transient transferred NOE experiments being the most prominent screening approaches.[1,6]

In STD experiments, a non-equilibrium magnetization is obtained via selective excitation of target resonances (the so called on-resonance experiment). Very quickly, the magnetization is distributed over the whole receptor by spin-diffusion, including bound ligand molecules. Medium to weak binders which are in fast exchange between the bound and the free state transfer the achieved magnetization into solution where it can be detected. By subtracting a reference spectrum from the on-resonant one which is recorded with an off-resonant irradiation one gains a difference spectrum that displays only those resonances which have been perturbed due to receptor binding. Due to ongoing developments, the fields of application of STD are very broad; for example, the mapping of receptor epitopes or the exact determination of binding constants is possible today by means of this approach.[6,47,48] If a target can not be selectively excited, the WaterLOGSY approach is the favored experiment. Here, bulk water is first excited; then, the magnetization is transferred to the target-ligand complex. Finally, the detection of ligands with receptor affinity depends on different cross-relaxation properties of binders and non-binding molecules.[49,50]

The concept of transient transferred NOEs also plays an important role for finding “early hits”. Whereas steady-state techniques rely on transfer of magnetic saturation, transient NOE experiments generate a non-equilibrium state which returns to equilibrium by T1 relaxation processes. The relaxation period is called mixing time. The most common experiments for detection of transient NOEs are NOESY- and ROESY-type measurements. During the mixing time, both intramolecular and intermolecular NOEs build up. In general, NOEs can be used for detecting binding events since their sign depends on the size of a molecule. Small ligands usually display positive NOEs whereas strongly negative NOEs arise in large molecules. Since a small binder adopts the properties of its receptor molecule and takes parts of these properties into solution after being released, the sign of the observed NOE within the ligand indicates binding. The utilization of intermolecular trNOEs not only allows for detecting ligand receptor interactions but also for the elucidation of the orientation of the ligand in a binding pocket. This concept relies on the fact that a ligand bound to a receptor not only alters its tumbling and diffusion times but also changes its relaxation behavior. Using different setups of diffusion and relaxation filters, binders can be distinguished from non-binders. The most

prominent representative of transient transferred NOE experiments are the so-called NOE pumping (based on intramolecular trNOEs) and reverse NOE pumping (relying on intermolecular trNOEs) approaches.[51-53]

Since ligand molecules change their translational diffusion and relaxation times, intermolecular interactions can directly be observed by NMR spectroscopy. Small molecules usually display fast diffusion and slow relaxation times; in contrast, when bound to the receptor, they adopt both the fast relaxation and slow diffusion of the target. Thus, simple one-dimensional spectra (gradient based pulse programs are used to cover diffusion effects and relaxation filters are applied to elucidate varying relaxation rates) can be recorded in order to trace the mentioned effects.[54,55] An enhancement of relaxation based screening techniques was the introduction of paramagnetic spin labels. Chemical groups with unpaired electrons drastically increase T_2 relaxation rates of those nuclei which are in close proximity to a paramagnetic agent. The method is of special interest when ligand molecules are screened for binding to a second active site at one receptor.[56] A ligand binding to a first active site is spin-labeled; if a molecule binds to a second site which is close to the first site, its NMR signals consequently vanish due to the fast relaxation induced by the paramagnetic agent.

However, ^{1}H-1D spectra often suffer from strong signal overlap, especially when huge compound libraries are screened. The concept of fluorine screening faces this drawback.[57] Here, the libraries to be screened contain molecules with fluorinated aromatic or trifluoro-methyl groups. Thus, each compound is represented by only one signal, making also one-dimensional spectra easily analyzable. Beside other advantages of detecting fluorine, tracing of intermolecular interactions is easy since binding events result in shifted resonances (perturbation of the chemical environment) or broadened line widths (enhanced T_2 relaxation rates).

Having found a first hit, the results of the library screening must be first confirmed by more sophisticated methods (hit validation) and finally optimized (hit optimization; this topic will not be further described here).[46] At this stage of drug research, informations regarding the target binding site are now of special interest. For this purpose, target based screening methods are well suited approaches if isotopically labeled receptors are available.[43-45] In principle, different labeling schemes have been developed. For small and medium-sized proteins, uniform ^{15}N-labeling is a good choice, thereby allowing

for the measurement of ^{15}N-HSQC-type spectra. When screening larger proteins, expensive ^{13}C-labeling and / or deuteration is often additionally required. Site- or amino acid selective labeling can extend the measurable size of the proteins as the spectral overlap is drastically reduced.[6] HSQC based screening - often shortly called SAR-by-NMR - usually relies on perturbations of the chemical shift pattern of a protein when a ligand binds.[58] This effect is caused by alteration of the magnetic environment of receptor atoms upon ligand binding.

In addition, another ligand-based screening method can be applied; by titrating a known binder with its receptor, it is even possible to calculate binding constants.[6] A further technique for validating intermolecular interactions is competition-based screening;[59] here, the potential of a newly found binder to replace a known ligand (with weak receptor affinity) from an active protein site is investigated. By help of this method, unspecific interactions or unknown binding sites (“second sites”) can be elucidated.

Figure 1.5: NMR methods designed for tracing intermolecular interactions. The main limitations and requirements as well as the potential of each technique are shortly summarized.

Method	Type	Limits and Requirements			Identification of		
		Target Mass Limit	Affinity Limit	Isotope Labels	Target Binding Site	Ligand Binding Epitope	Ligand Mixture
Diffusion Filtering	Lig	Lower	Upper & Lower	None	No	No	Yes
Relaxation Filtering	Lig	Lower	Upper & Lower	None	No	No	Yes
trNOE Methods	Lig	Lower	Upper & Lower	None	No	No	Yes
NOE Pumping	Lig	Lower	Upper & Lower	None	No	No	Yes
Rev. NOE Pumping	Lig	Lower	Upper & Lower	None	No	Yes	Yes
Water LOGSY	Lig	Lower	Upper & Lower	None	No	Yes	Yes
STD Methods	Lig	Lower	Upper & Lower	None	Quality	Yes	Yes
^{19}F Screening	Lig	None	None	^{19}F (Ligand)	No	No	Yes
Reporter Screening	Lig	None	None	None	Quality	No	No
CS-Perturbations	Tar	Upper	None	^{13}C/^{15}N (Target)	Yes	No	No

1.2 Utilization of Computational Tools in Medicinal Chemistry

During the last decades, computational techniques have become more and more popular for solving mathematical, physical, chemical, or biological problems. In general, computer aided approaches are used in order to reproduce and, sometimes, to predict experimental results or to verify theoretical considerations. Computational chemistry (CC) is one of the theoretical techniques that contributes to chemical questions with assistance of computers.[60] Results obtained via CC methods confirm or complement informations obtained via physical, (bio)chemical, or biological measurements. The methods used range from *ab initio* (entirely based on theory) to quantum-mechanical, empirical, or semi-empirical approaches (inclusion of experimental results). The power of CC is based on the fact that reliable predictions or explanations of chemical processes by computer programs could spare expensive and time-consuming experimental setups.[29] The ratio of computer performance to price increases every five to six years by a factor of about ten.[14] Thus, CC techniques became and will become more and more precise and can / will be used in more and more research fields. The most common virtual methods routinely applied in medicinal chemistry are listed below.

a) *Pharmacophore Models & Virtual Screening*
b) *Molecular Docking*
c) *Similarity Searching & QSAR*
d) *Molecular Mechanics*
e) *De novo Design*
f) *Property Prediction*

A subsection of computational chemistry is molecular modeling which reduces the complexity of a considered system by the description of only those parts that contribute most to the phenomena of interest.[60] The term molecular modeling sums up all theoretical methods and computational techniques that mimic the behavior of molecules. Compared to laboratory experiments, molecular modeling requires less time and is cheaper. For example, usually applied docking algorithms do not explicitly treat solvent molecules, and all degrees of freedom of a receptor are usually neglected which finally leads to a very fast computation of interactions but, of course, at the expense of a

reduced accuracy. Nowadays, modeling methods are often used in theoretical chemistry and biology. This also includes the simulation of the kinetic and thermodynamic behavior of molecules, the design of new drug compounds, and the examination of structure-activity relationships. However, experimental findings still often contradict computed outcomes; thus, virtual modeling techniques must be carefully applied and evaluated.[61]

In the following four sections, several computational chemistry techniques which are of interest for this work (see above, points a-d) and, especially, to medicinal chemistry, are presented in more detail.

1.2.1 Database Screening Based on Pharmacophore Models

Virtual screening belongs to the concept of computational database searching. In principle, the expression means “automatically evaluating very large libraries of compounds using computer programs“.[62] A database which contains molecules and their three-dimensional coordinates (virtual library) is filtered with respect to desired chemical and physical characteristics (pharmacophore model).[63,64] The method is routinely used in computer based drug discovery because it achieves a fast reduction of the number of potential drug candidates without the need of experimental measurements. The compounds obtained from a search, however, do not necessary bind to a specified target *in vitro* or *in vivo* since the algorithms present today are far away from being perfect. Therefore, experiments, e.g. via NMR spectroscopy, which elucidate the binding potential of virtually selected ligand molecules have to be performed after the computational procedure.

A possible (and the most common) way to select desired molecules from a virtual library is to use a pharmacophore model as shown in figure 1.6. IUPAC defines a pharmacophore as an “ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target“. Even if the original definition of a pharmacophore relates only to ligand molecules, computational tools nowadays also allow the construction of receptor (also called target) based models: a receptor derived pharmacophore is constructed by mapping the geometry and essential chemical groups of the binding site of a target. Thus, it can be considered as a combination of groups that represent a three-dimensional image of all features of an active site which are important for ligand binding. Receptor based

pharmacophore models are used when the active site of e.g. a protein is known and one is interested in finding a potential binder. In contrast, ligand based pharmacophore models are utilized when an interaction partner of a receptor is known but no information about the binding site is available.[60] In addition, both types of pharmacophore models can be combined. A molecule of a virtual compound library has to possess all operator-defined features or constraints (e.g. hydrogen bond acceptor / donor atoms, hydrophobic or aromatic groups, containing / excluded volumes) so that it is not discarded during the filtering runs (virtual screening). Both features and constraints can be provided with spatial tolerances; the stricter the tolerances are set, the less hits are found during the library search. To conclude, a computational pharmacophore model represents a reduced image of the active site of a receptor and essential groups of a ligand.

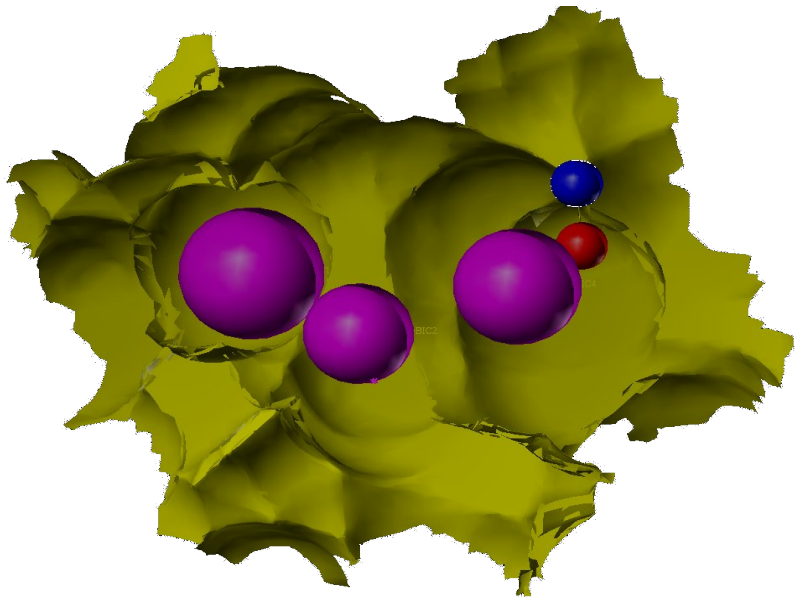


Figure 1.6: Presentation of a pharmacophore model which was designed by means of target and ligand properties. The greenish-yellow topology serves as a spatial constraint based on the surface of a receptor binding site. The purple, blue and red spheres represent physical features like hydrophobic groups or hydrogen donors and / or acceptors of a ligand. The volume of the spheres denote the spatial tolerance with which virtual databases are screened for new receptor binding ligands.

Even if the process of virtual database screening is highly automated, an elaborate and detailed preparation of search runs is indispensable. The more informations - if available - about natural interaction partners or known artificial ligands, about the character of the active site, or about essential amino acid residues of the receptor that take part in ligand binding are collected, the more precise a pharmacophore model can be constructed. Usually, it is suggested that the model is built in such a way that the natural binding mode is emulated; thus, those residues of a natural ligand and of an active site which are directly involved in binding are used as pharmacophore features or constraints.

In general, there are two ways for pharmacophore based searching of virtual databases. One can construct a pharmacophore model with all characteristics that seem to be essential for binding and start the screening process. This approach is quick and sophisticated but suffers from the fact that sometimes no hits are found due to the demanding character of the pharmacophore. Thus, it is advantageous to provide features and constraints with big enough spatial tolerances. An alternative procedure is to preselect potential hits. Here, only a part of essential features or constraints is used to screen the libraries. By iteratively increasing the numbers and the strictness of the pharmacophore model the quantity of hits is stepwise reduced.

Finally, the operator has to elect virtual databases that are suited for the kind of receptor that is studied. Some libraries contain only molecules for special receptor types, e.g. only compounds that can pass the blood-brain-barrier or which are assumed to bind proteases. Other databases comprise only molecules that, for example, match Lipinski's rule of five,[65] contain very similar compounds, or are built of complete diverse ligands. It always depends on the scientific question which kind of library is optimal for pharmacophore model based screening.

1.2.2 Docking for the Prediction of Molecular Interactions

Molecular docking is a computational method for finding an optimal receptor-ligand complex geometry (see Fig. 1.7).[63,66] Thus, the main task of the docking process is to minimize the intermolecular interaction energy (binding energy) between the two molecules of interest.[67] Usually, the binding energy is evaluated by a so-called scoring function. Scoring functions are fast but approximate mathematical descriptors used to calculate the strength of non-covalent intermolecular interactions after two molecules have been docked. Generally, the methods for exploring the conformational interaction space quite differ. The most prominent approaches - the fragment based method and the grid based technique - are treated below.[68,69]

The fragment based docking method uses a base fragment of a ligand (the “ligand core”) which is automatically selected and placed into the active site of a receptor. During the following steps the complete ligand is incrementally build up from the remaining fragments. A new fragment is added in all possible conformations to the fragments constituted in the previous iteration steps;

afterwards, the best placements are taken on to the next construction step. The conformational flexibility of the ligand is accounted for by generating multiple conformations for each fragment and including all possible constitutions in the ligand building steps. Finally, all created placements of the ligand are ranked according to the so-called score which is a measure of the "quality" of the receptor-ligand complex.[70] Usually, the extent of interactions is computed in terms of the free binding energy G. The free energy function contains terms for hydrogen bonding, ionic, aromatic or lipophilic interactions, each adjusted by a penalty function that depends on the deviation from the ideal complex geometry.[71]

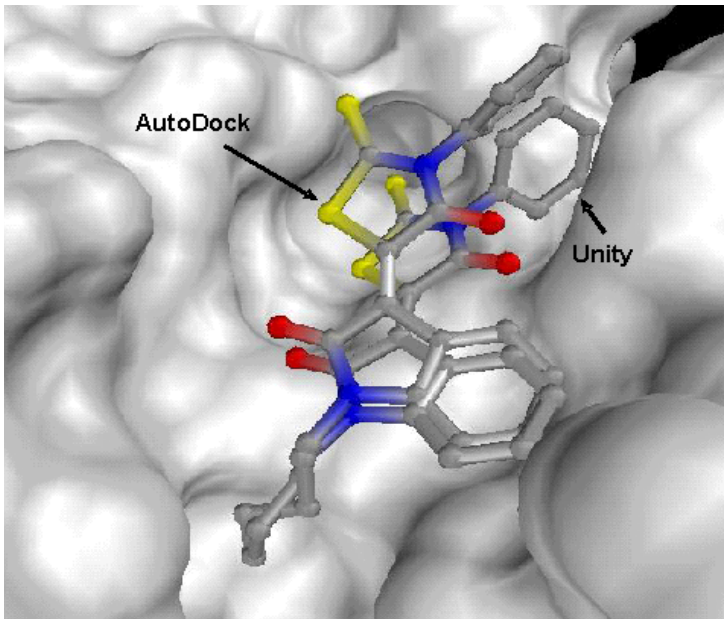


Figure 1.7: Illustration of a small organic ligand molecule that is virtually docked into the binding site of a receptor molecule. The receptor is shown with its Conolly surface. As the ligand was docked with two different algorithms (AutoDock(3) and Unity / FlexX), the binding modes vary. Thus, virtual docking should only be seen as an approach for estimating the binding affinity of a ligand to a given receptor.

Compared to the fragment based technique, grid based methods take advantage of a different concept.[69] First, a branch atom and active torsions (rotatable bonds) are defined for a ligand by the operator. During the docking run, the defined bonds around the branch atom of the ligand are rotated, thereby creating distinct conformations of the ligand in the active site. The active site of the receptor molecule is defined by a three-dimensional lattice of regularly spaced points (grid) where the ligand is allowed to change its conformation for an optimal interaction with the receptor. One grid map is calculated for each of the predefined atom types. This means that a ligand atom is placed at each grid point. The interaction energy of this atom with the receptor is assigned to the grid point. Grid size and spacing between the grid points is set by the user. Subsequently, a docking search algorithm (e.g. a

genetic algorithm as the best choice) creates different conformations of the ligand embedded in the grid.[72] For each resulting conformation, a scoring function evaluates the energy for the interaction, and outputs those conformations with lowest docking or binding energy, respectively. The binding energy is the sum of all intermolecular interactions plus torsional energies, whereas the docking energy is the sum of intermolecular and internal energies. All conformations can also be clustered due to their RMS deviations (RMSD). Here, the lowest-energy conformation is placed in the first cluster; if a conformation differs more than a predefined RMSD tolerance from the first one, it is placed into a new cluster. Otherwise, it is added to the first cluster.[69]

As mentioned above, the docking results are ranked according to their score and to a docking energy, respectively.[73-75] However, the conformations with the highest score / docking energy do not necessarily need to be the most realistic ones. Many docking tools like FlexX or AutoDock do not, for example, include bulk solvent molecules surrounding the receptor and the ligand in their calculations.[69,76] Whereas FlexX is only able to place water particles in the active site, AutoDock calculates at least solvation energy parameters for the receptor. According to this, the score / docking energy is only one of several criteria for the evaluation of docking results.[77] In general, the following criteria should also be taken into consideration:

Orientation and position of a ligand in the active site: the docked conformation has to be placed into the pocket with realistic orientations of its functional groups, e.g. hydrophobic parts should usually not be solvent-exposed.

Essential interactions of ligands with the residues of the receptor: the extent of hydrogen bonds, ionic, or hydrophobic interactions has to be considered.

The cluster size: big clusters usually contain the most meaningful and reliable complex conformations.

Finally, it is always needed to experimentally check the results obtained by docking runs, e.g. via NMR spectroscopy or isothermal calorimetry (ITC). As all present docking algorithms are not perfect, most of the potential docking hits do either not bind to the desired receptor or only bind with moderate binding constants in the millimolar range.[63,64]

1.2.3 QSAR as Virtual Similarity Search Method

Quantitative structure activity relationship (QSAR) is a technique where structural features of a molecule, represented by descriptors, are quantitatively correlated to (bio)chemical, physical, or biological properties.[78] The correlation is derived from a selected set of compounds with known structure and activity (training set). Mathematically depicted, n descriptor variables x are correlated to a dependent variable y (e.g. the biological activity of a certain compound; see also Equ. 1.1):

$$y = f(\sum_{1}^{n} x_n)$$

Equation 1.1: Variable y, e.g. an experimental value like a binding constant, as function of a number n of descriptor variables x; descriptors are physical properties or chemical groups of a ligand which are preferred for showing high receptor binding affinity.

Once having established such a correlation, it is theoretically possible to derive the dependent variable y of a novel compound by using the descriptor variables x yielded via the training set.[79] According to this, the main advantage of QSAR becomes obvious: no structural information about a biological receptor (e.g. a protein) is needed for the prediction of the biological activity of a set of ligand molecules.[60] However, a considerable number of QSAR studies incorporate information about the target in order to obtain a more significant correlation.

Since QSAR is a method for deriving correlations between structure and activity, it can be assigned to the field of similarity searching.[60] The objective of a similarity search is to retrieve molecules that are similar to a target compound of interest (e.g. one that is known to be biological active) in some way. The assumption that justifies such searches is embodied in the “similar property principle of reference” which states that structurally similar molecules are expected to possess similar properties.[80,81] Thus, a pairwise comparison of each compound with a target molecule is performed; the compounds are then ranked according to their similarity with the target. The most similar molecules may be used for further searches or structure-activity and pharmacophore studies. In the last decades, several methods for determining the similarity between two molecules have emerged,[82] with the most important ones are described below:

Fragment-based approach: this method evaluates the similarity of a pair of molecules by identical fragments. As this requires only two-dimensional analyses, the fragment based method is very fast. However, it lacks accuracy as no three-dimensional arrangements - which are crucial e.g. for the biological activity - are taken into consideration.

Distance-based approach: with increasing computational power, software packages like CONCORD (Tripos) have the potential to convert two-dimensional structures into the respective spatial conformations. These conformations can be compared to each other by calculating the distance between pairs of atoms. By means of this strategy, three-dimensional arrangements are amenable.

Surface-based approach: going further than the atomic level, molecular properties (such as electron density or a hydrophobic potential) can be projected onto the surface of a sphere. Thereby, the similarity of two molecules is estimated by the distance of the sphere surface points.

Field-based approach: the similarity is determined by a comparison of fields (e.g. steric, electrostatic, hydrophobic fields). The property of interest is sampled on each lattice intersection of a grid surrounding the molecule. The values of such grid points are then compared to the values of another molecule. Most of the recent QSAR studies make use of such field based methods, namely comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analyses (CoMSIA).[83,84]

In general, establishing a 3D-QSAR model requires several major steps. In the beginning, one has to consider the different types of 3D-QSAR methods and decide which one is best suited for the scientific question to be answered. The most frequently used methods are CoMFA and CoMSIA. After that, the training set molecules are chosen from a chemical library. This is the most crucial step in QSAR analyses, as the training set composition in particular influences the quality of the model, as further described below figure 1.8. In the following, the training set compounds have to be oriented uniformly in space (“alignment”) which is achieved either by pairwise alignment or determination of compound orientations in the active site of a receptor. In the next step, the similarity between aligned molecules is calculated and - as a result of this - according molecular fields can be created. Having confirmed the robustness and

predictability of the derived correlation, the activity of novel compounds (not present in the training set) can be finally predicted. A visual examination of QSAR fields usually reveals structural properties that are either favorable or unfavorable for biological activity as it is shown in figure 1.8.

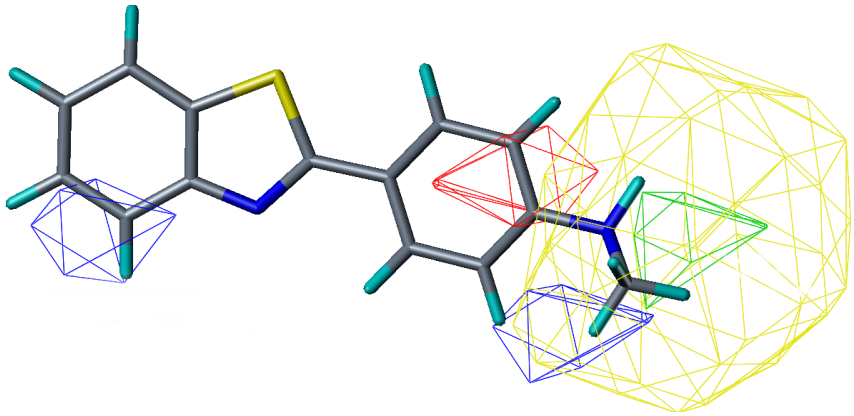


Figure 1.8: Illustration of QSAR derived CoMFA and CoMSIA fields which represent preferred or unfavored properties of a ligand molecule to be able to bind to a receptor of interest. For example, the blue, red, yellow, and green spheres denote areas where certain properties like steric bulk or a polar / apolar group are necessary prerequisites of a ligand to display high or low receptor affinity.

As mentioned above, the composition of the training set has enormous influence on the quality of the QSAR-model. For evaluation purposes, cross-validation is often used to estimate the external predictability of the created models.[85] This statistical method divides the training set into two groups. The QSAR model is re-derived from one group and used to predict the dependent variable for the compounds in the other group ("unknown" compounds). Then, predicted values are compared to experimental values, allowing a quantitative measure for the accuracy of prediction. This process is repeated several times with different group compositions. If one the groups consists of only one compound, cross-validation is referred to as leave-one-out (LOO) validation. However, the external predictability of a QSAR model is usually not determined only in this way. The term external predictability refers to the accuracy with which the activity of a novel compound which is not in the training set is predicted. An obvious and reliable way for assessing external predictability is to predict compounds with known biological activity that are not in the training set. The entity of these compounds is called the test set. The final output of the cross-validation are q^2 values between 0 and 1.0, where 1.0 indicates highest accuracy. Nevertheless, q^2 is only a rough indicator for the quality of a QSAR model and experimental validation is indispensable. In all cases, where activities of compounds which hardly resemble the molecular scaffolds present in the training set have to be predicted, q^2 has no explanatory power for the external predictability.[82,86]

Besides the cross-validated q^2 values, partial-least-squares algorithms additionally report r^2 parameters.[85] The value r^2 is a measure for the proportion of activity-change that can be explained by (a variation) of the descriptor variables (linear dependence). Again, a r^2 value of 1.0 represents maximum linear dependence of activity from descriptor parameters. If there is no linear dependence, r^2 values have absolutely no predictability power. Therefore, r^2 can be an additional measure for the external predictability of an established QSAR model.

1.2.4 The Concept of Molecular Dynamics Simulations

Among the techniques to predict the behavior of molecules or to complement experimental results, the concept of molecular dynamics (MD) simulations belongs to most popular computational chemistry approaches. During the last decades, MD calculations have become more and more important in chemistry and biophysics.[29,87,88] Both enhanced computer performance and improved molecular models enable realistic simulations of the behavior of different kinds of molecules at atomistic resolution.

In principle, those molecular processes which rely on weak, non-binding interactions between atoms (e.g. peptide and protein folding or intermolecular complex formation) are fundamental for the behavior of biomolecular systems as they determine the thermodynamic characteristics of the condensed phase. Since the energies that provoke such processes are in the range of 1-10 k_BT (according to a few ten kJ / mol; k_B is the Boltzmann constant), the laws of statistical mechanics, i.e. potential energy functions as part of classical Hamiltonian operators, can be applied to describe complete molecular systems.[29] Thus, it is possible to simulate the structure, the dynamics, and the motion of small, medium-sized, and even large molecules like peptides, proteins, nucleic acids, sugars, or lipids.[29] Compared to quantum-mechanical (QM) calculations, MD simulations often provide data that are of almost the same quality but can be gained in much shorter time, especially for more complex biomolecular systems.

Molecular dynamics considers a molecule as a collection of spheres (atoms with partial charges and a certain volume) which are connected by springs (bonds) with different elasticities (force constants). The forces underlying intra- and intermolecular interactions can be described by potential energy functions

of structural features like bond lengths, bond angles, non-bonded interactions, and several others. The combination of all potential energy functions is called “force field” (FF) which represents the backbone of all molecular dynamics simulations. By means of the force field parameters, Newton's equations of motion can be numerically solved, thereby resulting in modeling the dynamical behavior of a molecular system.[29]

Each molecular dynamics simulation is accompanied by the initial choice of the modeling level. First of all, the atomic or molecular degrees of freedom which should be explicitly treated in the model must be defined (e.g. simulating with or without solvent, with bond lengths vibrations or not). The next step is the selection of an appropriate force field for the calculation of the energy of the simulated system which depends on the before appointed degrees of freedom. Furthermore, one has to decide how the configurational space should be sampled with respect to all defined degrees of freedom. At last, the boundary conditions (e.g. system borders, temperature, pressure) and possible external forces like distance or dihedral restraints have to be adjusted. Based on this setup, biomolecular systems comprising 10^5 or 10^6 atoms or particles can be effectively simulated.[29] Compared to the Avogadro number which represents a scale for macroscopic systems this is only a small number of particles; however, the results of simulations of only a small part of a macroscopic system can often be used to explain its experimentally measurable properties.

As mentioned above, MD simulations provide very useful hints how molecules behave under certain boundary conditions and can, therefore, explain experimental results or guide new measurements. However, one has to keep always in mind that results from molecular dynamics simulations are only approximations.[29] Therefore, the next five sections (1.2.4.1 – 1.2.4.5) shortly deal with assumptions, prerequisites, and some of the major problems associated with MD calculations and how they can be solved or, at least, diminished. It should be noted that the following chapters represent not an exhaustive discussion about molecular dynamics simulations but give a more general overview.

Figure 1.9 shows the basic principle of a molecular dynamics simulation run. For more technical and practical details, the reader is referred to the tutorial given in chapter 1.3.3.

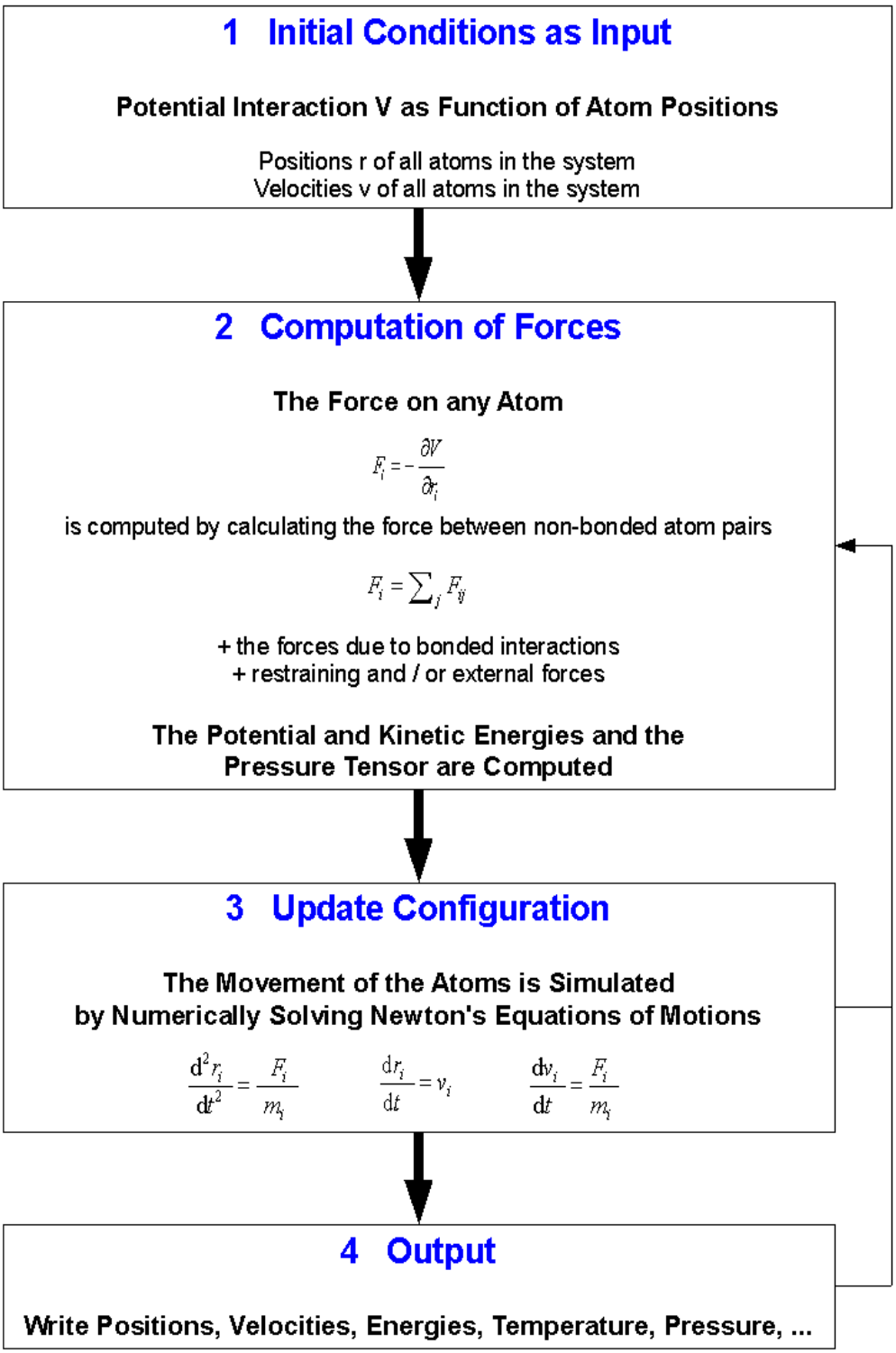


Figure 1.9: The principle of MD simulations. The potential interaction V as a function of atom positions r is provided with velocities v and serves as initial condition. Then, the force F on any atom i (the derivative of its potential V divided by the derivative of its position r) is calculated by summing up the arising forces between all non-bonded atom pairs ij. After that, the forces from bonded interactions plus possible external forces are included into calculations, resulting in computation of both potential and kinetic energies and the pressure tensor. Subsequently, the position of every atom is recalculated by integrating all Newton's equations of motions (for explanations of the formulas presented here, see reference [89]). The output of this computation step serves as updated input for the calculation of new forces. The atom positions and velocities as well as all kinds of energies, the temperature, or the pressure can be written out after each integration step; this option is, among many others, controlled by the MD setup parameters.

1.2.4.1 Force Fields as Backbone of MD Simulations

A force field, e.g. for biomolecules, consists of potential energy terms which form the basis of both covalent (bonds, angles, proper and improper dihedrals) and non-covalent interactions (Coulomb and van der Waals interactions).[14,90] For these terms, appropriate and consistent parameters must be defined. This

task is of high importance and, at the same time, very difficult since a small as possible set of parameters is usually used to describe the plethora of conceivable interactions. The idea behind this concept is that a simple and small set of force field parameters allows the transferability between different atoms or groups of atoms, thus making the MD approach very efficient. Geometric parameters are often found in crystal structures of small molecules, and the corresponding vibrational force constants are elucidated via infrared spectroscopic data on small molecules in the gas phase. Parameters for proper dihedrals are commonly obtained by fitting torsional energy profiles to quantum-mechanical data. Coulomb and van der Waals parameters are usually obtained by means of heats of vaporization, pure liquid densities as well as free solvation energies of small molecules in polar and apolar solvents. MD runs are then used for fine-tuning of the before determined force field values by comparison of calculated with experimental findings. Electron densities derived by QM methods are often taken as initial guess of partial atomic charges; due to the polarizability of electrons which are not taken into account in molecular dynamics simulations, their values must be finally empirically adjusted.[29,91,92]

Considering its parametrization, a force field can only be an approximation of "real" physical interactions. However, when parameters are carefully chosen on a thermodynamic basis, i.e. also entropic effects which often govern non-bonded interactions are considered for the FF parametrization, simulation results can reach the quality of those gained via experiments. For this to work, the accuracy of the individual force field terms in summation must be higher than the total non-bonded energy which is a great challenge since the energy differences that drive biomolecular processes are of the order of only 1-10 k_BT and result from thousands of (pairwise) interactions.[29]

1.2.4.2 Searching the Conformational Space

Biomolecular systems are characterized by a huge amount of degrees of freedom. The motions along these degrees of freedom are of harmonic, inharmonic, chaotic, or diffuse character, respectively, and there are spatio-temporal correlations. As a consequence, those regions of the energy hyper surface - determined by the potential energy functions of a force field - which contribute most to the free energy of the system are hardly traceable. If the energy hyper surface is expressed in terms of molecular conformations, it

becomes obvious that there is no single structural state of a system that corresponds to the global free energy minimum. In fact, by help of statistical mechanics, an ensemble of conformations has to be determined where the probability of occurrence of a single conformation is given by the Boltzmann factor (see Equ. 1.2). Due to the exponential weight in equation 1.2, regions of high energy on the hyper surface do not result in conformations that are relevant for the overall status of a system unless they arise very often (enthalpy-entropy compensation). In other words, the equilibrium properties of a system depend on those regions of the energy hyper surface where the conformational energies are low.

$$P(x) \approx \exp\left(\frac{-V(x)}{k_B T}\right)$$

Equation 1.2: The probability P of occurrence of a specific conformation x in an molecular ensemble is related to the exponent of its negative potential energy V divided by the product of the Boltzmann constant k_B and the absolute temperature T.

The challenge of advanced MD simulations is to identify those regions on the vast energy hyper surface which are of low energy. In principle, there are two approaches to face the problem of insufficient sampling: systematic and heuristic techniques.[93-95] Systematic MD algorithms usually search the whole conformational space for regions of low energy whereas heuristic ones are developed to only screen a small (predefined) region of the energy hyper surface. In order to overcome high energy barriers which separate local energy minima, three major techniques are often applied: deformation or smoothing of the potential energy surface (e.g. “local elevation”),[96,97] scaling of system parameters (e.g. “simulated annealing”),[98] and multicopy searching and sampling (e.g. “replica exchange”).[94] Even if advanced sampling methods are applied, it is difficult to judge if a simulated system is converged with respect to the properties of interest. Whereas present computer power limits the lengths of MD simulation runs to some hundreds of nanoseconds, time scales of relevant biomolecular dynamics range from picoseconds to seconds or even longer. Thus, MD trajectory averages are only representative when the relaxation time τ_{relax} (Q) of a property Q is both shorter than the equilibration period of a simulation τ_{equil} and much shorter than the sampling period τ_{sample} (see Equ. 1.3). If the conditions illustrated below are not fulfilled, the trajectory average with respect to property Q is not converged and thus shows a temporal drift.[29]

$$\tau_{equil} > \tau_{relax}(Q) \qquad \tau_{sample} >> \tau_{relax}(Q)$$

Equations 1.3: An average property Q of a simulation run is only meaningful if the equilibration time τ_{equil} is longer and the sampling time τ_{sample} is much longer than the relaxation period τ_{relax} of this property.

To get an indication of the magnitude of τ_{relax} (Q), one can calculate the decay time of the autocorrelation function or the build-up rate of the trajectory average of a property Q. In addition, if various MD simulations with different starting conditions do not result in the same trajectory average of a property Q, it becomes obvious that the equilibration and / or the sampling time is too short with respect to the relaxation time of this properties (“no convergence”).

1.2.4.3 The Calculation of Molecular Ensembles

As mentioned above, the behavior of biomolecular systems can only be characterized by the help of statistical mechanics. This means that elucidation of one minimum-energy conformation is not sufficient to describe a complete system. Instead, the state of a system at certain conditions (temperature, pressure, etc.) comprises a Boltzmann ensemble of structures. To allow for this fact, an additional term - the entropy S - is incorporated into the physical expression of the free energy. S is defined as the negative derivative of the free energy F with respect to the temperature. Thus, the presentation e.g. of a single molecular conformation to describe a system is only meaningful if one is interested in a structure at absolute zero temperature. Since this is usually not the case, not only energetic but also entropic effects must be taken into account when a system should be conformationally characterized. As a consequence, structural analysis of a molecular system should always, or better, must encompass the elucidation and discussion of molecular ensembles.[29] The rationale behind this concept becomes reasonable when the folding-unfolding equilibria of peptides and / or proteins are studied. At physiological temperatures, only parts of the molecules of a macroscopic system are usually in the folded state. Additionally, scaling of the temperature results in shifting of the folding equilibrium. Hence, it is senseless to describe such a system by presenting only the folded or the unfolded state of the molecule of interest.

The importance of considering both enthalpic and entropic effects, and thus, molecular ensembles, becomes even more obvious when experimental data should be explained by means of molecular dynamics simulations. For

example, peptides usually adopt several conformations which are in exchange in the solution state.[28] Since almost all experimental techniques like NMR spectroscopy provide only time and ensemble averages of an observable such as NOE peak volume intensities or scalar J-couplings, measured values often do not correspond to a single, realistic conformation of a molecule. In such a case, consistency with experimentally gained parameters is only reachable if a structural ensemble instead of a single conformation is calculated.[29]

1.2.4.4 Comparison of Experiments with Simulation Results

For a number of reasons, MD results are usually compared with experimental findings. First, the development of force fields or new MD algorithms is only possible when test runs are in agreement with experimental data. Secondly, experimental outcomes are often not easy to interpret, and MD simulations may help to understand them. Finally, the results of MD simulations have to be evaluated with respect to their reliability; thus, a careful comparison with experimental results is indispensable.

However, one of the major problems in this scope is the fact that almost all experiments deliver only averages over time (duration of a measurement), space (the sample volume), and molecular ensembles (a macroscopic system). Due to this, informations about the distribution of a property Q are not available. In addition, completely different distribution of Q can result in the same average (see Fig. 1.10A).[29] For example, if a molecule adopts two different shapes which are in fast exchange at the NMR time scale, the resulting NOE peak volumes represent a conformational average of differing distances between two atoms. Additionally, measured J-couplings are often very insensitive to underlying conformational distributions. As a consequence, distributions gained via MD simulations are not directly comparable with experimental outcomes.[30,99] Furthermore, even if a “simulated“ average is in agreement with experimental results, the MD distribution underlying the average could be erroneous due to error compensations.[29] Hence, it is advisable to use as many as possible experimental parameters - yielded by different experimental techniques - to evaluate MD results with respect to their reliability. Another problem of comparing experimental with simulated results is illustrated in figure 1.10B. If e.g. a conformation of a molecule in solution is analyzed by the help of MD simulations, the source of experimental results must be regarded. Whereas NMR spectroscopy data are recorded under

similar conditions, X-ray crystallography provides only insight in parts of the conformational space which is available in the solution state.[100-102] Therefore, the outcomes of an MD run should only be compared with experimental findings recorded under similar conditions. Moreover, MD results need not to be artificial if they do not match experimental data. If a calculation provides "solutions" which can not be explained by experiments, not only the simulations have to be checked for inconsistencies but also experimental data should be scrutinized.

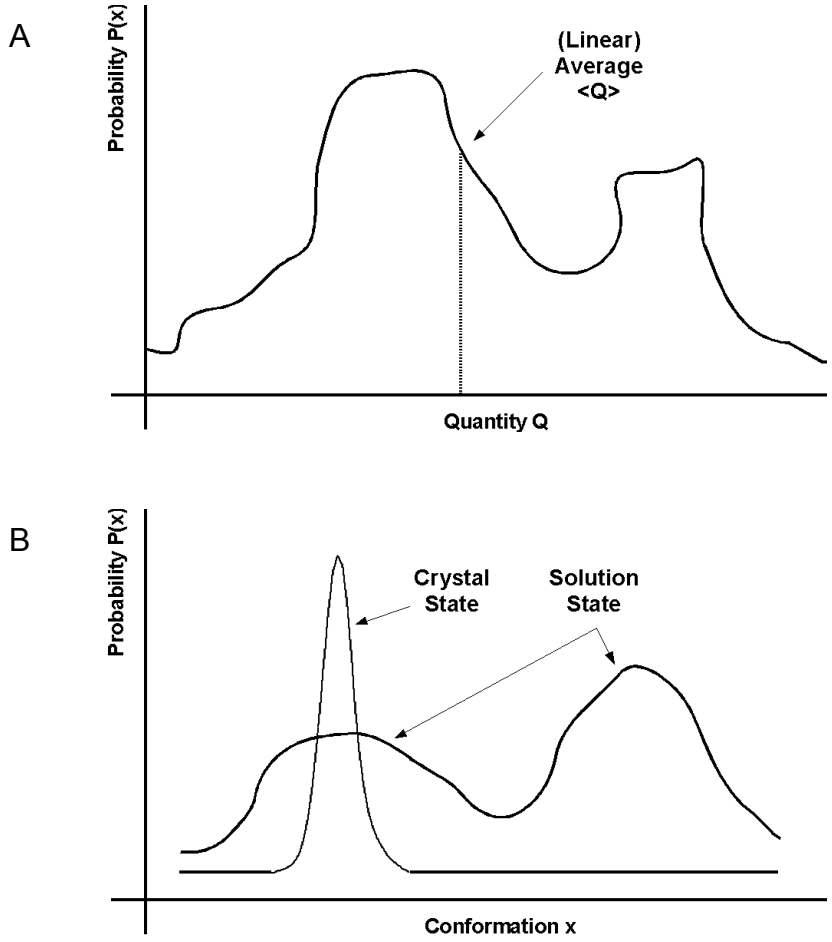


Figure 1.10: Illustration of two problems which guide the comparison of calculated data with measurement results.

A) NMR experiments only provide a (linear) average of a quantity or property Q. The underlying distribution of recorded averages can not be extracted via experiments and is, therefore, unknown. MD simulations which result in the same average of a property Q like a measurement need not to be „realistic" since the distribution that lead to the simulated average can be "wrong".

B) The comparison of experimental outcomes with MD simulation results is only meaningful if the boundary conditions used for both approaches are similar. For example, a conformation x elucidated by X-ray spectroscopy can not reflect the structure which a molecule adopts in solution. Therefore, MD simulations often provide 3D arrangements that are not identified by X-ray crystallography.

At least, discrepancies between experimental outcomes and MD simulation results can arise due to a insufficient number or inaccuracies of measured data. Compared to the number of degrees of freedom a flexible molecule intrinsically has, experimental results are often not sufficient to exhaustively determine a biomolecular system. In addition, experimental results are often afflicted by uncertainties and inaccuracies. Thus, the value of MD simulation results can only be adequately evaluated if enough experimental data are collected and, in particular, are precisely determined.[29]

1.2.4.5 The Operator Problem

Beside all assumptions and approaches, the main problem of the MD technique is an operator who is not familiar with a virtual technique like this. In contrast to experimental methods, molecular dynamics simulations in almost all cases produce an output; at a first glance, such results often seem reasonable and the calculated parameters are mostly presented as impressive and colorful pictures. This is especially true for the determination of three-dimensional structures of peptides and their analogs. Generally, it is not sufficient to only be able to start and analyze a simulation run; rather, the operator must be aware of the meaning of each MD parameter, how tuning of them may influence a trajectory, and how a run can be evaluated regarding its reliability. In other words, each MD calculation must be accurately planned and completely reflected, all MD parameters must be carefully adjusted, and all results must be exhaustively checked with respect to consistency and significance and compared with all experimental findings which are available. For this reason, it is a must to know the original literature where all aspects regarding force fields, adjustable parameters, interaction functions, boundary conditions, and potential vulnerabilities are explicitly discussed.[14,89,92,103]

In the following chapter 1.3 which comes in form of a tutorial, some of the most important sources of error in structure calculations are discussed and established procedures in MD aided calculation of peptide conformations are presented.

1.3 Determination of Dynamic Peptide Conformations: A Tutorial

NMR based elucidation of three-dimensional arrangements of peptides clearly differs in its strategy compared to protein structure determination. Even though the experimental parameters used in both approaches are, in principle, the same, their contribution to the elucidation process is quite different with respect to their emphasis, precision, and importance.[1] Since many peptide conformations in literature have been determined via typical protein structure elucidation procedures, their reliability seems to be questionable. In addition, even if obtained structures are well characterized, important aspects regarding

intrinsic flexibility are often neglected. Because of that, chapter 1.3 provides an overview how proper structures of peptides and peptide mimetics including their dynamical behavior can be calculated. The underlying NMR experiments, the most important experimental parameters, and the process how they are converted into dynamic three-dimensional arrangements are mainly focused. It will become obvious that only the interplay of NMR spectroscopic techniques and computational tools like distance geometry and molecular dynamics simulations result in both static and dynamic molecular models high quality.

1.3.1 NMR Spectroscopy as Backbone of Conformational Studies

In principle, there are several NMR spectroscopic parameters that could be used as for the calculation of proper molecular conformations. Compared to chapter 1.1.1, explicit details of the NMR experiments and the extracted parameters are presented in the following sections. In addition, arising problems and adequate solutions are discussed in more detail.

1.3.1.1 Elucidation of Homo- and Heteronuclear J-Couplings

Scalar couplings provide useful insight in the constitution and the conformation of peptidic molecules.[1] The most important J-couplings applied in structure elucidation of peptides are listed below:

a) *direct couplings:* $^{1}J_{CH}$, $^{1}J_{NH}$, $^{1}J_{CC}$, $^{1}J_{NC}$
b) *geminal couplings:* $^{2}J_{HH}$, $^{2}J_{NH}$, $^{2}J_{CH}$
c) *vicinal (long-range) couplings:* $^{3}J_{HH}$, $^{3}J_{NH}$, $^{3}J_{CH}$
d) *long-range couplings:* $^{n}J_{HH}$, $^{n}J_{NH}$, $^{n}J_{CH}$ *(with* $n \geq 3$*)*

Although easy to measure and generally useful, heteronuclear ^{1}J-coupling are barely applied as structural restraints; one exception are $^{1}J_{C\alpha H\alpha}$ couplings.[104] However, ^{1}J-couplings as well as geminal couplings can be used to identify atoms that are connected by a chemical bond (determination of the constitution). The vast majority of conformational restraints are dihedral angles derived from homonuclear ^{3}J-couplings; heteronuclear ^{3}J-couplings are rarely applied although they contain very important configurational and conformational information.[105-107] Practically all vicinal couplings obey the empirically derived Karplus relation (see Equ. 1.4).

$$^3J = A + B \cdot (\cos\varphi) + C \cdot (\cos^2\varphi)$$

Equation 1.4: The general form of the so-called Karplus equation. φ as the dihedral angle between the three bonds constituting the ^{3}J-coupling; A, B, and C are varying fitting constants which have to be adjusted for specific constitutional motifs. Typical fitting constants for $^3J_{HNHC}$ couplings can be found in [108] and for $^3J_{CH}$ couplings in [109].

For the measurement of long-range homonuclear and heteronuclear scalar couplings (including ^{3}J-couplings), a large number of experiments has been developed. The extraction of $^3J_{HH}$ couplings from 1D spectra is routinely used and often just the splittings in the multiplets are indicated as "coupling constants". It should be noted that this is only valid for first order spin systems. In case of signal overlap or the measurement of very small couplings, 2D methods exhibit distinct advantages. Even if multiplet components are not baseline separated, exact values for couplings can be fitted from COSY spectra calculated in absorption and dispersion mode.[110] E.COSY spectra are used for spin systems involving more than two spins.[111]

Long-range couplings (over more than three bonds) are only in rare cases directly readable from 1D spectra. If only qualitative values are needed, cross-peak intensities in heteronuclear long-range (e.g. HMBC) spectra yield sufficient information. The technology originally developed for carbon detected methods is transferred into one of the most often used methods for quantitative extraction of heterocoupling via measuring the build-up of cross peaks depending on coupling evolution times.[112-114] In systems with more than two coupled spins the heteronuclear analogues of E.COSY- type spectra, e.g. HETLOC, provide reliable values of scalar couplings.[115]

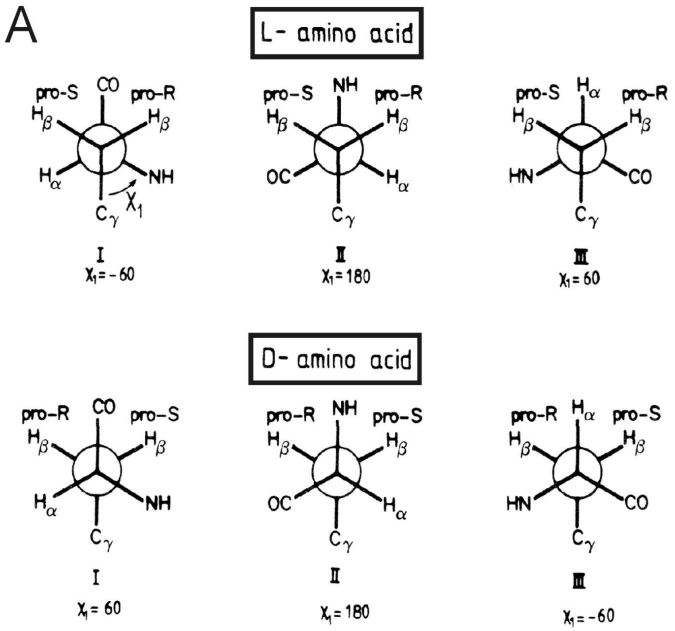
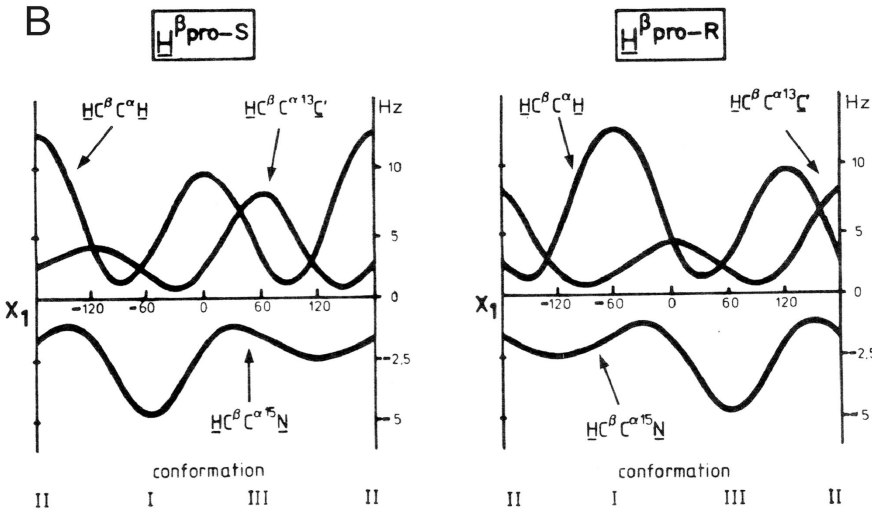


Figure 1.11: Example of how pro-R and pro-S of amino acid side chain methylene protons in the β-position of peptides are usually assigned.

A) Schematic presentation of the main conformations a methylene group adopts in L- and D-amino acids.

B) Illustration of Karplus relations for the corresponding $^3J_{H\alpha H\beta}$, $^3J_{H\beta C'}$, and $^3J_{H\beta N}$ scalar couplings. *(taken from reference [1])*

Continuation of Fig 1.11



Beside their application in determining molecular constitutions and backbone conformations, scalar couplings play an essential role in the elucidation of amino acid side chain rotamers and in the assignment of pro-chiral protons. Whereas proteins nowadays are almost exclusively studied as ^{13}C and ^{15}N labeled isotopomers, peptides usually have these isotopes only in natural abundance, i.e. the magnetically active heteronuclei are highly diluted. Most amino acids contain a methylene group at the *β*-position for which the χ angle can be determined via the conformation of the Cα-Cβ bond (see Fig. 1.11 A). Two vicinal J-couplings can be measured: Hα to $H\beta^{pro\text{-}R}$ and Hα to $H\beta^{pro\text{-}S}$. Usually, one can assume that the three staggered conformations are in an energetic minimum state. Two types of vicinal proton J-couplings have to be considered: the anti-periplanar (trans) arrangement leads to a large scalar coupling of about 12 Hz whereas the syn-clinal - gauche(+) or gauche(-) - results in 3.5 Hz (see Fig. 1.11 B). Hence, conformation III as shown in figure 1.11 A yields two small J-couplings and the sum of $^3J_{H\alpha H\beta R}$ and $^3J_{H\alpha H\beta S}$ is about 7 Hz. In conformations I and II of figure 1.11 A the sum is 15.5 Hz. If the three rotamers are in a fast exchanging equilibrium, the sum of both vicinal couplings is an indicator of the relative population of III in the equilibrium. In addition, two identical homonuclear couplings indicate that both conformations I and II are equally populated, whereas if they differ (in the most extreme case they are 12 Hz and 3.5 Hz) one of the conformations (I or II) dominates the equilibrium. Which of the two conformations is the higher populated one can only be decided from the assignment of the diastereotopic *β*-protons. In principle, this can be done via NOE / ROE spectra. However, it is strongly recommended not to perform the assignment by NOE parameters which are

later on used for the determination of the 3D structure; for this purpose, the measurement of heteronuclear ^{3}J-coupling should be preferred. If the population of conformation III can be neglected, only the *β*-pro-R in L-amino acids is in antiperiplanar arrangement to $^{13}C'$ and exhibits a strong cross-peak in the heteronuclear long-range correlation (which can be extracted by HMBC spectra). In conclusion, the sum of $^{3}J_{H\alpha H\beta R}$ and $^{3}J_{H\alpha H\beta S}$ yields the population of III, their difference the preference of I or II and the heteronuclear vicinal coupling decides if I or II is preferred. In ^{15}N-labeled proteins the same procedure can be applied to assign *β*-pro-R and *β*-pro-S using the large ^{15}NH-*β*-pro-S coupling in rotamer I.

Recently, it has been shown that couplings between ^{15}N and ^{13}C across hydrogen bond, e.g. in systems containing (^{15}N-H···O=^{13}C) units, can be directly detected and provide evidence for hydrogen bonds in biomolecules.[116] This method is used for peptides and larger molecules but is rarely applied for small organic compounds.[117]

1.3.1.2 Collecting Pairwise Distance Informations

Nuclear spins can be considered as dipoles that interact with each other via dipolar couplings. While this interaction leads to strongly broadened lines in solid-state NMR spectroscopy, it is averaged out in isotropic solution due to the fast tumbling of the solute molecules. In liquid-state NMR spectroscopy, the dipolar interaction can only indirectly be observed by relaxation processes where it represents the main source of longitudinal and transverse relaxation. For an isolated pair of ^{1}H nuclei with their magnetization oriented along the static magnetic field and considering only dipolar interactions, the cross-relaxation rate for the NOE is given by equation 1.5:

$$\sigma^{NOE} = \frac{\hbar^2 \mu_0^2 \gamma^4}{40\pi^2} \cdot \frac{\tau_c}{r^6} \left\{ -1 + \frac{6}{1 + 4\omega_0^2 \tau_c^2} \right\}$$

Equation 1.5: The cross-relaxation rate σ^{NOE} for the NOE. The parameter ħ as the Planck constant divided by 2π; μ_0 as the permeability of vacuum; γ as the gyromagnetic ratio, r the distance between the two relaxing nuclei, ω_0 the Larmor frequency of the nuclei, and τ_c as the rotational correlation time for the tumbling of the molecule.

The decisive part of the cross-relaxation rate is the r^{-6} dependence of σ^{NOE} with respect to the internuclear distance r. Since only τ_c is variable for a given spectrometer frequency and can be considered constant for a rigid molecule under defined conditions, measurement of the internuclear relaxation rate

directly provides distance information within the molecule of interest. The cross-relaxation rate can be quantified e.g. from cross peak intensities in two-dimensional NOESY spectra; here, the cross-relaxation induces a build-up of cross peak intensities as shown in figure 1.12A for a typical small peptide with a correlation time $\tau_c \approx 0.1$ ns for several internuclear distances at a spectrometer frequency of 600 MHz (which corresponds to the ^{1}H Larmor frequency ω_0).

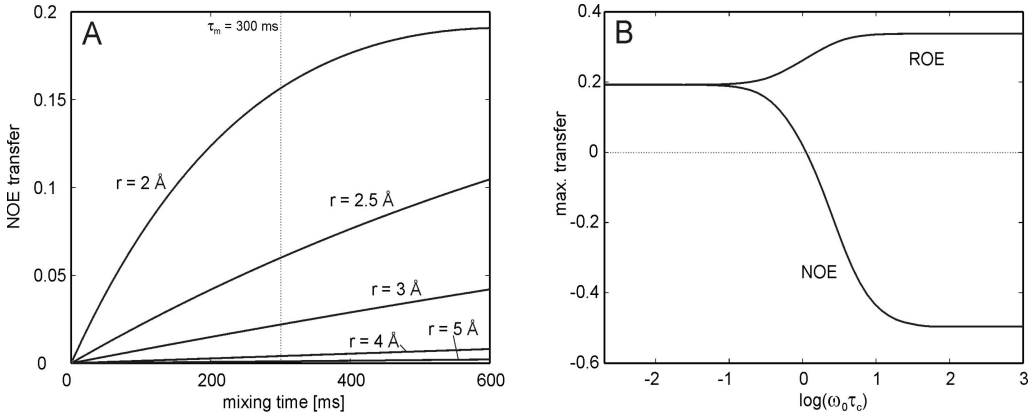


Figure 1.12: Transfer of magnetization via NOE / ROE cross-relaxation in an isolated spin-pair.

A) Build-up curves for the cross peak intensity in a 2D-NOESY experiment for various exemplary internuclear distances r. The dashed line indicates a mixing time of τ_m = 300 ms usually applied for small and medium-sized peptides. For the corresponding two protons at ω_0 = 600 MHz, a correlation time of τ_c = 0.1 ns was assumed.

B) The maximum magnetization transfer efficiency for an isolated proton spin pair calculated using exclusively dipolar relaxation processes. Note the sign-change for the NOE cross-relaxation at $\omega_0\tau_c \approx 1.12$. Also shown is the cross-relaxation of the NOE in the rotating frame (ROE). *(taken from reference [1])*

Integration of cross peaks at a certain mixing time τ_m, as for example at τ_m = 300 ms where the build-up curve still increases approximately linear, directly provides the r^{-6}-encoded distance information (linear two-spin approximation). With the calibration of experimental cross peak intensities by a known inter-proton distance, e.g. $r_{HH} \approx 1.78$ Å within a CH_2-group or $r_{HH} \approx 2.49$ Å for vicinal aromatic protons, distance restraints for structure calculations are readily available via the relationship ($\sigma^{NOE} \cdot r^6$ = const.).

For a reliable extraction of distances, it is important that dipolar relaxation is strongly dominating other relaxation processes. Hence, it is important to avoid paramagnetic ions, molecules such as transition metals, or (paramagnetic) oxygen. Therefore, sample solutions of small molecules should be always degased.

The method outlined so far is generally applicable for obtaining distance information, but several technical limitations have to be considered when adjusting the experimental setup and interpreting the cross peak integrals. The main limitation for medium-sized molecules like peptides is the dependence of σ^{NOE} on ω_0 and τ_c (see Fig. 1.12 B): For molecules for which $\omega_0 \cdot \tau_c \approx 1.12$, the cross relaxation rate σ^{NOE} is close to zero and magnetization transfer between

nuclei can not be observed. At a 600 MHz spectrometer this condition is fulfilled for globular molecules with a molecular weight of roughly ~ 500 g / mol in DMSO or ~ 2000 g / mol in $CDCl_3$ as solvent (at room temperature). A general estimate of τ_c can be gained via Stokes' law for spherical molecules (see Equ. 1.6):

$$\tau_c = \frac{4\pi\eta r_m^3}{3k_B T}$$

Equation 1.6: An estimate of the rotational correlation time τ_c. η as the viscosity of the used solvent, r_m as the effective hydrodynamic radius inclusive of a possible solvent shell, k_B as the Boltzmann constant, and T as the temperature.

The value of τ_c is very roughly $10^{-12} \cdot$ MW (molecular weight in Dalton) at room temperature in an low-viscosity organic solvent.[1] For the relatively viscous and widely used solvent DMSO at room temperature, the majority of small peptides falls in the τ_c range of weak NOE cross relaxation, and the extraction of distance information from NOESY spectra is limited or not possible. In such cases, a viable alternative is the measurement of the rotating frame NOE (also called ROE) which is based on cross-relaxation in the transverse plane perpendicular to the static magnetic field. The cross-relaxation rate for the ROE is given by equation 1.7:

$$\sigma^{ROE} = \frac{\hbar^2 \mu_0^2 \gamma^4}{40\pi^2} \cdot \frac{\tau_c}{r^6} \left\{ 2 + \frac{3}{1 + \omega_0^2 \tau_c^2} \right\}$$

Equation 1.7: The cross-relaxation rate σ^{ROE} for the ROE. The parameter $\hbar$ as the Planck's constant divided by 2π; μ_0 as the permeability of vacuum; γ as the gyromagnetic ratio of protons, r the distance between relaxing nuclei, ω_0 the Larmor frequency of corresponding protons, and τ_c the characteristic correlation time for the tumbling of the molecule.

The same r^{-6} dependence as in the NOE case applies and, as shown in figure 1.12B, the magnetization transfer in ROE-type spectra is always positive. For most small and medium-sized peptides the so-called ROESY experiment therefore is the preferred alternative for obtaining distance information.

In general, pulse sequences that provide conclusive distance restraints are the one-dimensional transient-NOE experiment with the selective inversion of a specific spin or, preferably, two-dimensional NOESY and ROESY experiments.[118-120] The central element in all NOE-type pulse sequences is the mixing time τ_m as shown in Figure 1.12A. Depending on the correlation time τ_c of a molecule, the mixing time should be chosen short enough to prevent errors from non-linear terms in the NOE / ROE build-up curves (spin diffusion) and

long enough to provide sufficient cross peak intensities. For small to medium-sized peptides, mixing times with typical standard values of 150-200 ms for ROESY and 300 ms for NOESY experiments are suggested. Pairwise atom distances are most reliably determined if a series of NOESY / ROESY experiments with different mixing times are recorded and used for fitting the build-up curves for every cross peak individually, but measurements for a single, carefully chosen mixing time also give precious results. Positive cross relaxation as shown in figure 1.12B leads to an increase in signal intensity of a 1D transient-NOE experiment. In two-dimensional NOESY and ROESY spectra, instead, it results in cross peaks with opposite sign relative to the diagonal peaks. Negative cross relaxation, as observed for NOESY spectra for large molecules with long correlation times, yields cross and diagonal peaks of identical sign. Potentially misleading cross peaks in NOESY and ROESY spectra as a result of chemical exchange, spin diffusion, or TOCSY transfer (*vide infra*) all have the same sign as the diagonal peaks. Therefore, such cross peaks can be distinguished in two-dimensional experiments with positive cross-relaxation. However, sometimes contributions to NOESY / ROESY cross peaks resulting from such sources lead to lowered intensities resulting in indistinguishable errors in the distance determination.

Particular care has to be taken when implementing ROESY experiments. The spin lock, which holds the spins along a defined axis perpendicular to the static magnetic field, can be realized in many different ways and is still an active field of research.[120,121] In most spin lock sequences the conditions for undesired TOCSY transfer are partially fulfilled and especially cross peaks close to the diagonal or anti-diagonal might not be accurately interpretable. Since in most cases the efficiency of the spin-lock also depends on the chemical shift offset, an offset dependent correction has sometimes to be applied to the measured ROE cross peak intensities.[121]

In addition to potential experimental errors, a number of systematic errors have to be considered when interpreting NOE / ROE data. As can be seen in figure 1.12A, the NOE / ROE build-up curve is not perfectly linear even in the case of two isolated spins. It is affected by the auto-relaxation rate and higher order transfers that relax magnetization forth and back to the original spin (see Fig. 1.13A). Especially at long mixing times and strong cross peaks for short inter-proton distances, cross relaxation rates are often underestimated. On the other hand, magnetization in larger spin systems can travel from one spin to

its neighboring spin and further to a third spin. This effect, which is called spin diffusion, leads to an overestimation of cross-relaxation rates for distant spins when a "shortcut" exists via intermediate spins (see Fig. 1.13B). Spin diffusion is least efficient at short mixing times which therefore must be considered more reliable for extracting distance restraints.

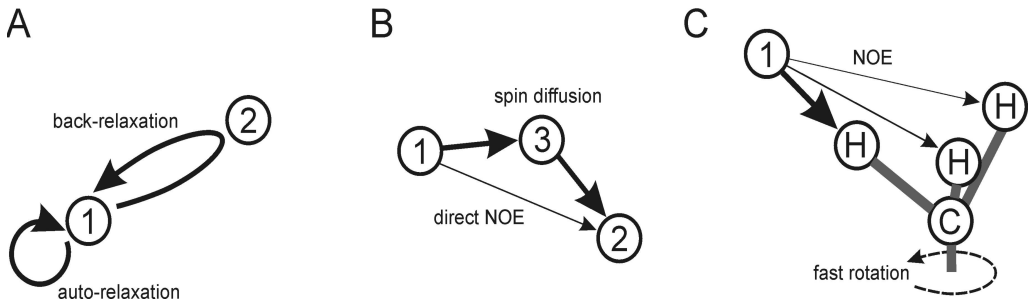


Figure 1.13: Schematic image of different NOE transfer pathways that can result in a misleading interpretation of σ^{NOE} relaxation rates.

A) Spin no. 1 experiences auto-relaxation and back-relaxation. In detail, the magnetization is transferred to spin no. 2 and finally relaxed back to 1.

B) A "shortcut" for the NOE relaxation via spin no. 3 close to spins 1 and 2 can lead to a significant contribution to the cross peak intensity for the direct NOE between spins no. 1 and no. 2.

C) For fast rotating methyl groups the intensity of NOE transfer from spin no. 1 is mainly determined by the closest distance to this spin. This effect is almost always observable if conformational averaging of distances is present. *(taken from reference [1])*

The typical time scale of an NMR experiment is in the millisecond range. All NMR parameters, including $\sigma^{NOE/ROE}$, are averaged over this time scale. If a molecule undergoes significant conformational changes, the r^{-6} dependence of cross relaxation rates will in turn lead to an overestimation of conformations with shorter inter-proton distances. One example is the fast rotation of methyl groups, where the measured cross relaxation rate to another proton is mainly determined by the closest distance between the protons (see Fig. 1.13 C). In such cases, the restraint limits for the upper distances must be increased accordingly in structure calculations.[122,123] Localized dynamics can also lead to variations in correlation times. Folded peptides with unfolded carboxy- or amino-terminal residues, for example, will have varying correlation times for the rigid and flexible parts of the molecule, resulting in different cross-relaxation rates. Such effects can usually be distinguished by the linewidths and intensities of the corresponding diagonal signals since the auto-relaxation rates also depend on the correlation time.[1] It should be noted that the internal distance calibration from one or more cross peaks between protons of known distances underlies the same experimental and systematic errors. Thus, the initial calibration should be used only as a first estimate. After first structure calculations, distances can be back-calculated and a larger number of distances can be used for recalibrating the restraints.

Considering all potential experimental and systematic errors of NOE / ROE cross peak intensities, it is remarkable how robust the derived distance restraints are. The reason for this lies in the r^{-6} dependence of the cross

relaxation rate: Even if a cross peak intensity is determined wrongly by a factor of two, the resulting distance restraint is only affected by the factor $\sqrt[6]{2} \approx 1.12$ which lies within the error range of distance restraints (± 10 %) usually used in structure calculations. It should be further noted that the quality of resulting structures is not so much determined by potential errors of a single distance restraint but rather by the total number of distance restraints per atom.

1.3.1.3 Measurement of Residual Dipolar Couplings

Residual dipolar couplings of a solute molecule can only be obtained if it is partially aligned with respect to the outer, static magnetic field. This means that an intermediate state between solid and liquid has to be reached. With the help of suitable alignment media, solute molecules are only oriented for a time average of about 0.05 % which, for example, reduces a 46 kHz dipolar coupling between a carbon and its directly attached proton to a residual dipolar coupling of only 23 Hz. This coupling adds or subtracts to the direct $^1J_{CH}$ coupling in the order of 130-160 Hz. Hence, a coupling of this size can accurately be measured and does not significantly decrease the spectral resolution. Mathematically, the dipolar coupling D_{IS} between two spins I and S is described by equation 1.8:

$$D_{IS} = 2 \cdot \frac{\hbar \gamma_I \gamma_S \mu_0}{16\pi^2 r_{IS}^3} (3\cos^2\Theta - 1)$$

Equation 1.8: Expression for the dipolar coupling D between two spins I and S. The parameter ħ as the Planck's constant divided by 2π; μ_0 as the permeability of vacuum; the angle Θ as the angle between the internuclear vector and the static magnetic field B_0, the gyromagnetic ratios γ_I and γ_S of the two spins and their distance r_{IS}.

Three different ways of introducing partial alignment are known today: (1) alignment by a liquid crystalline phase, typically a lyotropic mesophase, (2) alignment by a stretched gel, and (3) orientation via paramagnetic ions. Liquid crystals have been the first alignment media introduced and a large number of different systems is known.[124] They orient spontaneously in a magnetic field and the weak interaction with the solute produces the desired partial orientation. Liquid crystalline phases, however, have a first order phase transition and are therefore a limited to a minimum alignment. Stretched gels were introduced by Deloche and Samulski.[125] Their alignment strength is solely determined by the amount of mechanical stretching so that arbitrary scaling of residual dipolar couplings can generally be achieved. Paramagnetic

alignment requires a specific ionic binding site for the solute molecule which is either present naturally or in sometimes can be engineered by a designed tag.

If molecules are weakly aligned, RDCs simply contribute to the splitting of scalar couplings. The difference between couplings measured in an isotropic and the corresponding partially aligned sample directly results in the residual dipolar couplings. In principle, most experiments for measuring scalar couplings can be used this way to extract RDCs. In practice, however, the required precision for measuring residual dipolar couplings is higher than for scalar couplings where usually deviations on the order of 1 Hz are easily tolerated and the sign-information with respect to the coupling is of minor importance. Therefore existing methods have been strongly revised and extended during the last decade and further techniques can be expected in the future. Since peptides as well as small organic molecules are usually only available at natural abundance, we exclusively focus on this situation and neglect experiments specifically designed for labeled macromolecules like proteins or protein assemblies.

The most easily measured RDCs are along one-bond heteronuclear couplings like $^1J_{CH}$ or $^1J_{NH}$. If the alignment strength is correctly adjusted, RDCs are significantly smaller than the corresponding scalar couplings of known sign, so that a sign-sensitive measurement of RDCs is easily achieved. Since the distance between directly bound nuclei is usually well-known and fixed, RDCs can also directly be translated into relative angular information. Conventional HMQC / HSQC experiments can be recorded without heteronuclear decoupling during acquisition and the observed splitting gives the sum of scalar and dipolar couplings (see Fig. 1.14).[126]

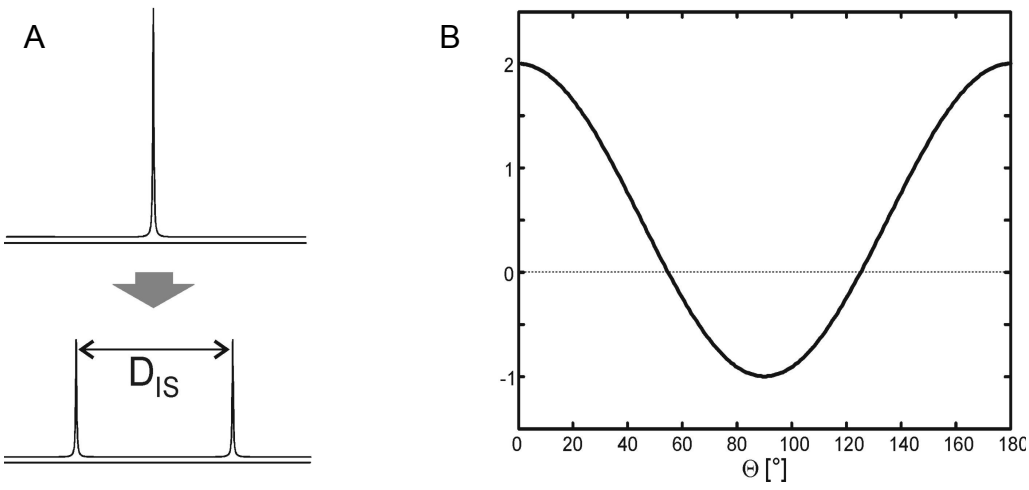


Figure 1.14: Illustration of the residual dipolar interaction as detectable via NMR spectroscopy.

A) The equal distribution of spins parallel and antiparallel to the magnetic field results in a signal splitting with the dipolar coupling D_{IS}.

B) The $(3\cdot\cos^2\Theta - 1)$ dependence of the dipolar coupling on the angle between the spin vector and the static magnetic field. *(taken from reference [1])*

A second class of easily measured RDCs with fixed geometry are along homonuclear $^2J_{HH}$ scalar couplings e.g. in methylene and methyl groups. A number of sophisticated experiments has been developed for the sign-sensitive measurement of these couplings,[127,128] with probably the P.E.HSQC as the most simple experiment especially suited for the investigation of drug-like molecules.[129] A problem with the extraction of homonuclear two-bond couplings are frequently observed second order artifacts which prohibit a simple extraction of J-couplings. Homonuclear RDCs can be measured using quantitative J- and COSY-type measurement techniques for protons and INADEQUATE for neighboring carbons at natural abundance.[130-132] In principle, E.COSY-type experiments provide sign-sensitive information for ^{1}H,^{1}H three-spin systems, but the increased linewidths of aligned samples makes the extraction of couplings very difficult. CT-COSY and quantitative J-techniques involve transfer steps to heteronuclei, which makes them insensitive at natural abundance. The insensitivity of INADEQUATE correlating two heteronuclei at natural abundance is well-known but might partially be overcome by the use of cryogenic probe-head technology.

RDCs for a given interatomic vector depend on the relative orientation of this vector with respect to the static magnetic field averaged over all dynamic processes, including Brownian motion, interactions with the solvent and the alignment medium, and conformational changes. If a molecule can be described by a single conformation, the dynamic averaging is identical for all interatomic connectivities and the interpretation of RDCs is straightforward. A very simple example is the distinction of axial versus equatorial protons in six-membered rings. Since all axial ^{1}H-^{13}C-vectors point into the same direction (or anti-direction), all RDCs being significantly different from these protons directly indicate an equatorial position.[133] More general configurational studies, as for example the identification of diastereotopic protons, can be achieved via the concept of the alignment tensor. The basic principle of this concept is easily understood: while the magnetic field points along the z axis in the laboratory frame and the orientation of the molecule changes over time, we can also place ourselves in a frame that is fixed with the orientation of the rigid molecule where the magnetic field changes its orientation relative to the fixed molecule. The alignment tensor then describes the distribution of the magnetic field in this molecular reference frame. From this alignment tensor, all RDCs of a given structural model are easily back-calculated and the comparison with experimental RDCs can verify or falsify the model. The

alignment tensor for conformational and configurational studies is widely used in the field of biomolecular NMR and in recent years has been also more and more applied to smaller molecules like peptides. Very nice examples include the assignment of diastereotopic protons in methylene groups or the determination of the relative configuration of remote parts in a molecule which are usually very difficult to obtain with conventional NMR methods.[134-138] In such cases, the alignment tensor can be fitted from experimental RDCs from at least five interatomic vectors of different orientation, out of which no more than three vectors lie in one plane.

The full potential of residual dipolar couplings, however, can be seen when they are incorporated into structural analysis procedures. Several programs like XPLOR-NIH, DISCOVER, or GROMACS allow the incorporation of RDCs as angular or combined angular and distance dependent restraints.[1] Due to the sensitivity to angular changes of the backbone relative to the static magnetic field as an external reference, a structural model of an average structure with previously unknown precision can be calculated. If structure calculations are combined with MD simulations, it can be expected that even dynamic structural models including conformational changes of the molecule of interest can be obtained with high precision.

1.3.2 Distance Geometry as Structure Generating Tool

The parameters extracted from NMR measurements (scalar couplings, nuclear Overhauser effect peak volumes, or residual dipolar couplings, etc.) have to be converted into a three-dimensional structure. Having established the constitution by J-couplings, informations about the spatial arrangement are now introduced. Here, mainly distances from NOE / ROE build-up rates are used to define the conformation. In contrast to the determination of protein structures, NOE / ROE based restraints should never be divided in distance classes for structural analysis of peptides;[1] since NMR measurements of smaller molecules like peptides result in less distance informations per amino acid compared to large biomacromolecules, the precision of all used restraints has to be as highest as possible. In general, the first step of the structure calculation is the application of distance geometry (DG) which does not use an energy function but only experimentally derived distances and restraints which follow directly from the constitution, the so-called holonomic constraints. Those constraints are e.g. the distance between geminal protons which normally are

in the range between 1.7 and 1.8 Å or the distance between vicinal protons in biogenic amino acids which can not exceed 3.1 Å when the protons are in anti-periplanar orientation.[1]

Distance geometry was primarily developed as a mathematical tool for obtaining spatial structures when pairwise distance informations are given.[139] The DG method does not use any force fields (see also chapter 1.2.4.1). Thus, the conformational energy of a molecule is neglected and all 3D structures which are compatible with the distance restraints are presented. Nowadays, it is often used in the determination of three-dimensional structures of small and medium-sized peptidic molecules. Compared to force field based methods, distance geometry is a fast computational technique in order to scan the conformational space. To get optimized structures, distance geometry calculations are usually associated with force field based simulations.

The procedure of distance geometry calculations can be subdivided in three separated steps (see below for detailed explanations).[140-142] At first, holonomic matrices with pairwise upper and lower distance limits are generated from the topology of the molecule of interest. These limits can be restrained by NOE derived distance information which are obtained via NMR experiments. In a second step, random distances within the upper and lower limits are selected and stored in a metric matrix. This operation is called metrization. Finally, all distances are converted into a complex geometry by mathematical operations; here, the matrix based distance space is projected into the Cartesian coordinate space (embedding).

1.3.2.1 Build-Up of Distance Matrices

The first step in the DG calculations is the generation of the holonomic distance matrix for all pairwise atom distances of a molecule.[142] Holonomic constraints are expressed in terms of equations which restrict the atom coordinates of a molecule. For example, two hydrogen atoms bound to neighboring carbon atoms have a maximum distance of 3.1 Å. As a result, parts of the coordinates become interdependent and the degrees of freedom of the molecular system are confined. The acquisition of these distance restraints is based on the topology of a model structure with an arbitrary but energetically optimized conformation. By taking the free rotatability of defined atom bonds into account, the upper and lower limits of the distances can be

determined with a tolerance of ± 1 %. If the lower distance limit is too short, the sum of the van der Waals radii of an atom pair is used for the following calculations. Normally, those limits are very wide. For a further restriction of the distance interval geometric considerations like the “triangle” or the “tetrangle” inequalities are applied. At this step, all experimentally derived informations like internuclear distances from NOESY / ROESY spectra are usually incorporated into the calculations. Due to experimental errors these restraints should be used with a tolerance of ± 5-10 %.[1]

1.3.2.2 The Metrization Process

The structure generating step during a distance geometry run is called metrization. Various methods are known to realize this operation but in most software tools the Havel random metrization is implemented.[143] In this process an atom pair in the holonomic matrix is arbitrarily chosen and a random distance between the upper and the lower limit is assigned. By solving the so-called triangle and tetrangle inequalities (see Fig. 1.15) the holonomic distance limitations of all other atoms inside the molecule are calculated. This procedure is repeated until all atom distances are allocated. In the end of the computation one obtains a symmetrical matrix of interatomic distances. Since the starting structure and the initial atom pair is casually selected, distance matrix generation and random metrization should be performed several times in order to get an ensemble of metric matrices. This means that a bundle of structures should be computed.[1]

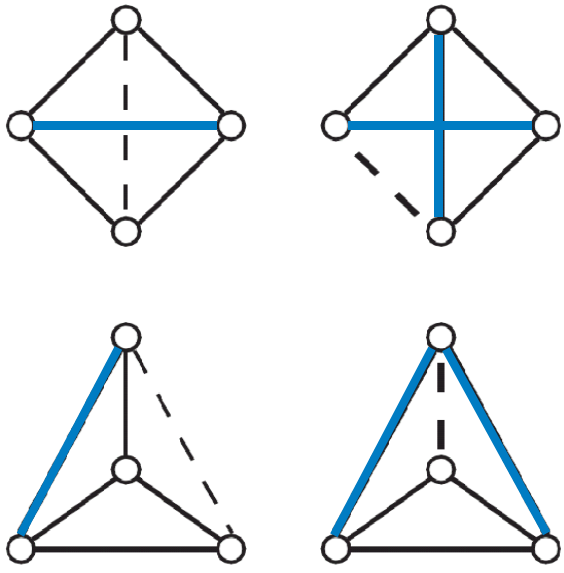


Figure 1.15: Illustration of the upper (left) and lower (right) tetrangle inequality distance limits. Blue solid lines denote the distance at its lower bound, while black solid lines denote the distance at its upper bound; dashed lines represent the associated limit. Tetrangle inequalities are used in DG for calculating upper and lower distances in DG-type structure calculations of small and medium-sized molecules.

1.3.2.3 The Embedding of Distance Matrices

The generation of random structures together with the associated metric matrices is followed by the last DG step, the so-called embedding. In this process the structures are projected ("embedded") from the distance space into the Cartesian space. For this purpose, the Eigensystems of the metric matrices are determined. The dimensionality of the Cartesian coordinate space can have any value but it suggests itself that a transformation into the three-dimensional space should be performed.[144] Working with higher dimensioned spaces has the advantage of bearing down energy barriers in order to find the global energy minimum, e.g. transforming enantiomeric molecules into four dimensions leads to a disappearance of chiral centers.[145] Therefore, no prior knowledge of three-dimensional stereochemical arrangements is needed. However, the indispensable final reduction of such spaces into the x, y, and z dimensions could produce physically meaningless structures.

In general, it is advisable to calculate between 50-100 structures per DG run which are finally aligned in the three-dimensional space. On one side, strongly distorted conformations are quickly identified if the majority of structures form a bundle of similar conformations. On the other side, the more structures are calculated the better are statistics which provide important hints with respect to the quality of the calculated structures and sometimes with respect to the presence of differing conformations. For example, if more than 90 out of 100 calculated structures exhibit similar conformations, it can be assumed that the molecule of interest is conformationally restricted to one distinct shape. The calculation of two structure bundles which show different conformations indicates that the studied system can adopt, at least, two conformational states. Finally, if only a small fraction of the calculated structures can be aligned, either the used pairwise distance informations are erroneous or high dynamics do not allow the determination of a preferred conformation. In this scope, it is also very important to evaluate the violations of distance restraints in the calculated structures. Deviations smaller than 0.1 Å indicate a well-defined structure whereas violations up to ca. 0.4 Å often occur when a molecule shows more than one conformation.[29] If the deviations are greater than 0.6 Å, there are in many cases errors in the pairwise distance restraints; however, DG calculations of highly flexible molecules also result in strong violations.

Since the distance ranges of holonomic restraints are very wide, embedded structures are often slightly distorted. To overcome this problem, further optimization is a must. This can be achieved by the utilization of methods based on classical force fields.

1.3.3 Molecular Dynamics Simulations as Final Step of Structure Determination

In principle, the MD aided structure elucidation can comprise three different procedures. First, and this is the most robust way, a coarse 3D conformation is provided before starting the simulations. The way such a model can be created is explained in sections 1.3.1 and 1.3.2. If the starting structure is already in good agreement with the final three-dimensional conformation, the MD simulations can be seen as a kind of structural refinement ("conformational smoothing"). A second way to get reliable models is to create an arbitrary conformation which is, during the MD simulation, enforced by the experimental restraints to adapt its natural 3D arrangement. However, problems can arise when very high energy barriers (e.g. a peptide bond in *cis* or *trans* orientation) separate the input structure from the desired, natural one. At last, an arbitrary starting structure could be exposed only to the potential of the force field; this means that neither a crude starting structure nor experimental restraints are used for structure elucidation ("*ab initio* approach"). This procedure, however, suffers from the same drawback as the one described before; in addition, if the utilized force field contains inaccurate parameters, the native structure could not be found via MD simulations. In conclusion, it is suggested to create a coarse starting conformation via tools like distance geometry. Thus, the main focus in the following sections lies on approaches which make use of this approach.

In general, experimental restraints gained by NMR measurements can be directly or indirectly used in MD computations: data can be utilized in order to restrain the degrees of motional freedom of defined atoms (restrained run, rMD) or to evaluate the quality of a free MD run (fMD). Before both procedures are described in detail, the preparation of an MD run and the design and equilibration of a simulation box are explained. Furthermore, important aspects of vacuum simulations and the utilization of advanced MD sampling techniques are treated. Finally, the principles of analyzing MD runs are discussed.

1.3.3.1 Preparing an MD Simulation Run

Since several different molecular dynamics software packages are offered one primarily has to decide which one should be used. While tools like NAMD and AMBER are designed for the simulation of biomolecules (peptides, proteins and nucleic acids), packages like CHARMM, GROMOS, and GROMACS have a wider range of application.[14,103,146-148] The second step is the selection of an adequate force field. For example, the OPLS FF was parametrized in particular for amino acids and nucleic acids.[149] Other often used force fields are AMBER, GROMOS, CHARMM, CFF, TRIPOS, or MM.[90,150-153] In all-atom force fields, all atoms are treated explicitly whereas in united-atom force fields non-polar hydrogens are combined with the carbon atoms they are attached to. This implicit way of parametrization saves a lot of computation time, especially for molecules with long aliphatic chains.

The next and probably most important step is the build-up of a molecular topology that includes all informations about both covalent (bond lengths, bond angles, proper and improper dihedral angles) and non-covalent interactions (Lennard-Jones and Coulomb terms). In addition, each atom type is assigned by partial charges, an atomic mass, and the van der Waals parameters. For common molecules like amino or nucleic acids, so-called building blocks are often available where all atom and interaction parameters are predefined.[14,89] Thus, to create the topology of a peptide, these building blocks only have to be combined with respect to the primary sequence of the molecule. When working with extraordinary molecules, the user has to assign atom types, the bonded-, and the non-bonded parameters manually. Care must be taken when the topology is constructed in such a way because an inaccurate parametrization may result in misleading molecular conformations and / or an incorrect modeling of the dynamical behavior.[154]

Now, a starting conformation of the molecule of interest must be provided. Usually, a crude molecular model from DG calculations acts as such. If no coarse structure exists, an arbitrary 3D structure has to be created. This can be done e.g. by commercial software tools like INSIGHT, SYBYL, or CHEMDRAW3D. However, as already discussed in detail above, this procedure could cause severe problems and is not suggested. As an alternative, density functional theory (DFT) could be applied which seems able to calculate starting structures of good quality (K. Kang, Chungbuk National University; *personal communication*).

After the decision for a program package, a convenient force field, the building of the necessary topology, and the design of a starting conformation, the MD run has to be prepared. In this scope, three parameter blocks are important:

a) *Run Control*
b) *Interaction Functions*
c) *Boundary Conditions*

In the following section, the above listed blocks are briefly explained and tips are given for a correct setup. By far not all parameters that can be adjusted are discussed but a short, general overview of the most important MD input values is presented.

The run control section could be seen as backbone of each simulation. First, the integrator of the Newton's equations of motions must be chosen. If a “regular” MD simulation should be performed, a so called leap-frog algorithm is the best choice since it is both robust and fast (see Fig. 1.16). If one intends to simulate *in vacuo*, a stochastic leap-frog integrator can be selected which allows the incorporation of friction and noise into the system.[155] At this stage, also the algorithms for energy minimization (EM) can be chosen (see also section 1.3.3.2). Of the same importance as the choice of the integrator is the correct election of the time step for integrating Newton's equations of motion. As a principle, this adjustment depends on the fastest dynamics that are present in the system to be simulated. To allow for the these motions, that are, in particular, vibrations of covalent bonds including hydrogen atoms, the integration time step must not exceed 0.5 fs.[89] Generally, the integration of Newton's equations of motion implies the usage of classical mechanics. This means that the motions of covalent bonds (and bond angles) are represented by harmonic oscillators. However, since vibrational modes mostly exceed the frequency limit of statistical mechanics, they are better described by a quantum oscillator. In order to avoid the incorporation of real quantum dynamics in MD simulations, it is common practice to constrain all bond lengths to physical mean values with an arbitrary tolerance. This approach is implemented by constraining algorithms like “Shake” or “Lincs”.[156,157] An advantageous side effect of eliminating the fastest motions in the system is that the time step of integration can be increased to 2 fs, making the simulations four times faster compared to unconstrained runs.

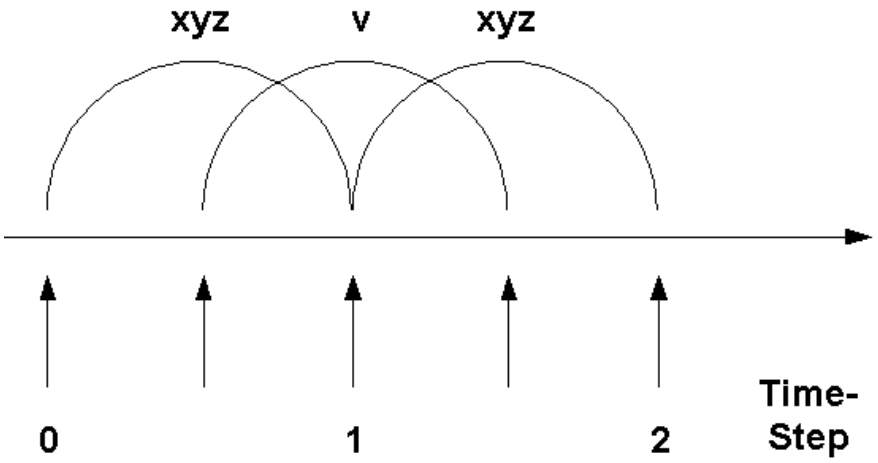


Figure 1.16: Schematic model of the MD leap-frog integration method. The algorithm utilizes positions xyz at time t and velocities v at time $(t - \Delta t/2)$; then, it continuously updates both positions and velocities by applying the forces F(t) determined by positions at time t. The algorithm is denoted leap-frog because xyz and v leap like frogs over each others back.

Finally, it must be decided which files should be written as output of the MD simulations and how often the calculated parameters are printed into these files. In general, the most important file is the produced trajectory. It includes the coordinates of the system or parts thereof and usually the atomic velocities. By help of the trajectory, all parameters based on coordinates like dihedral angles, backbone RMSD, side chain fluctuations, etc. can be extracted. Thus, writing of coordinates should be done frequently. However, it makes no sense to print the coordinates too often since, in such a case, all produced data are highly correlated and bear no additional information. For structure determination runs, a print frequency of 10-20 $(ns)^{-1}$ seems to be a good compromise. Another file of high interest is the energy output. Here, parameters like potential, kinetic, or total system energies are stored. In addition, temperature, density, volume, or pressure of the simulated box can be extracted. If one is interested in correlating coordinates with energetics, both output files must be written out with the same frequency.

The next important parameter block includes the adjustment of the interaction functions and the neighbor searching list. The potential functions can be subdivided into three parts: 1) Non-bonded interactions: Lennard-Jones and Coulomb interactions; 2) Bonded interactions: covalent bond stretching, angle bending, improper dihedrals, proper dihedrals; 3) Experimental restraints: position, angle, dihedral, distance, orientation restraints (for more details, see section 1.2.4).[14,29,89]

Since the bulk of the computation time to simulate a molecular system is used to calculate all non-bonded interactions (see Equ. 1.9), efficient tools have

been developed that speed up the simulation process. One of the most prominent approaches, the twin-range cut-off method, is presented below.[14]

Non-bonded interactions between atoms decrease with the distance between them. Thus, at first, only forces between atoms closer to each other than a certain cut-off radius r_{short} are calculated in MD simulations. To reduce integration errors at the cut-off radius, atoms are grouped into so-called charge-groups which usually have a zero net charge. The program searches, starting from one charge group, all neighbored groups that are located within the cut-off distance r_{short} and stores this information in a group pair list which is updated every n-th (5-10) integration step. To drastically reduce the computation time for non-bonded interactions, the mentioned cut-off radius is usually chosen relatively short (0.8-1.0 nm). To evaluate all non-bonded forces with sufficient accuracy, however, a second, long-range cut-off r_{long} must be defined (ca. 1.0-1.5 nm, depending on the type of force field; GROMOS FF usually use 1.4 nm).[91,158] The rationale behind this concept is that all interactions between groups within the short-range cut-off are evaluated at each integration time step whereas forces between charge groups at a distance longer than r_{short} but smaller than r_{long} are calculated with the same frequency with which the charge group pair list is updated. By means of this approach, the long-range interactions can be approximately treated without increasing the computational effort significantly.

The described scheme usually covers all Lennard-Jones interactions; for some force fields (e.g. OPLS) that use a shorter long-range cut-off, however, an energy and / or pressure correction term for the van der Waals dispersion potential must be included.[159] For electrostatics, the situation is more complex since the Coulomb interaction energy is inversely proportional to the distance between two particles, making it long-ranged. This means that Coulomb interactions can not be adequately covered when all electrostatics are cropped at r_{long}. Hence, further concepts have to be applied. One of the most often used methods is the incorporation of a Poisson-Boltzmann reaction-field (RF) force correction into the Coulomb potential energy term.[14] A reaction-field is constructed based on the assumption that a constant environment characterized by a uniform dielectric permittivity and ionic strength is present beyond the long-range cut-off r_{long}. Subsequently, such a dielectric continuum produces a reaction field in response to the charge distribution inside the cut-off sphere with radius R_{rf} (see Equ. 1.9).

$$V^{LJ}(r;C_{12},C_6) = \sum_{(ij)} \left[\frac{C_{12}(i,j)}{r_{ij}^{12}} - \frac{C_6(i,j)}{r_{ij}^{6}} \right]$$

$$V^{C}(r;q) = \sum_{(ij)} \frac{q_i q_j}{4\pi\,\varepsilon_0 \varepsilon_1} \cdot \frac{1}{r_{ij}}$$

$$V^{RF}(r;q) = \sum_{(ij)} \frac{q_i q_j}{4\pi\,\varepsilon_0 \varepsilon_1} \cdot \frac{(-\tfrac{1}{2} C_{rf} \cdot r_{ij}^2)}{R_{rf}^3}$$

Equations 1.9: Mathematical expressions (as implemented in GROMOS force fields) illustrating the non-bonding potentials that have to be calculated in MD simulations. V^{LJ} represents the Lennard-Jones potential, V^{C} the Coulomb potential, and V^{RF} the potential originating from a reaction-field to address for long-range electrostatics. The terms i and j describe atom pairs between which the potentials arise. r as the distance between atoms i and j, C_{12} and C_6 as constants which depend on the type of involved atoms; q as the atom charge, ε_0 as the dielectric constant under vacuum conditions, ε_1 as the dielectric constant of the medium where atoms i and j are located; C_{rf} as a reaction-field constant and R_{rf} as the radius of the spherical cavity within Coulomb interactions are explicitly treated.

With the presented approaches, short-range cut-off, long-range cut-off, and the introduction of a reaction-field, Coulomb interactions are treated quite well (but anyhow not perfectly). There are several other techniques to simulate the long-range electrostatics with the lattice-sum methods as the most prominent ones.[160-162] The idea here is to convert the slowly converging total electrostatic energy of all system particles into two fast converging terms and one constant term. One of the fast converging terms consists of a direct, the other of a reciprocal sum. Based on this mathematical strategy, short cut-offs for the direct (ca. 1.0 nm) and the reciprocal space sum (ca. 10 “wave vectors”) can be defined. Since only one cut-off in the direct space is applied, the method clearly differs from the twin-range cut-off methods including a reaction-field. Both techniques have advantages and disadvantages regarding precision and computational speed; thus, a distinct procedure is not favored.[161]

The bonded interactions are usually described by simple harmonic (bond length, bond angles, improper dihedral angles) or trigonometric (for proper torsional angles; the mathematical expression of dihedrals differ for distinct force fields) functions (see Equ. 1.10).[158,163] When the force field parameters have been adequately adjusted, the calculation of bonded potentials is straightforward. Additionally, the calculation of interactions between particles based on the parameters of the force field can be manipulated by experimental restraints. This is usually done via incorporation of penalty energy terms which force a system to adopt an experimentally determined property.[164] In detail, the potential energy is increased when the restrained properties, e.g. a distance between two atoms, exhibits values that are not in agreement with the predefined (experimentally obtained) ones.

$$V^{bond}(r;K_b,b_0) = \sum_{n-1}^{N_b} \tfrac{1}{4} \cdot K_{b_n} \cdot \left[b_n^2 - b_{0_n}^2\right]^2$$

$$V^{angle}(r;K_\theta,\theta_0) = \sum_{n-1}^{N_\theta} \tfrac{1}{2} \cdot K_{\theta_n} \cdot \left[\cos(\theta_n) - \cos(\theta_{0_n})\right]^2$$

$$V^{dihed}(r;K_\varphi,\delta,m) = \sum_{n-1}^{N_\varphi} K_{\varphi_n} \cdot \left[1 + \cos(\delta_n) \cdot \cos(m_n\varphi_n)\right]$$

$$V^{improp}(r;K_\xi,\xi_0) = \sum_{n-1}^{N_\xi} \tfrac{1}{2} \cdot K_{\xi_n} \cdot \left[\xi_n - \xi_{0_n}\right]^2$$

Equations 1.10: Mathematical expressions (as implemented in GROMOS force fields) describing bond interaction potentials. V^{bond} defines the potential between two atoms which are directly bonded, V^{angle} the potential between three atoms which form an angle, V^{dihed} the potential between four atoms that span a dihedral angle, and V^{improp} the potential that is utilized to maintain the planarity of e.g. aromatic groups or to keep a given chirality. r as the distance between involved atoms, K_b, K_θ, K_φ, and K_ξ as the force constant which constrain instantaneous calculation properties to its lowest-energy values b_0, θ_0, and ξ_0; an exception is the equation for the dihedral potential which is expressed in terms of a phase shift δ and a multiplicity m for distinct energetic minima with φ as (one of the) lowest-energy dihedrals.

Finally, in order to complete the list of essential MD parameters, the simulation boundary conditions are treated. The most important adjustable parameters are the temperature, the pressure, and the box wall properties. In order to maintain a certain temperature inside a simulation box, a system must be coupled to a temperature bath. In most cases, the utilization of a weak coupling algorithm (e.g. the Berendsen method) is sufficient to generate a stable average temperature.[165] It should be noted that not all atoms or molecules always exhibit the same kinetic energy; rather, the average temperature of a complete box (solute and solvent atoms) is preserved. The heat flow into or out of the system is effected by scaling the velocities of each particle. The frequency of the coupling can be adjusted, so that the simulated system is strongly or only hardly affected. A further way of temperature coupling is the utilization of Langevin dynamics.[155] Here, a friction and a noise term is incorporated intro Newton's equations of motion. If the friction constant is chosen to be very small, the system is stochastically coupled to a bath. In contrast to the Berendsen temperature coupling, the calculated ensemble is “known” a priori when using stochastic dynamics.

As next step, the operator has to decide if the box should be coupled to a pressure bath. Depending on the type of simulation - e.g. NVT or NpT ensemble calculations - a pressure algorithm has to be applied or not. The pressure effect is generated via scaling of the box dimensions and the coordinates of atoms.[165] If the pressure is isotropic, all mentioned parameters

are equally affected. If one models e.g. a membrane, the pressure bath can also be switched on for only one or two dimensions in order to avoid distortion of the membrane.

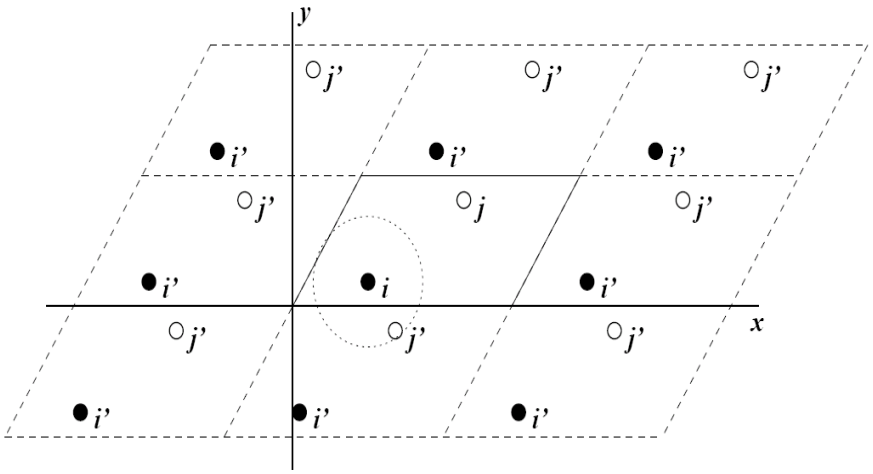


Figure 1.17: The concept of periodic boundary conditions as shown here in one dimension. An MD box is surrounded by identical copies of itself (in three dimensions). As a result, the particles i' and j' can interact with each other across the box walls although their distance inside the box is large (indicated by a dashed circle). (*taken from reference [168]*)

Finally, the treatment of the simulation box walls must be chosen. The most common way to minimize edge effects in a finite system is to apply periodic boundary conditions (PBC).[166,167] Here, the simulation box is surrounded by translated copies of itself (see Fig. 1.17). Thus, the problem of artificial box walls is relocated to infinity. For a crystal system, the choice of periodic boundary conditions is an excellent approach. Since one is mostly interested in the behavior of molecules in the solution-state (non-periodic systems), PBC also introduces artifacts which are, however, less severe than errors arising from imperious box walls.[29] As a rule of thumb, the errors caused by periodic boundary conditions are less dramatic the bigger a box and the longer the distance between solute atoms and the box boundaries is.

1.3.3.2 Generation and Equilibration of a Simulation Box

In the beginning of an MD simulation, the starting structure has to be positioned in the (mostly cubic or truncated octahedral, depending on the shape of the solute) simulation box. The minimum distance between the solute and the box walls must be long enough in order to avoid that the molecule of interest “feels” itself due to the periodic boundary conditions (see section 1.3.3.1). For most force fields, distances of 1.2-1.5 nm are useful values. Afterwards, the solute molecule is minimized *in vacuo* to rule out high-energetic distortions. For this purpose, different energy minimization

algorithms have been developed.[169] The “Steepest Descent” method should be applied when the starting structure is highly distorted. If the conformation of a molecule is already close to its energy minimum, the utilization of the “Conjugate Gradient” algorithm is recommended.

Solvent Boxes

In almost all MD calculations, a simulation box including a solute is filled with a desired type of solvent; thus, the molecule of interest is “soaked“ with solvent molecules until the box has reached approximately the correct experimental density (see Fig. 1.18). After that, the solvent is energy minimized while the coordinates of the solute are strongly restrained to their original positions.[170] This leads to an optimal arrangement of the solvent molecules around the starting structure. In the following, random velocities are assigned to all atoms according to a Maxwell distribution, and the MD equilibration runs can be started.

The whole system is heated by weakly coupling to a temperature bath until the temperature used for the experimental measurements is reached. Usually, one begins with a temperature of 50 K which is further on incrementally elevated in 50 K steps. During this process, the solute position restraints are slowly relaxed at each temperature equilibration step in order to “acclimate” the solute to its surroundings. Having reached the temperature at which the MD simulation should be carried out, the position restraints of the solute are switched off. The system to be simulated is now equilibrated with respect to its kinetic energy. To obtain a box that is also in equilibrium regarding its density, the system is coupled to a weak pressure bath. It should be noted that temperature and pressure equilibration should never be performed at the same stage; first, the temperature is adopted, then, the volume.[170] As soon as the box has adapted the correct experimental density the production runs (the “real” simulations) can be performed. In order to ensure that the whole system is well equilibrated, the first 50-1000 ps of the first production trajectory (in particular depending on the size of the system) should not be comprised in the calculations of the conformation of the solute.

The length of the production runs again depends on the size of the system and the properties one is interested in. Whereas a smoothing of structures can be reached by trajectories of 0.5-1 ns length, a converged conformational sampling usually requires much longer simulation times (> 100 ns).

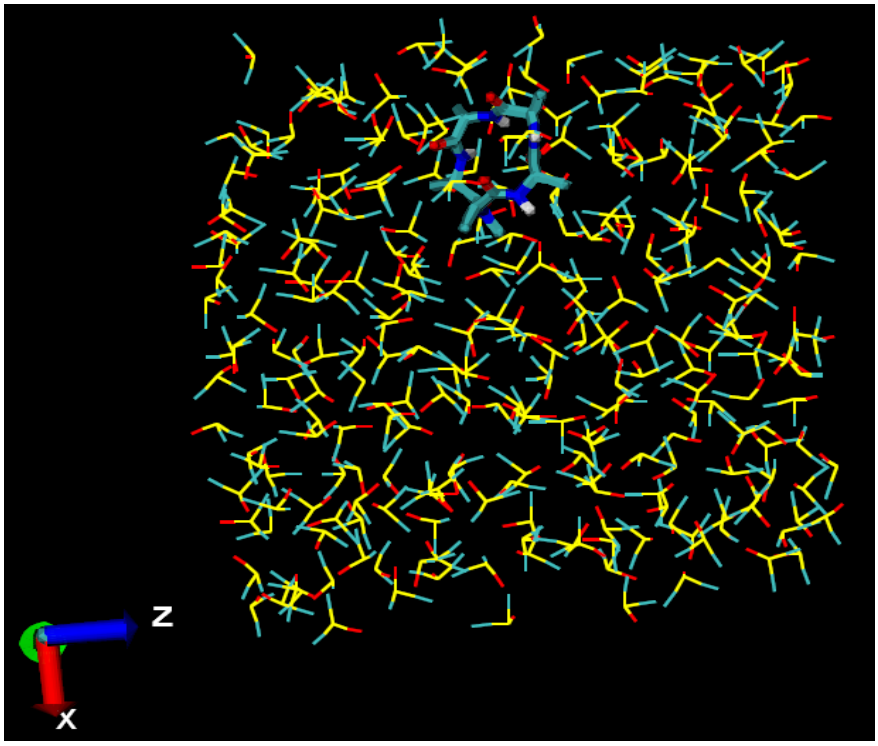


Figure 1.18: Illustration of an equilibrated MD box filled with one peptide as solute and several DMSO molecules as solvent. The usage of an united-atom force field results in the omission of apolar hydrogen atoms.

Vacuum MD Simulations

Molecular dynamics simulations *in vacuo* have the advantage of being much faster than calculations with explicit solvent because all solvent degrees of freedom can be neglected. However, such trajectories often lead to unreliable or incorrect results.[171] Especially for small molecules which possess a relatively large surface, the solvent may have strong effects on their global conformation. Furthermore, molecules simulated in vacuo tend to minimize their overall surface which possibly results in distorted three-dimensional structures.[172] Due to the fact that no dielectric shielding is present in vacuum, intramolecular interactions based on dipoles and charges are overestimated. At last, the formation of intermolecular interactions like hydrogen bonds between the solute and the solvent is completely disregarded. The situation is slightly better when experimental restraints are applied that pull the molecule closer to the real structure. However, to circumvent the above mentioned disadvantages, it is strongly recommended to perform MD simulations of small and medium-sized molecules with explicit solvent even if they require more computational time. If for some reasons the usage of vacuum simulations is inevitable, special vacuum force fields should be applied where partial atomic charges and polar interactions are downscaled.[14] In addition, it is recommended to make use of Langevin (see also section 1.3.3.1) dynamics. By introducing an arbitrary friction constant and a noise term into the potential

energy functions, collisions of a solute with surrounding solvent molecules is simulated; in contrast to pure vacuum simulations, both dynamics and kinetics of investigated molecules are reproduced in a much more realistic way.

When working with large molecules like proteins - this involves the usage of large box sizes and huge amounts of solvent molecules which again causes time-consuming calculations - it is feasible to make use of the “implicit solvent” concept (see e.g. reference [173]). Instead of using explicit solvent molecules a continuous medium according to the designated solvent is introduced. Thus, artificial molecular surface effects as well as unrealistic dipolar and charge-based interactions can be minimized.

1.3.3.3 Performance of Restrained and Free Molecular Dynamics Runs

In principle, experimentally (NMR) derived distance, dihedral, or orientation restraints can either be used only for validation of MD simulations or directly be applied to the calculations when a molecule should be enforced to adopt a distinct (“its natural”) conformation.[89] The latter mentioned procedure is advisable when a starting structure, e.g. obtained via DG calculations, strongly changes its conformation at the beginning of a free MD simulation run in order to stabilize the molecule. For very “stable“ starting conformations, free MD runs are sometimes sufficient for structure determination.

Restrained Molecular Dynamics

In restrained MD runs, the force field potential functions are extended with energetic penalty terms, thus preserving the given restrictions. However, experimental restraints always display averaged over both a tremendous number of molecules and certain time intervals. Since molecules are not static they can change their conformations and not all members of a molecular ensemble exhibit the same structure at a certain point in time. To allow for these aspects, experimental informations can be implemented as ensemble- or time-averaged restraints.[174] When using ensemble-averaging, single molecular representatives are allowed to violate the experimental restrictions as long as the whole ensemble is in agreement with the restraints. For this purpose, one has to simulate either a box with several molecules or many identical systems (containing only one molecule) with different starting conditions. On the other hand, the utilization of time-averaged restraints ensures that the distance, dihedral, or orientation information is maintained by

averaging these restrictions over a certain time period. This means that the instantaneous restraint values need not to fulfill the experimental ones as long as the average values over the complete trajectory agree with them.

To give an example, J-couplings may directly be used as constraints in MD computations for stabilizing certain dihedral angles, but they can also be utilized to check the reliability of the final results. If J-couplings (which usually represent more than one dihedral angle) are directly used as restraints in an MD run,[99] they must be "soft" enough (low force constant) to allow the molecule to pass restraint based energy barriers during the trajectory. In recent years, also orientation restraints (obtained from residual dipolar couplings and other anisotropic parameters) acquired by NMR experiments in anisotropic environment are utilized for the restrained structure elucidation of a molecule.[175] Generally, the first step in MD aided structure calculation is the performance of a restrained run. Already during equilibration, the experimental restrictions are allowed to influence the solute molecule. The goal of this approach is to force the solute to the experimental parameters. This procedure is especially important when using force fields that are not developed on a pure thermodynamic basis or if no (stable) DG derived starting structure is available.

Free Molecular Dynamics

The last step during the 3D structure generating process includes the simulation of a molecule in a free dynamics run and the comparison of calculated with experimentally derived structural parameters. The procedure explained above is exactly repeated but without restraining information effecting the force field. The final system / conformation computed in the rMD run usually serves as the starting box / structure in the free dynamics run. After representative conformations were established in a free run, one has to check if the structure matches the simulated one under restrictive distance conditions. Whereas small deviations are expected, bigger discrepancies between "restrained" and "free" conformations usually indicate erroneous results. In some cases, it is also possible to abandon a restrained run and only perform free dynamics. As a prerequisite for this approach, a well-defined starting structure must be available and thermodynamically parametrized force fields like GROMOS 53a6 must be utilized.[91,158] In addition, an exhaustive and strict comparison of simulated results with experimental outcomes must be performed.

Short & Long Sampling Runs

If molecular dynamics simulations are applied in order to only “smooth” a conformation obtained via DG calculations, short MD runs (0.1-1 ns) are usually sufficient for yielding reliable results.[176] However, almost all important dynamics parameters are not converged within this time. Thus, meaningful conformational ensembles can not be produced and extraction of parameters like pairwise distance informations (NOEs) or J-couplings for a reliability check becomes difficult or even impossible. By the use of long simulation runs (100-500 ns or even longer), one can not only overcome these obstacles but also gains much more important informations about e.g. intrinsic dynamics or conformational energy barriers.[22] Having produced long enough MD simulations, calculated NMR parameters are often in agreement with experimental ones and thus, errors resulting from experiments or from a wrong MD setup become obvious. Since the conformational space is better sampled by long MD runs, representative molecular conformations are of much higher quality (see e.g. section 3.1).

1.3.3.4 Simulations with Advanced Sampling Techniques

As it was described in section 1.3.3.1, an exhaustive sampling of the conformational space is in most cases not possible by standard MD simulations. The main reason for this is the presence of high energy barriers separating different low-energy conformations. Thus, alternative methodologies have been developed in order to efficiently sample most parts of the free energy hyper surface. The most prominent approaches, simulated-annealing (SA), local-elevation (LE), and replica-exchange (RE) are illustrated in more detail in the following. Other methods like coarse-grained (CG) simulations are also useful for enhancing the sampling efficiency of MD simulations but they have only barely used for the calculation of peptide structures.

Simulated Annealing

Simulated-annealing is a frequently used technique in order to screen the global conformational space of a molecule.[98] The method is also often applied even when only spare experimental structure informations are available. The preparation of such a simulation is identical with a normal MD run. The starting structure can be arbitrarily chosen, i.e. no crude model conformation must be

available. By heating up the simulation system to very high temperatures (1000-2000 K), all molecules are allowed to almost “freely” move on the energy hyper surface. Thus, molecules can partly overcome high energy barriers existent between different conformations. After it spends a certain time (100-500 ps) at high temperatures, the whole system is slowly and incrementally cooled down (ca. 100 K per 200 ps). Each cooling step is followed by a longer period of equilibration (about 200 ps). If the stepwise cooling process is slow enough, the molecule is assumed to approach the minimum of the energy hyper surface which may represent the thermodynamically most stable conformation. To make the procedure independent of an initial structure, the SA procedure is repeated several times with different starting conformations. Finally, all resulting structures are clustered with respect to the calculated potential energies and are compared with experimental structure information. In general, SA can be used as isolated method or as one preparation step for a regular MD run.

Local Elevation / Conformational Flooding

The local-elevation and conformational flooding techniques have been developed to improve the sampling of the configurational and conformational space of a molecule by systematically elevating the (local) potential energy of chosen degrees of freedom of a molecule.[90,97] Thus, the total potential energy function consists not only of the standard physical equations (the force field) but also of an additional, time-dependent local-elevation term. When the investigated solute molecule is trapped in a low-energy basin on the potential energy hyper surface during a trajectory, the LE algorithm gradually “floods” the energy valley by using Gaussian-shaped energy functions. The goal of this approach is to enforce the molecule out of the basin into a neighboring energy valley. If being trapped again, the procedure is automatically applied again until the whole accessible energy hyper surface is flattened. In this way, the energy surface is much more efficiently sampled than using standard MD approaches. If the simulation time is long enough, and all possible conformations are visited during the LE trajectory, the local-elevation potential energy term represents the negative of the free-energy surface (also called potential of mean force) for the LE degrees of freedom of a molecule. Thus, LE enhances not only the sampling of the conformational space but also allows the calculation of free-energy terms. In order to obtain data of sufficient accuracy, the LE technique should be combined with an umbrella sampling re-

weighting procedure. Here, the visited parts of the energy surface are related to their according energies. After localization of the conformations of interest on the energy hyper surface, corresponding free-energies can be calculated in a straight-forward and precise way.

Replica Exchange

Replica-exchange molecular dynamics (also called parallel tempering) is a technique for the enhancement of sampling the conformational space.[94,177] It is especially useful if conformations are separated by relatively high barriers as they are often present on the energy hyper surfaces of e.g. peptides and (other) drug-like molecules. In general, multiple replicas of the same system are simulated at different temperatures; subsequently, the complete state ("the whole box") of two replicas is randomly exchanged at regular intervals with an arbitrary probability (see Equ. 1.11).

$$P(1 \leftrightarrow 2) = \min\left(1, \exp\left[\left(\frac{1}{k_B T_1} - \frac{1}{k_B T_2}\right) \cdot \left(U_1 - U_2\right)\right]\right)$$

Equation 1.11: Mathematical principle of the replica-exchange concept. P as the exchange probability between the MD boxes 1 and 2; T_1 and T_2 as the reference temperatures, U_1 and U_2 as the potential energies of replicas 1 and 2, respectively; k_B is the Boltzmann constant.

After each exchange the velocities of the single simulation boxes are scaled by $(T1 = T2)^{\pm 1/2}$ and a neighbor search is performed the next step. This combines the fast sampling and frequent barrier-crossing in higher temperature boxes with a correct Boltzmann sampling at all distinct temperatures. As a consequence one finally obtains a physically meaningful trajectory at the temperature of interest (e.g. the room temperature at which the experimental parameters were elucidated) as regular MD simulations provide it; however, the sampling of the overall conformational space is dramatically improved. Thus, molecular arrangements that would appear either never or only after very long simulation times could be elucidated. The exchange of temperatures between two replicas is the most often found procedure of RE in literature. However, many other MD parameters could be exchanged, for example single terms of the Hamiltonian which describes the energetic state of the system.[178] Furthermore, only predefined parts of a simulation system could be addressed to the replica exchange procedure, e.g. a distinct dihedral angle that can adopt several values which are separated by relatively high energy barriers.[179]

1.3.3.5 Reliability Check and Analysis of MD Simulations

The last step of an MD simulation includes the check of results with respect to their reliability and the analysis of the output. Whereas the analysis procedure strongly depends on the type of simulation and the desired informations, the reliability check of results usually comprises two evaluation steps: the computational validation and the comparison of the simulated output with experimental results.

Computational Reliability

The validation of the computational reliability of MD simulations is of high importance since even “small” errors or inaccurate settings in the initial setup of a run can have a tremendous impact on the final results. Depending on the thermodynamic type of a simulation, certain parameters are kept constant or are allowed to slightly spread around an average value. For example, during the calculation of a canonical (NVT) ensemble, the density, the temperature, the potential and the total energy must not show a temporal drift (see Fig. 1.19 for the density of two boxes during a 1 ns MD simulation).

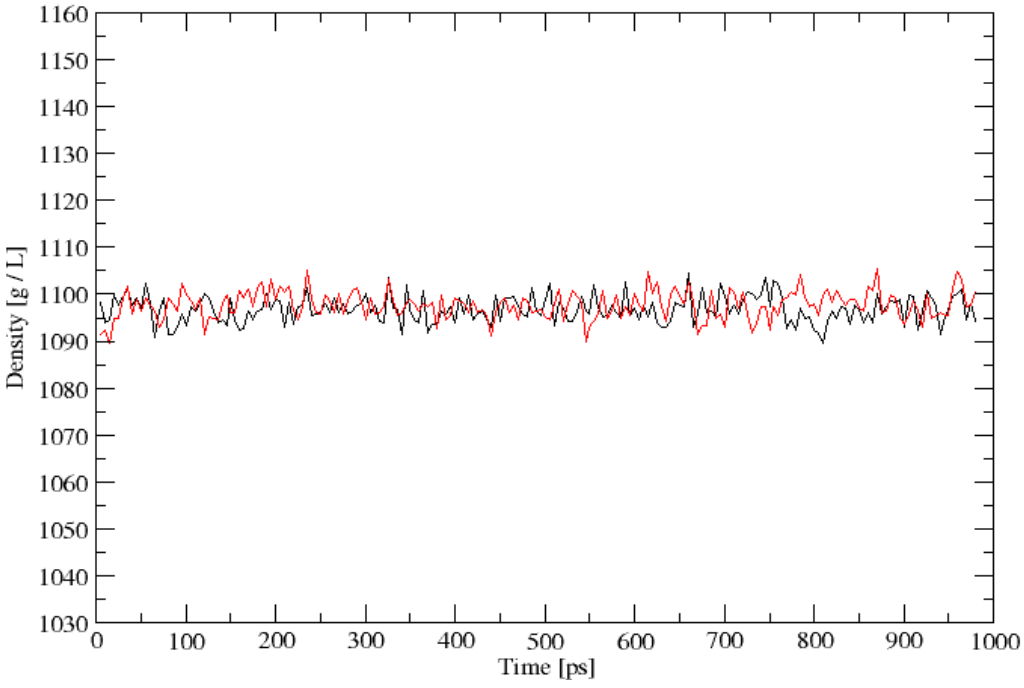


Figure 1.19: Distribution of densities of two (black and red) MD simulation boxes filled with DMSO over time. Both boxes are sufficiently equilibrated as no temporal drift of the density is visible; moreover, the average MD densities of 1.096 g / mL at room temperature correspond to the experimental value.

The presence of a drift means that the simulation box was not sufficiently equilibrated before a production run was started, or that there are errors in the molecular topologies or in the MD setup. Moreover, the average value of a

property like the density of a simulation box filled with a certain solvent must be in agreement with experimental values under the same conditions, e.g. when using DMSO as solvent, the average density at 300 K should be ca. 1096 g / L.[180] Furthermore, one has to check if all forces were treated correctly for the utilized type of force field (e.g. Lennard-Jones or Coulomb cut-offs). This can be done by evaluating the range of forces which have been calculated between two atoms. Only if computational errors can be excluded it is meaningful to also compare calculated with experimental data (for details, see next section).

Experimental Validation

The experimental validation consists of the comparison of calculated parameters with values yielded by experimental (NMR) measurements. This evaluation is only meaningful - especially in structure calculations - if MD simulations are produced with sufficient sampling since the relaxation times of e.g. NOEs, J-couplings, or RDCs are quite long. In principle, one should calculate average values of the properties of interest over a complete trajectory and compare them with the experimental results. If the validation comprises a restrained run, the calculated values must be in very good agreement with the measurement outcomes (see Fig. 1.20 for a comparison of experimental and calculated J-couplings for two different simulations).

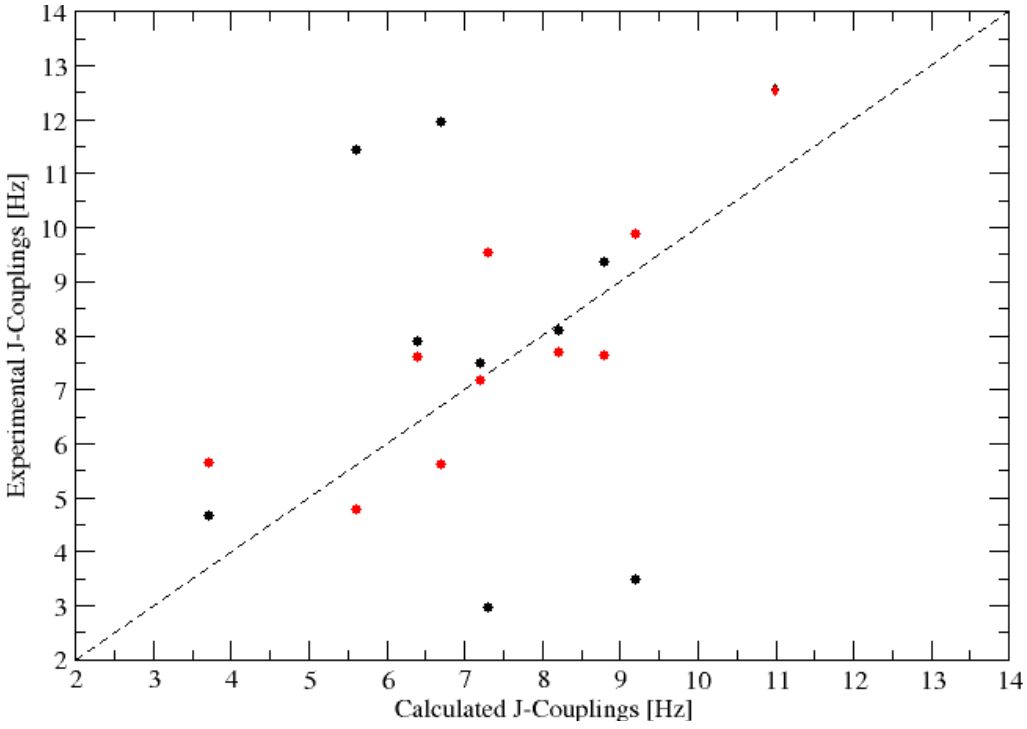


Figure 1.20: Comparison of simulated with measured J-couplings. The data points colored in black and red are the results from different MD calculations. Points which are located close to the dashed diagonal display a very good consensus between calculated values and experimental data. The J-couplings originating from the red colored simulation "set" are in better agreement with the experimental values compared to the "set" colored in black.

In cases where obtained deviations between measured and simulated properties are very high, either errors in the NMR parameters or in the simulation setup may be present; a possible explanation could be that the constraining algorithm is not able to force the simulated molecule to adopt an given structural parameter like a pairwise distance restraint or a certain dihedral angle value. For a free MD run, the evaluation is not so strict. A consensus of calculated and experimental data indicates that the modeled structure is most likely of good quality and should correspond to the “real, natural” one. However, deviations between measured and simulated data need not to be a sign of a wrong model. The presence of very high energy barriers separating several structures may result in an incomplete sampling of the accessible conformational space. In this case, the presentation of a structure is still possible if it is noted that it represents only parts of a molecular ensemble.

Parameter Extraction

There are several parameters which are of interest in conformational analysis of a peptide. First of all, the three-dimensional structure of the simulated molecule is usually presented.[1] It is important to mention that the elucidation of an average conformation (the mean over all calculated structures) is meaningless as it usually does not represent a reliable physical state; rather, 3D alignment, least-square fitting, and subsequent clustering of all produced conformations is suggested where a representative structure belonging to the most populated cluster serves as the structural model of the peptide.[181] Furthermore, a plethora of time distributions like root mean square deviations (RMSD) and fluctuations (RMSF) of backbone or side chain atoms (see Fig. 1.21 for backbone RMSD values as arising during a 150 ns MD trajectory), radial distribution functions (RDF), or dihedral angle and H-bond distributions can be determined.[89] In addition, also averages of a property can be calculated, e.g. Ramachandran plots, solvent accessible hydrophilic / hydrophobic surfaces, or root mean square fluctuations. By means of the listed parameters, one can identify regions of higher flexibility, dihedral angle flips, changes in the overall molecular shape, or the stability of H-bonds and secondary structural motifs. In this scope, the conformation of a peptide is not only described in a static way but a more dynamical model is presented, thereby accounting for the basic concept of molecular ensembles which are in a thermodynamic equilibrium.[29]

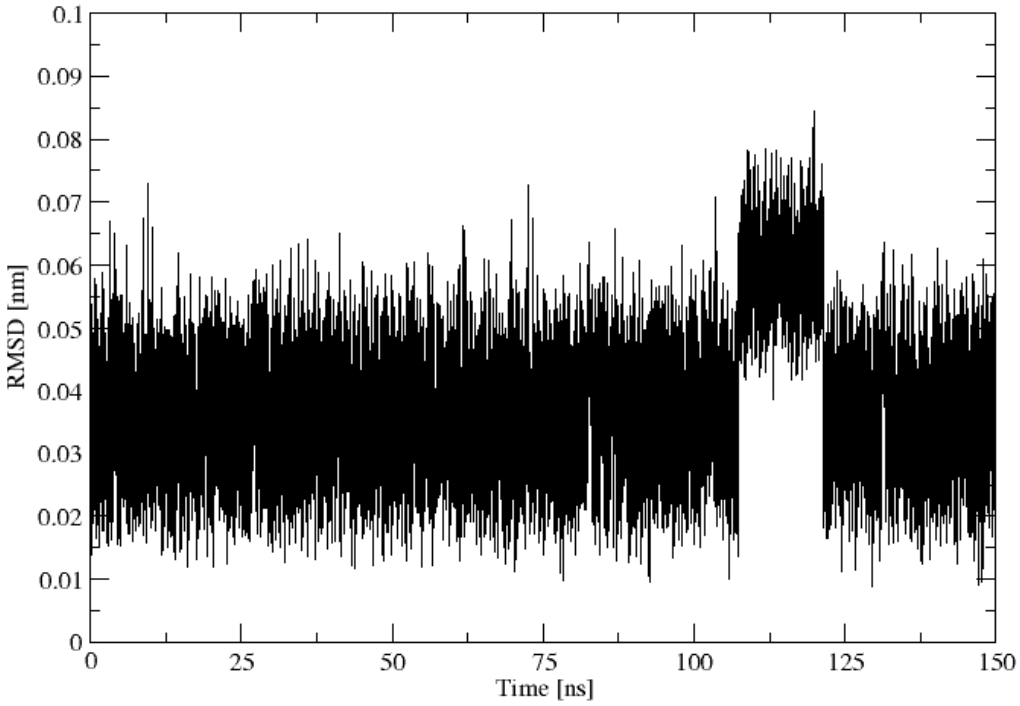


Figure 1.21: Distribution of the backbone root mean square deviations of a cyclic peptide over time. The most representative structure of the MD run served as reference conformation. Whereas slight to medium fluctuations are present throughout the whole trajectory, one bi-directional flip of a peptide bond around its adjacent φ and ψ dihedral angles occurs between 110 ns and 125 ns.

1.3.3.6 The Way from Experiments to Structure Presentation

A sophisticated determination and refinement of molecular conformations in solution comprehends three main steps: NMR experiments, distance geometry, and molecular dynamics. In principle, it is possible to exclusively make use of NMR and DG or only MD but normally it is strongly suggested to combine these methods in order to obtain robust and reliable structural models. Only if the computational results of different methods (at least partially) match the experimental outcomes, and no gained data contradict each other, a three-dimensional structure should be presented. There are various ways of combining the described techniques and single procedures may differ depending on what kind of molecules are investigated. However, with the flow chart in figure 1.22 an instruction how reliable structural models (especially of peptides) can be obtained by using the advantages of combining NMR spectroscopic studies with sophisticated tools of computational chemistry is given.[1]

The starting point of each structure elucidation is the collection of experimental data like NOE or ROE derived distances, torsional restraints from J-couplings or CCRs and RDCs as orientation restraints (**1**).

If NMR spectra contain a single set of signals, the molecule of interest has either a single preferred low energy conformation or several conformations

which are in fast equilibrium (maximal energy barrier height is below 60 kJ / mol). If no other information is available, one usually starts with the assumption that only one preferred structure (often called “rigid” conformation) is present. In the next step, experimental informations are converted into restraints (**2**). After a set of NOE and scalar / dipolar coupling based restraints is generated, the distance geometry computations are performed as described in section 1.3.2 (**3**). To get a “statistical bundle” of different conformations, this procedure is repeated several times. The best structural model, i.e. the structure which shows the best agreement with the experimental restraints, serves now as input for a restrained MD run (**4**). All produced conformations of the rMD run are subsequently clustered with respect to a predefined structural RMSD cut-off value; the most representative structure of the biggest cluster serves as conformational model of the rMD simulation. The restrained run is followed by a free molecular dynamics simulation by which the final conformational model is produced in the same way as described above (**5**).

At this point it is essential to compare the calculated “fMD”-structure with both the experimental data and the results of the rMD run (**6**). On one hand, the interatomic distances of the final model should be in good agreement with the NMR restraints; additionally, the “fMD”-averaged structure should correspond to the refined conformation obtained by the restrained molecular dynamics. Only if the rMD and the fMD simulations result in the same conformational model and no experimental restraints are severely violated the calculated structure can be presented as a three-dimensional image (**7**).

If the final structure either deviates from the refined model or does not match the NMR restraints (**8**) one has to revise the experimental data and the parameters used in the DG and MD computations (**9**). In many cases, mistakes are made when preparing and performing the computational processes (**10**) or sometimes experimental errors might be present (**11**). Those errors include a wrong NMR peak assignment, an inaccurate calibration of the NOE / ROE signals, erroneous extraction of scalar or dipolar couplings, an incorrect conversion of the experimental data to constraints, or a nonfactual parametrization of the rMD and fMD trajectories. In such cases, either new calculations or new experiments must be performed.

If the revision of the experimental and the calculative conditions yields neither errors nor inaccuracies one has to change the initial assumption that the molecule of interest has only one preferred low energy structure. Highly

flexible molecules like peptides often contain several conformations which are in exchange with each other (**12**). If the transition from one conformation to another is slow enough (at the NMR time scale) there are distinguishable sets of signals - one set for each conformation - in the NMR spectra. Thus, the structure elucidation is carried out by taking into account several independent sets of experimental restraints. However, the conformational exchange dynamics are often very fast. In this instance, only one averaged NMR signal set can be observed. Since those resonances represent multiple transient molecular conformations it is not possible to assign e.g. NOE / ROE signals to definite, unique interatomic distances.[182] In order to abolish this obstacle one can re-record NMR spectra under varied physical conditions (**13**). The utilization of different types of solvent, the usage of various pH values, and the measurement at lower or higher temperatures influence both conformational equilibria and exchange kinetics; in some cases, a molecule can be forced to exhibit one main structure (manipulation of equilibrium) or several clearly distinguishable conformations (biased exchange kinetics) which again results in only one or rather multiple NMR signal sets.

Another sophisticated possibility to handle the problem of conformation-averaged NMR signals is the implementation of time- or ensemble averaged restraints in MD calculations (**14**). The given distance informations are either distributed among several molecules or averaged over a certain time interval. Hence, a molecule is not forced to adopt one definite, potentially “meaningless” structure according to the conformation-averaged experimental data. Instead, all conformations which on average fulfill the restraints can possibly be observed in a trajectory (**15**).

If all described procedures fail, no structure determination is possible with the state-of-the-art methods demonstrated in this chapter (**16**). In the case of being successful in calculating multiple conformations by using time- or ensemble-averaged MD restraints the solved molecular structures are presented as 3D models and can be deposited in an electronic structure database (**17**). Finally, it is recommended to provide an accurate explanation of the procedures used for the structure elucidation because the application of different methods (NMR, DG, MD, SA, Monte-Carlo calculations, X-ray crystallography) may result in varying conformational models which do not need to display the real state of a molecule. This aspect should be always kept in mind when dealing with structure determination methods.

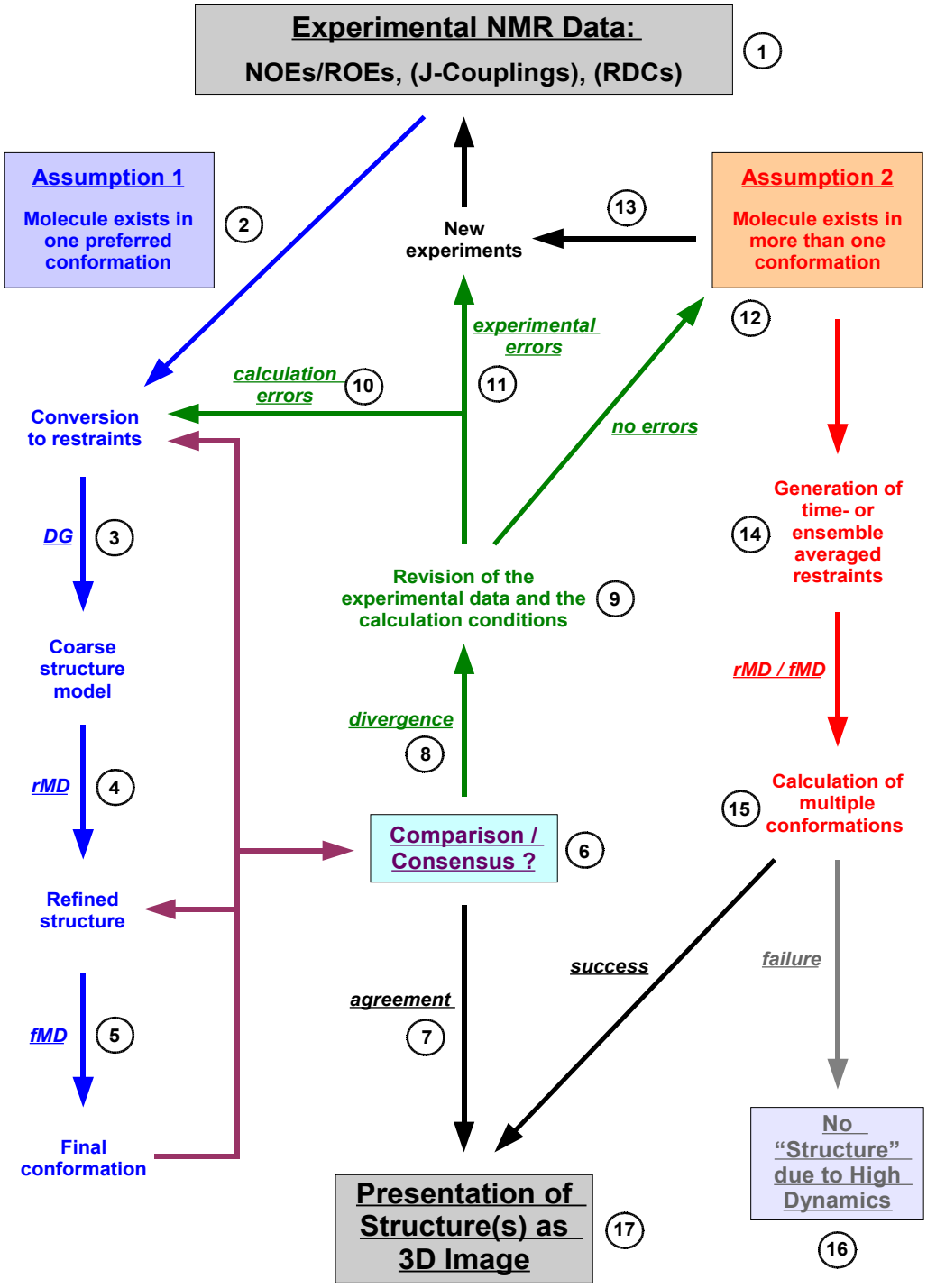


Figure 1.22: Flow chart presentation as summary of the procedure for the elucidation of peptide conformations. The stepwise concept includes NMR spectroscopic study of the molecules of interest in solution, distance geometry calculations, and molecular dynamics simulations. The single steps (as denoted by numbers) are described in the text above this figure. *(taken from reference [1])*

1.4 References

[1] B. Luy, A. Frank, H. Kessler, *Conformational analysis of drugs by NMR: Drug Properties: Measurement and Computation*, Wiley-VCH Weinheim, **2008**.

[2] A. Bax, *Protein Sci.* **2003**, *12*, 1-16.

[3] G. Kummerlöwe, B. Luy, *Trends Analyt. Chem.* **2009**, *28*, 483-493.

[4] J. R. Tolman, *Curr. Opin. Struct. Biol.* **2001**, *11*, 532-539.

[5] F. Kramer, M. V. Deshmukh, H. Kessler, S. J. Glaser, *Conc. Magn. Reson. A* **2004**, *21A*, 10-21.

[6] B. Meyer, T. Peters, *Angew. Chem. Int. Ed.* **2003**, *42*, 864-890.

[7] M. Billeter, G. Wagner, K. Wüthrich, *J. Biomol. NMR* **2008**, *42*, 155-158.

[8] D. D. Boehr, H. J. Dyson, P. E. Wright, *Chem. Rev.* **2006**, *106*, 3055-3079.

[9] A. J. Wand, *Nat. Struct. Biol.* **2001**, *8*, 926-931.

[10] J. M. Yon, D. Perahia, C. Ghélis, *Biochimie* **1998**, *80*, 33-42.

[11] S. C. H. J. A. McCammon, *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press Cambridge, **1987**.

[12] A. Mittermaier, L. E. Kay, *Science* **2006**, *312*, 224-228.

[13] O. F. Lange, N. Lakomek, C. Farès, G. F. Schröder, K. F. A. Walter et al., *Science* **2008**, *320*, 1471-1475.

[14] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger et al., *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, Vdf Hochschulverlag Zürich, **1996**.

[15] L. E. Kay, D. A. Torchia, A. Bax, *Biochemistry* **1989**, *28*, 8972-8979.

[16] G. Lipari, A. Szabo, *Biophys. J.* **1982**, *37*, A380-A380.

[17] G. Lipari, A. Szabo, *J. Am. Chem. Soc.* **1982**, *104*, 4546-4559.

[18] G. Lipari, A. Szabo, *J. Am. Chem. Soc.* **1982**, *104*, 4559-4570.

[19] G. Lipari, A. Szabo, R. M. Levy, *Nature* **1982**, *300*, 197-198.

[20] J. Chatterjee, D. F. Mierke, H. Kessler, *Chem. Eur. J.* **2008**, *14*, 1508-1517.

[21] D. F. Mierke, M. Kurz, H. Kessler, *J. Am. Chem. Soc.* **1994**, *116*, 1042-1049.

[22] X. Zhang, G. V. Nikiforovich, G. R. Marshall, *J. Med. Chem.* **2007**, *50*, 2921-2925.

[23] A. G. Palmer, J. Williams, A. McDermott, *J. Phys. Chem.* **1996**, *100*, 13293-13310.

[24] G. M. Clore, P. C. Driscoll, P. T. Wingfield, A. M. Gronenborn, *Biochemistry* **1990**, *29*, 7387-7401.

[25] N. R. Skrynnikov, F. W. Dahlquist, L. E. Kay, *J. Am. Chem. Soc.* **2002**, *124*, 12352-12360.

[26] O. Trott, A. G. Palmer, *J. Magn. Reson.* **2004**, *170*, 104-112.

[27] C. Y. Wang, A. G. Palmer, *Magn. Reson. Chem.* **2003**, *41*, 866-876.

[28] O. Demmer, A. O. Frank, H. Kessler, *Design of cyclic peptides: Design of Peptides and Proteins: Applications to Therapeutic Agents and Biomedical Research*, Wiley-VCH Weinheim, **2009**.

[29] W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen et al., *Angew. Chem. Int. Ed.* **2006**, *45*, 4064-4092.

[30] D. Trzesniak, A. Glattli, B. Jaun, W. F. van Gunsteren, *J. Am. Chem. Soc.* **2005**, *127*, 14320-14329.

[31] M. Coles, M. Heller, H. Kessler, *Drug Discov. Today* **2003**, *8*, 803-810.

[32] P. J. Hajduk, R. P. Meadows, S. W. Fesik, *Science* **1997**, *278*, 497-499.

[33] T. Diercks, M. Coles, H. Kessler, *Curr. Opin. Chem. Biol.* **2001**, *5*, 285-291.

[34] G. C. K. Roberts, *Drug Discov. Today* **2000**, *5*, 230-240.

[35] F. Ni, Y. Zhu, *J. Magn. Reson. B* **1994**, *103*, 180-184.

[36] F. Ni, H. A. Scheraga, *Acc. Chem. Res.* **1994**, *27*, 257-264.

[37] F. Ni, *Prog. Nucl. Mag. Res. Sp.* **1994**, *26*, 517-606.

[38] T. Peters, B. M. Pinto, *Curr. Opin. Struct. Biol.* **1996**, *6*, 710-720.

[39] J. Jimenez-Barbero, J. L. Asensio, F. J. Canada, A. Poveda, *Curr. Opin. Struct. Biol.* **1999**, *9*, 549-555.

[40] R. E. London, *J. Magn. Reson.* **1999**, *141*, 301-311.

[41] M. Mayer, B. Meyer, *Angew. Chem. Int. Ed.* **1999**, *38*, 1784-1788.

[42] L. Y. Lian, I. L. Barsukov, M. J. Sutcliffe, K. H. Sze, G. C. K. Roberts, *Meth. Enzymol.* **2004**, 239, 657-700.

[43] J. Feeney, G. C. K. Roberts, J. W. Thomson, R. W. King, D. V. Griffiths et al., *Biochemistry* **1980**, *19*, 2316-2321.

[44] J. Feeney, J. G. Batchelor, J. P. Albrand, G. C. K. Roberts, *J. Magn. Reson.* **1979**, *33*, 519-529.

[45] B. Birdsall, J. Feeney, C. Pascual, G. C. K. Roberts, I. Kompis et al., *J. Med. Chem.* **1984**, *27*, 1672-1676.

[46] J. Klages, M. Coles, H. Kessler, *Analyst* **2007**, *132*, 693-705.

[47] M. Mayer, B. Meyer, *J. Am. Chem. Soc.* **2001**, *123*, 6108-6117.

[48] A. O. Frank, H. Kessler, *Nature* **2008**, *452*, 822-823.

[49] G. Otting, E. Liepinsh, B. T. Farmer II, K. Wüthrich, *J. Biomol. NMR* **1991**, *1*, 209-215.

[50] G. Otting, E. Liepinsh, K. Wüthrich, *Science* **1991**, *254*, 974-980.

[51] A. D. Chen, M. J. Shapiro, *J. Am. Chem. Soc.* **2000**, *122*, 414-415.

[52] A. Chen, M. J. Shapiro, *J. Am. Chem. Soc.* **1998**, *120*, 10258-10259.

[53] M. J. Gradwell, J. Feeney, *J. Biomol. NMR* **1996**, *7*, 48-58.

[54] J. M. Moore, *Biopolymers* **1999**, *51*, 221-243.

[55] M. Lin, M. J. Shapiro, *J. Org. Chem.* **1996**, *61*, 7617-7619.

[56] W. Jahnke, L. B. Perez, C. G. Paris, A. Strauss, G. Fendrich et al., *J. Am. Chem. Soc.* **2000**, *122*, 7394-7395.

[57] C. Dalvit, P. E. Fagerness, D. T. Hadden, R. W. Sarver, B. J. Stockman, *J. Am. Chem. Soc.* **2003**, *125*, 7696-7703.

[58] S. W. Fesik, S. B. Shuker, P. J. Hajduk, R. P. Meadows, *Protein Eng.***1997**, *10*, 73-73.

[59] C. Dalvit, M. Flocco, M. Veronesi, B. J. Stockman, *Comb. Chem. High Throughput Screen.* **2002**, *5*, 605-611.

[60] W. L. Jorgensen, *Science* **2004**, *303*, 1813-1818.

[61] G. B. Barcellos, I. Pauli, R. A. Caceres, L. F. S. M. Timmers, R. Dias et al., *Curr. Drug Targets* **2008**, *9*, 1084-1091.

[62] W. P. Walters, M. T. Stahl, M. A. Murcko, *Drug Discov. Today* **1998**, *3*, 160-178.

[63] B. K. Shoichet, S. L. McGovern, B. Q. Wei, J. J. Irwin, *Curr. Opin. Chem. Biol.* **2002**, *6*, 439-446.

[64] J. F. Blake, E. R. Laird, *Annu. Rep. Med. Chem.* **2003**, *38*, 305-314.

[65] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Deliv. Rev.* **2001**, *46*, 3-26.

[66] R. D. Taylor, P. J. Jewsbury, J. W. Essex, *J. Comput. Aided Mol. Des*. **2002**, *16*, 151-166.

[67] T. Lengauer, M. Rarey, *Curr. Opin. Struct. Biol.* **1996**, *6*, 402-406.

[68] D. Hoffmann, B. Kramer, T. Washio, T. Steinmetzer, M. Rarey et al., *J. Med. Chem.* **1999**, *42*, 4422-4433.

[69] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew et al., *J. Comput. Chem.* **2009**, *in press.*

[70] M. Rarey, B. Kramer, T. Lengauer, *J. Comput. Aided Mol. Des.***1997**, *11*, 369-384.

[71] H. J. Böhm, *J. Comput. Aided Mol. Des.***1994**, *8*, 623-632.

[72] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart et al., *J. Comput. Chem.* **1998**, *19*, 1639-1662.

[73] P. S. Charifson, J. J. Corkery, M. A. Murcko, W. P. Walters, *J. Med. Chem.* **1999**, *42*, 5100-5109.

[74] C. Bissantz, G. Folkers, D. Rognan, *J. Med. Chem.* **2000**, *43*, 4759-4767.

[75] C. Perez, A. R. Ortiz, *J. Med. Chem.* **2001**, *44*, 3768-3785.

[76] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, *J. Mol. Biol.* **1996**, *261*, 470-489.

[77] M. N. Corbeil CR, *J. Chem. Inf. Model.* **2009**, *49*, 997-1009.

[78] C. Hansch, A. Leo, *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*, American Chemical Society Washington DC, **1995**.

[79] V. C. R. Todeschini, *Handbook of Molecular Descriptors*, Wiley-VCH Weinheim, **2000**.

[80] P. Gramatica, *A short history of QSAR evolution*, QSAR Research Unit in Environmental Chemistry and Ecotoxicology Italy, **2007**.

[81] N. B. A. Schuffenhauer, *Drug Discov. Today* **2007**, *3*, 387-395.

[82] P. Gedeck, R. A. Lewis, *Curr. Opin. Drug Discov. Devel.* **2008**, *11*, 569-575.

[83] N. L. A. U. Burkert, *Molecular Mechanics*, American Chemical Society Washington DC, **1982**.

[84] G. Klebe, *Three-Dimensional Quantitative Structure Activity Relationships: 3D QSAR in Drug Design: Recent Advances*, Springer Netherlands, **1998**.

[85] R. D. Cramer III, J. D. Bunce, D. E. Patterson, I. E. Frank, *Quant. Struct.-Act. Rel.* **1988**, *7*, 18-25.

[86] A. M. Doweyko, *J. Comput. Aided Mol. Des.***2008**, *22*, 81-89.

[87] J. M. Haile, *Molecular Dynamics Simulation: Elementary Methods*, John Wiley & Sons New York, **1992**.

[88] W. F. van Gunsteren, H. J. C. Berendsen, *Angew. Chem. Int. Ed.* **1990**, *29*, 992-1023.

[89] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark et al., *J. Comput. Chem.* **2005**, *26*, 1701-1718.

[90] C. Oostenbrink, A. Villa, A. E. Mark, W. F. van Gunsteren, *J. Comput. Chem.* **2004**, *25*, 1656-1676.

[91] C. Oostenbrink, A. Villa, A. E. Mark, W. F. van Gunsteren, *J. Comput. Chem.* **2004**, *25*, 1656-1676.

[92] M. Christen, P. H. Hünenberger, D. Bakowies, R. Baron, R. Burgi et al., *J. Comput. Chem.* **2005**, *26*, 1719-1751.

[93] W. F. van Gunsteren, T. Huber, A. E. Torda, *Proc. Eur. Conf.* **1995**, *330*, 253-268.

[94] Y. Okamoto, *J. Mol. Graphics Modell.* **2004**, *22*, 425-439.

[95] K. Tai, *Biophys. Chem.* **2004**, *107*, 213-220.

[96] T. Huber, A. E. Torda, W. F. van Gunsteren, *J. Comput. Aided Mol. Des.***1994**, *8*, 695-708.

[97] H. Grubmüller, *Phys. Rev. E* **1995**, *52*, 2893-2906.

[98] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, *Science* **1983**, *220*, 671-680.

[99] M. Karplus, J. N. Kushick, *Macromolecules* **1981**, *14*, 325-332.

[100] H. Kessler, *Fresenius J. Anal. Chem.* **1987**, *327*, 66-67.

[101] H. B. Yu, M. Ramseier, R. Burgi, W. F. van Gunsteren, *Chemphyschem* **2004**, *5*, 633-641.

[102] R. Burgi, X. Daura, A. Mark, M. Bellanda, S. Mammi et al., *J. Pept. Res.* **2001**, *57*, 107-118.

[103] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, *J. Chem. Theory Comput.* **2008**, *4*, 435-447.

[104] D. F. Mierke, S. G. Grdadolnik, H. Kessler, *J. Am. Chem. Soc.* **1992**, *114*, 8283-8284.

[105] W. A. Thomas, *Prog. Nucl. Mag. Res. Sp.* **1997**, *30*, 183-207.

[106] R. H. Contreras, J. E. Peralta, *Prog. Nucl. Mag. Res. Sp.* **2000**, *37*, 321-425.

[107] M. Eberstadt, G. Gemmecker, D. F. Mierke, H. Kessler, *Angew. Chem. Int. Ed.* **1995**, *34*, 1671-1695.

[108] R. Aydin, H. Günther, *Magn. Reson. Chem.* **1990**, *28*, 448-457.

[109] V. F. Bystrov, *Prog. Nucl. Mag. Res. Sp.* **1976**, *10*, 41-81.

[110] J. J. Titman, J. Keeler, *J. Magn. Reson.* **1990**, *89*, 640-646.

[111] C. Griesinger, , O. W. Sørensen, , R. R. Ernst, *J. Chem. Phys.* **1986**, *85*, 6837-6852.

[112] G. A. Zhu, A. Renwick, A. Bax, *J. Magn. Reson. A* **1994**, *110*, 257-261.

[113] H. Kessler, C. Griesinger, K. Wagner, *J. Am. Chem. Soc.* **1987**, *109*, 6927-6933.

[114] H. Kessler, C. Griesinger, J. Zarbock, H. R. Loosli, *J. Magn. Reson.* **1984**, *57*, 331-336.

[115] M. Kurz, P. Schmieder, H. Kessler, *Angew. Chem. Int. Ed.* **1991**, *30*, 1329-1331.

[116] S. Grzesiek, F. Cordier, V. Jaravine, M. Barfield, *Prog. Nucl. Mag. Res. Sp.* **2004**, *45*, 275-300.

[117] M. Heller, M. Sukopp, N. Tsomaia, M. John, D. F. Mierke et al., *J. Am. Chem. Soc.* **2006**, *128*, 13806-13814.

[118] D. Neuhaus, M. P. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, Wiley-VCH Weinheim, **2000**.

[119] J. Jeener, B. H. Meier, P. Bachmann, R. R. Ernst, *J. Chem. Phys.* **1979**, *71*, 4546-4553.

[120] A. A. Bothner-By, R. L. Stephens, J. M. Lee, C. D. Warren, R. W. Jeanloz, *J. Am. Chem. Soc.* **1984**, *106*, 811-813.

[121] T. E. Bull, *Prog. Nucl. Mag. Res. Sp.* **1992**, *24*, 377-410.

[122] T. M. G. Koning, R. Boelens, R. Kaptein, *J. Magn. Reson.* **1990**, *90*, 111-123.

[123] K. Wüthrich, M. Billeter, W. Braun, *J. Mol. Biol.* **1983**, *169*, 949-961.

[124] J. W. Emsley, J. Lindon, *NMR Spectroscopy using liquid crystal solvents*, Pergamon Press Oxford, **1975**.

[125] B. Deloche, E. T. Samulski, *Macromolecules* **1981**, *14*, 575-581.

[126] B. Luy, K. Kobzar, H. Kessler, *Angew. Chem. Int. Ed.* **2004**, *43*, 1092-1094.

[127] T. Carlomagno, W. Peti, C. Griesinger, *J. Biomol. NMR* **2000**, *17*, 99-109.

[128] P. Nolis, J. F. Espinosa, T. Parella, *J. Magn. Reson.* **2006**, *180*, 39- 50.

[129] P. Tzvetkova, S. Simova, B. Luy, *J. Magn. Reson.* **2007**, *186*, 193-200.

[130] C. Griesinger, O. W. Sørensen, R. R. Ernst, *J. Am. Chem. Soc.* **1985**, *107*, 6394-6396.

[131] C. Griesinger, O. W. Sørensen, R. R. Ernst, *J. Chem. Phys.* **1986**, *85*, 6837-6852.

[132] A. Bax, R. Freeman, T. A. Frenkiel, M. H. Levitt, *J. Magn. Reson.* **1981**, *43*, 478-483.

[133] J. L. Yan, A. D. Kline, H. P. Mo, M. J. Shapiro, E. R. Zartler, *J. Org. Chem.* **2003**, *68*, 1786-1795.

[134] C. M. Thiele, S. Berger, *Org. Lett.* **2003**, *5*, 705-708.

[135] L. Verdier, P. Sakhaii, M. Zweckstetter, C. Griesinger, *J. Magn. Reson.* **2003**, *163*, 353-359.

[136] C. Aroulanda, V. Boucard, F. Guibe, J. Courtieu, D. Merlet, *Chemistry* **2003**, *9*, 4536-4539.

[137] J. Freudenberger, P. Spitteler, R. Bauer, H. Kessler, B. Luy, *J. Am. Chem. Soc.* **2004**, *126*, 14690-14691.

[138] J. C. Freudenberger, S. Knör, K. Kobzar, D. Heckmann, T. Paululat et al., *Angew. Chem. Int. Ed.* **2005**, *44*, 423-426.

[139] L. M. Blumenthal, *Theory and Application of Distance Geometry*, Chelsea Publishing New York, **1970**.

[140] W. Braun, N. Go, *J. Mol. Biol.* **1985**, *186*, 611-626.

[141] J. Kuszewski, M. Nilges, A. T. Brünger, *J. Biomol. NMR* **1992**, *2*, 33-56.

[142] T. F. Havel, *Biopolymers* **1990**, *29*, 1565-1585.

[143] G. M. Crippen, T. F. Havel, D*istance Geometry and Molecular Conformation*, John Wiley & Sons New York, **1988**.

[144] G. M. Crippen, *J. Comput. Chem.* **1982**, *3*, 471-476.

[145] T. C. Beutler, W. F. van Gunsteren, *J. Chem. Phys.* **1994**, *101*, 1417-1422.

[146] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid et al., *J. Comput. Chem.* **2005**, *26*, 1781-1802.

[147] D. A. Case, T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo et al., *J. Comput. Chem.* **2005**, *26*, 1668-1688.

[148] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan et al., *J. Comput. Chem.* **1983**, *4*, 187-217.

[149] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **1996**, *118*, 11225-11236.

[150] J. W. Ponder, D. A. Case, *Adv. Prot. Chem.* **2003**, *66*, 27-85.

[151] A. D. MacKerell Jr, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck et al., *J. Phys. Chem. B* **1998**, *102*, 3586-3616.

[152] A. T. Hagler, C. S. Ewig, *Comput. Phys. Commun.* **1994**, *84*, 131-155.

[153] M. Clark, R. D. Cramer III, N. van Opdenbosch, *J. Comput. Chem.* **1989**, *10*, 982-1012.

[154] G. Kummerlöwe, S. Knör, A. O. Frank, T. Paululat, H. Kessler et al., *Chem. Commun.* **2008**, , 5722-5724.

[155] W. F. van Gunsteren, H. J. C. Berendsen, *Mol. Sim.* **1988**, *1*, 173-185.

[156] J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327-341.

[157] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, *J. Comput. Chem.* **1997**, *18*, 1463-1472.

[158] C. Oostenbrink, T. A. Soares, N. F. A. van der Vegt, W. F. van Gunsteren, *Eur. Biophys. J.* **2005**, *34*, 273-284.

[159] W. L. Jorgensen, J. Tirado-Rives, *Abstr. Pap. Am. Chem. Soc.* **1998**, *216*, U696-U696.

[160] B. A. Luty, W. F. van Gunsteren, *J. Phys. Chem.* **1996**, *100*, 2581-2587.

[161] P. H. Hünenberger, J. A. McCammon, *Biophys. Chem.* **1999**, *78*, 69-88.

[162] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089-10092.

[163] W. L. Jorgensen, J. D. Madura, C. J. Swenson, *J. Am. Chem. Soc.* **1984**, *106*, 6638-6646.

[164] A. E. Torda, R. M. Scheek, W. F. van Gunsteren, *Chem. Phys. Lett.* **1989**, *157*, 289-294.

[165] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Dinola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684-3690.

[166] H. Bekker, E. J. Dijkstra, M. K. R. Renardus, H. J. C. Berendsen, *Mol. Sim.* **1995**, *14*, 137-151.

[167] D. J. Adams, E. M. Adams, G. J. Hills, *Mol. Phys.* **1979**, *38*, 387-400.

[168] D. van der Spoel, E. Lindahl, B. Hess, A. R. van Buuren, E. Apol et al., *Gromacs User Manual Version 4.0*, www.gromacs.org, **2005**.

[169] K. Zimmermann, *J. Comput. Chem.* **1991**, *12*, 310-319.

[170] A. Glättli, X. Daura, P. Bindschädler, B. Jaun, Y. R. Mahajan et al., *Chemistry* **2005**, *11*, 7276-7293.

[171] J. Lautz, H. Kessler, W. F. van Gunsteren, H. P. Weber, R. M. Wenger, *Biopolymers* **1990**, *29*, 1669-1687.

[172] W. F. van Gunsteren, M. Karplus, *Nature* **1981**, *293*, 677-678.

[173] L. Wesson, D. Eisenberg, *Protein Sci.* **1992**, *1*, 227-235.

[174] A. E. Torda, R. M. Brunne, T. Huber, H. Kessler, W. F. van Gunsteren, *J. Biomol. NMR* **1993**, *3*, 55-66.

[175] B. Hess, R. M. Scheek, *J. Magn. Reson.* **2003**, *164*, 19-27.

[176] M. A. Dechantsreiter, E. Planker, B. Mathä, E. Lohof, G. Hölzemann et al., *J. Med. Chem.* **1999**, *42*, 3033-3040.

[177] Y. Sugita, Y. Okamoto, *Chem. Phys. Lett.* **1999**, *314*, 141-151.

[178] R. Affentranger, I. Tavernelli, E. E. Di Lorio, *J. Chem. Theory Comput.* **2006**, *2*, 217-228.

[179] M. B. Kubitzki, B. L. de Groot, *Biophys. J.* **2007**, *92*, 4262-4270.

[180] D. P. Geerke, C. Oostenbrink, N. F. A. van der Vegt, W. F. van Gunsteren, *J. Phys. Chem. B* **2004**, *108*, 1436-1445.

[181] A. E. Torda, W. F. van Gunsteren, *J. Comput. Chem.* **1994**, *15*, 1331-1340.

[182] H. Kessler, C. Griesinger, J. Lautz, A. Müller, W. F. van Gunsteren et al., *J. Am. Chem. Soc.* **1988**, *110*, 3393-3396.

2 Taking Focused Photographs

Determination of Molecular Arrangements

NMR spectroscopy is in addition to X-ray diffraction the only available method to elucidate three-dimensional structures of either small compounds or even very large biomacromolecules at atomistic resolution. The most prominent advantage of NMR based structural studies is that the molecule of interest is investigated in solution which usually represents - in a biological sense - its natural environment. Artificial packing effects of molecules with a small volume to surface ratio as they are often observed in X-ray crystallography structures do not arise in solution-state NMR spectroscopy. The determination of molecular structures can be of different character. In some cases, the elucidation of the constitution of a molecule is sufficient to answer a specific scientific question. For example, short or medium-sized linear peptide chains usually adopt no preferred conformations in solution. Hence, the calculation of their three-dimensional arrangement is not possible or not of concern. By contrast, molecules like cyclic peptides which are conformationally restricted often display more or less preferred structures. In such cases, not only informations about the constitution of a molecule is important but also its spatial assembly. For molecules which are of biological or medicinal interest, the knowledge of their three-dimensional arrangement(s) is even more essential since differing conformations usually exhibit different receptor affinities. Beside the constitution and the conformation of a molecule, its configuration is also often subject of structure determination procedures. In many cases, obscure stereocenters of molecules with known constitution must be elucidated via structure analysis approaches.

2.1 *N*-Methylated Amino Acids as Proline Substitutes in Conformational Design

The presence of a tertiary amide bond both in *N*-methylated amino acids and proline constituting peptides is tempting to consider that their corresponding stereo-impact on the backbone of peptides is similar. Although theoretical explanations on the factors influencing the conformation of proline or *N*-methylated amino acid containing peptides are available,[1-14] there is a lack of explicit experimental comparisons. It is known that proline markedly influences protein architectures due to its constrained five membered pyrrolidine ring and, in addition, acts as a reverse turn inducer in designed cyclic protein epitope mimetics.[15-19] On the other hand, *N*-methylated amino acids also have the potential to introduce turn structures into cyclic peptides.[[7,18-20]

Furthermore, it was shown that both proline and *N*-methylated amino acids can introduce *cis* peptide bonds into peptide sequences.[5,7,21-25] However, in contrast to L-proline in which the φ angle is constrained to ca. -60° due to the large pyrrolidine ring, *N*-methylated L-amino acids have much more flexibility as the φ angle can vary between -100° and +60°.[5] Hence, it is not a priori predictable if mutual substitution of both units has an identical influence on the overall peptide structure.

2.1.1 The Scientific Question

Although *N*-methylation of amide bonds has been used in peptide chemistry for nearly one hundred years,[26] introduction of proline or ring-derivated prolines in peptidic sequences is the most popular method until today to induce turns or *cis* peptide bonds in peptides since preparation of *N*-methylated amino acids on solid support came along with huge difficulties in their synthesis.[27,28] However, ring-derivated prolines often need extensive synthesis or do not force a molecule into a single preferred conformation.[29] Nowadays, the synthetic problems concerning *N*-methylated amino acids have been solved and peptides including them are easily accessible.[30,31]

Therefore, the question arose if it might be possible to replace a proline with an *N*-methylated amino acid without perturbing a favored molecular conformation.

If this holds true, one regains the side chain functionality which is lost in proline-type amino acids because of the five membered pyrrolidine ring. This side chain could subsequently be used to improve the activity, selectivity or can act as a pharmacophore in a peptide. Furthermore, it has been recently shown that *N*-methylation of peptide bonds could even increase the biological activity, receptor selectivity, enzymatic stability, and could turn orally unavailable peptides into bioavailable potential drug candidates.[6,20,32]

2.1.2 Alanine as Template Amino Acids

For the comparison of *N*-methylated and proline containing pentapeptides, alanine was chosen as amino acid template for *N*-methylation since its side chain has only minor influence on the overall backbone conformation and bears - in contrast to the smallest amino acid glycine - a chiral center.[33-37]

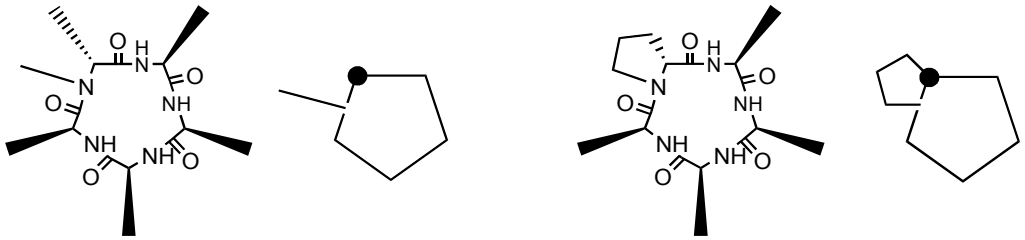


Figure 2.1: The template peptides *c(MeaAAAA)* and *c(pAAAA)* represented as both regular and schematic structures. The D-residue is marked as a black dot in the upper left corner.

Subsequently, the positions of *N*-methylated alanine or proline in asymmetric (a D-residue containing) cyclic peptides were spatially screened (see Fig. 2.1). Finally, thermodynamic equilibria of representative peptides were investigated via NMR spectroscopy.

2.1.3 Peptides Containing One *N*-Methylation or One Proline

The first comparison where only one *N*-methylation or one proline was varied led to the promising result that, in nearly every case, *N*-methylated alanine and proline show similar results with respect to the highest populated conformation and, in addition, are in good agreement concerning the *cis* / *trans* ratios of amide bonds (see Fig 2.2). Especially when introduced at the *i*-position (the position of the D-residue in the cyclic peptide; please note that "*i*" is also often used as starting residue in turn structures) Me-D-alanine and D-proline both revealed only one highly populated conformation and both

peptides are in the all-*trans* conformation. Both structures possess a *β*-turn in the upper part which is introduced by the D-residue and a *γ*-turn in the lower part of the molecule whereas the peptide bond between Ala2 and Ala3 is known to flip around its adjacent φ and ψ dihedral angles (flip of 180°). As both structures have been extensively discussed in literature, there is no need to go into further detail.[1-5,7]

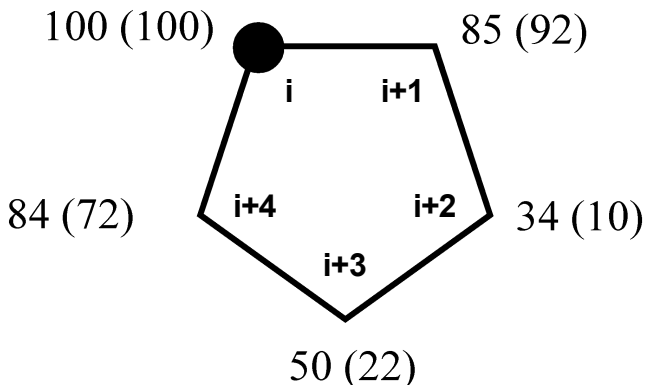


Figure 2.2: Populations of the all-*trans* conformations for the peptides with one Me-alanine or proline at different positions given in percent. Remaining percentages represent peptide populations containing a *cis* peptide bond between the *N*-alkylated amino acid and the preceding one. The D-residue is marked as a black dot in the upper left corner and is denoted as position "i".

Also when being introduced in the *i+1*-position, the all-*trans* conformation is preferred in both cases with a population of 85 % for the alanine and 92 % for the proline peptide. Both peptides show a nearly identical structure; this result shows that the replacement of one with another without changing the backbone conformation is possible (see superposition in Fig 2.3).

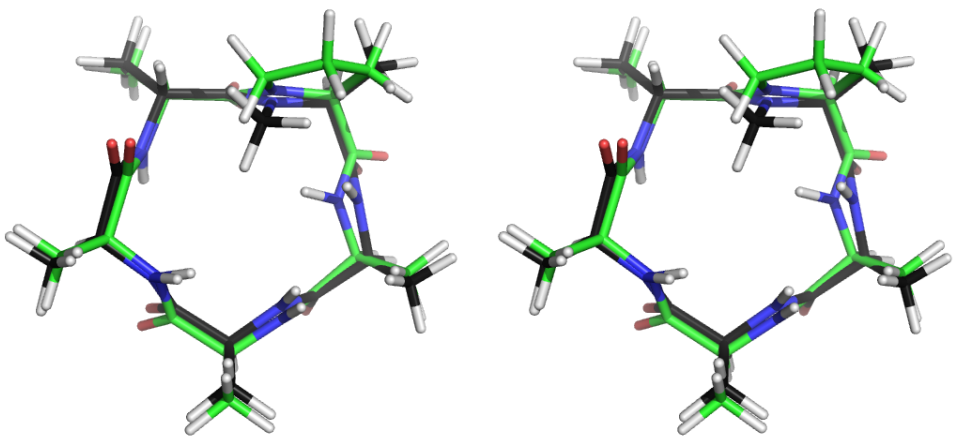


Figure 2.3: Stereoview of the superposition of the peptides *c(aMeAAAA)* and *c(aPAAA)*.

When the *N*-alkylated residue is introduced in the *i+2* or in the *i+3*-position, the all-*trans* conformation is disfavored. Especially for proline, a *cis* peptide bond is preferred at the position of the *N*-alkylation finally resulting in a conformation

with a population of about 90 % for *c(aAPAA)* (in general, a small letter indicates a D-residue) with a *cis* peptide bond at the position of proline. Being introduced at the *i+4*-position, the all-*trans* conformation is again favored. However, for *c(aAAAMeA)*, three conformations with a population of 84 / 13 / 3 are present where the two main conformations possess an all-*trans* conformation. *c(aAAAP)* also shows three conformations with a ratio of 75 / 19 / 6 where the main conformation also exhibits an all-*trans* arrangement.

To summarize, a MeAla has the ability to replace a single proline in cyclic pentapeptides without having an undesired impact on the backbone conformation. Furthermore, the *N*-methylated amino acid might be preferred because one can retain the side chain functionality which is lost when using proline. To prove this hypothesis, Me-D-alanine in *c(MeaAAAA)* was replaced by an *N*-methylated D-lysine which resulted in obtaining *c(MekAAAA)* with a population of 100 % (see Fig 2.4).

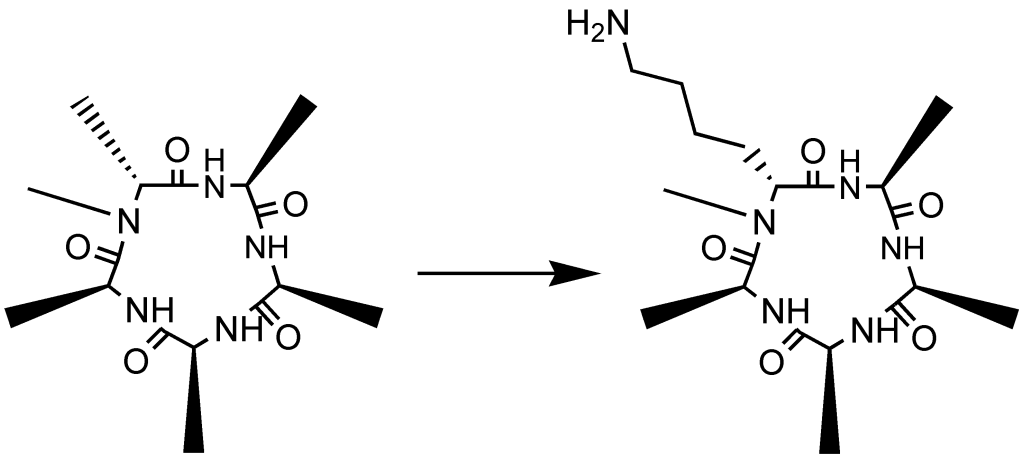


Figure 2.4: A schematic presentation of the peptide *c(MeaAAAA)* that exhibits an all-*trans* population of 100 %. The molecule was utilized as a template for *c(MekAAAA)* which also revealed a population of 100 %. It becomes obvious that alanine can be used as a template for amino acids with a longer side-chain.

2.1.4 Peptides Containing Two *N*-Methylations or Two Prolines

As the population differences between the all-*trans* conformations and the structures containing one *cis* peptide bond were the lowest in *c(aAAMeAA)* (1) and *c(aAAPA)* (2), it was decided to fix the *N*-alkylated amino acid at the *i+3*-position and to rotate another proline or MeAla through the different positions of the cyclic pentapeptide. Thus, the goal was to investigate the presence of a second *N*-alkylated amino acid alters the *cis* / *trans* equilibrium; finally, the possible occurrence of new preferred conformations was tested.

Since sequences containing two prolines are not well suited for an application in medicinal chemistry as two of the possible side chain functionalities are lost,

the incorporation of two *N*-methylated amino acids is usually preferred because all side chains can carry pharmacophoric functional groups. As can be seen in figure 2.5, there is a big population difference between *c(pAAPA)* (4) and *c(MeaAAMeAA)* (3) as the latter shows an all-*trans* arrangement with a population of about 70 % while in *c(pAAPA)* (4) the conformation with one *cis* and all other peptide bonds in *trans* configuration is the highest populated conformation (~ 65 %). For the two cyclic pentapeptides *c(aMeAAMeAA)* (5) and *c(aPAPA)* (6), the peptide with *N*-methylations prefers a highly populated conformation where one amide bond is in *cis* and the other *N*-alkylated amide bond is in *trans* conformation. In contrast, the cyclic peptide with two prolines shows a main population with two *cis* amide bonds at the positions of the prolines. In the case of *c(aAMeAMeAA)* (7) and *c(aAPPA)* (8), the all-*trans* structure is preferred for the peptide with *N*-methylations whereas for the proline containing peptide the conformation with one *cis* amide bond is favored.

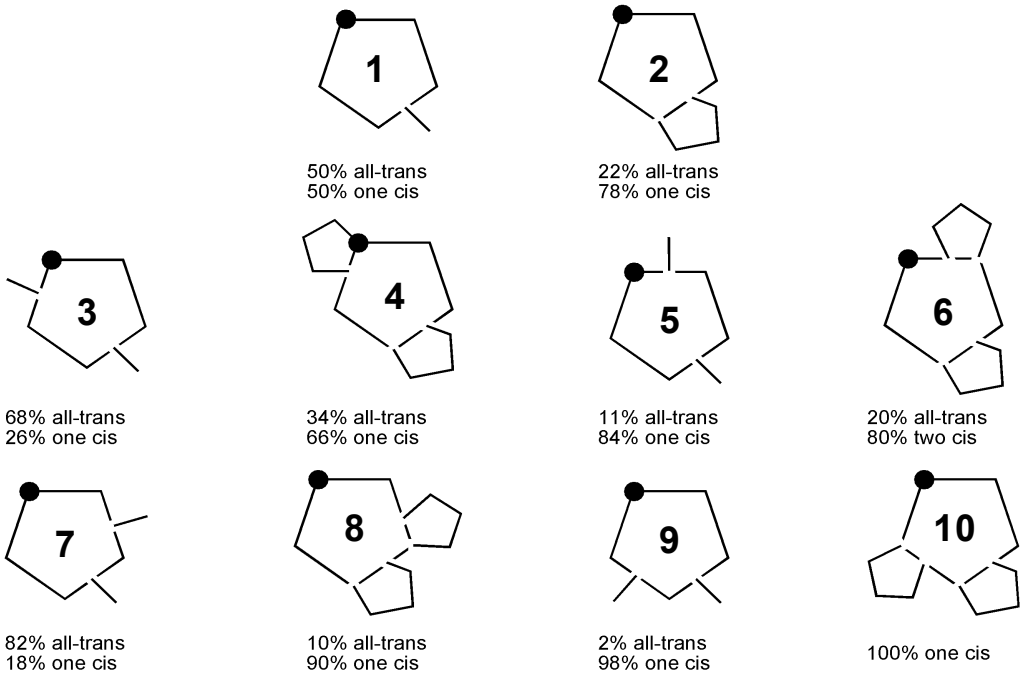


Figure 2.5: Comparison of *cis* / *trans* populations of different *N*-methylated and proline containing cyclic peptides. *Cis* amide bonds occur between *N*-alkylated amino acids and previous ones. Two of ten peptides revealed more than two conformations at the NMR time scale. In these cases, only the populations of the two major conformers are given, resulting in percentage statements that do not sum up to *100 %*.

For *c(aAAMeAMeA)* (9) and *c(aAAPP)* (10), both compounds show a preferred conformation which is highly populated bearing one *cis*-peptide bond between *i+3* and *i+4* (see superposition in Fig. 2.6).

In conclusion, structural features of a cyclic peptide are no longer predictable when two *N*-methylations or two prolines are present since peptides bearing more than one *N*-methylated amino acid often differ in their conformation

compared to the proline containing analogues. One benefit of this finding is the fact that replacement of proline by an *N*-methylated amino acid most often results in a different structure which could be useful in the search for candidates better fitting into a binding pocket.

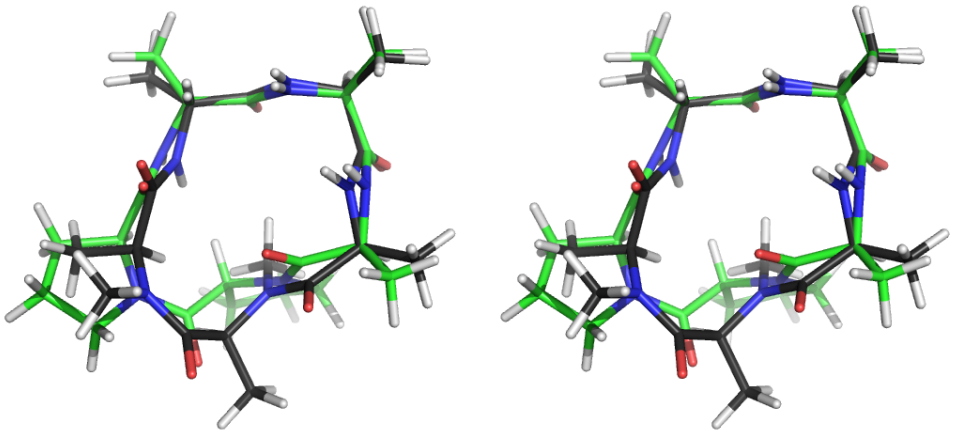


Figure 2.6: Stereoview of the superposition of the peptides *c(aAAMeAMeA)* and *c(aAAPP)*.

2.1.5 Peptides Containing One *N*-Methylation and One Proline

Finally, the conformation of cyclic peptides when both, a proline and an *N*-methylated amino acid, are incorporated, were investigated. In this study, one proline was fixed at the *i+3* position and the position of the *N*-alkylation was varied (and *vice versa*). The obtained set of compounds showed preferred conformations whereas the highest populated conformation of each peptide was in the range of 65-70 % (see Fig. 2.7).

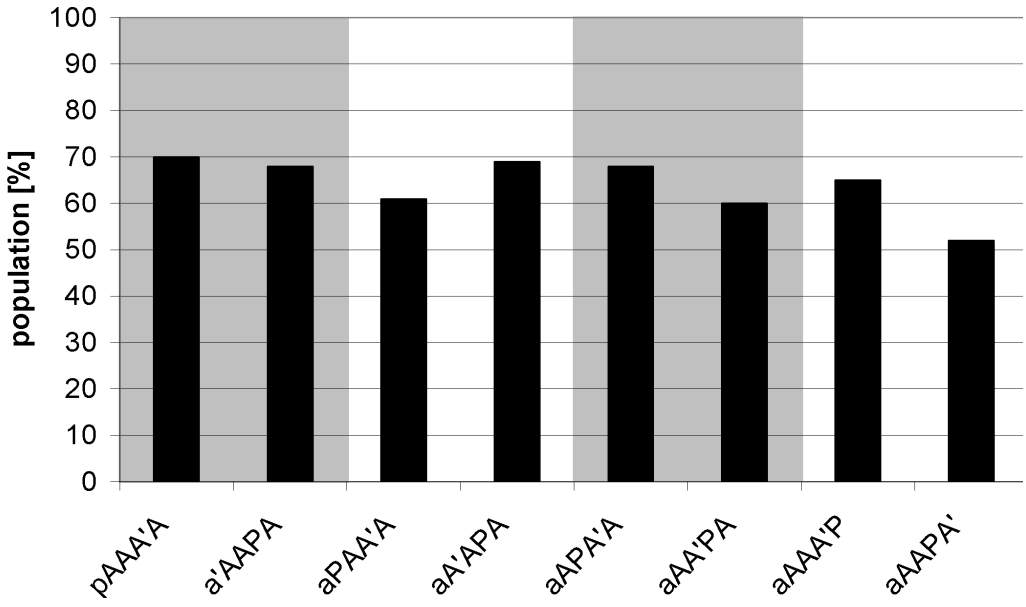


Figure 2.7: Comparison of peptides which contain one proline (or D-proline) and one *N*-methylated alanine (A' for Me-L-Ala and a' for Me-D-Ala). The peptides are grouped in pairs in which both units are interchanged. Populations of dominating conformers are given in percentages.

Hence, it was decided not to go into further detail about *cis* or *trans* amide bond distributions as peptides that show such poor population preferences are of limited interest in medicinal chemistry.

2.1.6 Conclusion

Summing up, it was shown that cyclic peptides containing one proline or one *N*-methylated amino acid reveal comparable conformational features. Both the ratios of all-*trans* and *cis* conformations and the three-dimensional structures are quite similar for *N*-methylated and proline containing peptides. This allows substituting a proline residue by an *N*-methylated amino acid without changing the overall conformation. Proline is often used as inducer of distinct conformations, e.g. turns or breaking helices. However, the position where a proline is located in a cyclic peptide can not be used in an easy way (with respect to synthesis procedures) to display a chemical group important for receptor binding. The use of *N*-methylated amino acids allows for bringing additional side chain functionality into conformationally restricted (cyclic) peptides and, therefore, opens new perspectives for drug design.

Furthermore, care should be taken when two prolines are to be replaced by *N*-methylated amino acids because the resulting structures usually differ in their conformation. Finally, incorporation of one proline and one *N*-methylated amino acid into cyclic pentapeptides seems not to be useful since no strong preference of one single conformation was observed in the here demonstrated cases, making the investigated candidates uninteresting template structures in medicinal chemistry.

2.1.7 Reagents, Methods, and Experiments

All informations regarding used reagents, peptide synthesis, and purification as well as NMR spectra can be found in literature.[5,7]

All spectra were recorded on a Bruker DMX 500 MHz spectrometer equipped with a TXI probe head. The measurements were performed at 300 K, and peptides were dissolved in d_6-DMSO. The assignment of proton and carbon resonances followed a strategy described in literature.[38] Sequential assignment was accomplished by through-bond connectivities from heteronuclear multiple bond correlation (HMBC) spectra whereas *N*-methyl groups served as

starting point.[39] Connectivities were proved by interresidual scalar couplings, e.g between carbonyl carbons and adjacent amide protons. TOCSY spectra were recorded with a mixing time of 60 ms, ROESY spectra with a mixing time of 150 ms, thus avoiding unwanted effects caused by spin diffusion.[40,41] Conformational exchange was proven by detection of inverted signal signs in ROESY spectra.[42] Proton distances were calculated according to the isolated two-spin approximation from volume integrals of ROESY spectra.[43] No ROE offset correction was performed since biasing offset effects at the field strength used in this study are rather small. Evidence of *cis* peptide bonds in proline containing peptides was achieved using proline Cβ and Cγ shifts in DEPT45 spectra.[44] The ratio of different conformational populations was determined via the integrals of amide and Hα-signals in ^{1}H-1D spectra.

The integrated volumes of ROE cross peaks were converted to proton–proton distances by the help of calibration to an averaged alanine Hα–Hβ* distance as reference (2.45 Å; including pseudo-atom correction). Upper and lower distance restraints were obtained by adding and subtracting 10 % to the calculated experimental values, thus accounting for experimental errors and simulation uncertainties. Metric matrix DG calculations were carried out with a home-written distance geometry program utilizing random metrization.[45] Experimental distance restraints which are more restrictive than the geometric distance bounds (holonomic restraints) were used to create the final distance matrix. All structure templates were first embedded in four dimensions and then partially minimized using conjugate gradient minimization followed by distance-driven dynamics (DDD) where only distance constraints were used.[46] The DDD simulation was carried out at 1000 K for 50 ps with a gradual reduction in temperature over the next 30 ps. The DDD procedure utilized holonomic and experimental distance constraints plus a chiral penalty function for the generation of violation energies and forces. A distance matrix was calculated from each structure, and the EMBED algorithm was used to compute Cartesian coordinates in three dimensions. 100 structures were calculated for each peptide, and > 90 % of the structure bundles of each peptide did not show any significant violations (> 0.2 Å). Molecular dynamics calculations were carried out using the program DISCOVER and the CVFF force field.[56] Structures resulting from DG calculations were placed in a cubic box with a vector length of 3.0 nm and soaked with DMSO. Intramolecular distances of all peptides were constrained according to the experimental values. After energy minimization using the steepest descent and the

conjugate gradient algorithms, the system was gradually heated in 50 K steps (equilibration time at each temperature was 2 ps) starting from 10 K, each by direct scaling of velocities. The system was equilibrated for 50 ps with weak temperature coupling at 300 K. Configurations in the subsequent production runs (150 ps) were saved every 100 fs. Finally, 150 ps of free MD simulations at 300 K were carried out in order to prove that no significant structural changes occur when no distance restraints are present during the simulation.

2.1.8 Declaration

The presented topic was conducted in collaboration with M.Sc. Burkhardt Laufer and Dr. Jayanta Chatterjee (Technical University Munich, Department Chemistry). B.L. and J.C. synthesized all shown molecules and calculated some conformations. The described results have been published as follows:

B. Laufer, J. Chatterjee, A.O. Frank, H. Kessler. Can *N*-Methylated Amino Acids Serve as Substitute for Prolines in Conformational Design? *J. Pept. Sci.* **2009**, *15*, 141-146.

2.2 An Alternative Approach of Configurational and Conformational Analysis

Residual dipolar couplings as anisotropic NMR parameters play an important role for the structure determination of biomolecules and small organic molecules in solution.[47,48] For the measurement of anisotropic parameters, liquid crystalline phases or stretched polymer gels can be used for achieving the necessary partial alignment, with gels having the advantage of arbitrary scalability of the alignment strength.[49-53] In general, RDCs are most often applied to refine spatial arrangements which were elucidated by means of NOE based pairwise atom distance informations and dihedral restraints yielded via J-couplings. Moreover, RDCs can also be utilized to ascertain molecular structures in addition to or even without the contribution of other NMR parameters.[54] In this respect, not only conformations can be elucidated but also the configuration of molecules with unknown stereochemistry is determinable.[55] For this purpose, however, it is indispensable to extract unambiguous RDCs from NMR spectra with sufficient accuracy.

2.2.1 The Scientific Question

For being of practical use, RDC data must be fitted onto a 3D structural model of the molecule under investigation. In this scope, several methods have been developed with the concept of singular value decomposition (SVD) as the most prominent approach.[56] Nevertheless, even the most powerful tools do not consider the flexibility of molecules.[56] For relatively rigid structures, this drawback does not noticeably affect the quality of yielded results; for particles which exhibit multiple degrees of freedom, however, often no adequate fitting of RDCs can be obtained.

Thus, the central goal of this project was to prove the applicability and accuracy of an RDC driven configurational and conformational fitting approach that includes the dynamical behavior of a medium-sized organic molecule.

For this purpose, it was tested if the known structure of the natural product staurosporine could be reproduced by means of RDC restrained molecular dynamics simulations. At first, deuterated polystyrene as new alignment medium was used to enable a precise measurement of residual dipolar couplings. Then, two computational approaches for the structural analysis of staurosporine were compared: the established but "static“ singular value decomposition that comes with the PALES program and orientation restraining during molecular dynamics simulations within the software program package GROMACS.[56-58]

2.2.2 Structural Aspects of Staurosporine

The natural product staurosporine (see Fig. 2.8) which is isolated from *Streptomyces staurosporeus* consists of an extensive aromatic ring system and a six-membered tetrahydropyran ring attached via two nitrogen atoms.[59] The tetrahydropyran ring, in principle, contains four stereogenic centers but cyclization reduces the degrees of freedom to three, resulting in $2^3 = 8$ possible configurations (four diastereomers SRRR, SRSR, SSRR, SSSR, and their enantiomers RSSS, RSRS, RRSS, and RRRS). NMR experiments are usually carried out in an achiral environment; hence, enantiomers result in identical spectra and only four diastereomers of staurosporine have to be

considered. Nevertheless - in addition to the differing configurations - chair or boat conformations of the six-membered ring must be taken into consideration, finally resulting in a total of eight distinguishable structures of staurosporine.

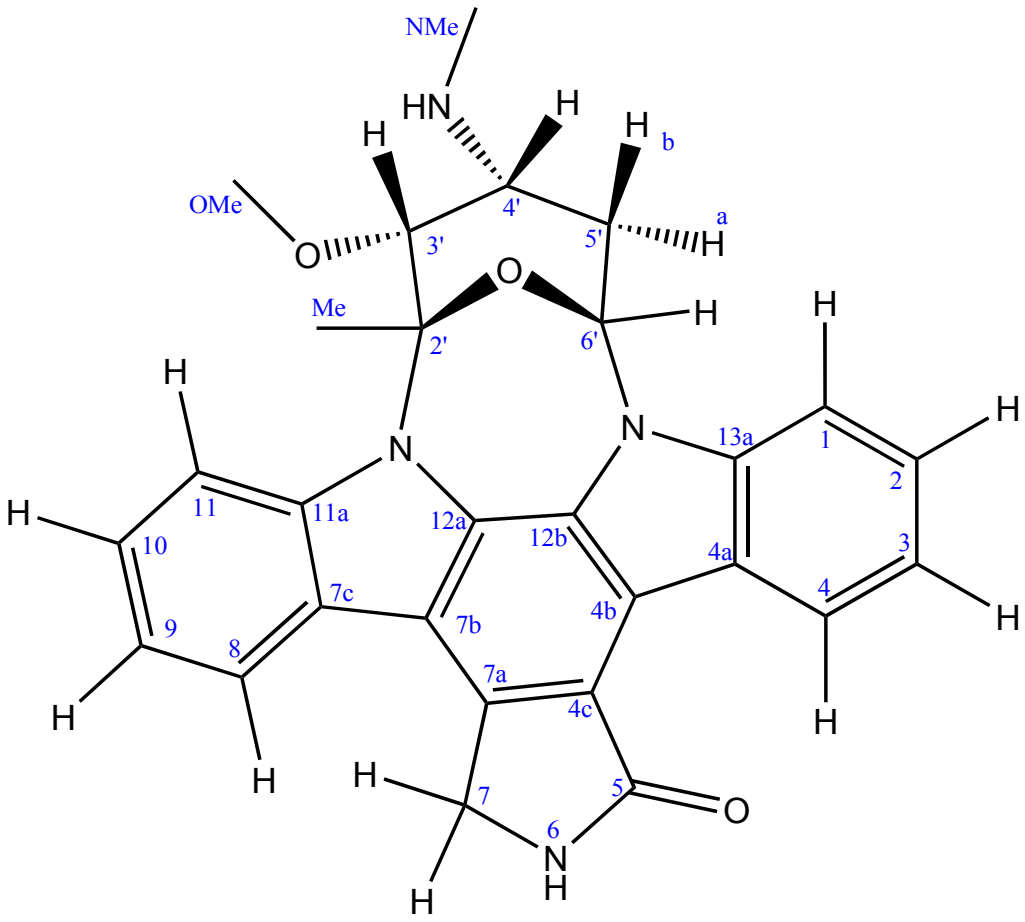


Figure 2.8: Presentation of the configuration of staurosporine. The natural product contains four stereocenters; due to the cyclic nature of the tetrahydropyran ring, the degree of configurational freedom is reduced to three independent stereogenic centers. The naturally found stereochemistry is the "SRRR" configuration; the lowest energy structure of the six-membered ring is the chair conformation.

2.2.3 Measurements of Staurosporine in Deuterated Polystyrene

As already addressed in the introductory section of this chapter, the configuration and the preferred conformation of staurosporine could only be verified via residual dipolar couplings if both a sufficient number and precise NMR data are available.[54] However, common alignment media generally introduce undesired NMR signals which might severely overlap with signals of the molecule of interest, and become especially annoying in polymer gels at low alignment strengths. Thus, a new, gel-based alignment medium was tested with *per se* full scalability of alignment strength and practically no additional interfering NMR signals: deuterated polystyrene in chloroform (dPS/$CHCl_3$).

After synthesizing and swelling cylindrically shaped, cross linked dPS sticks in $CDCl_3$ as well as addition of staurosporine, the corresponding $^1H,^{13}C$-CLIP-HSQC spectrum shows no overlap of the solute signals with the very small residual polymer signals.[60] The comparison of spectra recorded with and without alignment medium allows the extraction of thirteen one-bond heteronuclear RDCs which are (together with corresponding error-estimations) tabulated in figure 2.9.

Number	Group	$^1J_{CH}$ [Hz]	$^1(J+D)_{CH}$ [Hz]	$^1D_{CH}$ [Hz]
01	1	158.0 ± 1.5	130.8 ± 2.0	-27.2 ± 2.5
02	2	158.1 ± 3.0	147.0 ± 7.0	-11.1 ± 7.6
03	3	160.6 ± 3.0	162.3 ± 3.0	1.7 ± 4.2
04	4	165.0 ± 0.3	137.9 ± 1.0	-27.1 ± 1.0
05	11	162.5 ± 0.5	137.1 ± 1.0	-25.4 ± 1.1
06	10	159.6 ± 2.0	162.2 ± 3.0	2.6 ± 3.6
07	9	160.9 ± 2.0	148.0 ± 5.0	-12.9 ± 5.4
08	8	158.1 ± 0.3	132.8 ± 1.0	-25.3 ± 1.0
09	6'	159.3 ± 0.5	165.1 ± 1.0	5.8 ± 1.1
10	5'a	131.7 ± 1.0	139.2 ± 3.0	7.5 ± 3.2
11	4'	135.2 ± 0.3	131.5 ± 1.0	-3.7 ± 1.0
12	3'	140.3 ± 0.5	120.5 ± 5.0	-19.8 ± 5.0
13	Me	129.2 ± 0.2	126.0 ± 1.0	-3.2 ± 1.0

Figure 2.9: Table of the thirteen measured scalar couplings and RDCs for staurosporine. The nomenclature refers to the assignment which is given in figure 2.8. The obtained values were extracted from $^1H,^{13}C$-CLIP HSQC spectra with and without alignment in a deuterated polystyrene/$CDCl_3$ gel.

2.2.4 Singular Value Decomposition

With all of the RDCs in hand, a static configurational and conformational analysis of staurosporine was possible: the experimentally obtained thirteen residual dipolar couplings were fitted against the eight potential structures using the program PALES with the so-called “bestFit” option.[56] The resulting squares of the corresponding correlation factors (R^2) which are a measure of the quality of the fit as well as the alignment tensor for staurosporine in deuterated polystyrene / chloroform are illustrated in figure 2.10 and 2.11. It becomes obvious that the best correlation (R^2 = 0.996) is obtained for the staurosporine SRRR configuration with the tetrahydropyran ring in the chair conformation; this result supports the structure reported for the free base of the natural product.[61,62]

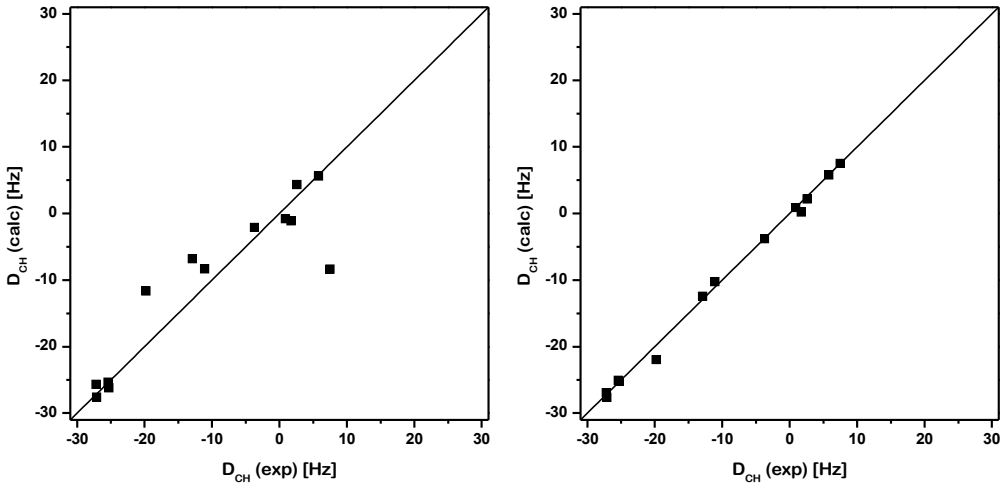


Figure 2.10: Comparison of experimentally determined RDCs of staurosporine and RDCs back calculated with the program PALES for the SRRR boat (left) and chair (right) arrangements. Based on all deviations between the measured RDCs and calculated ones, the correlation factor R can be computed. R^2, the square of the correlation factor, can have values between 1 (full consensus) and 0 (no correlation).

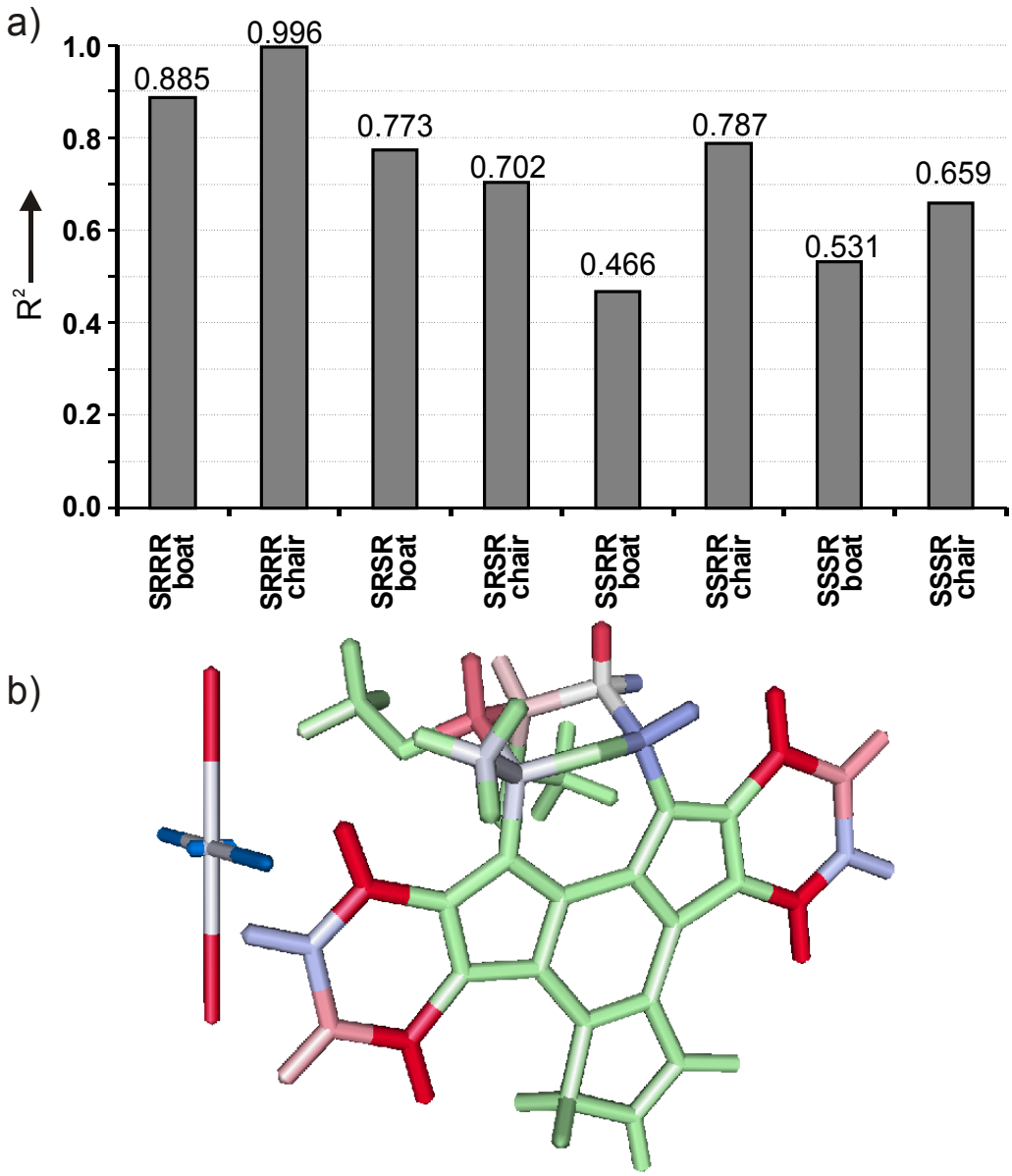


Figure 2.11: Comparison of experimental and back calculated RDCs of the SRRR boat and chair structures.

a) The square correlation factors R^2 for all potential arrangements of staurosporine.

b) The PALES fit derived R^2 factors clearly support the SRRR chair structure. The 3D structure of staurosporine is illustrated with color-coded atomic bonds representing negative (red) and positive (blue) $^1H,^{13}C$-RDCs as well as the axes of the corresponding alignment tensor drawn next to it. The $^1D_{CH}$ coupling of the methyl group (number 13 in figure 2.9) was utilized as converted $^1D_{CC}$ RDC in the PALES fittings.

2.2.5 Molecular Dynamics Simulations

By means of the "static" singular value decomposition approach, both the configuration and the preferred conformation of staurosporine could be verified. In order to test if a method that regards intrinsic molecular flexibility provides equal or even more convincing results than the PALES-SVD, residual dipolar couplings driven MD simulations were applied for confirming the configuration and the conformation of staurosporine (thereby following the method published by Hess and Scheek which is implemented in the GROMACS 3.3.2 simulation package).[57,58] Since the main interest of this work was to demonstrate the potential of RDC based MD calculations, the simulations focused only on the SRRR and SSSR configurations; in addition, the possible chair and boat conformations were also considered, resulting in four staurosporine structures to be tested. For the 5 ns long MD production runs, twelve of the 13 measured $^{1}D_{CH}$ couplings were taken into account (the RDC value of the methyl group "Me" was not included as the rotation of the three CH vectors would not converge to its average state in short trajectories). Furthermore, three different MD setups were tested: in run 1, staurosporine was not coupled to the experimental values but is only described by the parameters of the OPLS-AA force field. In the setups of simulations 2 and 3, the RDC coupling force constants were at each time 0.1 kJ / mol; in run 2, the orientation restraints were constantly coupled to the natural product whereas in run 3 time-averaging was applied with a coupling constant of 100 $(ps)^{-1}$.

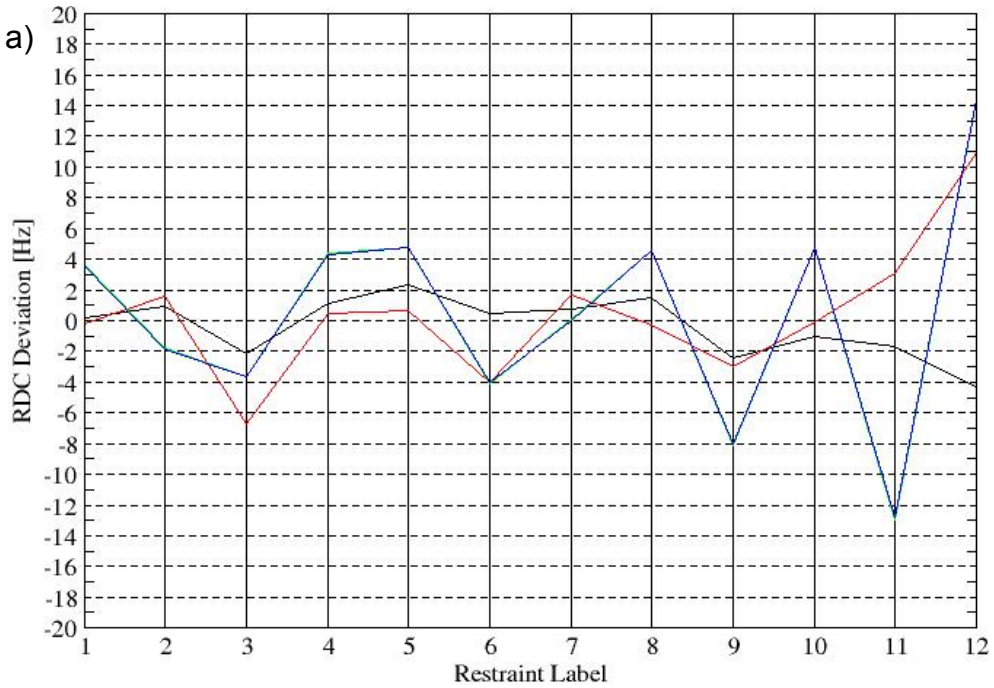
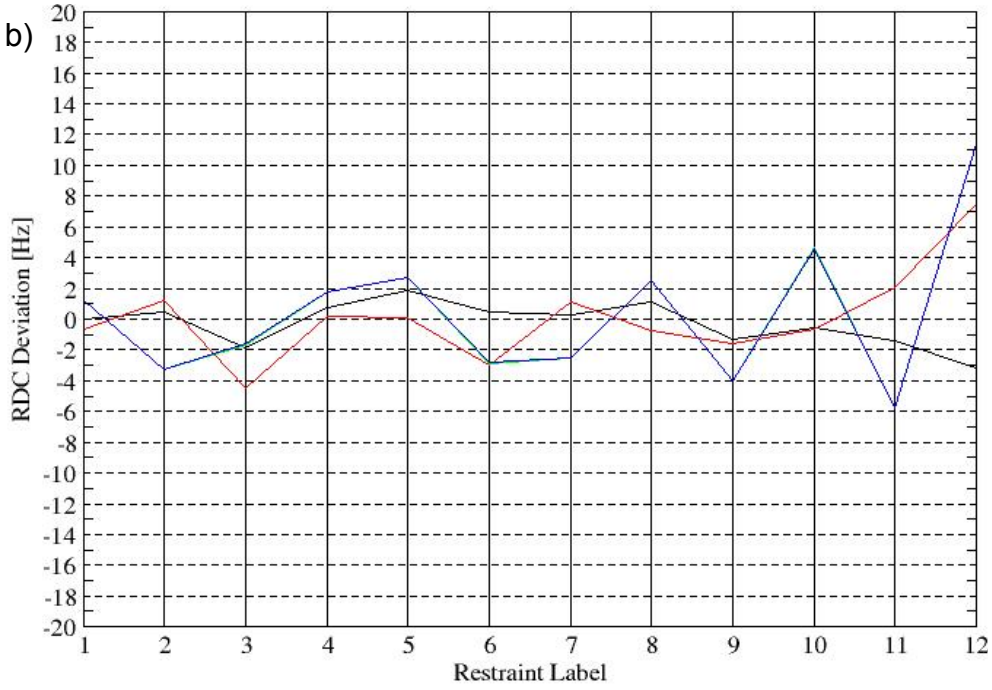
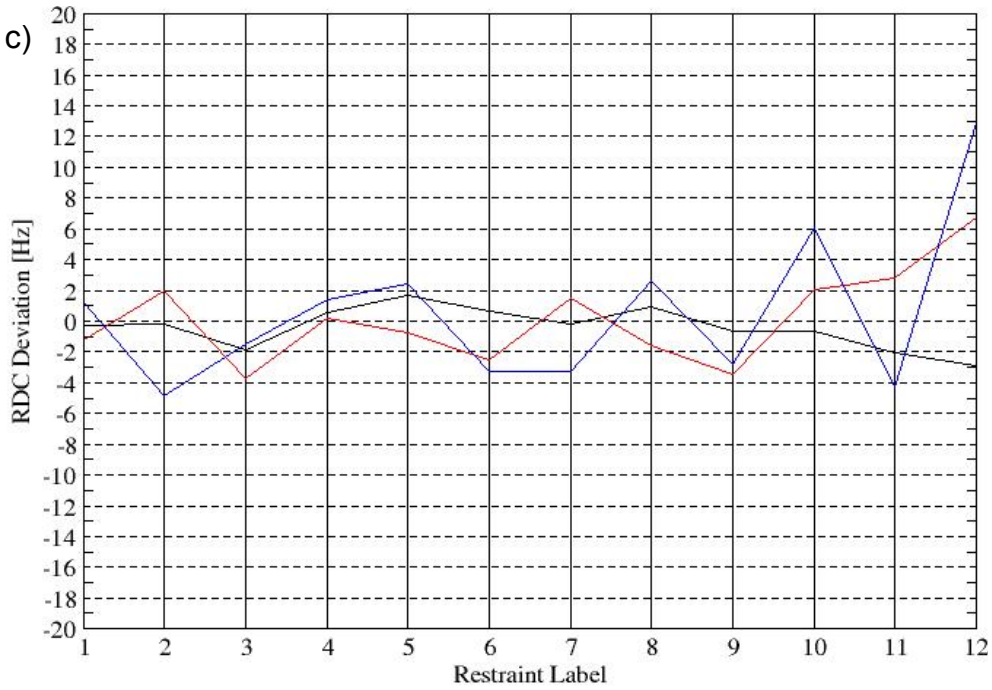


Figure 2.12: A graphical illustration of the results of a free and two RDC driven MD simulations. The three runs were performed with the following conditions: a) no coupling to recorded RDCs; b) with a constant restraint force of 0.1 kJ / mol; c) with a restraining force of 0.1 kJ / mol which is time averaged (1/100 ps). The labels correspond to the numbering shown in figure 2.9. The color code is as follows: black: SRRR chair, red: SRRR boat, blue: SSSR chair, green: SSSR boat (the SSSR boat and chair structures result in very similar deviations and are therefore hardly distinguishable in the diagram).

Continuation of Fig. 2.12: The RDC restraints 1-8 are located in the flat and rigid aromatic ring system of staurosporine. Thus, the MD deviations are small for all structures and all setups. Distinct deviations are visible for the restraints 9-12 which represent the vector orientation of the more flexible tetrahydro-pyran ring. In total, the smallest deviations from experimental RDC values are present in the SRRR chair structure (black line).





The deviations of the calculated RDCs from the measured ones are illustrated in figure 2.12 (with reference to the three tested simulation setups). The average variances as calculated for the four different configurations and conformations of staurosporine are tabularly summarized in figure 2.13; furthermore, the resulting restraint potential energies were computed and are also given in the table shown below.

RUN 1	SRRR chair	SRRR boat	SSSR chair	SSSR boat
Potential Restraint Energy (kJ / mol)	-----	-----	-----	-----
Restraint RMS Deviation (Hz)	1.944	4.303	7.193	7.151

RUN 2	SRRR chair	SRRR boat	SSSR chair	SSSR boat
Potential Restraint Energy (kJ / mol)	2.887	8.394	15.954	15.792
Restraint RMS Deviation (Hz)	1.429	2.960	4.734	4.717

RUN 3	SRRR chair	SRRR boat	SSSR chair	SSSR boat
Potential Restraint Energy (kJ / mol)	1.098	4.966	14.612	14.638
Restraint RMS Deviation (Hz)	1.345	2.998	5.121	5.125

Figure 2.13: Table of MD derived potential restraint energies as well as RMS deviations between the recorded and the calculated RDCs. In all MD runs, the SRRR configuration in the chair conformation led to the lowest potential restraint energies and the smallest RDC deviations. As run 1 was not coupled to the measured values, no potential restraint energies were obtained. In run 2, the restraint force was 0.1 kJ / mol and no time averaging was used. Run 3 was performed with a restraint force constant of 0.1 kJ / mol and included a time average coupling constant of 100 $(ps)^{-1}$.

By comparing the MD results of the four investigated arrangements of staurosporine, it becomes obvious that in all simulations the SRRR boat arrangement of staurosporine results in the best consensus. Compared to the other studied structures of the natural product, the restraint potential energies are the lowest ones; moreover, only the SRRR configuration with its six-membered ring in chair conformation reveals RDC deviations that are lower than 2 Hz in the case of a free (only force field based) simulation and lower than 1.5 Hz for RDC coupled runs. If the 3' CH-coupling which shows the strongest RDC violations in all three trajectories and has a relatively large experimental error of ± 5.0 Hz is neglected, the RMS deviation would be below 1.0 Hz; this demonstrates that MD simulations are as well suited as the PALES fitting procedure for elucidating and verifying 3D structures by help of accurate residual dipolar couplings. The comparison of the three MD setups - free dynamics and RDC coupled MD runs with and without time averaging - further reveals that the OPLS-AA force field in combination with a careful topology parametrization excellently reproduces the right configuration and conformation of staurosporine. The trajectories which are coupled to the measured RDCs result in lower RMS deviations with respect to the SRRR boat structure of the natural product compared to the free MD runs. However, due to the external forces which compel staurosporine to adopt a structure which is in agreement with the measured NMR data, the deviations between

recorded and computed RDCs for the three wrong structures are also lowered. On one side, a distinction between right and false structures is aggravated if only the average RMS deviations are regarded. On the other side, if no structure information is available *a priori*, an arbitrary starting structure can be enforced to adopt the desired natural arrangement by applying RDCs together with strong coupling forces (given that no high barriers separating local energy minima are present on the hyper surface). Finally, a significant difference between constant and time-averaged coupling runs could not be obtained for the relatively rigid staurosporine simulated in vacuum.

2.2.6 Conclusion

In summary, it could be demonstrated that RDC driven MD simulations are capable to verify known molecular arrangements, and thus should be able to contribute to the elucidation of unknown molecule structures. For this to show, the new developed alignment medium deuterated polystyrene that allowed the measurement of a sufficient number of precise residual dipolar couplings was utilized.[63] In general, two simulation procedures were tested: free MD calculations in combination with the application of a high-quality force field could produce structures which could be evaluated by means of measured RDCs. Alternatively, the MD runs can be directly coupled to the recorded RDCs; in the process the simulated molecule is enforced to adopt a structure which corresponds to the given RDCs. RMS deviations between provided and computed RDCs as well as the potential restraint energies of RDC restraints can be used as a quality criterion of the calculated structure. For (relatively) rigid molecules, MD simulations provide results of comparable quality as the PALES-SVD approach. For more flexible molecular arrangements, MD calculations might be even superior if the simulation setup is carefully adjusted.

2.2.7 Reagents, Methods, and Experiments

Informations about the new alignment medium and the NMR measurement parameters can be found in literature.[63]

Molecular dynamics calculations were carried out with the GROMACS 3.3.2 simulation package.[57] The OPLS-AA force field was used for parametrization of staurosporine.[64] Partial charges, molecular geometries and bonded and non-bonded interaction parameters were adopted from molecules with similar

chemical motifs compared to staurosporine. Starting coordinates of staurosporine in the different configurations and conformations were generated using the program SYBYL (Tripos). Each structure was placed in cubic periodic boxes with vector lengths of 5 nm. All molecules were first energy minimized using the Steepest Descent algorithm. The production runs were performed under vacuum conditions. A stochastic bath was applied with a reference temperature of 300 K using an atomic friction coefficient of 70 $(ps)^{-1}$. All bond lengths were constrained using the LINCS algorithm with a highest order parameter of 4.[65] Newton's equations of motions were integrated using a stochastic leap-frog algorithm and a time step of 0.002 ps.[65] Non-bonded interactions were calculated using a twin-range cut-off scheme. Interactions within a short-range cutoff of 0.9 nm were computed every time step from a pair-list which was updated every five steps. Interactions between 0.9 nm and 2.0 nm were only calculated at these time points. Since the simulations were performed in vacuo, the long-range cutoff was prolonged to 2.0 nm and long-range correction terms to electrostatic and van der Waals interactions were neglected. Non-bonded interactions between all atoms in the flat aromatic system of staurosporine were excluded. All simulations were performed for 6 ns; the first nanosecond was used for equilibration of the molecule. Energies, forces and coordinates were written out each picosecond. Twelve of the 13 measured one-bond ^{1}H-^{13}C RDCs were applied to staurosporine as orientation restraints. The aromatic ring system of the molecule served as fitting group for the determination of the internal orientation of flexible groups. Three simulations were performed where the force constants of the restraints was chosen to be 0 kJ / mol (no coupling) or 0.1 kJ / mol (weak coupling) whereas both constant (t = 0 ps) and time-averaged coupling (t = 100 ps) was utilized.

2.2.8 Declaration

The presented topic was conducted in collaboration with M.Sc. Grit Kummerlöwe and PD Dr. Burkhard Luy (Technical University Munich, Department Chemistry). G.K. contributed to the development of dPS and measured all residual dipolar couplings. B.L. performed the PALES calculations. The described results have been published as follows:

G. Kummerlöwe, S. Knör, A.O. Frank, T. Paululat, H. Kessler, B. Luy. Deuterated Polymer Gels for Measuring Anisotropic NMR Parameters with Strongly Reduced Artefacts. *Chem. Commun.* **2008**, *44*, 5722-5724.

2.3 Verifying the Constitution of Functionalized Biomolecules

In the last two chapters, a structural analysis of the conformation and / or configuration of peptide derivatives and a medium-sized natural product was presented. A necessary prerequisite of such studies is the knowledge of the constitution of the molecules under investigation. The utilization of electrospray ionization mass spectrometry (ESI-MS) reveals if a synthesized small molecule has the expected weight and therefore provides hints about the molecular constitution. However, the concrete connection of atoms or atom groups remains unknown. When dealing with regular peptides or common derivatives thereof which are in almost all cases synthesized by means of commercially available amino acids, a sequential NMR assignment - proving the covalent connection of the used building blocks - is usually sufficient to verify their constitution. In contrast, the composition and the arrangement of molecules that include atypical chemical groups (in a biological sense) or that are arranged in an unusual (not biogenic) way with respect to the connectivity of atoms, must exhaustively be elucidated via NMR spectroscopy or X-ray crystallography. If only the constitution and no conformational aspects of a molecule are of interest, NMR spectroscopy has distinct advantages over the latter mentioned method. For example, the often time-consuming search for adequate crystallization conditions of a molecule is not necessary; additionally, in opposition to conformational analysis, the type of solvent which results in the easiest interpretable NMR spectra can be freely chosen.[66,67]

2.3.1 The Scientific Question

Constitutional analysis is a standard procedure for describing new-found natural products. Furthermore, verifying the structure of designed molecules is also an essential prerequisite before their development can be published. If a molecule comprises a sufficient number of NMR active nuclei (e.g. labeled biomolecules or organic compounds available in high concentrations), solution-state NMR spectroscopy is a well suited method for structural analysis.[54]

In this scope, the goal of the following two projects was the NMR based elucidation of the constitutions of three compounds which may find applications as biosensors or biologically active agents.

The focus of both topics was on the determination of the constitution of two ferrocene containing and one ruthenium comprising amino acid derivative. Since the three compounds contain metal ions and non-biogenic atom groups, constitutional assignments were challenging. Due to the sufficient amounts of available substances, not only homonuclear but also heteronuclear spectra like HSQC or HMBC could be recorded and finally fully assigned.

2.3.2 Affinity Markers and the Concept of Biosensors

A relatively young method for characterizing the function of enzymes is the application of so-called affinity markers.[68] In principle, an affinity marker usually consists of a recognition sequence which can form a reversible complex with the enzyme of interest, and a reactive head group which is positioned in such a way that a irreversible, covalent bond is built with the enzyme during the catalytic transformation. In addition, the recognition element is mostly attached to a linker sequence which allows e.g. the immobilization of the affinity marker (see Fig. 2.14).

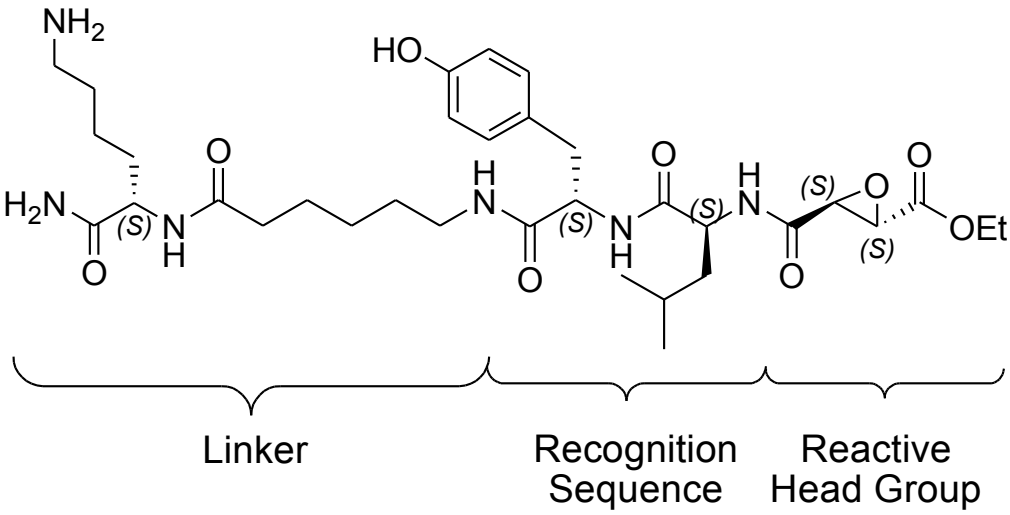


Figure 2.14: Presentation of a typical affinity marker for papain-like proteases. The molecule consists of a recognition sequence that is necessary for receptor binding, of a reactive head group that can build a covalent bond to the receptor, and usually of a long linker sequence.

Affinity markers are often one basic component of biosensors which are able to transform chemical informations (e.g. substance concentrations, sample composition) into an analytically detectable signal.[68] Whereas the affinity marker is responsible for the chemical recognition and binding, a physico-chemical transducer (the second part of a biosensor) converts the effect of an enzymatic reaction into e.g. a thermal, an electrochemical, or an optical signal that could be finally detected.

2.3.3 Verifying the Structures of Ferrocene Containing Affinity Markers

In order to create a new electrochemical biosensor, a known construct developed for studying the activity of cysteine-proteases was used as template structure (see Fig. 2.15).[69,70] It was decided to incorporate the inorganic ferrocene as redox active probe into the affinity marker scaffold since it displays several characteristics making it well suited for electrochemistry.[71-79]

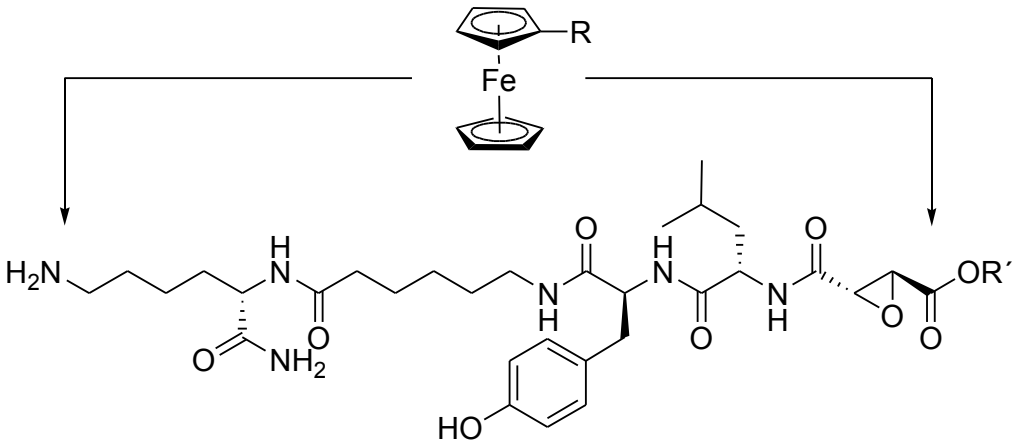


Figure 2.15: Illustration of a template affinity marker for cysteine-proteases and the two possibilities for attaching the redox active ferrocene to the scaffold.

After distinct modifications of the basic structure which were necessary to enhance e.g. the water solubility, two biosensor molecules were designed which further on had to be constitutionally analyzed; the expected structures of both compounds are illustrated in figure 2.16.

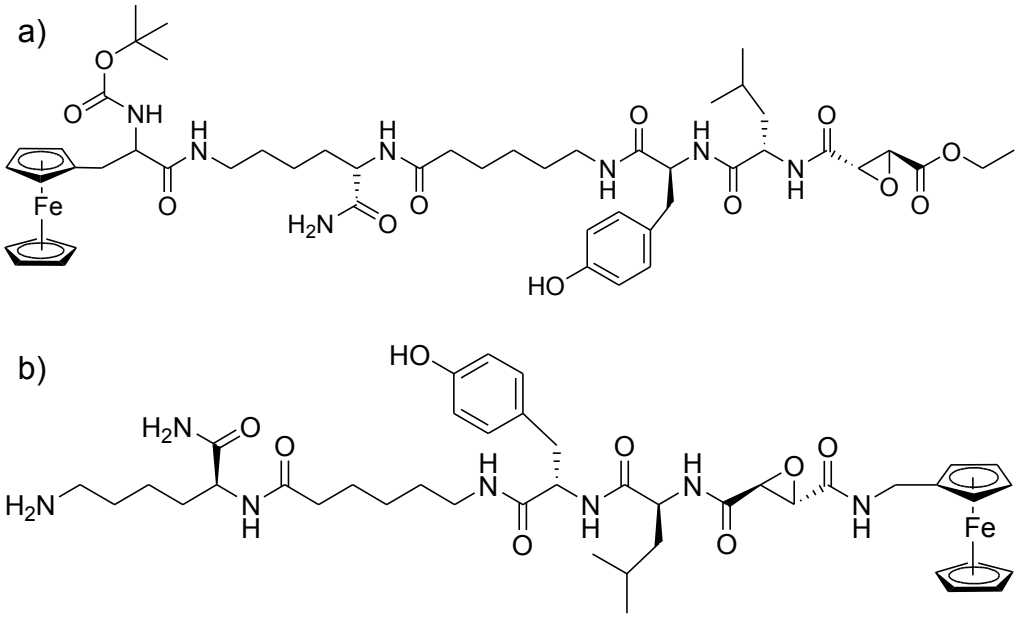


Figure 2.16: Presentation of the proposed structures of two molecules that can be utilized as biosensors. a) molecule 1 with the redox active ferrocene group covalently attached to the linker sequence; b) compound 2 with the ferrocene unit linked to the reactive head group.

For the elucidation of the constitutions, both molecules were dissolved in d_6-DMSO and multiple NMR spectra were recorded: ^{1}H-1D, ^{13}C-1D, ^{1}H,^{1}H-COSY, ^{1}H,^{1}H-TOCSY, ^{1}H,^{1}H-ROESY, ^{1}H,^{13}C-edHSQC, ^{1}H,^{15}N-edHSQC, and ^{1}H,^{13}C-HMBC. The ^{13}C-1D was used to check if all expected carbon resonances were present. Homonuclear J-couplings and signal intensities could be extracted from proton 1D and COSY spectra, respectively. Individual spin systems were verified based on their TOCSY pattern; distinct groups of the spin systems (especially the plethora of CH_2 groups) were assigned by the help of COSY and edited HSQC spectra. The “sequential” assignment (sequence blocks which are separated by an amide bond) was not performed via the popular TOCSY / ROESY approach but with the aid of heteronuclear long-range couplings provided by the HMBC spectrum.[80] ROESY data in combination with the TOCSY spectrum were used to evaluate the assignment.

First, the molecule shown in figure 2.16a was constitutionally investigated. All NMR active nuclei (^{1}H, ^{13}C, ^{15}N) with exception of degenerated proton and carbon signals were found in the spectra and could be assigned to the primary sequence. Homonuclear scalar ^{3}J couplings of amino acid amide protons (doublet splitting due to Hα protons) have values in the range between 7.8 and 8.5 Hz and are consistent with the “intraresidual“ assignment. The proton resonances of ferrocene exhibit the typical intensity pattern of 2:2:5. The chemical shifts of the cyclopentadiene rings (4.04-4.16 ppm) partially overlay with the Hα protons of the used amino acids (amino acid derivatives) which - as expected - arise in the spectra between 4.15 and 4.35 ppm. One of the amide protons (located in the main chain) close to the iron containing ferrocene group displays a regular shift of 7.72 ppm whereas the amide proton neighboring the Boc group is upfield shifted (6.67 ppm). All protons of the amino hexane acid are detected between 1.20 and 2.79 ppm; the two protons of the epoxide group arise at 3.72 and 3.60 ppm. The resonances of all carboxy C-atoms are in the range between 165 and 174 ppm. Carbon atoms of the ferrocene were assigned as follows: 84.9 ppm (C_{ipso}), 68.8 and 67.5 ppm (substituted cycle), and 69.5 ppm (not substituted ring). With exception of the Boc nitrogen which arises at 90.7 ppm and the free amide group arising at 104.8 ppm (negative signal due to the two attached protons in the edited HSQC spectrum; see overlay in figure 2.16), all ^{15}N are in the typical amide nitrogen region between 115.8 and 123.9 ppm. All further resonances were detected as expected. With the complete assignment of all atoms in hand, the connection of the individual spin systems could be determined by means of

^{1}H-^{13}C coupled resonances (heteronuclear ^{3}J couplings between amide protons and carboxy C-atoms) as arising in the HMBC spectrum. Therefore, it was possible to verify the constitution of molecule 1 (as shown in Fig. 2.15a).

For molecule 2 (see Fig. 2.16b), all atoms which have the same or a related chemical (and therefore magnetic) environment as in molecule 1 exhibit similar chemical shifts (shift ranges) and J-couplings. A first exception concerns the amide proton close to the ferrocene group which is - compared to its analog in molecule 1 - downfield shifted (8.57 ppm). In addition, no high field shifted nitrogen peak is visible in the $^{1}H,^{15}N$-HSQC spectrum due to the absence of the neighboring Boc group. Finally, no ^{15}N chemical shift information of the free amino group could be gained due to chemical exchange of attached protons with residual solvent water (see overlay in figure 2.17). Again, the connection of spin systems was proved via informations extracted from the $^{1}H,^{13}C$-HMBC spectrum. The full NMR assignment of both described molecules is given in the appendix of this work.

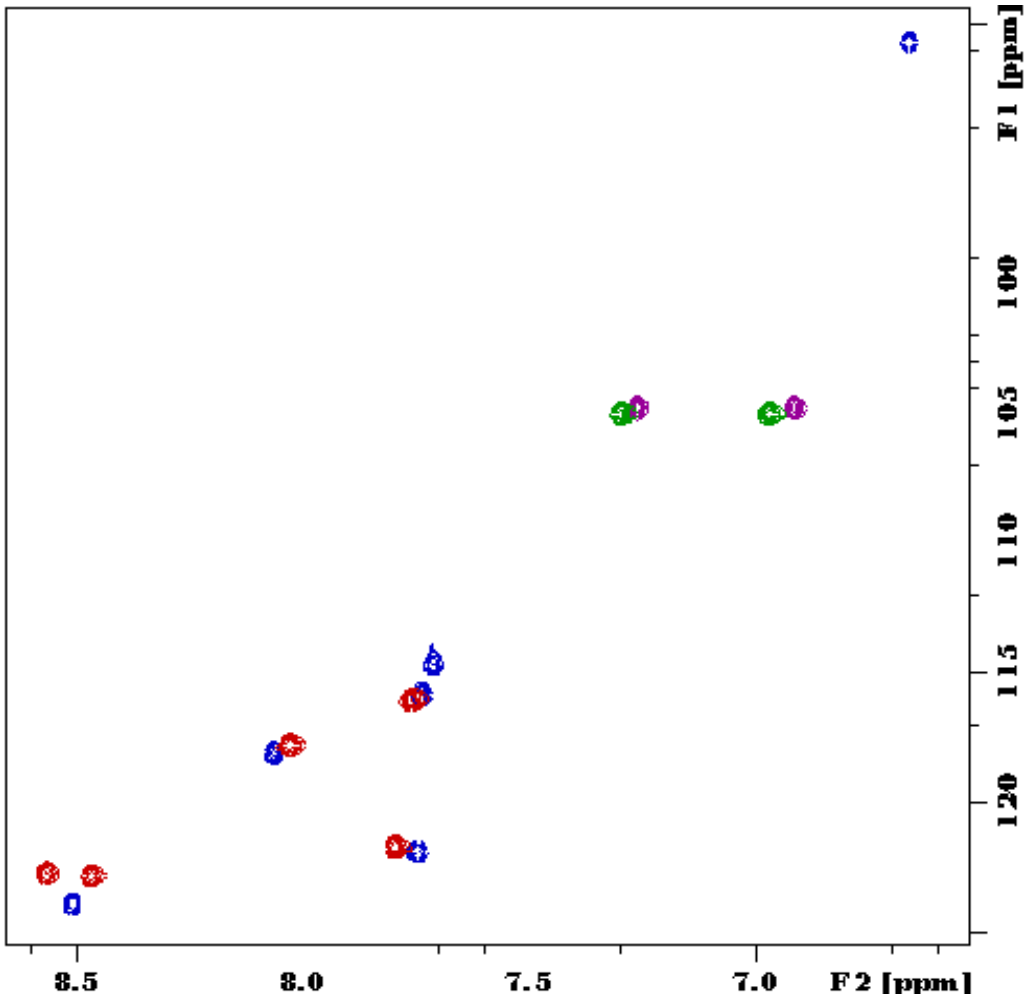


Figure 2.17: Overlay of the ^{15}N-HSQC spectra of molecules 1 and 2: Color code: blue: positive peaks (only one proton is bound to a nitrogen) for molecule 1; purple: negative peaks (two protons are attached to a nitrogen) for molecule 1; red: positive resonances for compound 2; green: negative resonances for compound 2. In F2, the ^{1}H NMR dimension is given; F1 is the ^{15}N dimension. A description of the spectrum and the resonances is given throughout the main text.

2.3.4 The Role of Sandwich and Half-Sandwich Complexes

The well established stability of *d6*-sandwich complexes towards oxygen, water, or even under physiological conditions lead to a wide spread interest into the application of such systems as biological probes and pharmacologically active compounds.[81-103] The transition to half-sandwich complexes adds an further dimension to the chemistry of these systems due to the introduction of potentially labile coordination sites. Specifically, half sandwich arene ruthenium(II) complexes conquered a plethora of applications. In particular, the *in vivo* and *in vitro* cytotoxicity of this class of compounds has recently triggered intense research activity targeting the design of selective organometallic anti cancer agents (see e.g. Fig. 2.18a).[104-112] They also have attracted considerable attention due to a wide range of applications including building blocks for supramolecular structures,[113] catalysts for transfer hydrogenations,[114-120] hydrogenations,[121] or C-C couplings.[122,123] Finally, half sandwich ruthenium complexes were successfully applied as catalytically active metal centers in organometallic enzyme hybrids.[124,125]

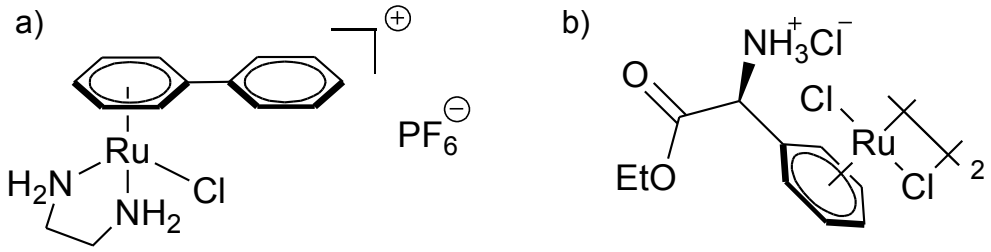


Figure 2.18: Presentation of published half-sandwich complexes which comprise ruthenium metal centers. a) Sadler agent and b) Beck compound. Both of the η^6-arene Ru(II) complexes have already successfully been applied in selective enzyme inhibition.

Ruthenium η^6-arene complexes with pendant carboxylate groups, alcohols, or even amino acids have successfully been designed and found important applications (see e.g. Fig. 2.18b).[93,126-137] Surprisingly, half sandwich ruthenium complexes of biogenic amino acids have yet not been reported. There is only one example where the synthesis non-biogenic amino acids comprising an η^6-arene ruthenium metal center was reported,[137] and descriptions of η^6-ruthenium complexes with biogenic amino acids are rare.[126-133] In addition, none of the published motifs exhibits an open coordination site which is required for a metal centered catalytic activity.

Therefore, it is of great interest to synthesize and characterize complexes that combine the chemical properties of biogenic amino acids and the versatile features of catalytically active metal centers.

2.3.5 Constitutional Analysis of an Amino Acid Based Ruthenium Complex

During the development of a biologically active half-sandwich compound based on a ruthenium complexed phenylalanine, four important intermediate products had to be analyzed by NMR spectroscopy as no crystal structure were present. At this point, the structural analysis of only one of the four complexes is presented in detail since all molecules exhibit a similar constitution. The compound of interest comprises the amino acid which is twofold complexed to the metal ion; a tosylated ethylenediamine group serves as a second ruthenium ligand (see Fig. 2.19).

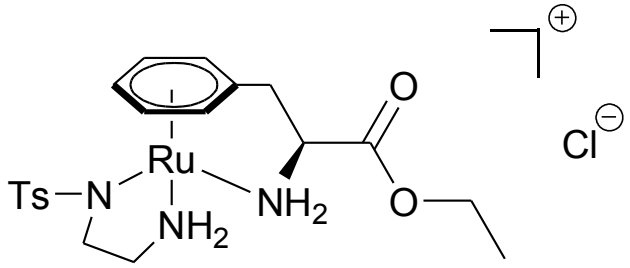


Figure 2.19: Illustration of the proposed structure of a new ruthenium consisting half-sandwich complex (Ts stands for tosylate). The molecule is built of the amino acid phenylalanine and ethylenediamine that are always twofold co-ordinated to the metal ion.

The constitutional NMR study of the Ru-complex shown above came along with two challenges. First, as the complexes have different, not interconnected spin systems, no complete assignment of both the (S)-Ru and (R)-Ru complex is possible; thus, an elucidation of the molecular constitution - in particular, the determination of the coordination of the phenylalanine ester and the tosylated ethylenediamine to the metal center - could only indirectly be obtained. Further on, the bioinorganic complex contains two chiral centers (α-CH carbon and the ruthenium center), resulting in diastereomeric molecules. Hence, two signal sets were expected for the system to occur in the NMR spectra.

For the structural analysis, the molecule was dissolved in d_6-DMSO and multiple NMR spectra were recorded: ^{1}H-1D, ^{13}C-1D, ^{1}H,^{1}H-COSY, ^{1}H,^{1}H-TOCSY, ^{1}H,^{13}C-edHSQC, ^{1}H,^{15}N-HSQC, and ^{1}H,^{13}C-HMBC. The proton 1D spectrum reveals that the diastereomers of the complex were distributed in a ratio of 3:2; one of the diastereomers might be disfavored since the bulky ethyl ester group disfavors one of the conformations. By means of the 2D spectra, all nuclei of the single spin systems (the ethyl ester group, phenylalanine, ethylenediamine, and the tosyl group) could be assigned. The ^{1}H,^{13}C coupled spectra reveal that all atoms of the phenylalanine aromatic ring system display

differing chemical shifts (e.g. ^{1}H: 5.08, 5.45, 5.73, 5.95, and 6.57 ppm for one of the diastereomers) which are - compared to the free amino acid - strongly upfield shifted (e.g. ^{13}C: 105.4-70.1 ppm for one of the diastereomers). In addition, the $^{1}H,^{15}N$-HSQC spectrum also shows drastically upfield shifted nitrogen and proton resonances of the phenylalanine and of one of the ethylenediamine amino groups (with respect to not-complexed molecules; see Fig. 2.20 as an example of the arising "atypical" chemical shifts of the two amino groups).

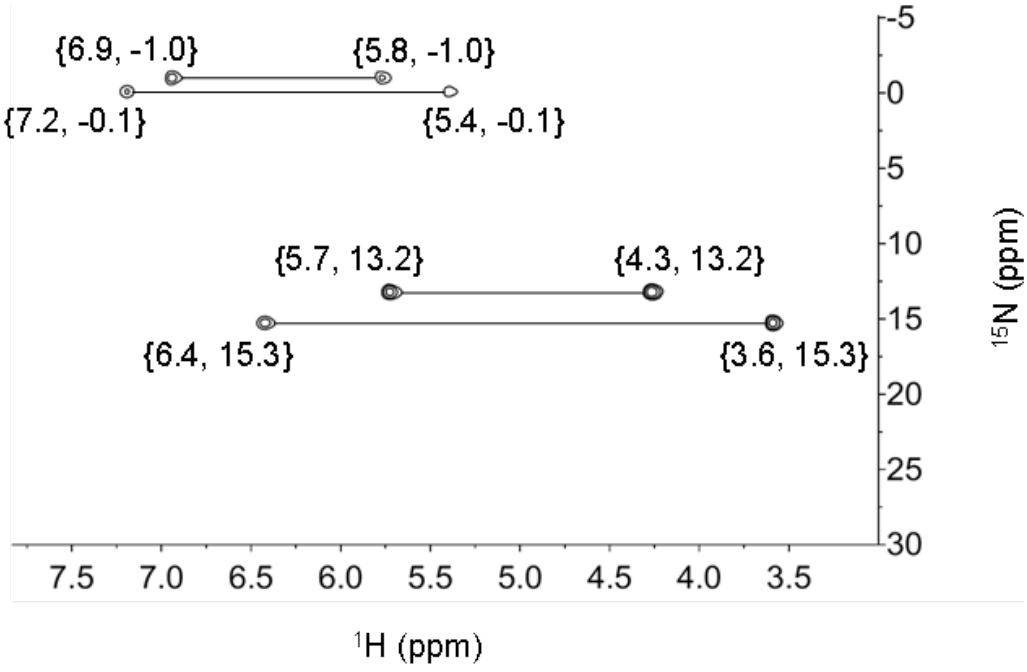


Figure 2.20: Section of the $^{1}H,^{15}N$-HSQC spectrum of the half-sandwich complex to be elucidated. Due to the presence of two diastereomers, four different NH_2 groups were detected. The Ru(II) ion is responsible for the dramatic upfield shift of the resonances in the nitrogen dimension. Depending on the spatial orientation of the two protons bound to nitrogen atoms with respect to the metal center, each time one of the ^{1}H nuclei is strongly upfield shifted in the ^{1}H dimension. Further details about the analysis of the $^{1}H,^{15}N$-HSQC spectrum are given in the main text.

Thus, it becomes obvious that both phenylalanine and ethylenediamine form a complex with the ruthenium ion, and that this coordination results in a strong structural tense of the complex - otherwise, the aromatic carbon and proton nuclei would exhibit the same chemical shifts for each of the two atoms in the ortho and para positions. Further on, four inequivalent nitrogen resonances are detectable in the $^{1}H,^{15}N$ coupled spectrum; two peaks arise from the "S_{ru} / $S_{\alpha\text{-}CH}$", the other two from the "R_{ru} / $S_{\alpha\text{-}CH}$" diastereomers. Each ^{15}N nucleus is coupled to two inequivalent protons which are shifted by values of up to 2.8 ppm in the proton dimension (see again Fig. 2.20); this finding could be attributed to different orientations of the amino group atoms with respect to the localization of the ruthenium ion. In total, all atoms with exception of the not-proton bonded nitrogen atom could unambiguously be assigned. The complete list of chemical shifts is given in the appendix of this work.

2.3.6 Conclusion

By the help of NMR spectroscopy, the constitution of three functionalized bioinorganic compounds which can serve as biosensors and biologically active agents, respectively, could be elucidated. In all three cases, the proposed (expected) structures were unequivocally verified. The constitutional analysis of both ferrocene containing molecules could be obtained by an assignment strategy using the homonuclear TOCSY and COSY and the heteronuclear edHSQC and HMBC spectra. Due to the presence of not-interconnected spin systems, the constitution of the Ru(II) containing half-sandwich complex could only indirectly be verified by excluding other structure solutions. With the aid of TOCSY and HMBC spectra (together with COSY and HSQC spectra), the nuclei localized in the different spin systems were assigned; the characteristic chemical shift data of nuclei that are coordinated to or in the neighborhood of the metal ion finally allow the elucidation of the molecular constitution.

2.3.7 Reagents, Methods, and Experiments

All informations regarding the design of the described molecules, their synthesis, and the biological analysis can be found in literature.[138,139]

All spectra (^{1}H-1D, ^{13}C-1D, ^{1}H,^{1}H-TOCSY, ^{1}H,^{1}H-ROESY, ^{1}H,^{1}H-COSY, ^{1}H,^{13}C-edHSQC, ^{1}H,^{13}C-HMBC, ^{1}H,^{15}N-(ed)HSQC) were recorded on a Bruker 900 MHz or 600 MHz spectrometer, both equipped with a TXI cryo probe. The temperature was 300 K in all cases. Standard Bruker pulse programs were applied.[140-145] d_6-DMSO was used as solvent and for the lock signal. Spectral widths were 10 ppm in all ^{1}H dimensions, 140 ppm (^{1}H,^{13}C-edHSQC; center frequency: 70 ppm) and 200 ppm (^{1}H,^{13}C-HMBC; center frequency: 100 ppm) in ^{13}C dimensions, or 60 ppm (^{1}H,^{15}N-(ed)HSQC; center frequency: 110 ppm for the ferrocene containing molecules and 0 ppm for the ruthenium complex) in the ^{15}N dimension. HSQC spectra were recorded with heteronuclear composite-pulse decoupling. The ^{13}C-1D spectrum was recorded with standard proton decoupling. In all 2D spectra, 1k (HSQC spectra) or 4k (all other spectra) data points were recorded in the direct dimensions; in the indirect dimensions, 512 data points were recorded. Exponential, sine or square sine window functions were used for apodization of the spectra. The spectrometer frequencies (^{1}H: 900.13 MHz / 600.13 MHz, ^{13}C: 226.34 MHz / 150.90 MHz; ^{15}N: 91.21 MHz / 60.81 MHz) were used for internal calibration of all spectra.

2.3.8 Declaration

The presented topic was conducted in collaboration with M.Sc. Alice Schlichtiger and M.Sc. Thomas Reiner (Technical University Munich, Department Chemistry). A.S. and T.R. designed and synthesized all molecules, and prepared all NMR samples. The described results have been submitted for publication as follows:

A. Schlichtiger, C. Baier, A.O. Frank, U. Stimmig, J. Eppinger. Synthesis and application of water-soluble ferrocene affinity labels for cysteine proteases. *Dalton Trans.* **2009**.

T. Reiner, A.O. Frank, G. Raudaschl-Sieber, F. Kiefer, E. Herdtweck, J. Eppinger. Employing L-phenylalanine as flexible ligand for η^6- and η^6:η^1- coordination to ruthenium transition metal centers. *Dalton Trans.* **2009**.

2.4 References

[1] D. F. Mierke, M. Kurz, H. Kessler, *J. Am. Chem. Soc.* **1994**, *116*, 1042-1049.

[2] G. V. Nikiforovich, K. E. Kövér, W. Zhang, G. R. Marshall, *J. Am. Chem. Soc.* **2000**, *122*, 3262-3273.

[3] M. Heller, M. Sukopp, N. Tsomaia, M. John, D. F. Mierke et al., *J. Am. Chem. Soc.* **2006**, *128*, 13806-13814.

[4] X. Zhang, G. V. Nikiforovich, G. R. Marshall, *J. Med. Chem.* **2007**, *50*, 2921-2925.

[5] J. Chatterjee, D. Mierke, H. Kessler, *J. Am. Chem. Soc.* **2006**, *128*, 15164-15172.

[6] J. Chatterjee, O. Ovadia, G. Zahn, L. Marinelli, A. Hoffman et al., *J. Med. Chem.* **2007**, *50*, 5878-5881.

[7] J. Chatterjee, D. F. Mierke, H. Kessler, *Chem. Eur. J.* **2008**, *14*, 1508-1517.

[8] E. Benedetti, A. Bavoso, B. Di Blasio, V. Pavone, C. Pedone et al., *Biopolymers* **1983**, *22*, 305-317.

[9] K. A. Williams, C. M. Deber, *Biochemistry* **1991**, *30*, 8919-8923.

[10] M. W. Macarthur, J. M. Thornton, *J. Mol. Biol.* **1991**, *218*, 397-412.

[11] A. Yaron, F. Naider, *Crit. Rev. Biochem. Mol. Biol.* **1993**, *28*, 31-81.

[12] M. P. Williamson, *Biochem. J.* **1994**, *297*, 249-260.

[13] D. Pal, P. Chakrabarti, *J. Mol. Biol.* **1999**, *294*, 271-288.

[14] H. Reiersen, A. R. Rees, *Trends Biochem. Sci.* **2001**, *26*, 679-684.

[15] J. F. Brandts, H. R. Halvorson, M. Brennan, *Biochemistry* **1975**, *14*, 4953-4963.

[16] I. L. Karle, *J. Am. Chem. Soc.* **1979**, *101*, 181-184.

[17] D. K. Chalmers, G. R. Marshall, *J. Am. Chem. Soc.* **1995**, *117*, 5927-5937.

[18] Y. Takeuchi, G. R. Marshall, *J. Am. Chem. Soc.* **1998**, *120*, 5363-5372.

[19] G. D. Rose, L. M. Gierasch, J. A. Smith, *Adv. Protein Chem.* **1985**, *37*, 1-109.

[20] J. Chatterjee, C. Gilon, A. Hoffman, H. Kessler, *Acc. Chem. Res.* **2008**, *41*, 1331-1342.

[21] D. A. Torchia, F. A. Bovey, *Macromolecules* **1971**, *4*, 246-251.

[22] H. Kessler, *Angew. Chem. Int. Ed.* **1982**, *21*, 512-523.

[23] H. Kessler, A. Müller, *Liebigs Ann. Chem.* **1986**, *1986*, 1687-1704.

[24] C. Francart, J. Wieruszeski, A. Tartar, G. Lippens, *J. Am. Chem. Soc.* **1996**, *118*, 7019-7027.

[25] A. Tonelli, *Biopolymers* **1976**, *15*, 1615-1622.

[26] E. Fischer, W. Lipschitz, *Ber. Dtsch. Chem. Ges.* **1915**, *48*, 360-378.

[27] M. Keller, C. Sager, P. Dumy, M. Schutkowski, G. Fischer et al., *J. Am. Chem. Soc* **1998**, *120*, 2714-2720.

[28] Y. Che, G. R. Marshall, *Biopolymers* **2006**, *81*, 392-406.

[29] N. W. Owens, C. Braun, J. D. O'Neil, K. Marat, F. Schweizer, *J. Am. Chem. Soc.* **2007**, *129*, 11670-11671.

[30] E. Biron, H. Kessler, *J. Org. Chem* **2005**, *70*, 5183-5189.

[31] E. Biron, J. Chatterjee, H. Kessler, *J. Pept. Sci.* **2006**, *12*, 213-219.

[32] E. Biron, J. Chatterjee, O. Ovadia, D. Langenegger, J. Brueggen et al., *Angew. Chem. Int. Ed.* **2008**, *47*, 2595-2599.

[33] S. J. Russell, T. Blandl, N. J. Skelton, A. G. Cochran, *J. Am. Chem. Soc.* **2003**, *125*, 388-395.

[34] R. Mahalakshmi, A. Sengupta, S. Raghothama, N. Shamala, P. Balaram, *Biopolymers* **2007**, *88*, 36-54.

[35] H. Y. Meng, K. M. Thomas, A. E. Lee, N. J. Zondlo, *Biopolymers* **2006**, *84*, 192-204.

[36] C. Toniolo, G. M. Bonora, A. Fontana, *Int. J. Pept. Protein Res.* **1974**, *6*, 371-380.

[37] C. Toniolo, G. M. Bonora, M. Mutter, *J. Am. Chem. Soc.* **1979**, *101*, 450-454.

[38] H. Kessler, W. Schmitt, *Encyclopedia of nuclear magnetic resonance*, Wiley-VCH Weinheim, **1995**.

[39] A. Bax, M. F. Summers, *J. Am. Chem. Soc.* **1986**, *108*, 2093-2094.

[40] L. Braunschweiler, R. R. Ernst, *J. Magn. Reson.* **1983**, *53*, 521.

[41] A. A. Bothner-By, R. L. Stephens, J. M. Lee, C. D. Warren, R. W. Jeanloz, *J. Am. Chem. Soc.* **1984**, *106*, 811-813.

[42] H. Kessler, M. Gehrke, C. Griesinger, *Angew. Chem. Int. Ed.* **1988**, *27*, 490-536.

[43] H. Kessler, C. Griesinger, R. Kerssebaum, K. Wagner, R. R. Ernst, *J. Am. Chem. Soc.* **1987**, *109*, 607-609.

[44] O. W. Howarth, D. M. Lilley, *Prog. NMR Spec.* **1978**, *12*, 1-40.

[45] T. F. Havel, *Prog. Biophys. Mol. Biol.* **1991**, *56*, 43-78.

[46] D. F. Mierke, A. Geyer, H. Kessler, *Int. J. Pept. Protein Res.* **1994**, *44*, 325-331.

[47] A. Bax, *Protein Sci.* **2003**, *12*, 1-16.

[48] C. M. Thiele, *Conc. Magn. Reson. A* **2007**, *30A*, 65-80.

[49] R. Tycko, F. J. Blanco, Y. Ishii, *J. Am. Chem. Soc.* **2000**, *122*, 9340-9341.

[50] H. J. Sass, G. Musco, S. J. Stahl, P. T. Wingfield, S. Grzesiek, *J. Biomol. NMR* **2000**, *18*, 303-309.

[51] B. Luy, K. Kobzar, H. Kessler, *Angew. Chem. Int. Ed.* **2004**, *43*, 1092-1094.

[52] P. W. Kuchel, B. E. Chapman, N. Müller, W. A. Bubb, D. J. Philp et al., *J. Magn. Reson.* **2006**, *180*, 256-265.

[53] G. Kummerlöwe, F. Halbach, B. Laufer, B. Luy, *Open Spec. J.* **2008**, *2*, 29-33.

[54] B. Luy, A. Frank, H. Kessler, *Conformational analysis of drugs by NMR: Drug Properties: Measurement and Computation*, Wiley-VCH Weinheim, **2008**.

[55] C. M. Thiele, S. Berger, *Org. Lett.* **2003**, *5*, 705-708.

[56] M. Zweckstetter, A. Bax, *J. Am. Chem. Soc.* **2000**, *122*, 3791-3792.

[57] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark et al., *J. Comput. Chem.* **2005**, *26*, 1701-1718.

[58] B. Hess, R. M. Scheek, *J. Magn. Reson.* **2003**, *164*, 19-27.

[59] S. Omura, Y. Iwai, A. Hirano, A. Nakagawa, J. Awaya et al., *J. Antibiot.* **1977**, *30*, 275-282.

[60] A. Enthart, J. C. Freudenberger, J. Furrer, H. Kessler, B. Luy, *J. Magn. Reson.* **2008**, *192*, 314-322.

[61] P. D. Davis, C. H. Hill, W. A. Thomas, I. W. A. Whitcombe, *J. Chem. Soc.* **1991**, *3*, 182-184.

[62] I. Takahashi, Y. Saitoh, M. Yoshida, H. Sano, H. Nakano et al., *J. Antibiot.* **1989**, *42*, 571-577.

[63] G. Kummerlöwe, S. Knör, A. O. Frank, T. Paululat, H. Kessler et al., *Chem. Commun.* **2008**, , 5722-5724.

[64] W. L. Jorgensen, J. Tirado-Rives, *Abstr. Pap. Am. Chem. Soc.* **1998**, *216*, U696-U696.

[65] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, *J. Comput. Chem.* **1997**, *18*, 1463-1472.

[66] M. Billeter, G. Wagner, K. Wüthrich, *J. Biomol. NMR* **2008**, *42*, 155-158.

[67] A. A. Yee, A. Savchenko, A. Ignachenko, J. Lukin, X. Xu et al., *J. Am. Chem. Soc.* **2005**, *127*, 16512-16517.

[68] K. R. Rogers, Affinity Biosensors: *Techniques and Protocols*, Humana Press Totowa, **1998**.

[69] D. C. Greenbaum, W. D. Arnold, F. Lu, L. Hayrapetian, A. Baruch et al., *Chem. Biol.* **2002**, *9*, 1085-1094.

[70] A. J. Barrett, A. A. Kembhavi, M. A. Brown, H. Kirschke, C. G. Knight et al., *Biochem. J.* **1982**, *201*, 189-198.

[71] G. R. Knox, P. L. Pauson, *Proc. Chem. Soc.* **1958**, 289-289.

[72] T. J. Kealy, P. L. Pauson, *Nature* **1951**, *168*, 1039-1040.

[73] G. Wilkinson, M. Rosenblum, M. C. Whiting, R. B. Woodward, *J. Am. Chem. Soc.* **1952**, *74*, 2125-2126.

[74] R. B. Woodward, M. Rosenblum, M. C. Whiting, *J. Am. Chem. Soc.* **1952**, *74*, 3458-3459.

[75] E. O. Fischer, W. Pfab, *Z. Naturforsch., B, Biosci.* **1952**, *7*, 377-379.

[76] G. Jaouen, A. Vessieres, I. S. Butler, *Acc. Chem. Res.* **1993**, *26*, 361-369.

[77] R. H. Fish, G. Jaouen, *Organometallics* **2003**, *22*, 2166-2177.

[78] N. Metzler-Nolte, *Angew. Chem. Int. Ed.* **2001**, *40*, 1040-1043.

[79] S. A. Miller, J. A. Tebboth, J. F. Tremaine, *J. Chem. Soc.* **1952**, 632-635.

[80] M. F. Leopold, J. L. Urbauer, A. J. Wand, *Mol. Biotechnol.* **1994**, *2*, 61-93.

[81] G. Winkhaus, H. Singer, *J. Organomet. Chem.* **1967**, 7, 487-491.

[82] G. Winkhaus, H. Singer, M. Kricke, *Z. Naturforsch.* **1966**, 21b, 1109-1110.

[83] R. Zelonka, M. C. Baird, *J. Organomet. Chem.* **1972**, *44*, 383-389.

[84] R. A. Zelonka, M. C. Baird, *Can. J. Chem.* **1972**, *50*, 3063-3072.

[85] M. A. Bennett, G. B. Robertson, A. K. Smith, *J. Organomet. Chem.* **1972**, *43*, C41-C43.

[86] M. A. Bennett, A. K. Smith, *J. Chem. Soc. Dalton Trans.* **1974**, 233-241.

[87] A. N. Nesmeyanov, A. Z. Rubezhov, *J. Organomet. Chem.* **1979**, *164*, 259-275.

[88] T. Arthur, T. A. Stephenson, *J. Organomet. Chem.* **1981**, *208*, 369-387.

[89] E. L. Muetterties, J. R. Bleeke, E. J. Wucherer, *Chem. Rev.* **1982**, *82*, 499-525.

[90] R. M. Moriarty, U. S. Gill, Y. Y. Ku, *J. Organomet.. Chem.* **1988**, *350*, 157-190.

[91] M. A. Bennett, *Coord. Chem. Rev.* **1997**, *166*, 225-254.

[92] D. B. Grotjahn, *Coord. Chem. Rev.* **1999**, *190-192*, 1125-1141.

[93] R. Krämer, *Angew. Chem. Int. Ed.* **1996**, *35*, 1197-1199.

[94] K. Severin, R. Bergs, W. Beck, *Angew. Chem. Int. Ed.* **1998**, *37*, 1634-1654.

[95] G. Jaouen, A. Vessières, *Acc. Chem. Res.* **1993**, *26*, 361-369.

[96] P. Köpf-Maier, H. Köpf, *Chem. Rev.* **1987**, *87*, 1137-1152.

[97] R. Alberto, *J. Organomet. Chem.* **2007**, *692*, 1179-1186.

[98] D. S. Williams, G. E. Atilla, H. Bregman, A. Arzoumanian, P. S. Klein et al., *Angew. Chem. Int. Ed.* **2005**, *44*, 1984-1987.

[99] H. Bregman, P. J. Carroll, E. Meggers, *J. Am. Chem. Soc.* **2006**, *128*, 877-884.

[100] J. É. Debreczeni, A. N. Bullock, G. E. Atilla, D. S. Williams, H. Bergman et al., *Angew. Chem. Int. Ed.* **2006**, *45*, 1580-1585.

[101] C. Streu, E. Meggers, *Angew. Chem. Int. Ed.* **2006**, *45*, 5645-5648.

[102] P. Haquette, B. Talbi, S. Canaguier, S. Dagorne, C. Fosse et al., *Tetrahedron Lett.* **2008**, *49*, 4670-4673.

[103] W. H. Ang, E. Daldini, L. Juillerat-Jeanneret, P. J. Dyson, *Inorg. Chem.* **2007**, *46*, 9048-9050.

[104] R. H. Fish, G. Jaouen, *Organometallics* **2003**, *22*, 2166-2177.

[105] R. E. Morris, R. E. Aird, P. del S. Murdoch, H. Chen, J. Cummings et al., *J. Med. Chem.* **2001**, *44*, 3616-3621.

[106] R. E. Aird, J. Cummings, A. A. Ritchie, M. Muir, R. E. Morris et al., *Br. J. Cancer* **2002**, *86*, 1652-1657.

[107] S. J. Dougan, P. J. Sadler, *Chimia* **2007**, *61*, 704-715.

[108] W. F. Schmid, R. O. John, V. B. Arion, M. A. Jakupec, B. K. Keppler, *Organometallics* **2007**, *26*, 6643-6652.

[109] A. F. A. Peacock, A. Habtemariam, R. Fernandez, V. Walland, F. P. A. Fabbiani et al., *J. Am. Chem. Soc.* **2006**, *128*, 1739-1748.

[110] A. F. A. Peacock, S. Parsons, P. J. Sadler, *J. Am. Chem. Soc.* **2007**, *129*, 3348-3357.

[111] H. Kostrhunova, J. Florian, O. Novakova, A. F. A. Peacock, P. J. Sadler et al., *J. Med. Chem.* **2008**, *51*, 3635–3643.

[112] J. C. Gray, A. Habtemariam, M. Winnig, W. Meyerhof, P. J. Sadler, *J. Biol. Inorg. Chem.* **2008**, *13*, 1111-1120.

[113] K. Severin, *Chem. Commun.* **2006**, 3859-3867.

[114] W. Baratta, W. A. Herrmann, R. M. Kratzer, P. Rigo, *Organometallics* **2000**, *19*, 3664-3669.

[115] W. Baratta, E. Herdtweck, W. A. Herrmann, P. Rigo, J. Schwarz, *Organometallics* **2002**, *21*, 2101-2106.

[116] S. Gladiali, E. Alberico, *Chem. Soc. Rev.* **2006**, *35*, 226-236.

[117] K. J. Haack, S. Hashiguchi, A. Fujii, T. Ikariya, R. Noyori, *Angew. Chem. Int. Ed.* **1997**, *36*, 285-288.

[118] J. Hannedouche, G. J. Clarkson, M. Wills, *J. Am. Chem. Soc.* **2004**, *126*, 986-987.

[119] A. M. Hayes, D. J. Morris, G. J. Clarkson, M. Wills, *J. Am. Chem. Soc.* **2005**, *127*, 7318-7319.

[120] D. J. Morris, A. M. Hayes, M. Wills, *J. Org. Chem.* **2006**, *71*, 7035-7044.

[121] C. Daguenet, R. Scopelliti, P. J. Dyson, *Organometallics* **2004**, *23*, 4849-4857.

[122] L. Ackermann, A. Althammer, R. Born, *Angew. Chem. Int. Ed.* **2006**, *45*, 2619-2622.

[123] A. Fürstner, L. Ackermann, *Chem. Commun.* **1999**, 95-96.

[124] M. Creus, A. Pordea, T. Rossel, A. Sardo, C. Letondor et al., *Angew. Chem. Int. Ed.* **2008**, *47*, 1400-1404.

[125] J. Steinreiber, T. R. Ward, *Coord. Chem. Rev.* **2008**, *252*, 751-766.

[126] W. S. Sheldrick, A. Gleichmann, *J. Organomet. Chem.* **1994**, *470*, 183-187.

[127] A. J. Gleichmann, J. M. Wolff, W. S. Sheldrick, *J. Chem. Soc., Dalton Trans.* **1995**, 1549-1554.

[128] J. M. Wolff, A. J. Gleichmann, C. Schmidt, W. S. Sheldrick, *J. Inorg. Biochem.* **1995,** *59*, 219.

[129] J. M. Wolff, W. S. Sheldrick, *J. Organomet. Chem.* **1997**, *531*, 141-149.

[130] J. M. Wolff, W. S. Sheldrick, *Chem. Ber. Recueil* **1997**, *130*, 981-988.

[131] D. A. Herebian, C. S. Schmidt, W. S. Sheldrick, C. van Wüllen, *Eur. J. Inorg. Chem.* **1998**, 1991-1998.

[132] R. Stodt, S. Gencaslan, A. Frodl, C. Schmidt, W. S. Sheldrick, *Inorg. Chim. Acta* **2003**, *355*, 242-253.

[133] R. Stodt, S. Gencaslan, I. M. Müller, W. S. Sheldrick, *Eur. J. Inorg. Chem.* **2003**, 1873-1882.

[134] J. Soleimannejad, A. Sisson, C. White, *Inorg. Chim. Acta* **2003**, *352*, 121-128.

[135] G. Marconi, H. Baier, F. W. Heinemann, P. Pinto, H. Pritzkow, et al., *Inorg. Chim. Acta* **2003**, *352*, 188-200.

[136] J. Soleimannejad, C. White, *Organometallics* **2005**, *24*, 2538-2541.

[137] H. Dialer, P. Mayer, K. Polborn, W. Beck, *Eur. J. Inorg. Chem.* **2001**, 1051-1055.

[138] A. Schlichtiger, Dissertation (Ph. D. Thesis), Technische Universität München, **2009**.

[139] T. Reiner, Dissertation (Ph. D. Thesis), Technische Universität München, **2009**.

[140] W. Willker, D. Leibfritz, R. Kerssebaum, W. Bermel, *Magn. Reson. Chem.* **1993**, *31*, 287-292.

[141] A. L. Davis, J. Keeler, E. D. Laue, D. Moskau, *J. Magn. Reson.* **1992**, *98*, 207-216.

[142] A. A. Shaw, C. Salaun, J. F. Dauphin, B. Ancian, *J. Magn. Reson. A* **1996**, *120*, 110-115.

[143] B. Ancian, I. Bourgeois, J. F. Dauphin, A. A. Shaw, *J. Magn. Reson.* **1997**, *125*, 348-354.

[144] J. Cavanagh, M. Rance, *J. Magn. Reson. A* **1993**, *105*, 328-328.

[145] J. Cavanagh, M. Rance, *J. Magn. Reson.* **1990**, *88*, 72-85.

3 Imaging the Real World

Simulation of Molecular Dynamics

The determination of three-dimensional structures allows insight into the architecture of single molecules and molecular complexes. Informations obtained in this way are crucial e.g. for rationally designing or optimizing receptor inhibitors, for studying structural consequences of amino acid mutations in proteins, or for elucidating the binding modes of interaction partners. However, images of 3D structures only represent a snapshot and can be regarded as low (potential) energy conformations which are frozen at absolute zero temperature. This fact implies that entropic and therefore dynamic effects are completely neglected. As the entropy plays often an essential role in structure formation and intermolecular recognition, respectively, it is indispensable for understanding bio-chemical processes to embed static molecular arrangements into a dynamical framework. This is especially true for small and medium-sized peptides since they usually display more degrees of freedom per amino acid residue compared to proteins or their complexes. As it is often very difficult or even not possible to trace fast dynamics via experimental methods, computational tools have been developed which are utilized to simulate the dynamical behavior of all kinds of molecules. In addition, by means of these virtual approaches, informations about thermodynamically preferred states or molecular motions like free or restricted rotational tumbling and translational diffusion can also be gained. In principle, molecular dynamics simulations can be applied in two different ways: on one side, they can serve as tools to predict the behavior of molecules under certain conditions; on the other side, they can be used to complement or explain experimental outcomes.

3.1 Conformational Dynamics of an Integrin Inhibitor in Various Environments

The strength of biomolecular interactions depend on the mutual geometry of the receptor and ligand atoms which participate in the binding.[1-3] Only when groups that attract or repulse each other are optimally oriented in space, a stable intermolecular complex arises. In the context of rational drug design, this implies that a ligand displays a high target affinity only if its three-dimensional structure allows an ideal orientation of the essential pharmacophore groups.[4] In other words, the formation of a receptor ligand complex requires that a designed molecule (the key) adequately fits into the binding pocket (the lock) of e.g. a protein.[5] These considerations are particularly of great importance when one deals with the development of peptidic drugs. Due to the presence of multiple degrees of freedom (which are also present in conformationally restricted cyclic species), peptides can adopt several conformations that differ not only in their 3D shape but also in their affinity to a target molecule of interest.[1] It generally pertains that the receptor affinity of a (cyclic) peptide is the more decreased the larger the number of differing (backbone) conformations is. Therefore, the aim in rational peptide drug development is the design of structurally stable - often called “rigid” - scaffolds which enable an optimal spatial orientation of those atoms that take part in the intermolecular recognition.

3.1.1 The Scientific Question

One outstanding example of the successful design of a peptidic drug is the $\alpha_v\beta_3$ integrin binding peptide *cyclo(RGDf(N^{Me})V-)*, usually called Cilengitide, which is currently in clinical phase III for treating brain tumors and in clinical II for curing other cancer types.[6-10] The three-dimensional arrangements of the cyclic pentapeptide in water and in complex with its integrin receptor were elucidated a couple of years ago;[11] a comparison of both structures reveals slight differences in the backbone conformations. In this scope, it became of interest if Cilengitide can adopt varying 3D structures when being located in different environments. Obtaining insight into the conformational stability of the peptidic drug would provide not only beneficial informations for the rational design of further peptidic integrin inhibitors but might also allow for drawing conclusions with respect to the extremely high $\alpha_v\beta_3$ receptor affinity of Cilengitide.

Thus, the central goal of this project was to investigate the conformational stability and intrinsic dynamics of the $\alpha_v\beta_3$ integrin binder Cilengitide in different environments.

For this purpose, the structure of the cyclic pentapeptide was studied in both protic and aprotic, polar and apolar solvents: water (H_2O), dimethyl sulfoxide (DMSO), methanol (MeOH), and chloroform ($CHCl_3$). By means of solution-state NMR spectroscopy and DG / MD computations, the 3D structures of Cilengitide in various environments were determined and compared both with each other and with the peptide conformation when bound to its receptor. Finally, correlations between the dynamical behavior of Cilengitide and its $\alpha_v\beta_3$ integrin binding mode are elucidated.

3.1.2 Cilengitide as Peptidic Integrin Inhibitor

Integrins are heterodimeric receptors found on the surface of most cell lines that facilitate attachment to other cells and the extracellular matrix.[12] Moreover their bidirectional signaling pathways can affect cell growth, differentiation, division, migration, and apoptosis.[13,14] Apart from these functions integrins are responsible for the binding of certain viruses to the cell and take part in blood coagulation. These characteristics also make them targets for pathological processes like inflammation, cardiovascular disorders, thrombosis, restenosis, vascular homeostasis, osteoporosis, cancer invasion, metastasis and tumor angiogenesis.[15-19] The challenge when targeting integrins is to develop selective drugs that discriminate between the receptor subtypes and thereby cause fewer side effects.

The elucidation of the minimal integrin recognition motif, the RGD amino acid sequence,[20] served as initial basis for the development of Cilengitide. Further on, the central RGD motif was first elongated by a *C*-terminal valine and an *N*-terminal phenylalanine and finally cyclized in a head-to-tail manner. By means of the so-called spatial screening technique, the stem sequence *c(RGDfV-)* was obtained; for this to work, each amino acids was once incorporated into the cycle in its D-configuration.[21-24,24,25] From the resulting five peptides, the one which contained a D-phenylalanine turned out to have ca. 1000 times higher activity for the $\alpha_V\beta_3$ receptor and about ten times reduced activity for the platelet receptor $\alpha_{IIb}\beta_3$ (compared to the linear reference peptide *GRGDSPK*).

[23,24] Hence, a super-active and selective integrin inhibitor was yielded.[23,26,27] Later, the new-found peptide was optimized by using another technique, named *N*-methylation scan.[6] The most promising peptide, *c(RGDf(N^{Me})V-)*, was elected by the Merck KGaA (Darmstadt) as drug candidate (see Fig. 3.1).

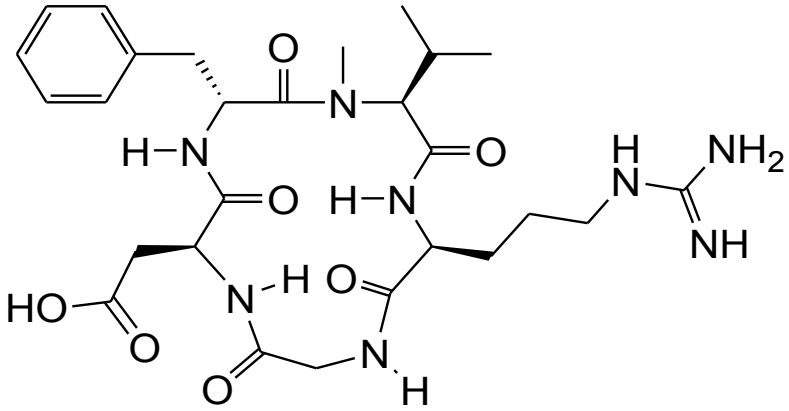


Figure 3.1: Presentation of the chemical constitution of Cilengitide. The shown structure does not imply the presence of secondary turn structures and hydrogen bonds, respectively.

Cilengitide has already entered the clinical phases II and III for the treatment of *glioblastoma multiforme,* metastatic prostate cancer, and lymphoma.[28-30] The cyclic pentapeptide is completely stable *in vivo* (it is excreted without metabolic degradation and shows a half life time of ca. four hours in man). As a drawback, Cilengitide is not orally available and has to be injected. Thus, further developments, also based on insights in conformational preferences, are necessary to convert promising peptides like this into orally available drug compounds.

3.1.3 Structural Studies of Cilengitide in Various Environments

After synthesis and purification of Cilengitide, the peptide was dissolved in the various four solvents and subsequently studied via NMR spectroscopy. Intra- and interresidual assignments were obtained by ^{1}H,^{1}H-COSY, ^{1}H,^{1}H-TOCSY, ^{1}H,^{13}C-HSQC, and ^{1}H,^{13}C-HMBC spectra. Distance restraints were yielded by ^{1}H,^{1}H-ROESY measurements; ^{1}H-1D and ^{1}H,^{1}H-E.COSY spectra provided several homonuclear J-couplings. For the creation of template structures, only ROE based distance restraints were used in DG computations. All MD simulations were conducted without constraining forces; the DG derived 3D arrangements of Cilengitide served as starting structures. To evaluate the results of the 150 ns molecular dynamics calculation runs, the experimentally gained distance and dihedral informations were utilized (see Fig. 3.2).

Parameters	H_2O	DMSO	MeOH	$CHCl_3$
NMR ROEs	59 / 23 / 17	55 / 18 /12	60 / 22 / 17	47 / 15 / 9
NMR ^{3}J-Couplings	12 / 5	12 / 5	12 / 5	12 / 5
DG Distance Restraints	51 / 23 / 17	44 / 18 / 12	49 / 22 / 17	37 / 15 / 9
DG Dihedral Restraints	0	0	0	0
MD Distance Restraints	0	0	0	0
MD Dihedral Restraints	0	0	0	0
Evaluation of Distances	56 / 21 / 15	52 / 18 / 10	58 / 20 / 15	47 / 15 / 9
Evaluation of Dihedrals	10 / 5	10 / 5	10 / 5	10 / 5
ROE Violations (1 ns)	6 (0.15 Å)	10 (0.24 Å)	4 (0.12 Å)	2 (0.09 Å)
ROE Violations (150 ns)	5 (0.15 Å)	6 (0.20 Å)	1 (0.01 Å)	1 (0.09 Å)
^{3}J Violations (1 ns)	1.1 Hz	2.5 Hz	1.1 Hz	3.1 Hz
^{3}J Violations (150 ns)	1.2 Hz	1.2 Hz	1.3 Hz	0.9 Hz

Figure 3.2: Table of extracted and utilized experimental parameters for the elucidation of the 3D structures of Cilengitide in four different types of solvent.

NMR section: number of experimental total / backbone / (interresidual backbone) distance and dihedral informations; DG / MD section: number of total / backbone / interresidual backbone distance and dihedral restraints; Evaluation section: number of applied total / backbone / (interresidual backbone) distance and dihedral restraints; Violation section: number and RMSD values of distance and dihedral restraint violations of short and long MD calculations.

An overlay of the most representative four structures reveals that all backbone conformations have (almost) the same shape (see Fig. 3.3). In all cases, no structural elements like regular turns or H-bonds can be found. The only deviation arising in the overlay is the orientation of the ^{Me}Val5-Arg1 peptide bond of the MeOH model which is flipped by 180° around its adjacent φ and ψ dihedrals compared to the bond geometry of the other structures.

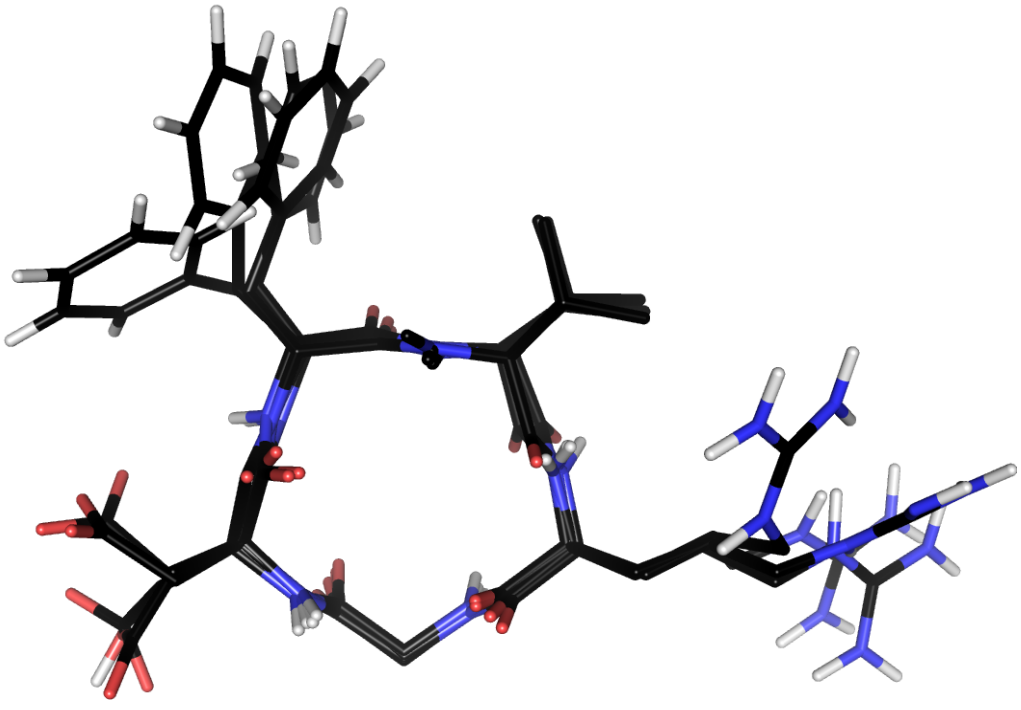


Figure 3.3: Illustration of an overlay of the computed structures of Cilengitide measured in H_2O, DMSO, MeOH, and $CHCl_3$. In all cases, an all-*trans* arrangement is present. The cyclic peptide backbone conformations almost perfectly fit on each other. One deviation is present in the MeOH derived model where the peptide bond between ^{Me}Val5 and Arg1 is flipped by 180° around its adjacent φ and ψ backbone dihedral angles.

These findings imply that both the three-dimensional arrangement of Cilengitide is determined by steric characteristics (no stabilizing H-bonds) and that the conformation is - even in different, both polar and apolar, environments - extremely robust. If the most representative structure of the cyclic peptide matches the active conformation (the 3D arrangement in the receptor bonded state) its structural stability would be one of the reasons of the very high $\alpha_v\beta_3$ integrin affinity. To prove this hypothesis, the elucidated models were compared with the (crystal) conformation of Cilengitide complexed to its target.[11]

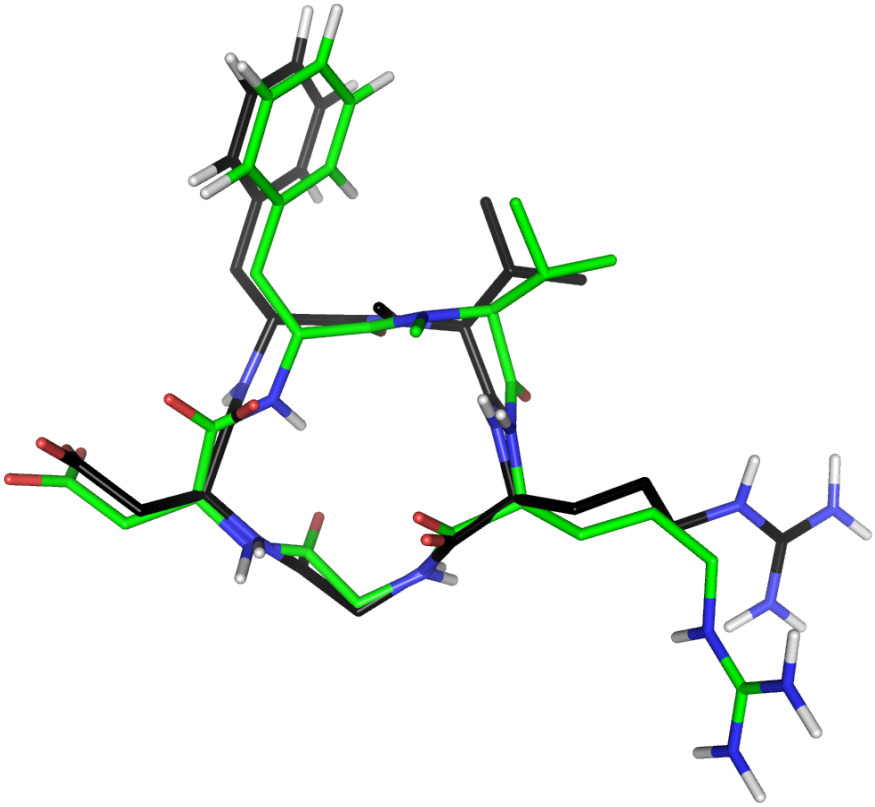


Figure 3.4: Presentation of an overlay of two Cilengitide conformations: the black structure represents the free conformation in water; the green colored structural model was extracted from the crystal structure of Cilengitide bound to its receptor. The backbone atoms of the most important pharmacophore groups display a good fitting. Structural deviations are visible for phenylalanine and valine.

The overlay of the water derived structure of Cilengitide and its structure in the complex reveals that the backbone atoms of the most important pharmacophore groups (the RGD sequence) fit very well on each other; for example, the positions of the three RGD Cα atoms and the orientations of their Cα-Cβ bond vectors of both structures are nearly identical (see Fig. 3.4). Deviations between the conformations are present for the phenylalanine and valine backbone atoms. The strongest variance is visible for orientation of the peptide bond between Asp3 and D-Phe4 (“phe4”). Phenylalanine plays an important role in the interaction of Cilengitide and the integrin. As the type of the interaction is of hydrophobic nature, however, a “perfect“ orientation (with

respect to an optimal binding mode) of the aromatic Phe side chain is not that crucial for yielding high affinity as it would be e.g. for a group exhibiting an H-bond (its strength strongly depend on the distance and angle of the participating groups). As the crystal complex structure that served as source for the bound peptide conformation has a resolution of solely 3.2 Å, a closer analysis would only provide speculative results.[7,11] The good consensus of the overlayed RGD backbone atoms, however, shows that the structure of Cilengitide in its "free" state adequately matches the biologically active conformation.

To give an estimation of the similarity of the four solvent based structures and the conformation of Cilengitide in the receptor bound state, the distances between the important Cα and Cβ atoms of the pharmacophore carrying Arg1 and Asp3 residues as well as the angular orientation of two planes spanned by the Arg1/Asp3 Cα-Cβ bond vectors are shown in figure 3.5.

Environment	^{Arg1}Cα - ^{Asp3}Cα Distance [Å]	^{Arg1}Cβ - ^{Asp3}Cβ Distance [Å]	^{Arg1}Cα/β - ^{Asp3}Cα/β Plane Angle [°]
H_2O	6.2	8.4	173.2
DMSO	5.9	8.4	159.0
MeOH	6.3	8.7	153.1
$CHCl_3$	6.0	8.3	176.6
Bound *	6.4	8.9	178.6

Figure 3.5: Table of the distances between Arg1 and Asp3 Cα and Cβ atoms and of the angle of the planes spanned by the four atoms. As Arg1 and Asp3 carry the pharmacophore groups, the orientations and distances of their Cα and Cβ groups provide hints about the similarities of the different Cilengitide conformations and thus of differing receptor binding modes. It becomes obvious that all four structures have highly similar distances and plane angles of Cα and Cβ atoms which furthermore reveal only small differences compared with the 3D geometry in the bound state.

* the atomic resolution of the Cilengitide structure in the bound state is only 3.2 Å; thus, it is not possible to give accurate values.

3.1.4 The Dynamical Behavior of Cilengitide

After the 3D structures of Cilengitide in different solvent types were elucidated and compared with each other and with its biologically active conformation, the dynamical behavior of the peptide freely tumbling in solution was analyzed. As can be seen in figure 3.2, the best agreement between experimental and calculated results is obtained for methanol and chloroform studies. Thus, dynamical characteristics of Cilengitide were investigated by the help of the MeOH (more similar to water than $CHCl_3$) based MD calculations. A good indicator for molecular flexibilities is the analysis of the RMS deviations of backbone atoms (see Fig. 3.5).[31] It becomes clear that not only positional fluctuations are present; rather, RMSD "jumps" were detected which represent 180° flips of φ and ψ dihedral angles. Three of the five peptide bonds display such motions: ^{Me}Val5-Arg1, Arg1-Gly2, and Gly2-Asp3 (see e.g. Fig. 3.6).

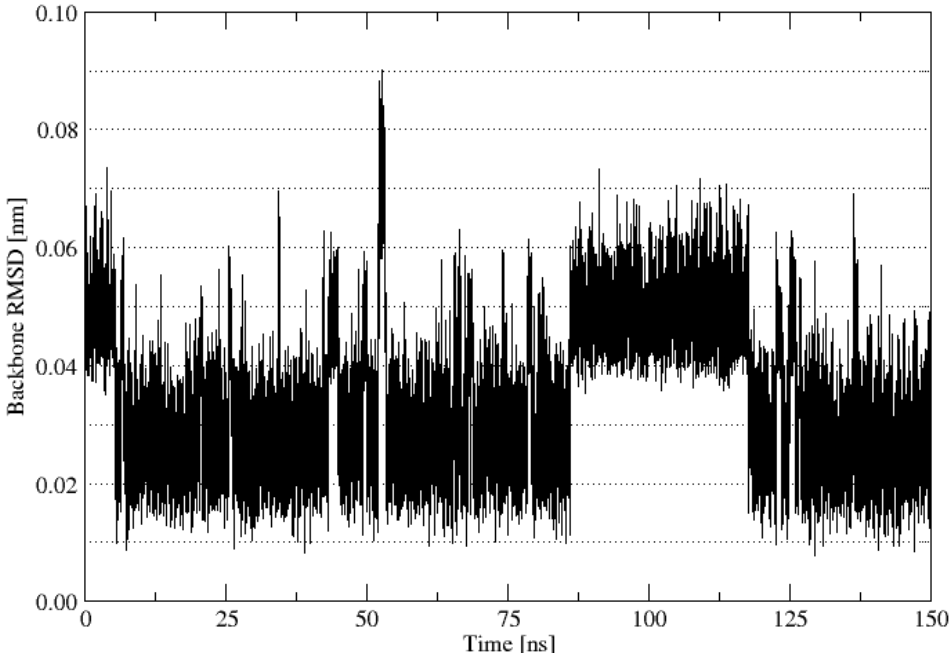


Figure 3.6: Illustration of the MD simulation derived backbone RMSD of Cilengitide in methanol. The distribution of points reflect intrinsic dynamics of atoms. At some time steps, RMSD jumps are visible that represent 180° flips of peptide bonds around adjacent φ and ψ backbone dihedral angles, e.g. the bidirectional jump at ca. 52 ns results from a peptide bond flip between Gly2 and Asp3.

Since the observed 180° peptide bond flips only lead to different orientations of the afflicted C-O and N-H bond vectors, the backbone structures of the cyclic peptide remains - except small deviations - the same during the MD simulations. A detailed study of the different peptide bond orientations (not *cis-trans* populations!) showed that all distributions are very similar, except for the ^{Me}Val5-Arg1 bonds; whereas the preferred orientations of this peptide bond in water, DMSO, and chloroform are in the same range - 90-100 % with respect to the “basic structure” shown for the H_2O, DMSO and $CHCl_3$ studies in figure 3.3 -, the distribution in the “MeOH” case (which is the only run that displays almost “converged” distance restraints; see Fig. 3.2) is ca. 1:4 (see Fig. 3.7).

Figure 3.7: Table of MD preferred orientations of Cilengitide peptide bonds in four different solvents. The H_2O derived structure as illustrated in figures 3.3 and 3.4 served as reference conformation. The MD simulations from which the shown results were extracted display a not-converged sampling of conformational space for H_2O, DMSO, and $CHCl_3$ runs.

Peptide Bond	H_2O [%]	DMSO [%]	MeOH [%]	$CHCl_3$ [%]
^{Me}Val5-Arg1	100 / 0	91 / 9	21 / 79	100 / 0
Arg1-Gly2	94 / 6	99 / 1	93 / 7	100 / 0
Gly2-Asp3	100 / 0	100 / 0	99 / 1	100 / 0
Asp3-phe4	100 / 0	100 / 0	100 / 0	100 / 0
phe4-^{Me}Val5	100 / 0	100 / 0	100 / 0	100 / 0

However, it must be stated that the remaining distance violations in the H_2O, DMSO, and $CHCl_3$ studies (as given in figure 3.2) in particular involve Arg1-NH protons; this implies that the conformational space of the mentioned atom was

not exhaustively sampled during the 150 ns MD simulations due to the presence of high energy barriers.[31] Thus, it is likely that converged "water, DMSO, and chloroform" run would result in ^{Me}Val5-Arg1 peptide bond distribution values that are more shifted to the preferred "MeOH" orientation. Summarizing the data given in figure 3.7, it becomes clear that the peptide bonds ^{Me}Val5-Arg1 and Arg1-Gly2 are the most flexible ones with respect to 180° reorientations. In contrast, the C-O and N-H vector orientations of Asp3-phe4 and phe4-^{Me}Val5 are very stable. The peptide bond between Gly2 and Asp3 reveals an extremely small tendency to flip around its adjacent φ and ψ dihedral angles.

3.1.5 The Binding Mode of Cilengitide and its Receptor Integrin

By means of the crystal complex structure of Cilengitide and the $\alpha_v\beta_3$ integrin, the binding mode of the ligand to its receptor could be elucidated at atomistic resolution. As it is expected, the charged / polar side chains of Arg1 and Asp3 - the pharmacophore part of the RGD motif - are involved in essential contacts to the protein (H-bonds and a manganese ion complex). Moreover, D-Phe4 further stabilizes the complex via hydrophobic interactions of its aromatic side chain. Finally, the amide proton of Asp3 exhibits a hydrogen bond to the integrin (see Fig. 3.8 and 3.9).[11]

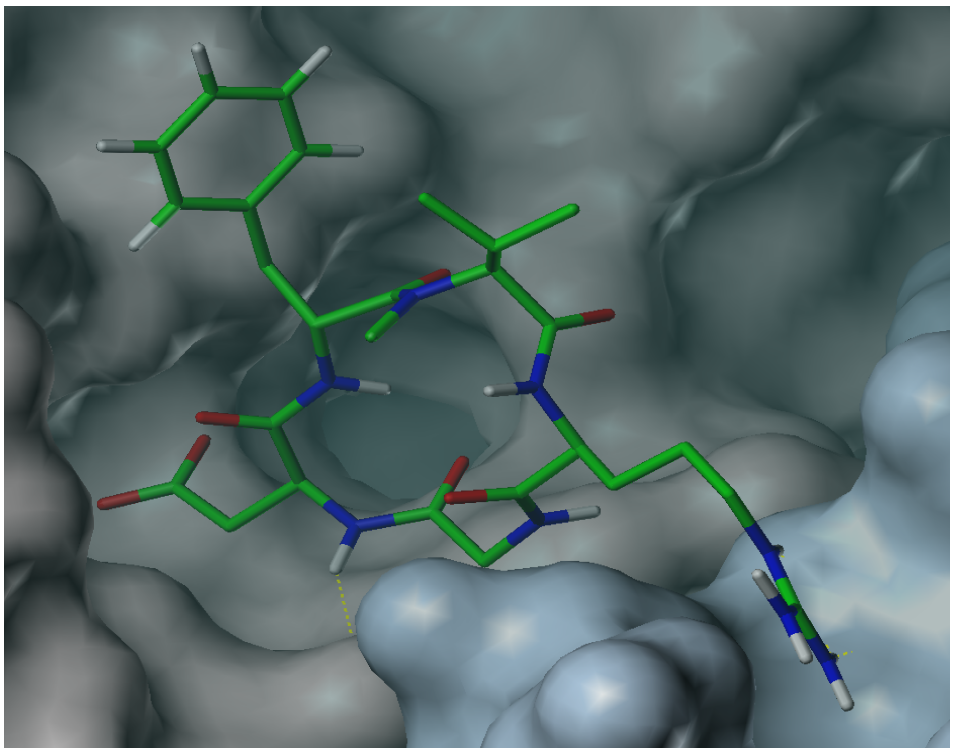


Figure 3.8: Illustration of the crystal structure of the extracellular segment of the $\alpha_v\beta_3$ integrin in complex with Cilengitide. The image section displays the most essential features of the binding mode. The carboxy group of Asp3 binds to a manganese ion; the guanidinium group of Arg1 is involved in hydrogen bonds with the protein. As visualized by the dashed yellow line, the amide proton of Asp3 also exhibits an intermolecular H-bond.

With the data about the conformational flexibility of Cilengitide in hand, it becomes interesting to compare the obtained results with the binding mode of the integrin ligand. Since the overall backbone structure is very stable in both polar and apolar solvents, the already optimally oriented Cα-Cβ bond vectors of Arg1 and Asp3 in the unbound state of Cilengitide (see Fig. 3.4) are not subjected to any, with the binding mode interfering, dynamical rearrangements.

Moreover, a steady orientation of the Asp3 amide hydrogen is important for the stability of the receptor ligand complex as the strength of a hydrogen bond strongly depends on the distance and, especially, the angle between the participating groups. In this scope, it could be demonstrated for the free Cilengitide that the Gly2-Asp3 peptide bond is adequately oriented at ca. 99-100 % of time and that disturbing 180° flips occur - if anyway - very rarely. The reduced flexibility of the mentioned peptide bond in the unbound state contributes to the free binding energy of the complex as the loss of entropy due to H-bond forming is "*a priori*" diminished. In contrast, the peptide bonds ^{Me}Val5-Arg1 and Arg1-Gly2 which are not involved in the intermolecular interaction show a much higher tendency to flip around their adjacent φ and ψ dihedral angles (see Fig. 3.9).[7]

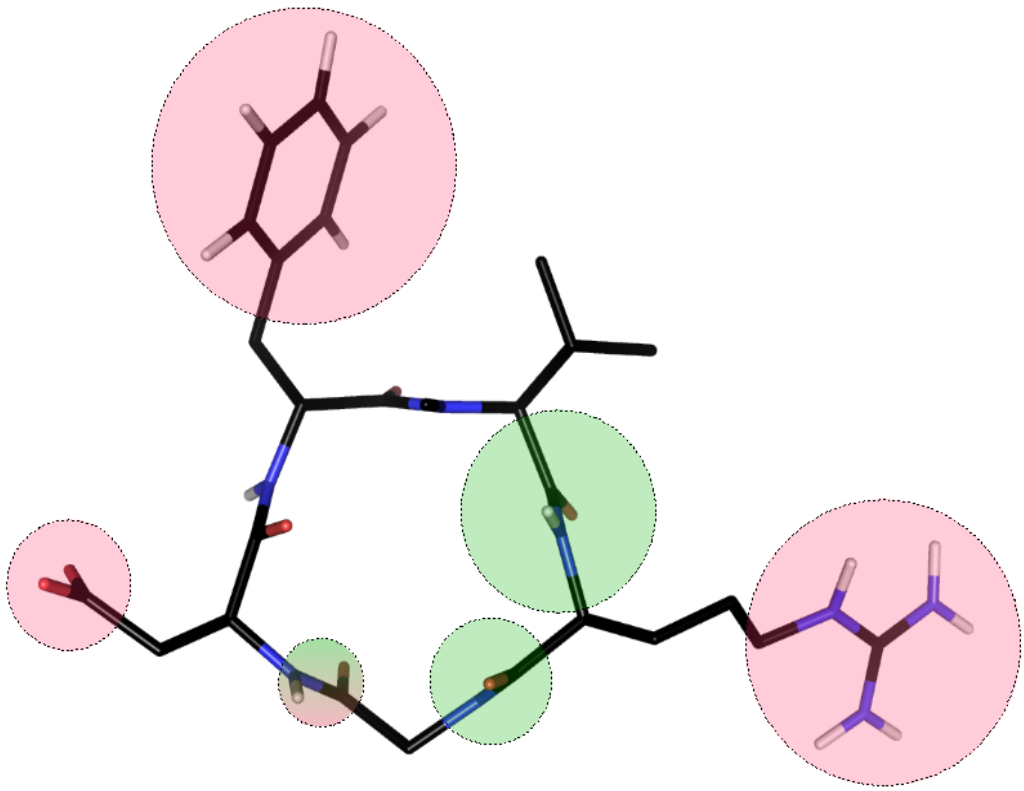


Figure 3.9: Overview of all relevant pharmacophore groups (slight red spheres) and the regions displaying the highest MD backbone flexibilities of Cilengitide (green spheres). While the ^{Me}Val5-Arg1 peptide bond shows several and the Arg1-Gly2 peptide bond some 180° flips around adjacent φ and ψ dihedral angles, the Gly2-Asp3 C-O and N-H vectors are at 99-100 % of time in the orientation illustrated here (red-green sphere).

3.1.6 Conclusion

By the help of NMR spectroscopy, DG calculations, and MD simulations, the three-dimensional structures of the $\alpha_v\beta_3$ integrin inhibitor Cilengitide in four different solvents were determined. A comparison of the most representative conformations revealed that the overall backbone arrangements were practically the same and, in addition, extremely stable. Small variances are present for the average peptide bond orientations. However, the MD simulations for deriving 3D structures in water, DMSO, and chloroform are not completely converged regarding the sampling of the conformational space; thus, precise data about the distribution of correlated C-O and N-H vector orientations could only be given for the calculations in MeOH. In total, the backbone structure (inclusive the important Cα-Cβ bond vectors) of Cilengitide in the free state excellently matches the biologically active conformation. An analysis of conformational dynamics with respect to the protein ligand binding mode reveals that the backbone atoms responsible for orienting the pharmacophore groups exhibit restricted flexibilities. All these findings provide very interesting insights into the origin of the high $\alpha_v\beta_3$ integrin affinity of Cilengitide.

3.1.7 Reagents, Methods, and Experiments

All informations regarding used reagents, peptide synthesis, and HPLC purification can be found in literature.[32]

All NMR spectra were recorded on a Bruker 600 MHz spectrometer equipped with a QXI probe head. For the lock signal, the following solvents were used: d_3-acetic acid in H_2O + 5 % D_2O (pH 4.5; less chemical exchange)), d_6-DMSO, d_3-MeOH, and d_1-$CDCl_3$. Due to the poor solubility of Cilengitide in pure chloroform, 10 % methanol was added to the solution. The measurements for water and ethanol were recorded at 280 K (less chemical exchange), the other two at 300 K. The following NMR experiments were performed: ^{1}H-1D, ^{1}H,^{1}H-E.COSY, -TOCSY, -ROESY, ^{1}H,^{13}C-HSQC, and -HMBC.[33-38] Mixing times for TOCSY spectra were 80 ms, for ROESY spectra 150 ms. HSQC spectra were recorded with a direct proton carbon coupling constant of 140 Hz, and HMBC spectra with a long-range ^{1}H-^{13}C coupling constant of 7 Hz. For all HSQC spectra, a ^{13}C composite pulse decoupling was utilized. For the measurements in H_2O, a Watergate solvent suppression scheme was applied. 8k (except

HSQC: 1k) data points were recorded in the direct dimension, 256 (homonuclear spectra except of E.COSYs: 2k) and 512 (heteronuclear spectra) in the indirect dimension. For all spectra, a recovery delay of 1.5 s was used. Exponential / square sine window functions were used for spectra apodization.

The integrated volumes of ROE cross peaks were converted to proton–proton distances by the help of calibration to three different distances as references. Upper and lower distance restraints were obtained by adding and subtracting 10 % to the calculated experimental values, thus accounting for experimental errors and simulation uncertainties. Metric matrix DG calculations were carried out with a home-written distance geometry program utilizing random metrization.[39,40] Experimental distance restraints which are more restrictive than the geometric distance bounds (holonomic restraints) were used to create the final distance matrix. All other details of the DG computations can be found elsewhere.[41] 100 structures were calculated for each system.

All MD simulation were carried out with the GROMACS 3.3.3 and GROMOS96 software packages.[42-44] The DG structures served as starting conformations for the following 16 MD runs: Cilengitide as uncharged and as zwitter ion molecule at 280 K and 300 K (for all four solvent types). The GROMOS 53a6 force field was utilized for parametrization of the cyclic pentapeptide and all solvent molecules.[45,46] Cilengitide was first energy minimized *in vacuo*, then placed in truncated octahedral boxes with a minimum distance of 1.5 nm between solute atoms and the box walls. After the boxes were filled with solvent molecules, the systems were equilibrated as published elsewhere.[47] A triple-range cutoff for Coulomb interactions including a reaction-field was used (0.8 and 1.4 nm). Van der Waals interactions were calculated with a short-range cutoff of 0.8 nm and a long-range cutoff of 1.4 nm. A atom pair-list was used with a cutoff of 0.8 nm and was updated each five integration steps. All bonds were constrained with the Shake algorithm. The integration time step was 2 fs. Production runs had a length of 150 ns.

3.1.8 Declaration

The presented topic was conducted in collaboration with Dr. Florian Manzenrieder (Technical University Munich, Department Chemistry). F.M. synthesized the Cilengitide peptide. The described results are in preparation for publication.

3.2 Orientational Preferences of Small Molecules in Alignment Media

Residual dipolar couplings play an important role in NMR based elucidation of molecular conformations, dynamics, and interactions.[48-51] Whereas the structure refinement of proteins via RDCs has already become a standard technique, the anisotropic parameters are less frequently applied in conformational analysis of small organic molecules or peptides.[52] In order to be able to measure residual dipolar couplings by NMR spectroscopy, the solute of interest must be partially aligned in solution so that dipolar couplings are not completely averaged to zero. Ongoing developments of suitable alignment media make the utilization of RDCs more and more applicable to a wide range of interesting molecules and solvent types.[53] Compared to other alignment media (e.g. liquid crystalline phases), stretched polymer gels have the advantage of an arbitrary scalability of the alignment strength.[54,55] Since anisotropic parameters like RDCs are proportional to the time averaged orientation of a molecule relative to a static magnetic field (as it is present in an NMR spectrometer), both short- and long-range correlations within a molecule can be obtained which allow the refinement of three-dimensional structures, the elucidation of conformational flexibilities, and the proof of intermolecular binding events.[52,53]

3.2.1 The Scientific Question

Beside their application in the above mentioned research fields, RDCs are also subject of multiple, more theoretical studies. To name but a few, methods have been developed in order to evaluate the reliability and accuracy of recorded anisotropic parameters (e.g. the PALES-SVD algorithm) or to extract informations of correlated dynamics in secondary structural elements via RDCs.[51,56-61] Concerning theoretical / computational investigations, prediction of residual dipolar couplings seems to be the most ambitious scheme.[57,62,63] There are several reasons why an accurate estimation of RDCs is of distinct interest; for example, prediction of RDCs via simplified models might provide insight in the interactions of alignment media and solutes which lead to the occurrence of residual dipolar couplings. In addition, differing RDC signal sets measured for mixtures of enantiomers in chiral alignment media (e.g. gelatine or PBLG) could be stereospecifically assigned to distinct chiral molecules, thereby allowing an NMR based discrimination of enantiomers.[64,65]

Therefore, the aim of the study was to investigate if molecular dynamics simulations of highly simplified alignment models allow an estimation of residual dipolar couplings.

For this purpose, a molecular dynamics calculations test set consisting of polystyrene as alignment medium and strychnine as solute was established. Various MD simulation setups were examined with respect of their practicability, performance, reliability, and precision. Furthermore, the obtained results of the MD study were compared with the established steric PALES RDC prediction technique.[63] Finally, the power and the drawbacks as well as ways to improve the designed approach were evaluated and discussed.

3.2.2 Choice of a Robust Test System

The first task was to select a proper simulation test system. Referring to this, the most important prerequisites were the availability of precise experimental RDC values, an accurately force field parameterizable solute molecule exhibiting minimal intrinsic dynamics (to exclude errors or uncertainties due to an improper modeling of internal flexibilities), and an experimentally and computationally well investigated polymer. After exhaustive considerations and practical MD test runs of norcamphor, strychnine, and menthol as solutes and polystyrene (PS) and polydimethylsiloxane (PDMS) as stretched polymers, it was decided to concentrate on a “polystyrene-strychnine“ system.[59,66]

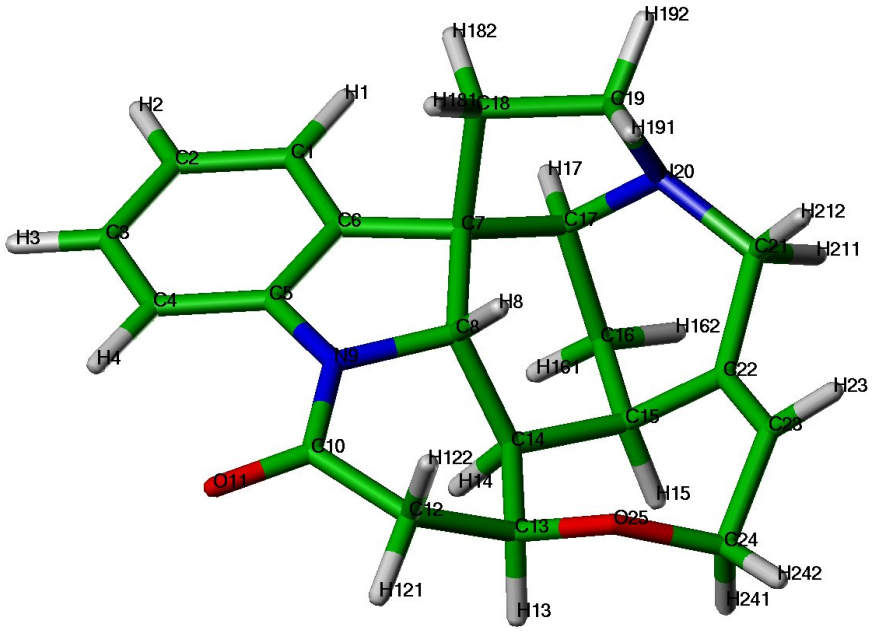


Figure 3.10: Illustration of the 3D structure of strychnine that was used as test molecule for the MD based prediction of residual dipolar couplings. All atoms are labeled with names as used in the simulations.

The usage of polystyrene was preferred as its conformational and dynamical characteristics are experimentally investigated in more detail than PDMS; in addition, there is a plethora of computational studies of polystyrene described in literature which could be used as precious information source for MD based modeling of the polymer.[67-84] The reasons for choosing strychnine (see Fig. 3.10) as solute molecule are explained throughout the following chapters.

NMR measurements of strychnine with stretched polystyrene in chloroform as alignment medium yielded 19 accurate RDCs which are listed in figure 3.11.[85] The nomenclature of strychnine atoms is illustrated in figure 3.10.

Vector	$^1D_{CH}$ [Hz]	Vector	$^1D_{CH}$ [Hz]	Vector	$^1D_{CH}$ [Hz]
C1-H1	-9.3	**C13-H13**	11.4	**C19-H191**	0.2
C2-H2	0.8	**C14-H14**	-3.6	**C19-H192**	-4.8
C3-H3	---.	**C15-H15**	12.0	**C21-H211**	6.9
C4-H4	-10.5	**C16-H161**	-9.6	**C21-H212**	-5.4
C8-H8	3.6	**C16-H162**	-2.1	**C23-H23**	-6.6
C12-H121	-4.5	**C17-H17**	1.8	**C24-H241**	13.5
C12-H122	7.5	**C18-H181 / 2**	---	**C24-H242**	2.2

Figure 3.11: Table of the 19 measured $^1D_{CH}$ RDCs for strychnine. The given vector names correspond to the labeling as shown in figure 3.10.

3.2.3 Search for an Adequate MD Simulation Setup

The greatest challenge of this project was to find optimal MD conditions for realistically simulating the interactions of polystyrene and strychnine. In order to be able to evaluate the calculation outcomes, three prerequisites must be fulfilled: 1) the results of the MD calculations have to converge (temporal drifts of RDC values must vanish after a certain period of simulation time); 2) the preferred orientations of strychnine (and therefore the simulated RDCs) arise only due to the interaction of the solute with the stretched polymer; 3) the MD derived RDCs must be in good agreement with the measured values (to be more precise, the calculated values must be proportional to the experimental RDCs as the strength of alignment is not simulated; conveniently, the calculated values are called “simulated / virtual“ RDCs).

Regarding point 1, it was tested if simulations of strychnine freely tumbling in a chloroform solution result in virtual RDCs which are averaged out to zero after an acceptable calculation time. It became obvious that 500 ns trajectories with

an explicit treatment of solvent are not sufficiently long enough to provide RDCs which converge to zero; thus, it was decided not to use solvent molecules. Instead, stochastic dynamics (SD; also called Langevin dynamics: LD) were utilized where a friction and a noise term is included in Newton's equation of motions for mimicking collisions of strychnine with the solvent.[86-88] By means of this change in the MD setup, the simulation of strychnine under vacuum SD conditions results in virtual RDCs (19 $^{1}D_{CH}$ couplings) which are averaged to zero after about 200-300 ns of simulation time (see Fig. 3.12).

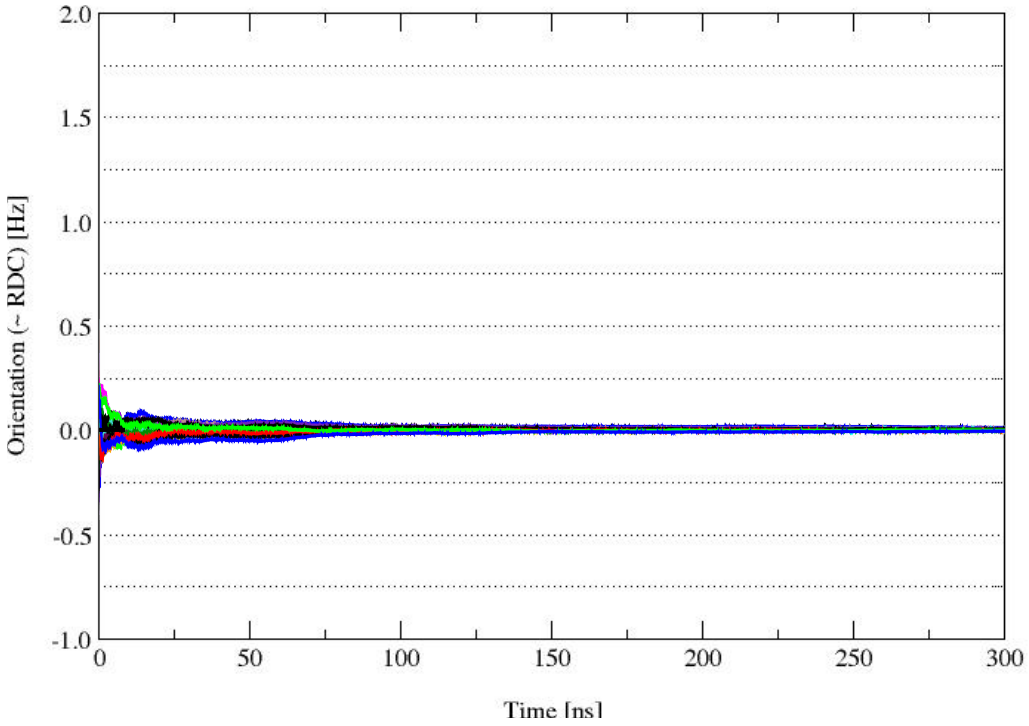


Figure 3.12: Presentation of the MD derived preferred orientations of the 19 used C-H bond vectors of strychnine, expressed as $(3 \cdot \cos^2 \theta - 1)$ dependent values where θ is the angle between the bond vectors and the z-axis of the MD simulation box (see also chapter 1.3.1.3). In the absence of the oriented polystyrene strand, no favored orientations of strychnine C-H vectors are present. After about 200-300 ns, the calculated RDC values converge to zero. If the PS polymer strand is not oriented along the z-axis, the same result is observed (see also chapter 3.2.4).

The next and probably biggest problem was to imitate the characteristics of the stretched polymer/solvent gel. First attempts relied on long PS strands - up to 100 units in order to attribute for differing localized polymer subunit arrangements - which were cross-linked via divinylbenzene and finally oriented along the z-axis of the simulation box (in average over the whole strands). However, it appeared that the calculation time was drastically prolonged due to the long polymer chains until convergence of RDCs is reached; in addition, no setup for adequately treating attractive forces under the used vacuum conditions could be found, resulting in “sticky” PS strands. Based on these observations, only one very short (6-10 units) linear and strictly z-oriented PS chain was designed and further applied. To keep the orientation of the polymer strand along the z-axis, different setups were tested:

1) the PS chain ends where connected across the box walls with concurrent application of periodic boundary conditions (PBC), leading to simulations of an infinite polymer that can transversally move in the x,y-dimension; 2) the complete polystyrene backbone (or at least all PS Cα atoms) were positionally restrained; 3) only the end points (the first and the last atom) of the PS backbone were fixed, thereby allowing oscillating dynamics of the polymer strand. Exhaustive testing of the three setups showed that only the last mentioned approach was successful. Whereas the simulations of an infinite strand produced a lot of calculation errors due to Shake/Lincs algorithm bond length constraining (resulting in return in additional artificial noise), the fully restrained PS backbone lacks of a realistic dynamical behavior of the polymer; as a result, strychnine - once interacting with the chain - is stuck to the polystyrene strand and stays there for the rest of the simulation. To circumvent this artifact, very high temperatures have to be applied for which the used force field entries were not parametrized. Finally, it had to be decided which type of PS optimally reflects the experimentally used polymer: isotactic, syndiotactic, or atactic PS. As no special synthetic procedures were applied for synthesizing the polystyrene/chloroform gel, an atactic arrangement of the aromatic side chains was favored.

The next test simulations exercises the optimal number of strychnine molecules in the simulation box. It was assumed that the utilization of several solutes should increase the accuracy of both ensemble characteristics (as always measured in an NMR experiment) and the MD based orientational sampling (the scanning of the available orientation space in the presence of a z-oriented polymer strand). However, it was found that many strychnine molecules obviously affect the behavior of the short and almost freely oscillating polymer chain under vacuum conditions. Thus, the best results originate from calculations where only one solute molecule is present in the simulation box. Nevertheless, this setup leads to prolonged trajectories (until convergence is reached) compared to systems that rely on simulating ensemble characteristics.

At last, reasonable simulation boundary conditions had to be elected. In principle, two type of systems were tested: 1) the box was provided with long xyz edges and no periodic boundary conditions were applied; 2) the x and y box vector lengths were chosen to be very short (the z-vector length was adjusted to the length of the z-oriented polymer strand) and PBCs were

utilized. Several short MD simulations showed that setup 1 was not suitable for MD system as constructed by the so far yielded test results. Due to the utilization of only one polymer strand and the exclusion of solvent molecules, strychnine - once moved out of the radius where interactions with the polymer are regarded – is not affected any more by the polystyrene strand; this results in an undesired isotropic tumbling of the solute. In addition, contacts of strychnine with the impervious box walls artificially contributes to the final alignment of the molecule. In contrast, the use of short box edges (together with PBCs) enforces the solute to constantly interact with the PS chain, thereby enhancing the sampling of the accessible orientational space (resulting in shorter simulation times).

Figure 3.13: Table of all failed and successful MD simulation setups for an adequate and converged modeling of interactions between z-oriented polystyrene and strychnine that lead to the occurrence of residual dipolar couplings.

Problems	Success
Consideration of solvent molecules	Vacuum stochastic dynamics
Long polystyrene chains	Short polymer strands
Branched polymer strands	Linear, z-oriented polystyrene
Infinite or backbone restrained PS	Only end points fixed PS
Isotactic / syndiotactic polystyrene	Atactic polystyrene
Multiple strychnine molecules	Only one strychnine solute
Long x/y box edge lengths	Short x/y box vectors

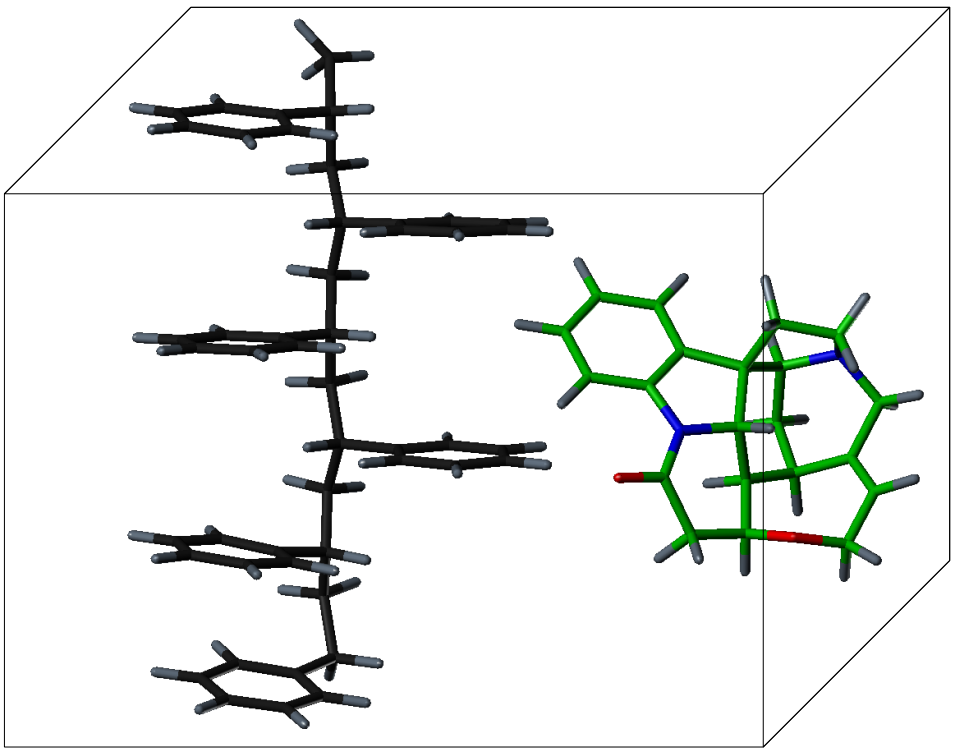


Figure 3.14: Illustration of the final MD simulation test system for the computational prediction of residual dipolar couplings. A short and linear, atactic polystyrene polymer strand with fixed backbone end atoms is oriented along the z-axis of the simulation box. Only one strychnine molecule is placed into the small box which is treated with periodic boundary conditions. Stochastic MD runs were used instead of calculations with explicit solvent molecules.

3.2.4 Computational Prediction of Residual Dipolar Couplings

After having found robust conditions for an MD test system (see Fig. 3.13 and Fig. 3.14), the calculation of “virtual“ RDCs was more or less straightforward. Because of the above described requirements being essential for an adequate modeling of the interactions between the stretched polystyrene strand and strychnine, the simulation parameters (especially the treatment of Coulomb and van der Waals forces) had to be adjusted in a way atypical for e.g. calculations of three-dimensional conformations in solvent containing MD boxes (see section 3.2.8 for more details).[89,90]

Several long Langevin dynamics runs (≥ 500 ns) were conducted and analyzed with respect to the reliability criteria presented at the beginning of section 3.2.3: 1) convergence of simulation results, 2) a solute alignment that originates only from the interaction with the stretched, along the z-axis of the simulation box oriented polymer chain, and 3) an accurate consensus between experimental and virtual outcomes. As already observed for strychnine freely tumbling in the stochastic MD bath, the virtual RDCs converge after ca. 200-300 ns; however, in contrast to the calculations without the z-oriented polystyrene chain, the 19 simulated C-H vectors of strychnine are not averaged to zero but exhibit preferred orientations (see Fig. 3.15; compare the convergence and alignment of virtual RDCs with those shown in figure 3.12).

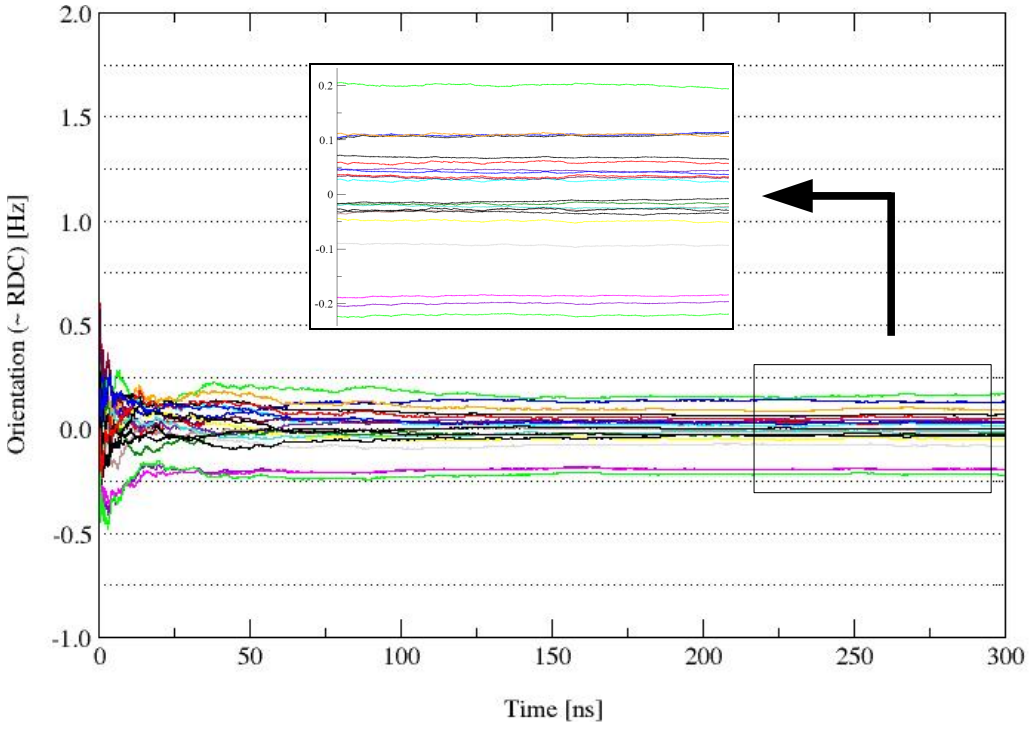


Figure 3.15: Presentation of the MD derived preferred orientations of the 19 used C-H bond vectors of strychnine, expressed as $(3 \cdot \cos^2 \theta - 1)$ dependent values where θ is the angle between the bond vectors and the z-axis of the MD simulation box (see also chapter 1.3.1.3). Due to presence of the z-oriented polymer chain, the virtual RDCs are not averaged to zero but converge to differing values.

To evaluate if the origin of the observed alignment exclusively arose from interactions between the z-oriented polymer and the solute, polystyrene was positionally fixed at one backbone atom in the middle of the short strand; consequently, the PS chain could not translationally move but freely rotate in all three dimensions ("no z-orientation"). As expected, all virtual RDCs are averaged out to zero after ca. 300 ns. This outcome confirms one of the prerequisites: virtual RDCs only occur when the polymer is oriented along the z-axis of the box, and there is no other source that contributes to the preferred orientation of strychnine.

Finally, the calculated RDCs are compared with the experimental results (RDCs for strychnine in a polystyrene/chloroform gel);[91] in addition, the 19 $^{1}D_{CH}$ RDCs of strychnine were estimated with the aid of the so- called rod model as implemented in the steric prediction algorithm of PALES.[63] The outcomes of both approaches, the MD based and the PALES rod based predictions, are shown in figure 3.16.

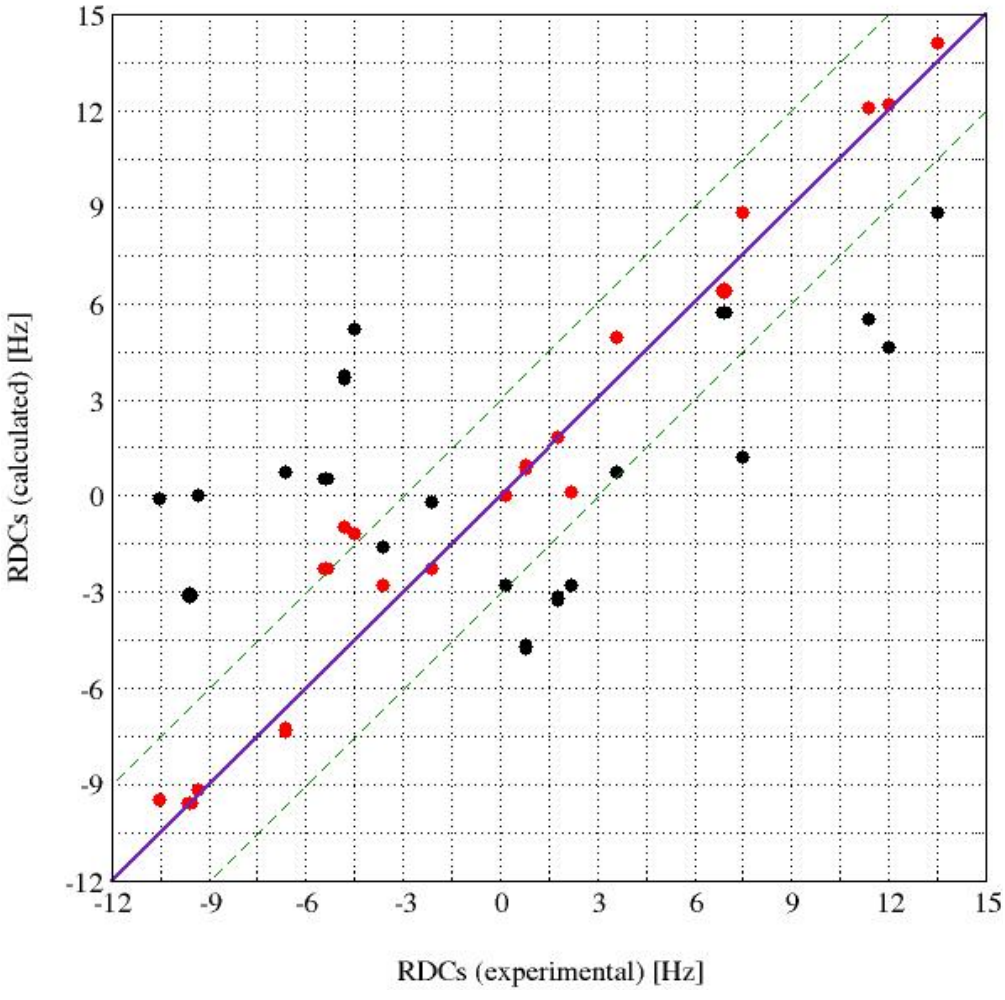


Figure 3.16: Illustration of a graphical comparison of PALES (black) and MD derived (red) virtual RDCs for strychnine in a polystyrene based alignment medium with experimental values. Whereas the violet diagonal represents a perfect consensus, the dashed green lines attribute to a deviation of 3 Hz. It is obvious that the MD technique is superior to the steric PALES approach.

MD based RDC values were obtained by converting the preferred C-H bond vector orientations via equation 1.8 (see section 1.3.1.3) plus a constant factor to correct effects of different alignment strengths.

Detailed values of virtual RDCs from the PALES and the MD approach are deposited in the appendix.

In order to give an estimation of the quality of both approaches, the correlation factors R^2 (while 0 means absence of a systematic relationship, 1 corresponds to a perfect systematic relationship of data; thus, R^2 is is a measure of how well the linear regression lines represent the data) were compared; whereas the PALES algorithm results in an R^2 value of 0.51, the MD driven method has a correlation factor of 0.98. Thus, both the plot shown in figure 3.16 and the statistical analysis reveal that molecular dynamics simulations are capable to predict residual dipolar couplings via the utilization of a very simple alignment model. By assigning the strongest deviations between measured and predicted RDCs to strychnine bond vectors, it becomes obvious that the three virtual RDCs which display a variance higher than 3 Hz all belong to CH_2-groups (C12-H121, C19-H192, C21-H212). If this observation arises by chance or could be attributed to inaccuracies in methylene force field parameters or to experimental errors (strong scalar couplings which often occur between CH_2 group protons may disturb the splitting of signals) is not clear.[53] However, it seems that a systematical problem is present as the utilization of other solute molecules (norcamphor and menthol) which contain a lot of CH_2 groups suffer from similar drawbacks. The problems regarding an accurate modeling of methylene groups was therefore the main reason why norcamphor and menthol were discarded during the search of an robust MD test system (see section 3.2.2).

3.2.5 Determination of the Origin of the Alignment

As it was mentioned in the introductory section of this chapter, a precise MD based prediction of RDCs also provides insight in the interactions between the alignment medium (here: the stretched polymer strand) and the solute (here: strychnine) that result in the occurrence of preferred compound orientations. As thousands of different orientations of strychnine contribute to the final alignment, all individual time frames (≥ 2,500,000) written out during the MD trajectory had to be examined which seemed not to be meaningful. However, it is easy to elucidate those regions in the simulation box where the probability of finding the strychnine molecule during the trajectory is high. With the aid of this approach, two- or three-dimensional density maps can be calculated which show the preferred areas for strychnine to stay. By correlating the structure of the polystyrene chain with the probability of presence of the solute, the origin of the alignment could be determined (see Fig. 3.17).

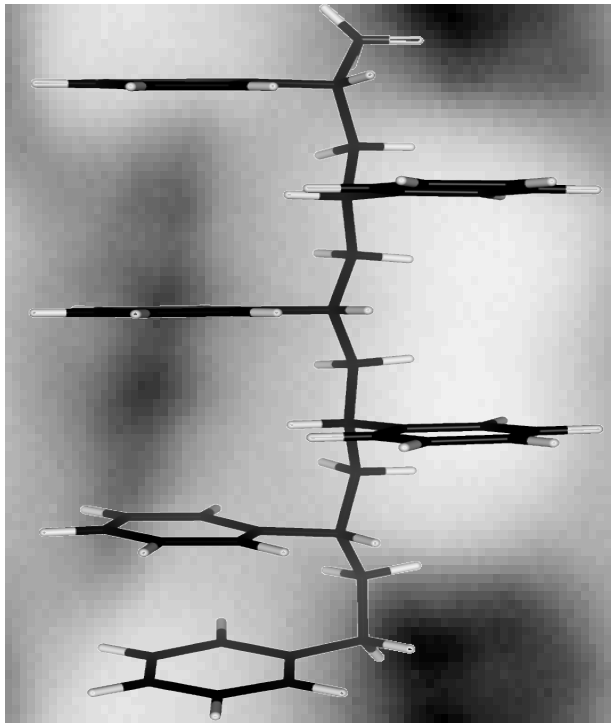


Figure 3.17: Illustration of an overlay of the two-dimensional density map of strychnine and the polystyrene strand. The image displays the yz plane as it is visible from the x direction; the black-white gradient indicates the probability of presence of strychnine in the molecular dynamics simulation box (during the 500 ns long trajectory).

White regions correspond to zero and black areas to a high density (probability).

An analysis of figure 3.17 reveals that strychnine particularly interacts with polystyrene at two sites of the polymer: the most frequently “visited“ area is present at the top and lower ends of the PS backbone (the atactic part of the polymer), thereby being localized between two PS aromatic groups that are separated by three side chains being oriented to the opposite side (plus the small space between the PS chain ends); in addition, strychnine often stays close to the polystyrene side chains in the isotactic middle part of the polymer, with its benzene ring being perpendicularly oriented to the aromatic side chains. The latter described interaction can be assumed to be “realistic“ as it is known that aromatic systems sometimes exhibit so-called T-stackings. In contrast, the most often observed intermolecular interaction seems to be questionable since the MD setup of the polymer does not (necessarily) reflect macroscopic characteristics of the PS/$CHCl_3$ alignment medium as present in an NMR tube: first, polymer chain ends are expected to only rarely exist in a PS gel; moreover, the positional fixing of the polymer end atoms is artificial. To evaluate the reliability of the observed favored interaction, a robust system had to be designed with connected polystyrene chain ends - construction of an fully reliable infinite polymer - which was not possible with the available software. However, the density map provides first hints that the PS side chain orientation pattern and therefore the solute accessibility of the polymer backbone might play an important role in the evolution of the molecular alignment.

3.2.6 Weaknesses of the MD Simulation Approach

Many different MD setups were tested in order to optimize the simulation conditions for adequately treating intermolecular interactions between polystyrene and strychnine. By the help of the final test system, the obtained RDCs are in very good agreement with the experimental ones. However, the setup reveals several weaknesses. Slight variations of the boundary conditions, e.g. enlarging of box edges or utilization of plain interaction cutoffs result in virtual RDC values which display no or only a poor correlation with the measured couplings. Moreover, calculated RDCs for other solutes (norcamphor and menthol) than strychnine show a by far not as good consensus with experimental data. One of the reasons of this observation - the presence of multiple methylene groups which revealed in all simulations higher deviations - was already mentioned. Further explanation attempts rely on an inaccurate force field parametrization of the solutes or, which seems to be much more problematic, that the MD setup (the construction) of the polymer strand is not robust (representative) enough to cover also interactions being highly relevant for the alignment of norcamphor or menthol (and thus in return being not important for the polystyrene-strychnine system). Regarding this, the current investigations in particular deal with the construction of more robust PS models (e.g. infinite strands, alternative side chain orientations). In addition, possible effects of solvent molecules on favored solute orientations were neglected in the presented model.

3.2.7 Conclusion

To summarize, the design of a very simple MD model of the alignment medium polystyrene/chloroform allowed for an accurate prediction of $^{1}D_{CH}$ RDCs for the natural product strychnine. In detail, one short atactic, z-oriented PS chain with fixed backbone end atoms served as mimic of a stretched polymer gel. By the use of Langevin dynamics, MD simulations of sufficiently long duration could be performed; this enabled the calculation of converged virtual RDC values which display a better consensus with experimentally derived RDCs compared to the commonly used steric PALES prediction approach. The simulated interactions showed that the alignment of strychnine arises due to different orientations of PS side chains. As the created test system lacks adequate robustness, further optimizations are necessary until the new technique could be applied as a reliable standard tool for predicting RDCs.

3.2.8 Reagents, Methods, and Experiments

Detailed informations about the preparation of polystyrene gels, the soaking of strychnine, and the measurement of RDCs can be found in literature.[91]

The MD calculations were carried out with the Gromacs 3.3.2 and 3.3.3 simulation packages.[42] In all simulations, the OPLS all-atom force field was used.[89,90] The lengths of all strychnine and polystyrene bonds were constrained with a tolerance of 0.001 Å by the Shake algorithm. The temperature was kept constant at 300 K by means of a weak coupling scheme or with the aid of a stochastic bath. No center-of-mass motion removal was applied. Newton's equation of motion were integrated every 2 fs. A pair-list was constructed and updated every 10 integration steps. For dispersive van der Waals interactions, an energy correction term was utilized. Different MD setups were tested with respect to convergence and reliability of virtually obtained RDCs. For the simulations of boxes filled with chloroform solvent molecules, a short-range cutoff of 0.9 nm and a long-range cut-off of 1.1 nm together with the PME scheme for treating long-range electrostatics was used.[92] In all vacuum calculations, a stochastic leap-frog integrator inclusive of varying friction coefficients was applied.[86-88] The cut-offs were adjusted to the lengths of the box vectors; the best MD run was performed with the following setup: the cutoff for the pair-list was 0.8 nm; Coulomb and Lennard-Jones interactions were regularly treated until 0.6 nm and slowly switched off between 0.6 and 0.7 nm via an interaction shift function; electrostatic contributions to the intermolecular forces behind the cutoffs were neglected; all three box vectors were 1.65 nm long. Periodic boundary conditions were switched on for all dimensions. For the six styrene units containing polymer chains, force field parameters were used as described already cited above. Strychnine (and the other tested solutes) was parametrized by adaption of force field data of similar chemical and structural motifs as found in amino acid based molecules.

3.2.9 Declaration

The presented topic was conducted in collaboration with PD. Dr. Burkhard Luy (Technical University Munich, Department Chemistry). B.L. measured all residual dipolar couplings and programmed the molecular dynamics RDC analysis algorithm. The described results have not yet been published as optimization of the test system is ongoing.

3.3 Relating Molecular Characteristics to the Cellular Uptake of Cyclic Peptides

The use of peptides in medicinal research provides a lot of advantages compared to organic molecules of artificial origin (see the Frost & Sullivan study in *Chem. Eng. News*, **2006**, *83*, 17-24). Considering pharmacotoxicity, enzymatic degradation of peptidic drugs results in biogenic molecules which are either excreted or further used in metabolism. From a chemical point of view, it is relatively simple to create large libraries of peptides by automated synthesis. Modifications of active ligands can easily be introduced, e.g. when selectivity is lacking, because peptide chemistry is well advanced (there are several journals which exclusively deal with peptides). If cellular protein-protein interactions are known in atomistic detail, analogies to peptides rapidly provide conceptions to create first hits that can further be developed. Since peptides consist of 20 natural building blocks that could be (almost) stochastically combined, even small compounds (connected via regular peptide bonds) exhibit a great conformational variety that seems to be inexhaustible. Despite this high potential, peptides are often considered as poor drugs because they are usually cleaved by enzymes *in vivo* within a short time. Additionally, the active or passive transport from gut into blood is normally very restrictive. Hence, in contrast to the rapidly increasing number of peptides with interesting biological properties, their application as promising drug candidates in pharmacological research is still limited.[93-96]

3.3.1 The Scientific Question

Very recently, it was shown that an *N*-methylated derivate of the cyclic Veber-Hirschmann hexapeptide - a highly active but orally not available somatostatin agonist - exhibits not only a comparable biological affinity to its target (receptor subtype 2) but also a paracellular uptake of ca. 10 % (and extremely high metabolic stability in rat).[97] As Veber et al. have pointed out that increased numbers of rotatable bonds inside a molecule have an adverse effect on its oral bioavailability,[98] the conformational stability of the compound was investigated via NMR spectroscopy and MD simulations. The studies proved that the described ligand is rather rigid (one distinct conformation is strongly preferred) whereas all amide groups exposing their protons to the solvent surroundings in the parent compound are *N*-methylated.[97] Due to this outcome, the question arose if membrane permeability (through tight-junctions) can be

quantitatively correlated to the extent of molecular dynamics and / or structural features like the solvent accessibility of H-bond donors and acceptors.

Thus, it was investigated if structural and dynamical characteristics of *N*-methylated cyclic peptides can be correlated to the propensity for passing cellular membranes through tight-junctions.

For this purpose, a library of cyclic pentapeptides with different numbers and pattern of *N*-methylations was constructed and subsequently tested for the cellular permeability of its members.[99-102] Then, molecular dynamics simulations of the most interesting candidates were performed and analyzed with respect to the following characteristics: intrinsic flexibilities, solvent accessible H-bond donors and acceptors, radii of gyration, and polar / apolar surfaces. Finally, it was evaluated if the outcomes of the MD calculations can be correlated to the experimentally yielded cell passing rates.

3.3.2 Studying the Cellular Transport of Test Set Molecules

The main purpose for the design of *N*-methylated, “alanine-only“ template molecules was to investigate structural and dynamical requirements of cyclic pentapeptide backbones being important for oral availability, thereby excluding additional influences of amino acid side chain groups.[100,103,104] In total, 31 peptides (out of 32 possible compounds; the permethylated peptide could not be obtained) were synthesized of which 13 molecules were interesting candidates for template based structural design (the main conformation is highly populated at the NMR time scale): seven of the 13 peptides were conformationally homogeneous (one preferred conformation ranging from 98 - 100 % at the NMR time scale is found; they are named P01 to P07) and six were conformationally inhomogeneous (peptides resulting in two NMR signal sets but with one preferred conformation ranging from 80 - 90 %; they are named S08 to S13).[99,102] Subsequently, the so-called Caco-2 *in vitro* model was employed which is commonly used for screening intestinal drug permeability (the peptides permeate the membrane via paracellular pathways, i.e. the transport is not triggered by proteins, as they lack transport in PAMPA based assays; one exception is P07 that exhibits some permeability across the PAMPA, owing to its enhanced lipophilic profile).[105,106] The biological *in vivo* studies exhibited partially very different permeabilities of the individual

library members. As a general tendency, the conformationally homogeneous peptides reveal a better transport than the inhomogeneous ones; the only exception is the threefold methylated peptide S13 that shows a surprisingly high membrane passing propensity although existing in a mixture of three conformers (85 / 10 / 5 %). Due to this result, it was decided to concentrate on the seven conformationally homogeneous templates and the inhomogeneous peptide S13 as the permeability of the remaining cyclic penta-alanines were too low to be of interest (see Fig. 3.18a/b).

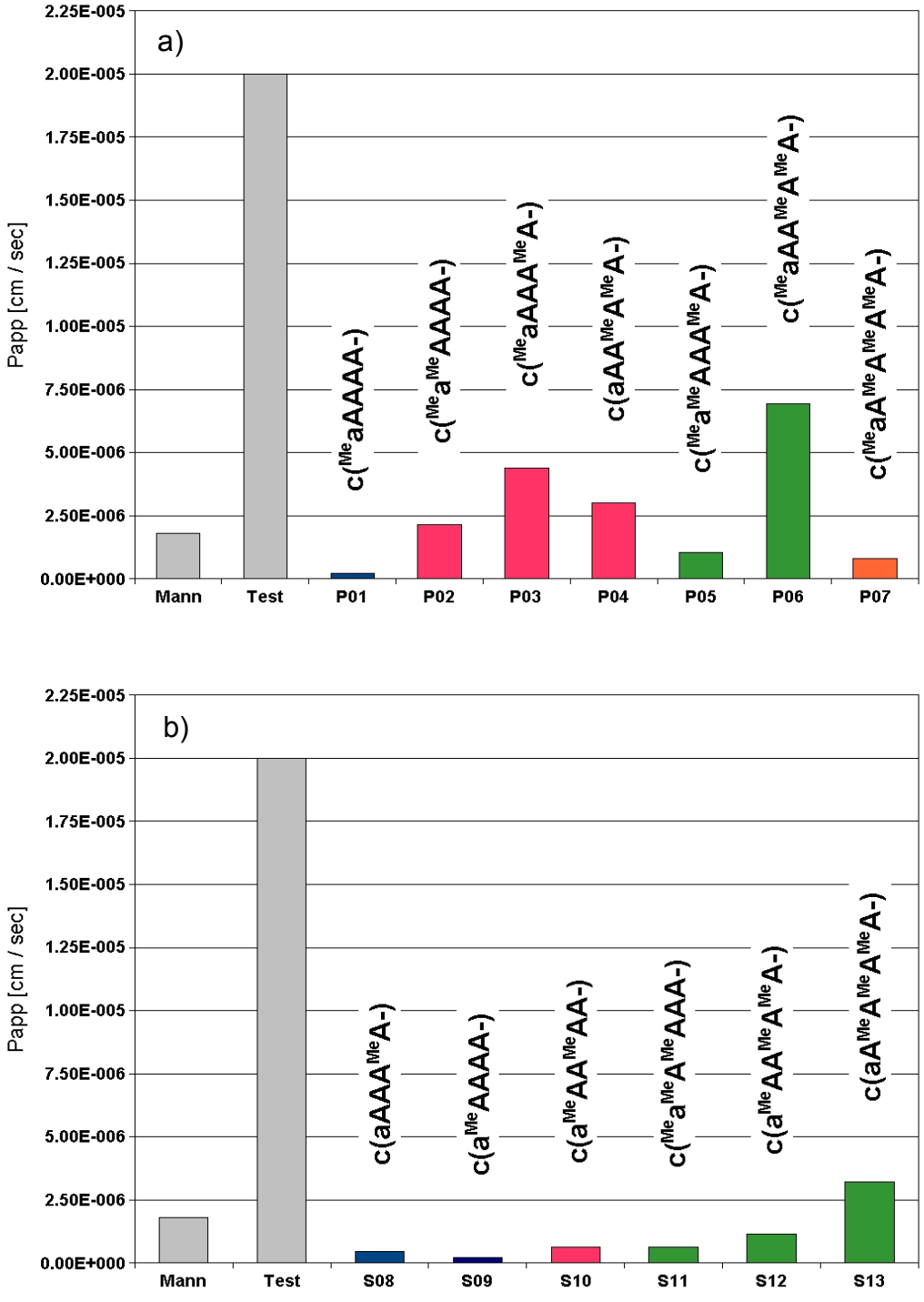


Figure 3.18: Presentation of Caco-2 based membrane permeabilities (Papp permeability coefficient given in cm/sec) of seven conformationally homogeneous (a) and six inhomogeneous (b) cyclic penta-alanines carrying different numbers and patterns of *N*-methyl groups.

Two molecules (gray) were included into the studies as reference compounds: the sugar mannitol (Mann), a common marker for paracellular permeability and the steroid hormone testosterone (Test), a marker for transcellular permeability.

The color code was used as follows: 1 (blue), 2 (red), 3 (green), and 4 (orange) *N*-methylations.

Peptide denotation: "Me" stands for an *N*-methylated backbone amide group; a small "a" stands for an D-Ala, an "A" for an L-Ala.

All Caco-2 permeability coefficients and conformer populations are explicitly given in the appendix of this work.

Among the conformationally homogeneous peptides, the mono-*N*-methylated peptide P01 has the lowest permeability. The permeability is drastically enhanced by a subsequent second *N*-methylation of the Ala1-Ala2 peptide bond giving rise to P02. P03, obtained by exchanging the site of *N*-methylation from Ala1-Ala2 to Ala4-Ala5, has a further enhanced permeability than P02. P04, the only peptide in the series of homogeneous peptides which is not methylated at its D-residue, displays a slightly decreased permeability compared with P03. The tri-*N*-methylated peptide P05 which is basically an extension of the *N*-methylation of either P02 or P03 has a strongly reduced permeability than either of these two. The most surprising result is comes for P06 which is thrice as much permeable than the standard compound mannitol. S13, the only inhomogeneous peptide, has a comparable permeability to the twofold *N*-methylated P04. Finally, coming to the tetra-*N*-methylated peptide P07, there is a strong decrease in the Caco-2 permeability.

The obtained results make clear that the measured permeability rates can not be directly correlated with the number of *N*-methyl groups. As the influence of side chain functionalities was excluded, the reason for this unexpected behavior must be based on different structures and dynamics of the peptide backbone or exposure of free NH groups. Consequently, conformationally flexibilities of the cyclic peptides were investigated via MD simulations in order to determine the origin of the observed membrane permeability variances.

3.3.3 Elucidation of Molecular Flexibilities

Molecular dynamics of (cyclic) peptides can be referred to different kind of flexibilities like angle bending, torsional degrees of freedom, or overall conformational stability. In addition, their extent can be quantified (especially via MD simulations) in several ways. For example, if one is interested in the dynamics of dihedral angles, Ramachandran plots for all structures arising during an MD trajectory can be created and subsequently analyzed with respect to the spreading of data points. In order to evaluate the total conformational flexibilities, usually RMS deviations are calculated;[31] here, the variance of each computed structure from the starting or the most representative conformation is quantified after a least-square fitting procedure.

For the present study, it was decided to quantify all peptide dynamics in three different ways: 1) all conformations occurring during the MD trajectories which exhibit mutual RMS deviations smaller than a certain cutoff were clustered into

the same structure families; the higher the number of clusters, the more different conformations are present, and thus the higher is the extend of molecular dynamics; 2) the RMS fluctuations of all backbone atoms (N, Cα, C) around their lowest potential energy positions taking place during the MD runs were recorded (see e.g. Fig. 3.19); 3) the number of 180° flips of peptide bonds around adjacent φ and ψ dihedral angles were monitored.

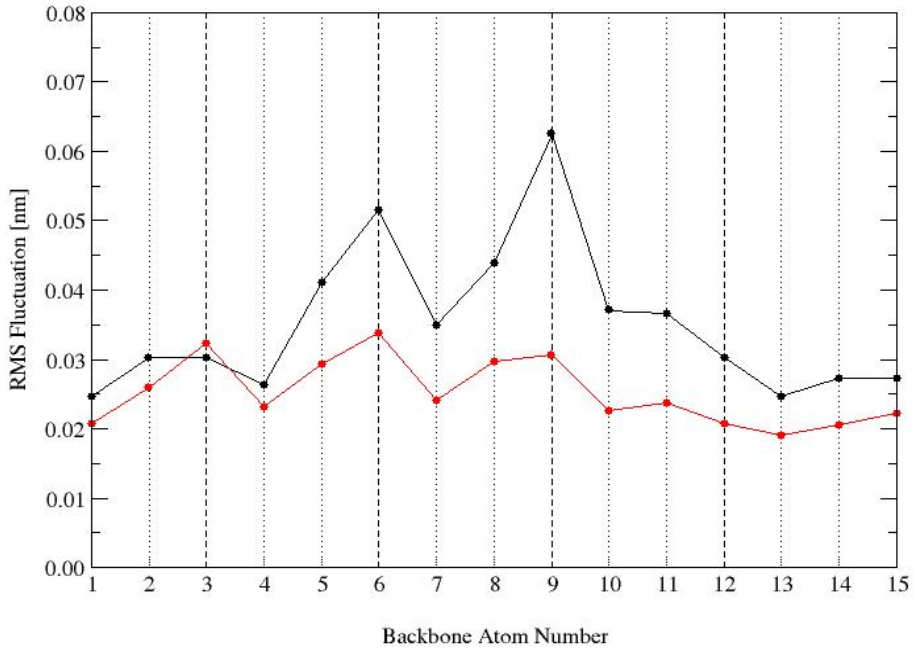


Figure 3.19: Presentation of an overlay of P03 (black) and P04 (red) backbone root mean square (RMS) fluctuations. The numbers 1, 4, 7, etc refer to nitrogen atoms, 2, 5, 8, etc to Cα atoms, and 3, 6, 9, etc. to carbonyl carbon atoms. All backbone atoms of P04 except the CO carbon of residue 1 less fluctuate than those of P03. The strong dynamics of atoms number 6 and 9 correspond to 180° flips of peptide bonds around adjacent φ and ψ dihedrals.

The MD values obtained for the three utilized categories describing different types of molecular dynamics are listed (physical units and the meaning of the abbreviations are given in the figure legend) for the eight investigated cyclic penta-alanines in figure 3.20. It should additionally be noted that all simulations were conducted at two different temperatures: 300 K (values are given in figure 3.20) and 340 K (enhancing the conformational sampling).[31]

MD 300 K	P01	P02	P03	P04	P05	P06	S13	P07
Cluster	12	18	8	4	10	6	2	1
RMSF	0.50	0.54	0.36	0.32	0.42	0.35	0.25	0.21
Flip-51	0	0	0	0	0	0	0	0
Flip-12	0	0	0	0	0	0	0	0
Flip-23	6	> 20	2	0	> 20	(1)	0	0
Flip-34	3	6	1	0	6	0	0	0
Flip-45	0	6	0	0	0	0	0	0

Figure 3.20: Table of outcomes from long MD calculations of the preselected cyclic penta-alanines. The three chosen characteristics describe the dynamical behavior of the peptides.

Cluster: number of 3D structure families; RMSF: RMS fluctuations of backbone atoms [nm]; Flip: number of 180° peptide bond flips around adjacent φ and ψ dihedral angles.

Regarding the number of clustered structure families, it becomes obvious that P01 and P02 adopt the most varying conformations whereas S13 and P07 display only two and one, respectively, 3D arrangements. Peptides P03 and P06 which permeate the Caco-2 membranes at best have structure clusters of medium size. The increased temperature runs result in the very similar cluster values except for P01 (23 structure families) and P03 (16). An analysis of the backbone RMS fluctuations yields no additional informations as the obtained values correlate with the number of conformational clusters. Usually, 180° flips of amide bonds contribute at most to the overall dynamics of cyclic pentapeptides.[99,107,108] The evaluation of backbone dihedral angles reveals that P04, S13, and P07 exhibit no peptide bond flips during the trajectories, both at 300 K and at 340 K. For P06, one 180° flip is observed which, however, occurs only once (no bidirectional flip as it is observed for all other peptide bonds); thus, a slow conformational exchange between both orientations can be assumed. P03 shows rare flips of two peptide bonds. The highest torsional dynamics are present in P01, P02, and P05; whereas two more or less frequent flips are detected for P01, the peptide bonds between Ala2 and Ala3 of P02 and P05 (both peptides are *N*-methylated at Ala1 and Ala2) change their orientation more than 20 times. Aside, it is interesting that the peptide bonds neighboring the D-Ala residue never flip during the MD runs.

3.3.4 Examination of Structural Features

In principle, there are a lot of structural characteristics which could affect the paracellular transport of the cyclic peptides. Common parameters which are tested are the solvent accessibilities of H-bond donors and acceptors, the spatial expansion of molecules, the hydrophilic / hydrophobic character of the peptide surface, and the presence of intramolecular hydrogen bonds or salt bridges.[98] In addition, moments of inertia, the order and distribution of solvent molecules around the solute, or RMS deviations between different conformations might be of particular interest.

The analysis of the MD runs concentrated on three structural characteristics of the eight pentapeptides: 1) the average radius of gyration of the cyclic peptide backbones was calculated; with the aid of these parameters, it could be studied how voluminous the distinct peptides are and if there is a correlation between bulkiness and permeability rates; 2) the overall hydrophobic surfaces of the cyclic peptides was computed and also averaged over the simulation

times; this feature is of interest since the polarity of molecular surfaces plays often an important role for transport processes;[98] 3) the average solvent accessibility of amide protons was quantified (in terms of radial distribution functions; RDFs) as intermolecular hydrogen bonds may play an important role in the efficiency of membrane passing (see e.g. 3.21).[98]

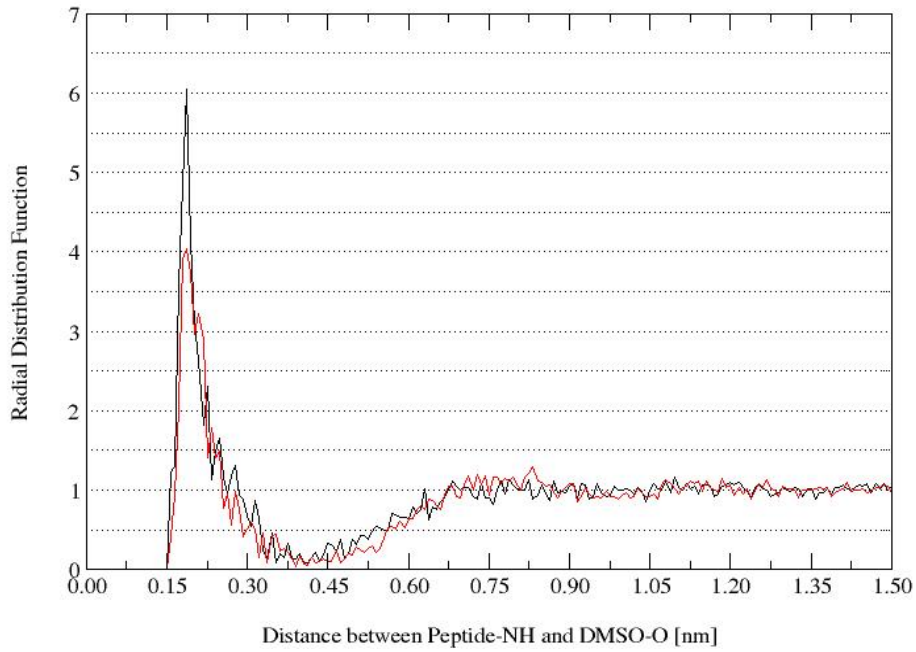


Figure 3.21: Illustration of an overlay of distances between peptide amide bond protons and DMSO oxygens in terms of radial distribution functions. RDFs are a measure of solvent accessibilities of hydrogen bond donors or acceptors. The overlay shows that the amide proton H3 of P06 (black) is more exposed to the solvent than H3 of P07 (red). The found distance between the peptide hydrogens and the DMSO oxygen acceptors lies in the typical range between 1.6 and 2.0 Å.

The MD data evaluated for the three used structural features representing electronic and steric characteristics are listed in figure 3.22 (physical units and the meaning of the abbreviations are given in the figure legend) for the eight studied pentapeptides. As already mentioned above, all simulations were conducted at two different temperatures: 300 K (values are given in figure 3.22) and 340 K (enhancing the conformational sampling).

MD 300 K	P01	P02	P03	P04	P05	P06	S13	P07
Gyration	0.290	0.292	0.293	0.287	0.292	0.287	0.282	0.285
H'phoSurf	3.45	3.70	3.75	3.74	3.98	3.90	3.98	4.17
RDF-H1	---	---	---	6.9	---	---	7.6	---
RDF-H2	5.9	---	5.3	4.2	---	6.0	4.9	4.0
RDF-H3	6.9	7.0	7.5	8.1	5.2	7.8	---	---
RDF-H4	5.0	4.7	5.0	---	3.7	---	---	---
RDF-H5	3.1	4.5	---	---	---	---	---	---

Figure 3.22: Table of outcomes from long MD calculations of the preselected cyclic penta-alanines. The three chosen characteristics qualify the conformational features of the peptides.

Gyration: molecule backbone radius of gyration [nm]; H'phoSurf: overall hydrophobic surfaces of the peptides [nm^2]; RDF: measure for the extend of the solvent accessibility of amide bond hydrogens.

The radii of gyration were extracted with respect to the spatial expansion of the peptide backbones, thus the *N*-methyl groups and the side chains were not regarded. For P01-P06, the computed values are very similar; only P07 and especially S13 exhibit smaller backbone volumes than the other peptides; it seems that an increasing number of *N*-methylations decreases the expansion of the peptide backbones. The hydrophobic character of the pentapeptide surfaces was computed regarding the complete molecules; thus, it is only meaningful to compare peptides carrying the same number of *N*-methyl groups. Whereas P03 and P04 exhibit almost the same hydrophobicities, their analog compound has an decreased apolar surface (all three peptides have two *N*-methylations). A comparison of P05, P06, and S13 (all three molecules carry three *N*-methyl groups) reveals that the surface of the best permeable peptide P06 is less hydrophobic than the others. The solvent accessibilities of the amide protons of the eight peptides are very different (here, RDF values higher than ca. 6-7 correspond to a good accessibility): whereas P02 (H3), P03 (H3), and S13 (H1) have only one NH proton being solvent accessible, two amide protons of P01 (H2 and H3), P04 (H1 and H3), and P06 (H2 and H3) are exposed to the DMSO oxygens. The only amide proton of P07 is not accessible for the solvent. The simulations at 340 K show very similar results as the discussed calculations at 300 K; hence, their outcomes are not explicitly presented.

3.3.5 Correlating Dynamics and Structures with Measured Permeabilities

With the outcomes of the MD studies in hand, it was tried to find possible correlations between the Caco-2 absorption rates and characteristics of the cyclic template peptides. Whereas RMSF and radii of gyration values show no clear relation to the bioassay results, the number of structure clusters and peptide bond flips as well as the radial distribution functions indicate some correlations with the extends of membrane passing; regarding this, it seems that a good paracellular transport can be expected when the peptide is not too flexible and has at least one amide bond proton being accessible to its surroundings while its spatial orientation remains stable over a longer time period. In the following, this observation is discussed by means of the distinct peptides (Caco-2 permeability rankings are given in parenthesis, MD results in brackets).

P06 (1st): There are two NH protons (H2/H3) having high DMSO accessibilities [6.0/7.8]; whereas H3 changes its orientation only once (slow conformational exchange), H2 is fully stable. The peptide exhibits only six structure families.

P03 (2nd): H3 is strongly exposed to its surroundings [7.5] and only rarely changes its orientation [2]; H2 becomes also solvent accessible [6.0] at 340 K (enhanced sampling) and does not flip at all. Dynamics are slight as many of the eight structure clusters result from rare flips of two amide bonds.

S13 (3rd): The H1 proton shows a very good solvent accessibility [7.6] and never changes its orientation; in addition, the peptide exhibits only two structure families. As S13 displays two different conformations on the NMR time scale and only the major conformer was investigated, it can be assumed that the minor conformation decreases the overall permeability.

P04 (4th): The peptide has two protons (H1/H3) that display a very good exposure to the solvent [6.9/8.1]; in addition, the corresponding peptide bonds never flip. There are only four structure clusters present. However, the enhanced sampling run at *340 K* shows that both protons display (in part drastically) decreased DMSO accessibilities [6.5/5.9]. This could be a reason why P04 has a lower permeability rate than peptide P03.

P02 (5th): H3 is the only proton that has a good solvent accessibility [7.5]; however, it changes its orientations more than 20 times during the trajectory. The other two amide protons show at both temperatures very poor DMSO exposures. In addition, 18 conformational clusters arise in the MD run. An explanation could not be found why the peptide shows nevertheless a modest membrane absorption. It could be speculated that the high frequency of peptide bond flipping does not affect binding to tight junction membrane proteins (see Fig 3.23).

P05 (6th): The peptide has one amide proton (H3) which has a moderate solvent exposure [5.2]; as it is the case for P02, the corresponding peptide bond flips very often (> 20 times). Possible explanations for the poor but still detectable cellular passing might be the partially restricted dynamics as only 10 structure families were found.

P07 (7th): The only existing amide proton (H2) of P07 is barely solvent accessible [4.0] and does not flip at all, resulting in only one structure family; thus, a more favorable orientation of the proton is not possible which might be the reason for the slight membrane passing rate.

P01 (8th): Two (H2/H3) of the four amide protons show a good solvent accessibility [5.9/6.9] but frequently change their orientation (6/3 times); at 340 K, the DMSO exposure of H2 is clearly diminished [5.0]. This result and the moderate / high number of conformational clusters (12 at 300 K, 23 at 340 K) might be the explain the very poor membrane absorption of the peptide.

In general, the solvent accessibilities of amide protons and their dynamics might be the most critical point (regarding the here studied template peptides) for a good paracellular permeability via passing tight-junctions (TJ; see Fig. 3.23). TJs are areas between the membranes of two vertebrate epithelia cells that, amongst other functions, control the particle flow through the paracellular space (diffusion barrier). They are flanked by several protein complexes like claudines, occludines, or junctional adhesion molecules.[109,110] Although the role of these proteins in transporting molecules is not fully understood, it is assumed that in particular claudines create the cellular barrier of tight-junctions.[110] Thus, the presence or absence (e.g. solvent accessible H-bond donors), the type (e.g. H-bonds), and the strength of intermolecular interactions between solvated molecules and claudines might be responsible for the rate of uptake. In general, the obtained observations are in agreement with Veber`s rules which correlate the number of rotatable bonds (which could be related to the discussed peptide bond flips) and the overall polar surface (that in particular corresponds to the number and orientation of solvent accessible H-bond donors and acceptors) with oral availability.[98]

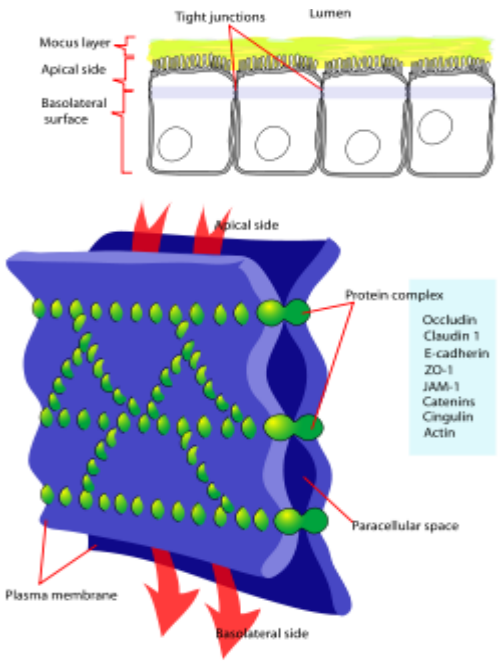


Figure 3.23: Schematic presentation of tight junctions that control the paracellular transport of particles across the intestinal vertebrate epithelial cell barrier. The absorption of molecules is in particular regulated by protein complexes like junctional adhesion molecules (JAMs), claudines, and occludines. (*taken from the WWW: http://www.wikipedia.org/wiki/Tight_junctions*)

3.3.6 Possible Ways to Improve the MD Calculations

As written above, only a potential correlation between molecular dynamics and structural characteristic on one side and the extent of membrane passing of the studied peptides could be found. However, the findings are not robust enough to postulate a general correlation model. To improve the quality of the investigations, several things can be done: 1) the 3D conformations of the eight cyclic peptides as present in water (not DMSO) should be recalculated since Caco-2 tests were also performed in an aqueous environment; 2) by means of the new structures, all MD simulations should be carried out in water as conformational and dynamical equilibria, e.g. of peptide bond flips, might deviate from those obtained for DMSO based calculations; in addition, the yielded values for the hydrophobicity of surfaces or the radii of gyration might be different in the more polar and protic solvent H_2O; 3) the MD calculations should be prolonged and carefully evaluated with respect to an exhaustive sampling of the accessible conformational space (compare with the procedure described in chapter 3.1.4); 4) all bioassay data should be checked for their reliability; as explained in section 1.2.4.4, experimental data are sometimes not sufficiently precise enough to find correlations with results of MD runs.

3.3.7 Conclusion

Experimental investigations of cyclic penta-alanines with varying numbers and patterns of *N*-methyl groups revealed that distinct molecules have different membrane permeability rates. Thus, MD simulations of eight (out of 13 tested) selected compounds were analyzed with respect to conformational flexibilities and structural characteristics. An unambiguous, general correlation of these parameters with the extend of cellular permeabilities could not be elucidated; however, the comparison of calculated and experimental data showed that there might be a relation between the solvent accessibilities of amide protons, their flexibilities, and the membrane passing rates; it seems that the most important prerequisites of the studied cyclic peptides for showing an adequate absorption is (at least) one NH proton that is highly exposed to its surroundings - therefore being able to exhibit a hydrogen bond to cellular proteins - and does not significantly change its orientation over a certain period of time. (The here found importance of intermolecular hydrogen bonds is, however, not in agreement with previously published studies that relate solvent shielding of amide hydrogens with good cellular absorption.[111,112]) As not all studied

peptides showed the discussed kind of relation, further examination under optimized conditions must be performed in order to be able to postulate a generally valid correlation rule.

3.3.8 Reagents, Methods, and Experiments

All informations regarding used reagents, peptide synthesis, and HPLC purification can be found in literature.[99,102] To prove if the further used structures (elucidated in DMSO) exhibit the same conformational preferences in water (the solvent which was used in biological testings), the chosen eight pentapeptides were re-synthesized and studied via NMR spectroscopy on a Bruker 600 MHz spectrometer, equipped with a QXI probe head. ^{1}H-1D and several two-dimensional spectra (see section 3.1.7 for more information) were recorded in water plus 5 % D_2O.

All molecular dynamics simulations were carried out with the GROMACS 3.3.3 software package.[42] The parametrization of peptides were performed via the GROMOS 53a6 force field.[45,113] A rigid four-point model of DMSO was applied as solvent.[114] The already published structures of the eight cyclic penta-peptides served as starting conformations for the following MD runs. All peptides were energy minimized *in vacuo* and placed in truncated octahedral boxes with a minimum distance of 1.5 nm between solute atoms and the box walls. After the boxes were filled with DMSO molecules, the systems were equilibrated as published elsewhere.[47] Temperatures in all runs were 300 K and 340 K, respectively. A triple-range cut-off for Coulomb interactions including a reaction-field was used (0.8 and 1.4 nm). VdW interactions were calculated with a short-range cut-off of 0.8 nm and a long-range cut-off of 1.4 nm. A atom pair-list was used with a cutoff of 0.8 nm and was updated each five integration steps. All bonds were constrained with the Shake algorithm.[115] The integration time step was 2 fs. Production runs had a length of 250 ns.

3.3.9 Declaration

The presented topic was conducted in collaboration with Dr. Jayanta Chatterjee (Technical University Munich, Department Chemistry) and Oded Ovadia (The Hebrew University, Jerusalem). J.C. synthesized all peptides and calculated several 3D structures. O.O. was responsible for all bioassays. The described results have not yet been published.

3.4 References

[1] H. Kessler, *Angew. Chem. Int. Ed. Engl* **1982**, *21*, 512-523.

[2] E. Fischer, *Ber. Dtsch. Chem. Ges.* **1894**, *27*, 2984–2993.

[3] D. E. Koshland, *Proc. Natl. Acad. Sci.* **1958**, *44*, 98-104.

[4] J. H. Pöch, *Wirkungen von Pharmaka*, Thieme New York, **1990**.

[5] R. Flaumenhaft, *Chemical Biology in Comprehensive Medicinal Chemistry,* Elsevier Amsterdam, **2007**, p. 129-149.

[6] M. A. Dechantsreiter, E. Planker, B. Mathä, E. Lohof, G. Hölzemann et al., *J. Med. Chem.* **1999**, *42*, 3033-3040.

[7] K. E. Gottschalk, H. Kessler, *Angew. Chem. Int. Ed.* **2002**, *41*, 3767-3774.

[8] A. E. Aplin, A. K. Howe, R. L. Juliano, *Curr. Opin. Cell Biol.* **1999**, *11*, 737-744.

[9] R. Stupp, C. Ruegg, *J. Clin. Oncol.* **2007**, *25*, 1637-1638.

[10] D. A. Reardon, L. B. Nabors, R. Stupp, T. Mikkelsen, *Expert Opin. Investig. Drugs* **2008**, *17*, 1225-1235.

[11] J. P. Xiong, T. Stehle, R. Zhang, A. Joachimiak, M. Frech et al., *Science* **2002**, *296*, 151-155.

[12] J. A. Elbe, K. Kühn, *Integrin-Ligand Interaction*, Springer Heidelberg, **1997**.

[13] M. J. Humphries, P. A. McEwan, S. J. Barton, P. A. Buckley, J. Bella et al., *Trends Biochem. Sci.* **2003**, *28*, 313-320.

[14] M. J. Humphries, *Biochem. Soc. Trans.* **2000**, *28*, 311-339.

[15] T. Arndt et al., *Integrins in Angiogenesis: Implications for Tumor Therapy, Cancer Therapy: Molecular Targets in Tumor-Host Interactions*, Horizon Bioscience Norfolk, **2005**.

[16] K. J. Clemetson, J. M. Clemetson, *Cell. Mol. Life Sci.* **1998**, *54*, 502-513.

[17] H. Jin, J. Varner, *Br. J. Cancer* **2004**, *90*, 561-565.

[18] H. Yusuf-Makagiansar, M. E. Anderson, T. V. Yakovleva, J. S. Murray, T. J. Siahaan, *Med. Res. Rev.* **2002**, *22*, 146-167.

[19] A. I. Rojas, A. R. Ahmed, *Crit. Rev. Oral Biol. Med.* **1999**, *10*, 337-358.

[20] M. D. Pierschbacher, E. Ruoslahti, *Nature* **1984**, *309*, 30-33.

[21] L. Tranqui, A. Andrieux, G. Hudry-Clergeon, J. J. Ryckewaert, S. Soyez et al., *J. Cell Biol.* **1989**, *108*, 2519-2527.

[22] R. Haubner, R. Gratias, B. Diefenbach, S. L. Goodman, A. Jonczyk et al., *J. Am. Chem. Soc.* **1996**, *118*, 7461-7472.

[23] M. Aumailley, M. Gurrath, G. Muller, J. Calvete, R. Timpl et al., *FEBS Lett.* **1991**, *291*, 50-54.

[24] T. Weide, A. Modlinger, H. Kessler, *Top. Curr. Chem.* **2007**, *272*, 1-50.

[25] O. Demmer, A. O. Frank, H. Kessler, *Design of cyclic peptides: Design of Peptides and Proteins: Applications to Therapeutic Agents and Biomedical Research*, Wiley-VCH Weinheim, **2009**.

[26] M. Gurrath, G. Müller, H. Kessler, M. Aumailley, R. Timpl, *Eur. J. Biochem.* **1992**, *210*, 911-921.

[27] M. Pfaff, W. Göhring, J. C. Brown, R. Timpl, *Eur. J. Biochem.* **1994**, *225*, 975-984.

[28] K. W. Beekman, A. D. Colevas, K. Cooney, R. Dipaola, R. L. Dunn et al., *Clin. Genitourin. Cancer* **2006**, *4*, 299-302.

[29] D. A. Reardon, L. B. Nabors, R. Stupp, T. Mikkelsen, *Expert Opin. Investig. Drugs* **2008**, *17*, 1225-1235.

[30] D. A. Reardon, K. L. Fink, T. Mikkelsen, T. F. Cloughesy, A. O'Neill et al., *J. Clin. Oncol.* **2008**, *26*, 5610-5617.

[31] W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen et al., *Angew. Chem. Int. Ed.* **2006**, *45*, 4064-4092.

[32] F. Manzenrieder, Dissertation (Ph. D. Thesis), Technische Universität München, **2009**.

[33] W. Willker, D. Leibfritz, R. Kerssebaum, W. Bermel, *Magn. Reson. Chem.* **1993**, *31*, 287-292.

[34] A. L. Davis, J. Keeler, E. D. Laue, D. Moskau, *J. Magn. Reson.* **1992**, *98*, 207-216.

[35] A. A. Shaw, C. Salaun, J. F. Dauphin, B. Ancian, *J. Magn. Reson. A* **1996**, *120*, 110-115.

[36] B. Ancian, I. Bourgeois, J. F. Dauphin, A. A. Shaw, *J. Magn. Reson.* **1997**, *125*, 348-354.

[37] J. Cavanagh, M. Rance, *J. Magn. Reson. A* **1993**, *105*, 328-328.

[38] J. Cavanagh, M. Rance, *J. Magn. Reson.* **1990**, *88*, 72-85.

[39] T. F. Havel, *Biopolymers* **1990**, *29*, 1565-1585.

[40] T. Havel, K. Wüthrich, *Bull. Math. Biol.* **1984**, *46*, 673-698.

[41] D. F. Mierke, A. Geyer, H. Kessler, *Int. J. Pept. Protein Res.* **1994**, *44*, 325-331.

[42] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark et al., *J. Comput. Chem.* **2005**, *26*, 1701-1718.

[43] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger et al. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, Vdf Hochschulverlag Zürich, **1996**.

[44] M. Christen, P. H. Hünenberger, D. Bakowies, R. Baron, R. Burgi et al., *J. Comput. Chem.* **2005**, *26*, 1719-1751.

[45] C. Oostenbrink, T. A. Soares, N. F. A. van der Vegt, W. F. van Gunsteren, *Eur. Biophys. J.* **2005**, *34*, 273-284.

[46] C. Oostenbrink, A. Villa, A. E. Mark, W. F. van Gunsteren, *J. Comput. Chem.* **2004**, *25*, 1656-1676.

[47] A. Glättli, X. Daura, P. Bindschädler, B. Jaun, Y. R. Mahajan et al., *Chemistry Eur. J.* **2005**, *11*, 7276-7293.

[48] A. Saupe, G. Englert, *Phys. Rev. Lett.* **1963**, *11*, 462-464.

[49] A. Bax, *Protein Sci.* **2003**, *12*, 1-16.

[50] B. Simon, M. Sattler, *Angew. Chem. Int. Ed.* **2002**, *41*, 437-440.

[51] O. F. Lange, N. Lakomek, C. Farès, G. F. Schröder, K. F. A. Walter et al., *Science* **2008**, *320*, 1471-1475.

[52] G. Kummerlöwe, B. Luy, *Trends Analyt. Chem.* **2009**, *28*, 483-493.

[53] B. Luy, A. Frank, H. Kessler, *Conformational analysis of drugs by NMR: Drug Properties: Measurement and Computation*, Wiley-VCH Weinheim, **2008**.

[54] B. Luy, K. Kobzar, S. Knör, J. Furrer, D. Heckmann et al., *J. Am. Chem. Soc.* **2005**, *127*, 6459-6465.

[55] B. Luy, K. Kobzar, H. Kessler, *Angew. Chem. Int. Ed.* **2004**, *43*, 1092-1094.

[56] M. Zweckstetter, A. Bax, *J. Biomol. NMR* **2002**, *23*, 127-137.

[57] H. F. Azurmendi, C. A. Bush, *J. Am. Chem. Soc.* **2002**, *124*, 2426-2427.

[58] F. Kramer, M. V. Deshmukh, H. Kessler, S. J. Glaser, *Conc. Magn. Reson. A* **2004**, *21A*, 10-21.

[59] C. M. Thiele, S. Berger, *Org. Lett.* **2003**, *5*, 705-708.

[60] S. Marcelja, *J. Chem. Phys.* **1974**, *60*, 3599-3604.

[61] U. M. Reinscheid, J. Farjon, M. Radzom, P. Haberz, A. Zeeck et al., *Chembiochem* **2006**, *7*, 287-296.

[62] M. Zweckstetter, G. Hummer, A. Bax, *Biophys. J.* **2004**, *86*, 3444-3460.

[63] M. Zweckstetter, A. Bax, *J. Am. Chem. Soc.* **2000**, *122*, 3791-3792.

[64] M. Panar, W. D. Phillips, *J. Am. Chem. Soc.* **1968**, *90*, 3880-3882.

[65] K. Kobzar, H. Kessler, B. Luy, *Angew. Chem. Int. Ed.* **2005**, *44*, 3145-3147.

[66] B. L. Marquez, W. H. Gerwick, R. T. Williamson, *Magn. Reson. Chem.* **2001**, *39*, 499-530.

[67] A. L. Frischknecht, J. G. Curro, *Macromolecules* **2003**, *36*, 2122-2129.

[68] S. W. Sides, J. Curro, G. S. Grest, M. J. Stevens, T. Soddemann et al., *Macromolecules* **2002**, *35*, 6455-6465.

[69] R. M. Sok, H. J. C. Berendsen, W. F. van Gunsteren, *J. Chem. Phys.* **1992**, *96*, 4699-4704.

[70] H. Sun, D. Rigby, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **1997**, *53*, 1301-1323.

[71] F. Müller-Plathe, *Chem. Phys. Lett.* **1996**, *252*, 419-424.

[72] F. Müller-Plathe, *Macromolecules* **1996**, *29*, 4782-4791.

[73] G. Milano, F. Muller-Plathe, *J. Phys. Chem. B* **2005**, *109*, 18609-18619.

[74] M. Mondello, H. J. Yang, H. Furuya, R. J. Roe, *Macromolecules* **1994**, *27*, 3566-3574.

[75] H. Furuya, M. Mondello, H. J. Yang, R. J. Roe, R. W. Erwin et al., *Macromolecules* **1994**, *27*, 5674-5680.

[76] S. R. Quake, H. Babcock, S. Chu, *Nature* **1997**, *388*, 151-154.

[77] Q. Sun, R. Faller, *Macromolecules* **2006**, *39*, 812-820.

[78] Q. Sun, R. Faller, *J. Phys. Chem. B* **2005**, *109*, 15714-15723.

[79] H. D. H. Stöver, J. M. J. Frechet, *Macromolecules* **1991**, *24*, 883-888.

[80] Y. Tamai, M. Fukuda, *J. Chem. Phys.* **2004**, *121*, 12085-12093.

[81] L. Zhao, Y. G. Li, J. G. Mi, C. L. Zhong, *J. Chem. Phys.* **2005**, *123*, 124905-124912.

[82] V. A. Harmandaris, N. P. Adhikari, N. F. A. van der Vegt, K. Kremer, *Macromolecules* **2006**, *39*, 6708-6719.

[83] W. L. Jorgensen, D. L. Severance, *J. Am. Chem. Soc.* **1990**, *112*, 4768-4774.

[84] C. Ayyagari, D. Bedrov, G. D. Smith, *Macromolecules* **2000**, *33*, 6194-6199.

[85] G. Kummerlöwe, S. Knör, A. O. Frank, T. Paululat, H. Kessler et al., *Chem. Comm.* **2008**, , 5722-5724.

[86] W. F. van Gunsteren, H. J. C. Berendsen, *Angew. Chem. Int. Ed.* **1990**, *29*, 992-1023.

[87] W. F. van Gunsteren, H. J. C. Berendsen, *Mol. Sim.* **1988**, *1*, 173-185.

[88] W. F. van Gunsteren, H. J. C. Berendsen, J. A. C. Rullmann, *Mol. Phys.* **1981**, *44*, 69-95.

[89] W. L. Jorgensen, J. Tirado-Rives, *Abstr. Pap. Am. Chem. Soc.* **1998**, *216*, U696-U696.

[90] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **1996**, *118*, 11225-11236.

[91] P. Tzvetkova, S. Simova, B. Luy, *J. Magn. Reson.* **2007**, *186*, 193-200.

[92] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089-10092.

[93] A. S. Dutta, Small Peptides, *Chemistry, Biology and Clinical Studies in Pharmacochemistry*, Elsevier Amsterdam, **1993**.

[94] R. A. Conradi, A. R. Hilgers, N. F. Ho, P. S. Burton, *Pharm. Res.* **1992**, *9*, 435-439.

[95] R. A. Conradi, A. R. Hilgers, N. F. Ho, P. S. Burton, *Pharm. Res.* **1991**, *8*, 1453-1460.

[96] G. M. Pauletti, S. Gangwar, B. Wang, R. T. Borchardt, *Pharm. Res.* **1997**, *14*, 11-17.

[97] E. Biron, J. Chatterjee, O. Ovadia, D. Langenegger, J. Brueggen et al., *Angew. Chem. Int. Ed.* **2008**, *47*, 2595-2599.

[98] D. F. Veber, S. R. Johnson, H. Y. Cheng, B. R. Smith, K. W. Ward et al., *J. Med. Chem.* **2002**, *45*, 2615-2623.

[99] J. Chatterjee, D. F. Mierke, H. Kessler, *Chemistry Eur. J.* **2008**, *14*, 1508-1517.

[100] J. Chatterjee, C. Gilon, A. Hoffman, H. Kessler, *Acc. Chem. Res.* **2008**, *41*, 1331-1342.

[101] J. Chatterjee, O. Ovadia, G. Zahn, L. Marinelli, A. Hoffman et al., *J. Med. Chem.* **2007**, *50*, 5878-5881.

[102] J. Chatterjee, D. Mierke, H. Kessler, *J. Am. Chem. Soc.* **2006**, *128*, 15164-15172.

[103] W. L. Cody, J. X. He, M. D. Reily, S. J. Haleen, D. M. Walker et al., *J. Med. Chem* **1997**, *40*, 2228-2240.

[104] F. E. Ali, D. B. Bennett, R. R. Calvo, J. D. Elliott, S. M. Hwang et al., *J. Med. Chem.* **1994**, *37*, 769-780.

[105] P. Artursson, K. Palm, K. Luthman, *Adv. Drug Deliv. Rev.* **2001**, *46*, 27-43.

[106] M. Kansy, F. Senner, K. Gubernator, *J. Med. Chem.* **1998**, *41*, 1007-1010.

[107] D. F. Mierke, M. Kurz, H. Kessler, *J. Am. Chem. Soc* **1994**, *116*, 1042-1049.

[108] X. Zhang, G. V. Nikiforovich, G. R. Marshall, *J. Med. Chem.* **2007**, *50*, 2921-2925.

[109] B. E. A. Alberts, *Molecular Biology of the Cell*, Garland Science London, **2002**.

[110] J. M. Anderson, *News Physiol. Sci.* **2001**, *16*, 126-130.

[111] T. Rezai, B. Yu, G. L. Millhauser, M. P. Jacobson, R. S. Lokey, *J. Am. Chem. Soc.* **2006**, *128*, 2510-2511.

[112] T. Rezai, J. E. Bock, M. V. Zhou, C. Kalyanaraman, R. S. Lokey et al., *J. Am. Chem. Soc.* **2006**, *128*, 14073-14080.

[113] C. Oostenbrink, A. Villa, A. E. Mark, W. F. van Gunsteren, *J. Comput. Chem.* **2004**, *25*, 1656-1676.

[114] D. P. Geerke, W. F. van Gunsteren, *Chemphyschem* **2006**, *7*, 671-678.

[115] J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327-341.

4 Searching the Holy Grail

Elucidation of New Drug Candidates

There are many reasons why scientists are interested in small molecules which are capable to bind to a specific protein receptor. On one side, biochemical processes can be blocked by small inhibitors, and thus, the function of proteins can be elucidated by monitoring the inhibitory effects on the cellular behavior. On the other side, metabolic pathways which run out of control due to protein mutations or other misregulations (e.g. of gene expression) can be suppressed; this procedure is essential for the treatment of several diseases like cancer or Alzheimer. For that reason, a plethora of methods with NMR and bioassay based target and / or ligand screening as the most prominent experimental techniques have been developed to identify small molecules that interact with the receptor molecules under investigation. In order to save time and to evade expensive experimental setups, virtual screening approaches have been designed; here, huge electronic libraries comprising the three-dimensional coordinates of small organic compounds are screened for molecules which exhibit favored chemical and physical characteristics (pharmacophore models). Having identified molecules that match a pharmacophore model, the receptor affinity of the potential binders can be roughly estimated via virtual docking approaches which express the strength of a receptor-ligand complex in terms of binding energies or scoring values. In cases where no information about the three-dimensional structure of a receptor is available but ligand affinities are known, QSAR based methods can be utilized to optimize the compound scaffolds with respect to higher target affinities.

4.1 Phosphorus NMR Spectroscopy as New Screening Technology

As it was extensively described in the introductory chapter 1.1.3, NMR spectroscopy is a well established technique for screening of compound libraries.[1] In general, one of the biggest advantages of NMR in relation to other screening methods is that it directly detects even weak interactions between ligand and target molecules. In addition, the number of false positive hits, often obtained in bioassays, is minimized.[2,3] Among the variety of NMR screening approaches, methodologies based on exclusively tracing ligand signals are the most powerful tools to identify binders in compound libraries.[4] In this scope, standard and group selective STD spectroscopy or fluorine (^{19}F) screening are prominent and well established examples.[5-8] Together with recently developed high-throughput approaches, e.g. target immobilized NMR screening (TINS),[9] ligand-based screening is a potent technology in the field of drug science.[1,10-12] Having identified a small molecule in a compound library which exhibits receptor affinity, target-based screening techniques can be applied in order to get informations about binding sites and modes.

4.1.1 The Scientific Question

Each ligand based NMR screening method has distinct benefits and drawbacks. Approaches using proton resonances only are very simple and quick but most often suffer from severe signal overlap. Detection of heteronuclei like fluorine which results in easily interpretable NMR spectra requires in almost all cases chemical modifications (incorporation of ^{19}F) of the ligands to be screened. Thus, it would be desirable to have recourse to a method that combines the advantages of both homo- and heteronuclear based NMR ligand screening.

Accordingly, the goal of this project was to enhance the applicability of NMR spectroscopy in biological and pharmaceutical research by introducing phosphorus as new nucleus for compound library screening of protein inhibitors.

In principle, phosphorus reveals several physical and chemical properties making it well suited for the utilization in library screening. For example, the

problem of overlapping resonances normally arising in proton detected spectra is reduced because each compound is represented by only one ^{31}P NMR signal in most cases. Moreover, phosphorus is often an intrinsic element of compounds mimicking the tetrahedral intermediate of a peptide bond hydrolysis (often called “transition-state analogs”, see also Fig. 4.1), e.g. in protease inhibitors.[13-17] Hence, interactions with proteases can be detected without auxiliary ligand labeling.[7]

4.1.2 The Relevance of Phosphorus in Chemistry and Medicine

Phosphorus is a spin-½ nucleus with a gyromagnetic ratio of 17.24 MHz / T and has a natural abundance of 100 %; therefore, it is widely used in magnetic resonance techniques, especially in *in vivo* NMR applications. Common organic phosphorus compounds show a wide chemical shift dispersion of about 100 ppm, thus allowing the screening of huge compound libraries. As ^{31}P nuclei exhibit large chemical shift anisotropy (CSA), even weak binding events can easily be traced due to a strong T2 relaxation rate dependent line broadening of the affected ligand signals. As already mentioned above, protease inhibitors often contain motifs including phosphorus; according to the oxidation state of ^{31}P and the nature of direct bound atoms, such motifs can be structurally and physically very different, resulting in diverse compound libraries whose representatives do not need to be supplementary labeled with an NMR active nucleus (see Fig. 4.1).

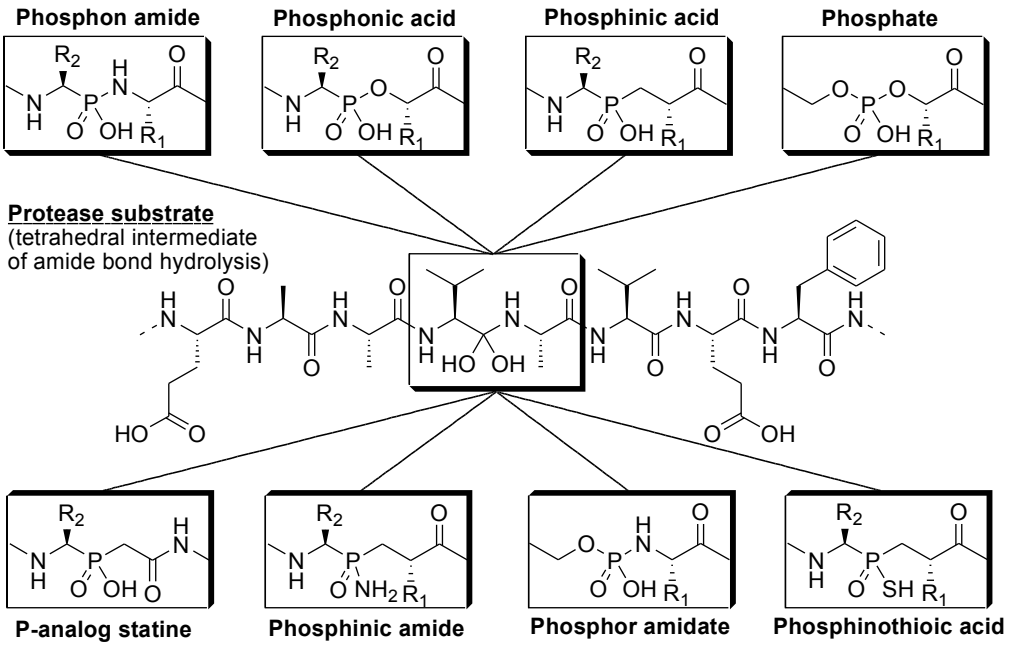


Figure 4.1: Presentation of selected isosteres that comprise phosphorus as mimic of the hydrolyzable peptide bond in the rational design of protease inhibitors.

Phosphorus comprising molecules do not only play an important part with respect to proteases but also in the wide range of biochemical pathways, e.g. in phosphoryl transfer reactions. After identification of a small phosphorylated substrate, this ligand or a more stable derivate, respectively, may serve as reporter molecule in the screening for new inhibitors. Figure 4.2 gives some examples for this strategy.[18-21]

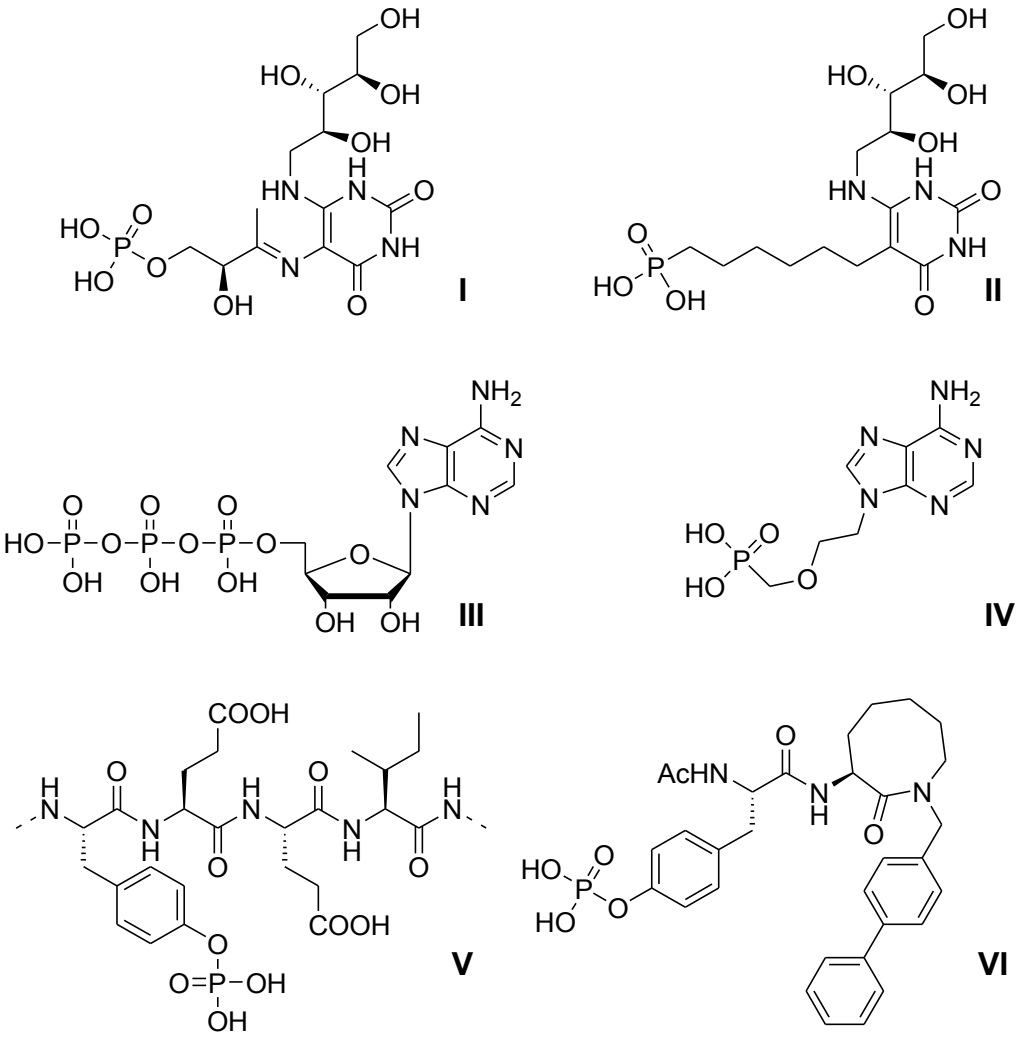


Figure 4.2: Illustration of three selected natural phosphorylated substrates (PS) and their phosphorus containing analogs (PCA) as possible reporter molecules for the screening of compound libraries: PS of lumazin synthase (I) and PCA (II) as inhibitor of lumazin synthase. ATP (III) and its nucleotide analogs such as AMP-PNP (not shown) or Hepsera® (IV) that is a bioavailable drug used for treatment of hepatitis B by inhibition of reverse transcriptase. PS of Src protein (V) and PCA (VI) as inhibitor of the Src SH2 kinase domain.

In addition, various phosphorus containing drugs were approved by the FDA in recent years, e.g. Cidofavir®, IFEX®, Monopril®, or Foscavir® (see Fig. 4.3). To break through the obstacle of poor oral availability, several ^{31}P comprising agents were developed in form of prodrugs (e.g. Monopril®) which develop their full activity after enzymatic hydrolysis. Because of this it becomes obvious that ^{31}P based NMR screening could fruitfully contribute to the process of drug development.

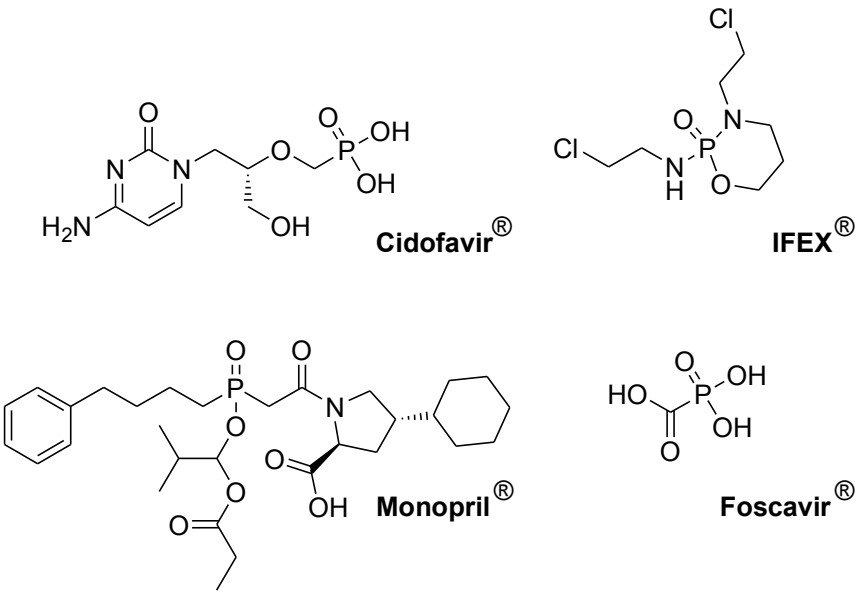


Figure 4.3: Examples of approved phosphorus containing drugs: Cidofavir is an antiviral agent for the treatment of Cytomegalovirus retinitis. IFEX is a nitrogen mustardalkylating drug used in the treatment of cancer. Monopril is an inhibitor of the angiotensin converting enzyme and is used for medicating hypertension and chronic heart failure. Foscavir is an antiviral agent to treat herpes virus diseases.

4.1.3 Screening of a Small Compound Library

The fundamental soundness of the presented method should be established by a series of different NMR experiments. For this purpose, the exhaustively investigated thermolysin-phosphoramidon system was chosen.[22] Thermolysin (~ 35 kDa) is a thermostable, calcium-binding zinc endopeptidase isolated from *Bacillus thermoproteolyticus*. Phosphoramidon, a natural product first isolated from *Streptomyces tanashiensis*,[23] contains a phosphor amidate unit as isostere of the hydrolyzable peptide bond and binds to thermolysin with a K_i of 28 nM.[24] To complete the representative small molecule library, five substances displaying some of the isosteres shown in figure 4.1 were selected (see Fig. 4.4).

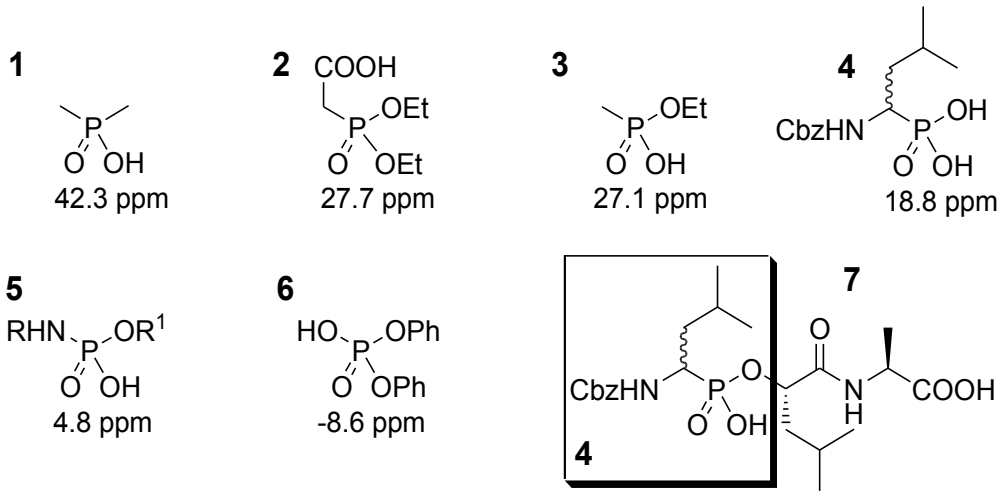


Figure 4.4: Presentation of structures as well as NMR chemical shifts of the used compounds (1-6) and the $CbzL^{p}(O)LA$ molecule (7) which contains 4 as fragment.

R: L-leucyl-L-tryptophan, R^1: L-rhamnopyranosyloxy (hydroxyphosphinyl).

A proton-decoupled ^{31}P-1D NMR spectrum of an equimolar mixture of 1-4 and 6 (the concentration of each compound was 0.5 mM) was recorded and is shown in figure 4.5A. Hydrogen phosphate was utilized as buffering agent and internal pH control. Since both the resonances exhibit large chemical shift dispersion and each compound is represented by only one signal, the spectrum of the library is simple and every peak was easily assigned. Upon addition of thermolysin to a final concentration of 0.25 mM, the signal of compound 4 almost completely vanishes whereas all other resonances remain unaffected (see Fig. 4.5B). Hence, a new ligand for thermolysin was identified. Due to the ligand-receptor interaction, the small compound adopts the fast T2 relaxation time of the protein. This causes a strong signal line broadening, finally leading to the disappearance of the ligand resonance. The identified binder turned out to be a fragment of the molecule CbzL^{P}(O)LA (7) that was introduced by Bartlett et al. as a thermolysin inhibitor (see Fig. 4.4).[25] Hence, the described screening technology is able to detect small binders ideally suited for fragment based drug design (FBDD).[26,27]

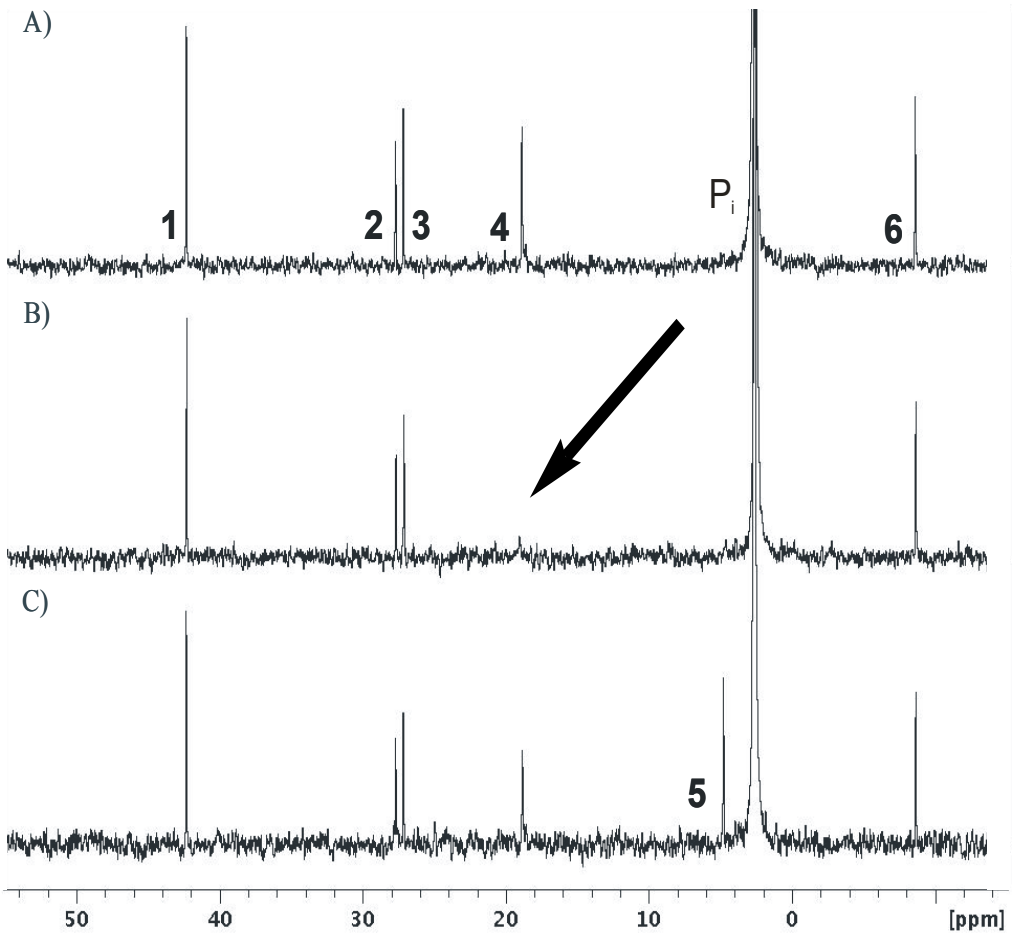


Figure 4.5: Illustration of ^{31}P-NMR spectra of a small ligand library (each compound: 0.5 mM) in phosphate buffer (pH 7.5) containing 3 M KBr for keeping thermolysin in solution A) before and B) after addition of thermolysin (final concentration: 0.25 mM). The signal of **4** disappears in the presence of protein. C) Recovery of the vanished signal upon addition of the tight binder thermolysin (0.5 mM). Effects on the signal line width to identify binders are already visible with less than 1 % of the protein concentration used here (see also Fig. 4.6).

4.1.4 Recovery of a Ligand with Medium Receptor Affinity

Owing to the large CSA of phosphorus nuclei that significantly contributes to the transverse relaxation time of the fraction of bound ligand, the new approach may be vulnerable to false positive hits when unspecific interaction between a compound and the target is present. In order to overcome this drawback or to rule out ligand binding to an undiscovered second site of thermolysin, respectively, the signal of (4) must be restorable by adding a known, high affinity substance to the library. Here we used the strongly binding phosphoramidon (5) as competing molecule. As can be seen in figure 4.5C, the addition of (5) upon a final concentration of 0.5 mM results in an almost complete recovery of (4) whereas the signal of free phosphoramidon appears at 4.8 ppm. Hereby, it is clearly shown that both (4) binds exclusively at the active site of thermolysin and that (5) competes with (4) for a specific interaction with the protein. It should be additionally noted that this approach can also be used for reporter based screening. This means that a phosphorus containing molecule that weakly binds to a protein is monitored by ^{31}P NMR while a substance library of any constitution is screened with regard to competition. As soon as the resonance of the reporter molecule is recovered, a ligand with equal or higher affinity to the target is found.

4.1.5 Titration of a New-Found Receptor Ligand

To evaluate the sensitivity of phosphorus based NMR screening, the dependence of the peak height of the binding compound on various protein concentrations was examined. Although the ligand (0.5 mM starting concentration) is in large excess over the protein, the observed resonance of (4) completely vanishes upon addition of small amounts of thermolysin (see Fig. 4.6).

To be more precise, the presence of only 6 µM protein leads to a reduction of the ligand signal to half its original height. As a result of the large CSA of phosphorus nuclei, a strong line broadening of the compound signal is induced by complex formation. Thus, solely low micromolar or even submicromolar protein concentrations are required for the facile identification of an unknown binder with ^{31}P NMR screening. Considering the titration plot, it becomes clear that ligand based phosphorus NMR screening is a very sensitive method for probing intermolecular interactions between proteins and

small or medium sized ligand molecules. Therefore, the NMR screening process is enhanced by both shortening the time and lowering the costs of protein production.

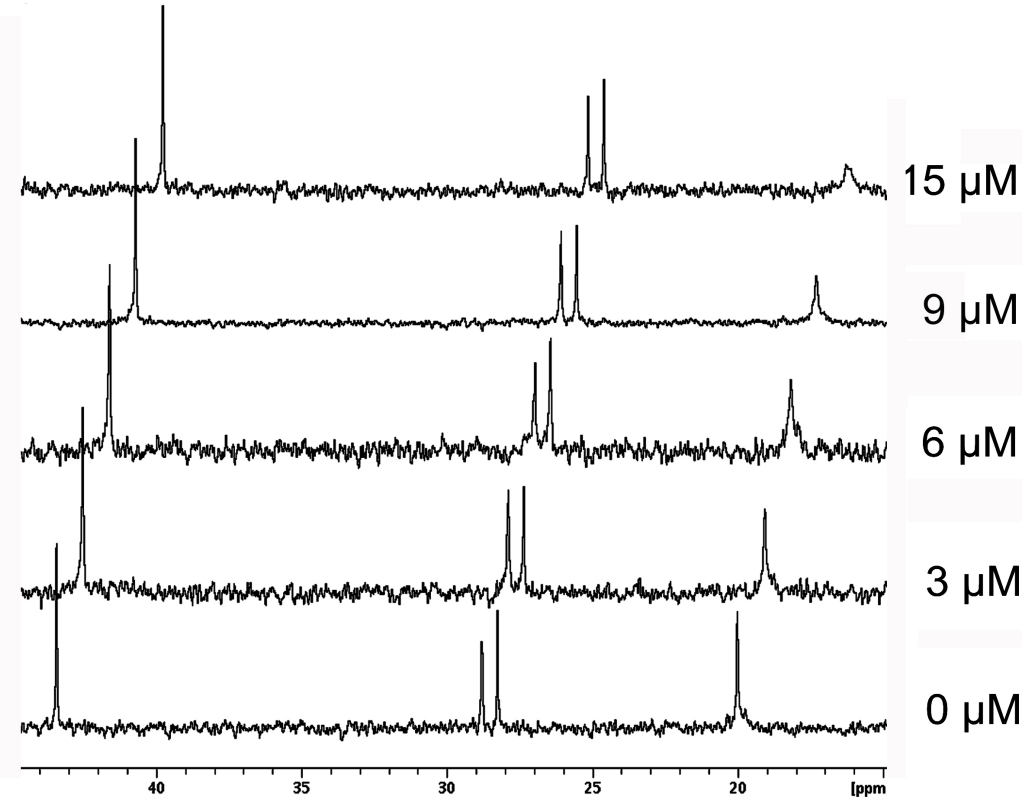


Figure 4.6: Section of a shifted overlay of several ^{31}P-1D NMR spectra. Compound library (1-4, 0.5 mM); without protein, after addition of 3 µM, 6 µM, 9 µM and 15 µM thermolysin. An increase of protein concentrations results in a strong line broadening of the affected resonance (4).

4.1.6 Two-Dimensional Experiments

Although ^{31}P resonances show a wide chemical shift range of more than 100 ppm, signal overlap may arise when working with libraries consisting of similar substances. One way to reduce or even overcome this problem is the performance of heteronuclear two-dimensional measurements. To give an example, we recorded 2D-^{1}H,^{31}P-COLOC experiments of our molecule library, [28] both with and without thermolysin (see Fig. 4.7). Considering only the ^{31}P dimension, the signals of (2) and (3) are close to each other. Assigning those resonances may be difficult in cases when induced shift changes occur. However, the signals are clearly separated in the 2D spectrum. As it is the case in the one-dimensional spectrum, the disappeared NMR peak of (4) indicates binding. In addition, it is evident that signals with similar chemical shifts in both dimensions can also be discriminated by their cross-peak intensity which is almost linearly dependent on the number of ^{31}P coupled protons via vicinal scalar couplings.

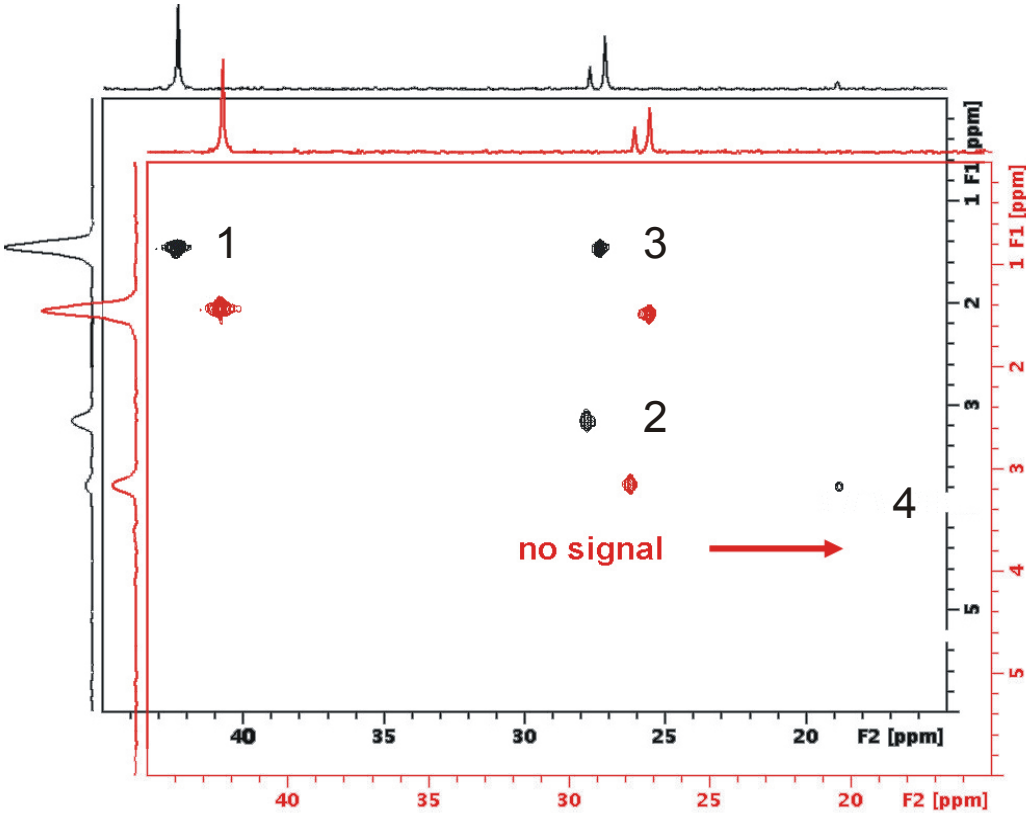


Figure 4.7: Section of an overlay of two $^{1}H,^{31}P$-2D COLOC spectra. The compound library (1-4, 0.5 mM) without (black) and with (red) thermolysin (0.25 mM). Addition of the protein leads to a disappearance of the cross peak of molecule (4). The two phosphorus signals which are close to each other at ca. 27 ppm are well separated in the proton dimension.

4.1.7 Conclusion

In conclusion, it was demonstrated that phosphorus NMR enhances and even extends ligand-based screening of compound libraries. To show the broad applicability of ^{31}P NMR screening, not only a proof of principle of the basic concept but also valuable extensions of the method like recovery experiments and heteronuclear 2D NMR measurements were presented. Since a lot of substances that mimic the tetrahedral intermediate of peptide bond hydrolysis contain phosphorus, ^{31}P NMR may especially be applied to screen large mixtures of protease inhibitors. Furthermore, stable analogs of naturally phosphorylated substrates constitute powerful starting points for the design of ^{31}P containing compound libraries. However, mixtures that comprise non-phosphorus substances can also be screened by utilizing a ^{31}P containing reporter ligand. In contrast to other screening techniques, e.g. bioassay-based methods that are exclusively focused on finding strong binders, phosphorus NMR screening also allows for identifying ligands with medium or even weak affinity to a target molecule. To diminish the number of false positive hits, utilization of low magnetic fields where unspecific binding causes only weak T2 relaxation dependent line broadening as well as recovery experiments are suggested.

4.1.8 Reagents, Methods, and Experiments

All informations regarding used reagents, compound synthesis, and purification as well as additional experimental informations can be found in literature.[29]

Thermolysin was dissolved in phosphate buffer containing 3 M KBr at a pH of 7.5. Protein concentrations were determined via UV / VIS spectroscopy, using an extinction coefficient ε_{280} of 66300 L mol^{-1} cm^{-1}. A small amount of the aqueous protein solution was added to the NMR sample upon a final thermolysin concentration of 0.25 mM. Phosphoramidon was also dissolved as highly concentrated solution in DMSO. Addition of phosphoramidon from this stock solution resulted in a total amount of 0.5 mM of (5) in the NMR sample. Titration studies were performed by stepwise addition of compounds (1-4) and (6) from highly concentrated DMSO stock solutions. 50 mM phosphate buffer containing 3 M KBr at a pH of 7.5 was used for the NMR sample, resulting in a final volume of 400 µL. In all measurements, the total amount of DMSO in solution was less than 5 %.

All NMR experiments were carried out on a Bruker Avance 250 MHz spectrometer equipped with a 5 mm QNP probe head. The sample temperature in all measurements was 300 K and 10 % D_2O was used as lock signal. ^{31}P-1D spectra were recorded with a standard Bruker pulse program, using a 30° excitation pulse, a 1.5 s recycle delay and a WALTZ-16 power-gated composite-pulse proton decoupling. For all one-dimensional experiments except the fourth titration step (9 µM thermolysin, see figure 4.6), 4k (16k) FIDs were accumulated, resulting in a total measurement time of ca. 2 hrs (8 hrs). 128 dummy scans were performed in order to prevent proton decoupling dependent temperature artifacts. The transmitter frequency was placed at 20 ppm whereas the phosphorus resonance frequency was at 101.26 MHz in all experiments. In all experiments, the spectral width was 70 ppm. Each spectrum was Fourier transformed after apodization with an exponential window function and a line broadening factor of 3. For heteronuclear 2D experiments, the COLOC pulse program as published by Kessler et al. was utilized.[28] The delay for evolution of long range couplings was 25 ms. For the 2D-COLOC spectra, 4k data points in the direct and 128 data points in the indirect dimensions were recorded. The number of FIDs per data point was 32 and the total measurement time was ca. 2.5 hrs. All two-dimensional spectra were Fourier transformed in both dimensions. For the F2

dimension, an exponential window function with a line broadening factor of 3 was applied. The F1 dimension was apodized with an QSINE function including a sine bell shift of 2. The ^{31}P transmitter frequency was again placed at 20 ppm.

4.1.9 Declaration

The presented topic was conducted in collaboration with Dr. Florian Manzenrieder (Technical University Munich, Department Chemistry). F.M. chose the test system, synthesized ligand (4), and prepared all solutions. The described results have been published as follows:

F. Manzenrieder*, A.O. Frank*, H. Kessler. Phosphorus NMR spectroscopy as versatile tool for compound library screening. *Angew. Chem. Int. Ed.* **2008**, *47*, 2608-2611. (* equally contributing authors)

4.2 Virtual Screening for and Docking of Potential Hop-Protein Ligands

In chapter 4.1, a new NMR screening technique was presented. Although being powerful and robust, the method is not suited to screen compound mixtures comprising thousands of molecules. Thus, libraries to be investigated by NMR spectroscopy often contain only a subset of compounds that were preselected usually based on a rough idea of how a molecule has to be structured to exhibit receptor affinity.[30,31] A very popular way to make such a preselection is to draw on the concepts of virtual screening and docking.[32] Since the speed of computer processors has dramatically improved and screening algorithms have been optimized during the last years, electronic libraries with more than one million members can be quickly searched for potential receptor ligands by means of pharmacophore models (see chapter 1.2.1). To evaluate the quality of screening outcomes, new-found ligands can be virtually docked to the receptor of interest. Subsequently, the most promising molecules are usually synthesized and (NMR) experimentally tested for receptor affinity. Nowadays, this procedure is a standard strategy in pharmaceutical research and also found its way into functional studies of biomacromolecules.[33,34]

4.2.1 The Scientific Question

A highly interesting and elaborately studied biochemical pathway is the Hsp70 / Hsp90 substrate cycle which plays an essential role in the folding and maturation of important regulatory proteins in eukaryotes, e.g. estrogen receptors or proto-oncogenic protein kinases.[35] Hop, an adaptor protein that belongs to the group of co-chaperones, mediates the association of Hsp70 and Hsp90. Since Hsp90 - a key protein in cancer formation - receives its substrates from Hsp70 in a Hop dependent transfer reaction, breaking of the Hsp90-Hop complex may block the passage of Hsp90 client molecules, thus being a route to novel anti-cancer agents.[36-40]

Hence, the aim of this project was to find a small organic compound that exhibits affinity to the Hop protein and could therefore be used as a first fragment for the development of a Hop-Hsp90 complex breaker.

For this purpose, a pharmacophore model based on the crystal structure of the Hsp90 binding domain of Hop complexed with a known peptidic binder was created and further on applied in screening of huge virtual compound libraries. Finally, interesting candidates were evaluated via a virtual docking approach and also by NMR spectroscopy.

4.2.2 The Hsp90 Binding Mode of Hop

The ca. 60 kDa Hop protein is composed of three tetratricopeptide repeat (TPR) domains: TPR1, TPR2A, and TPR2B.[41,42] Each of the TPR domains which are assumed to show a very similar 3D structure consisting of three or more tandem repeat amino acid sequences.[43][44] TPR2A (amino acids 223-352 in human Hop) which is the natural binding domain of the highly conserved C-terminal Hsp90 sequence MEEVD (k_D = 11 µM for the isolated peptide) folds in a meander of seven α-helices which are arranged in a head-to-tail manner.[45,46] The crystal structure of TPR2A (complexed with MEEVD; PDB code: 1ELR) reveals that the meander forms a cradle shaped groove, and that the HSP90 derived peptide MEEVD binds to this groove in an ordered extended conformation (see Fig. 4.8).[35]

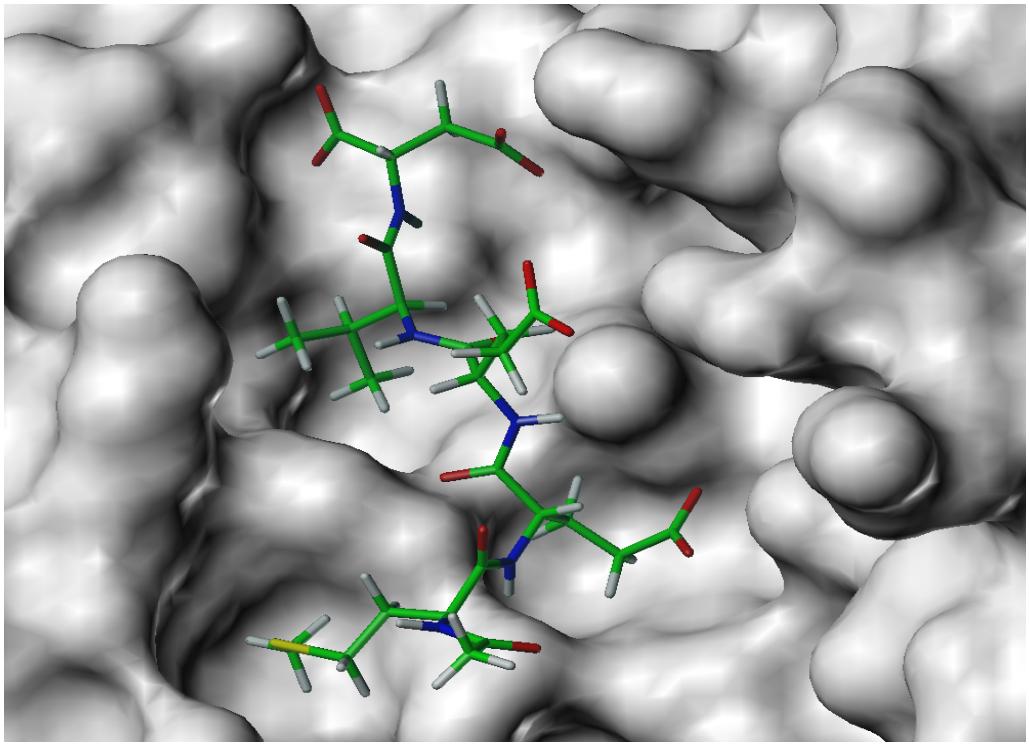


Figure 4.8: Section of the crystal structure of the TPR2A domain of HOP in complex with the HSP90 derived peptide MEEVD. The MEEVD peptide binds in an extended backbone conformation to the Hop-TPR2A domain.

Due to the stretched structure of the Hsp90 derived peptide in the active site of TPR2A, all MEEVD side chains (except Glu2) interact with the protein. One essential structural motif of MEEVD is the so-called two-carboxy-clamp since all residues of TPR2A (Lys229, Asn233, Asn264, Lys301, Arg305) which interact with the aspartic acid terminal and side chain carboxy groups are highly conserved. Whereas the two-carboxy-clamp seems to be essential for the intermolecular recognition, the amino acids upstream from Asp0 (the first residue) are responsible for selectivity and also trigger affinity.[47] In figure 4.9, all important interactions between TPR2A and the MEEVD peptide are listed.

No.	Protein Residue	Protein Atom	Peptide Residue	Peptide Atom	Interaction Type
01	Asn264	*sc*-NH2	Asp0	*bb*-CO	electrostatic
02	Asn264	*sc*-CO	Asp0	*bb*-NH	electrostatic
03	Asn233	*sc*-NH2	Asp0	*bb*-CO	electrostatic
04	Lys229	*sc*-NH2	Asp0	*bb*-CO	electrostatic
05	Lys301	*sc*-NH2	Asp0	*sc*-CO	electrostatic
06	Gln298	*sc*-NH2	Asp0	*sc*-CO	electrostatic
07	Arg305	*sc*-NH2	Glu2	*bb*-CO	electrostatic
08	Arg305	*sc*-NH2	Glu3	*sc*-CO	electrostatic
09	Asn308	*sc*-NH2	Glu3	*sc*-CO	electrostatic

Figure 4.9: Table of all important interactions between the Hop protein and the MEEVD peptide.

sc: side chain; bb: backbone; vdW: van der Waals; h'phobic: hydrophobic.

Continuation of Fig. 4.9

No.	Protein Residue	Protein Atom	Peptide Residue	Peptide Atom	Interaction Type
10	Tyr236	*sc*-OH	Glu3	*bb*-CO	electrostatic
11	Glu271	*sc*-CO	Glu3	*bb*-NH	electrostatic
12	Thr260	---	Asp0	---	h'phobic / vdW
13	Asn233	---	Val1	---	h'phobic / vdW
14	Asn264	---	Val1	---	h'phobic / vdW
15	Ala267	---	Val1	---	h'phobic / vdW
16	Tyr236	---	Met4	---	h'phobic / vdW
17	Phe270	---	Met4	---	h'phobic / vdW
18	Glu271	---	Met4	---	h'phobic / vdW

4.2.3 Creation of a Pharmacophore Model

Virtual libraries usually contain only small organic compounds. Thus, it is not possible to find a ligand that “mimics“ the complete binding mode of the MEEVD peptide. As a consequence, only a fraction of the Hsp90 derived amino acid sequence had to be elected which subsequently serves as basis for parts of the pharmacophore model.

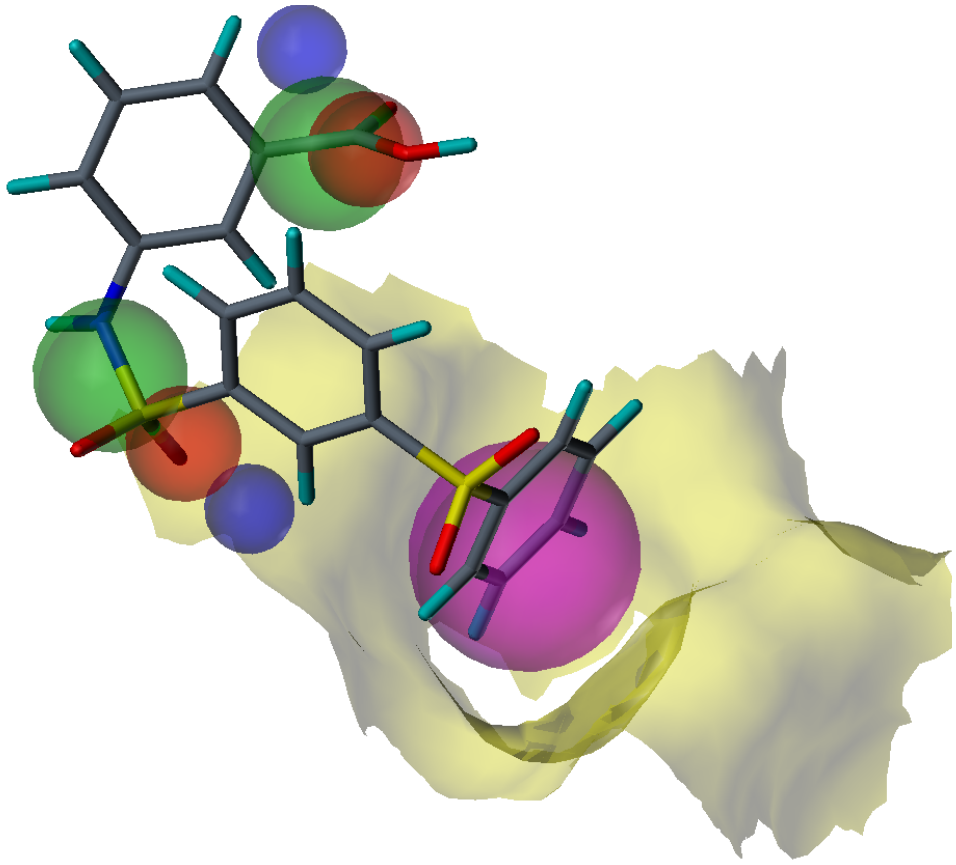


Figure 4.10: Illustration of the created pharmacophore model for the search of new Hop TPR2A domain ligands. One of the new-found compounds is fitted into the 3D model.

Blue spheres: hydrogen bond donor sites (spatial tolerances: 0.7 Å); green spheres: negative centers (1.2 Å); red spheres: H-bond acceptor atoms (0.9 Å); violet sphere: hydrophobic center (1.8 Å); light yellow topology: section of the Hop receptor surface used as a spatial constraint (provided with a van der Waals radius tolerance).

Since the two-carboxy-clamp seems to be the central binding motif, it was decided to concentrate on Asp0 and its interactions with the Hop protein. Furthermore, additional informations about the contacts between Val1 and the receptor molecule were included into the pharmacophore model in order to extend the demands of the search mask. At last, a part of the surface of the Hop binding site was utilized as target based constraint; this means that those ligands which match the ligand based features defined before (Asp0 and Val1 interactions with Hop) but contain additional groups which might clash with the receptor scaffold are discarded. After checking out various setups, a final pharmacophore model was created and further on used in the computational screening process. Details of the applied features and constraints are shown in figure 4.10 and explained in the figure caption.

4.2.4 Virtual Screening of Compound Libraries

For the virtual screening, several commercial libraries or a preselection thereof which contain typical drug-like organic molecules were utilized (see Fig. 4.11). In total, more than 380,000 compounds were tested for matching the requirements of the created pharmacophore model. During the three-dimensional screening process, no additional filtering was used; this means that e.g. molecules which are not in agreement with Lipinski's rule of five were not rejected.

Library Provider	Members	Library Provider	Members
ACD Labs	110,135	ChemStar	32,479
Ambinter	59,549	InterBioScreen	54,937
Asinex	96,756	Tripos	26,582

Figure 4.11: Table of the six used compound databases and the number of totally screened molecules per library.

The computational database search revealed 70 molecules (~ 0.02 % of all screened compounds) that both display all features and do not violate the constraints defined in the pharmacophore model. In the next step, the selected compounds were visually inspected; those molecules which have nearly identical scaffolds or display structures which seem not to exhibit receptor affinity were discarded. Finally, 36 compounds were accepted as potential Hop binders and further on virtually docked to the active site of the TPR2A domain.

4.2.5 Docking of Selected Screening Hits

For the docking runs, the structure file of the Hop TPR2A domain had to be prepared. First, “broken” side chains as they often appear in crystal structures were repaired, apolar hydrogens were removed, and all atoms were provided with Kollman-type (partial) atomic charges. Then, a virtual grid was placed into the active site of the protein domain which was afterwards used to calculate all possible interactions between the receptor and the ligands to be docked. Further on, all 36 preselected compounds were also prepared; root atoms as well as rotatable bonds and Gasteiger-type (partial) charges were defined. Finally, solvation energy parameter were calculated in order to estimate the role of solvent molecules in the binding process.

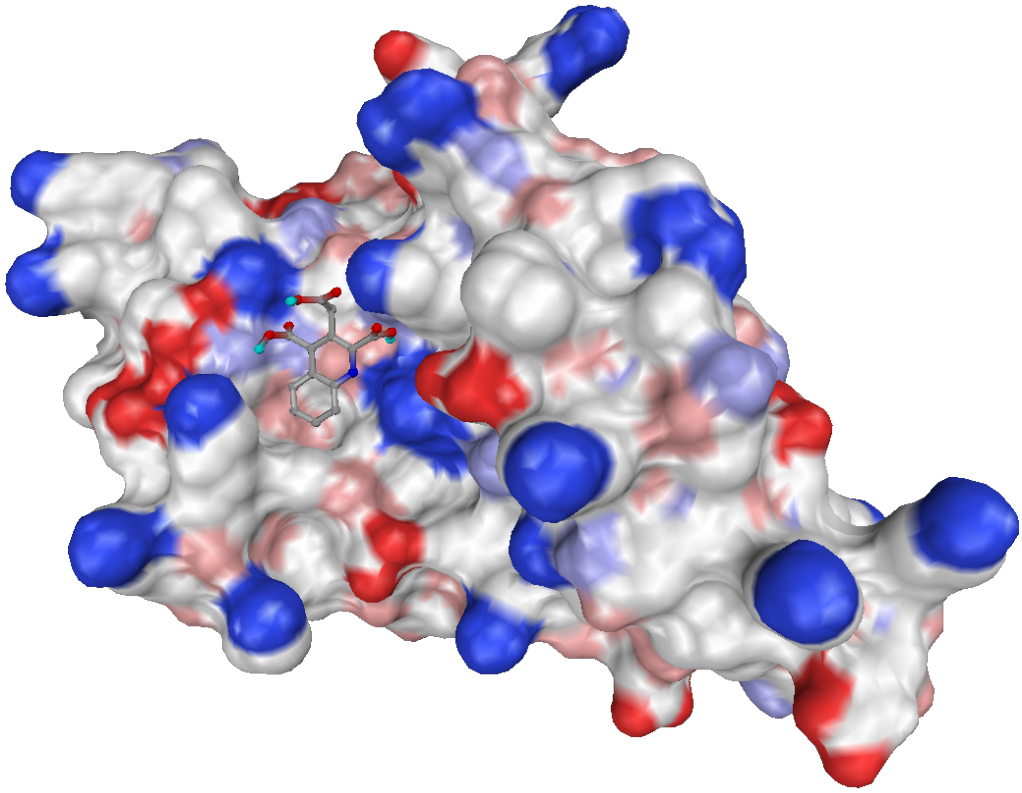


Figure 4.12: Presentation of one of the 36 screening hit compounds docked to the TPR2A domain of Hop. The ligand interacts with the MEEVD binding site and exhibits those contacts which were proposed via the designed pharmacophore model (the aromatic ring system shows hydrophobic interactions with an apolar groove of Hop and two of the three carboxy groups contact positively charged Hop residues).

The Connolly surface of the Hop protein is colored according to its electrostatic potential: blue color: a positive potential; red color: a negative potential.

All 36 compounds which were previously selected could be docked to the Hop binding domain (see Fig. 4.12 as an example of one of the docked compounds). To evaluate the quality of the outcomes, three procedures were applied: comparison of binding energies, RMS deviation based clustering of docked structures, and visual inspection of the computed intermolecular binding modes. In addition, another docking algorithm which makes use of the fragment based approach was utilized to check the reliability of the output of

the grid based technique. After a careful analysis, six molecules (out of the 36 docked potential ligands) that exhibited a reliable binding mode, highly populated RMSD clusters, and low interaction energies were chosen for experimental NMR binding studies (see Fig. 4.13 and Fig. 4.14).

Compound Label	Mean / Lowest Energy [kJ / mol]	RMSD Cluster Size	Visual Inspection
A	-6.39 / -6.89	68 / 100	very good
B	-8.56 / -8.94	97 / 100	very good
C	-5.39 / -5.55	78 / 100	good
D	-7.32 / -7.81	67 / 100	good
E	-7.53 / -8.42	88 / 100	very good
F	-7.15 / -7.64	97 / 100	good

Figure 4.13: Table of the virtual docking results of six elected compounds. The more negative computed binding energies are, the stronger the receptor ligand interactions are proposed to be. Furthermore, large RMSD cluster sizes also provide evidence of ligand receptor affinities. The 2D structures of the compounds are shown in figure 4.14.

It should be noted that some of the virtually tested compounds revealed better or equal docking results than the six elected compounds shown in figure 4.13. However, they could not be experimentally tested since the commercial companies that provide the utilized screening databases were not able to deliver the ordered molecules.

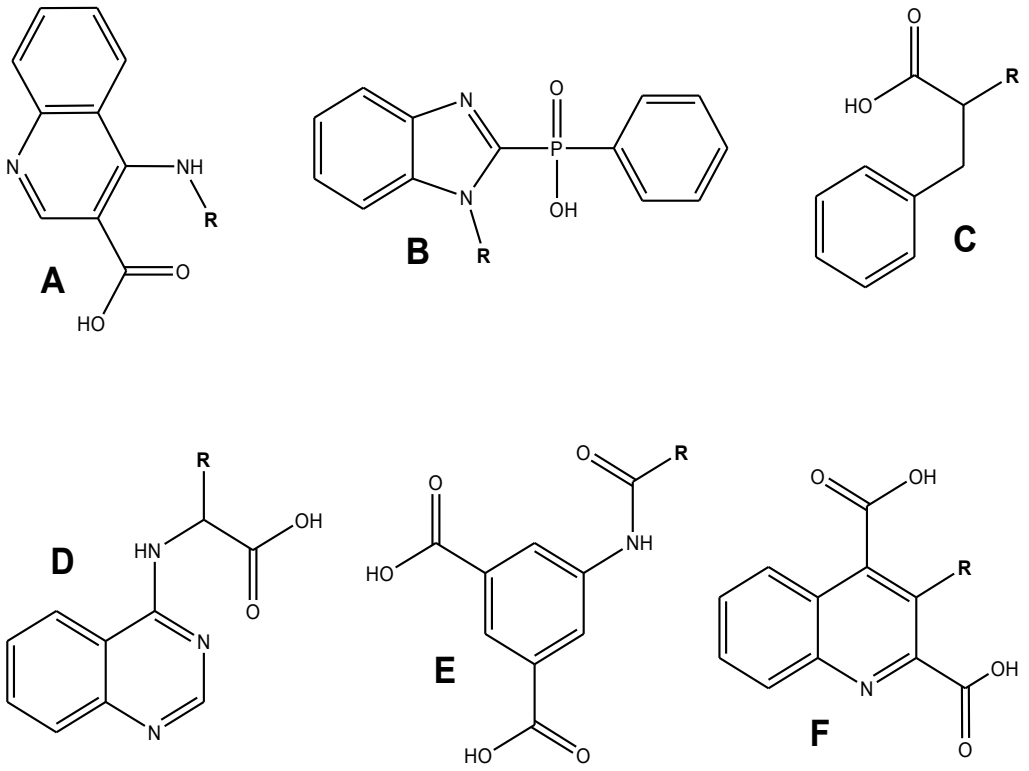


Figure 4.14: Illustration of the chemical structures of the six ligands that were selected after virtual compound library screening, grid-based docking runs, and visual inspection for experimentally testing of Sti1TPR2B affinity.

4.2.6 Experimental Binding Studies

As already mentioned in chapter 1.2, computational tools designed for the elucidation of new drug molecules are not perfect since they usually neglect molecular dynamics, especially of the receptor molecule, do not explicitly include the participation of solvent molecules or ionic particles in the intermolecular binding, and dependent on the quality of the provided three-dimensional target structures. Hence, the outcomes of the virtual screening and docking was experimentally evaluated. For this purpose, target based NMR binding studies of the six chosen compounds (as illustrated in figure 4.13) were performed. As a sufficient batch of the TPR2A domain of human Hop which served as basis for the designed pharmacophore model was not available, the affinity of the ligands was tested by help of the isolated, ^{15}N labeled TPR2B (see also chapter 4.2.2) domain of Sti1, a Hop related yeast protein.[42] Two of the compounds to be tested (A and D) are neither soluble in water nor in DMSO and could therefore not be screened via solution-state NMR spectroscopy. The remaining molecules were dissolved in highly concentrated stock solutions of DMSO and titrated in two steps (1 mM and 10 mM final solution concentrations) to the TPR2B domain (200 µM). Whereas addition of all compounds upon a concentration of 10 mM results in chemical shift perturbations in $^{1}H,^{15}N$-HSQC spectra, only molecule F effects significant changes in the spectrum at a final concentration of 1 mM. Thus, it was decided to perform an exhaustive NMR binding study (seven titration steps) for ligand F which allows the estimation of a protein-ligand dissociation constant.

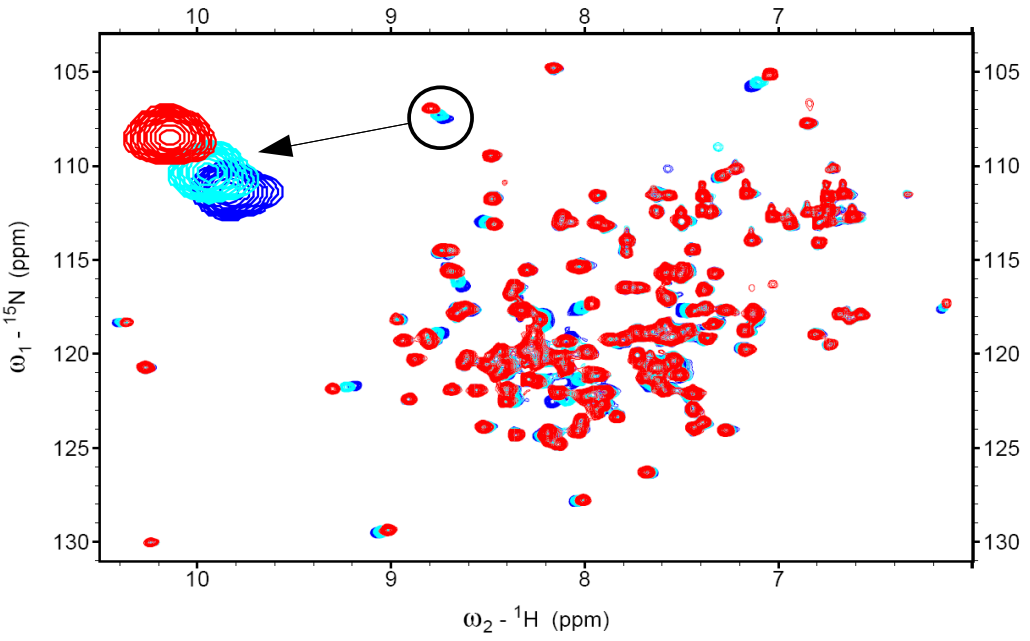


Figure 4.15: Overlay of three $^{1}H,^{15}N$-HSQC NMR spectra. It is obvious that addition of ligand F to the TPR2B domain of Sti1 results in moderate shifts of several protein signals (see e.g. the enlarged Sti1 TPR2B peak).

Blue: reference spectrum (only 200 µM of protein is present); cyan: first ligand titration step (200 µM of receptor and 0.2 mM of ligand F); red: sixth ligand titration step (200 µM of protein and 1.2 mM of ligand F).

The Sti1 TPR2B concentration (200 µM in phosphate buffer) was kept constant in each titration step; the final ligand concentrations were: 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, and 1.4 mM. To exclude artifacts caused by DMSO, the same amounts of solvent as used in the ligand titration steps were added to a new protein sample; the resulting chemical shift perturbations were subsequently applied as correction terms. In figure 4.15, an overlay of three $^{1}H,^{15}N$-HSQC spectra is shown (for explanations, see the figure caption). Due to the arising of several ligand induced signal shifts, it became obvious that ligand F is able to interact with the TPR2B domain of Sti1. Since no NMR backbone assignment of the utilized protein was present, and thus no informations about the origin of shift perturbations - direct binding or allosteric effects - were available, only a qualitative analysis of the titration steps and an estimation of the protein-ligand dissociation constant was possible. Addition of molecule F led to chemical shift perturbations of about 20 HSQC peaks from which 16 resonances could be used for the computation of a mean k_D. For this purpose, the absolute values of the chemical shift difference vectors which represent a function of the dissociation constant were calculated. Finally, the resulting 16 kD values were arithmetically averaged.[48]

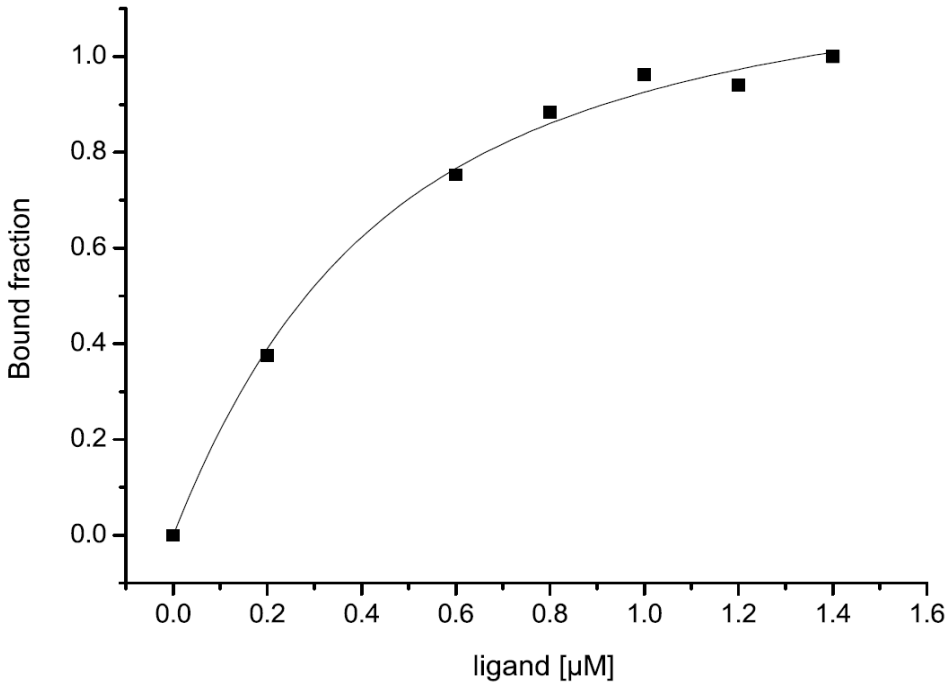


Figure 4.16: Illustration of the binding curve for one arbitrarily selected $^{1}H,^{15}N$-HSQC protein peak. Seven ligand titration steps were performed. For a ligand F concentration of 0.4 µM, no precise data point could be extracted due to NMR experimental uncertainties. By means of the shown curve, a k_D value of ca. 300 µM could be extracted. The detailed measurement conditions are described in the main text and in the caption of figure 4.15.

The estimations for some residues revealed k_D values of about 900 µM, others result in ca. 170 µM. The majority of evaluated shifts have k_D values of about 300 µM. The mean dissociation constant is averaged to a value of 435 µM.

4.2.7 Conclusion

By means of the three-dimensional structure of the HSP90 derived MEEVD peptide complexed to the Hop TPR2A domain, a pharmacophore model could be designed which served as search mask for screening several virtual compound libraries with in total more than 380,000 members. Out of this huge number of molecules, 70 compounds matched all requirements of the pharmacophore model. Since both redundant and improperly seeming scaffolds were discarded, 36 molecules remained which were virtually docked to the TPR2A receptor. By means of a visual inspection of the docking modes, an energetic evaluation, and RMSD clustering, six compounds were elected for NMR spectroscopic binding studies. Since two molecules were not soluble in water and DMSO, only four ligands could be tested. All of them revealed a weak affinity to the utilized Sti1 TPR2B protein domain. For compound F, the best binder, a dissociation constant of ca. 435 µM was estimated.

It can be concluded that virtual screening is able to identify molecules which exhibit (weak) affinity to a target molecule of interest. As this method is accompanied by computational approximations, e.g. no explicit consideration of protein dynamics or solvent effects, and depends on the quality of the provided 3D structures, the estimated affinity of the new-found Sti1 ligand is higher than it is usually expected and in a similar range like compounds recently published for Hop binding.[49] Possibly, the affinity of ligand F may be even higher if the TPR2A domain of Hop is used for experimental studies.

For the creation of the pharmacophore model, only parts of the MEEVD TPR2A interaction mode was taken into consideration. Thus, a further virtual search that is based on so far not utilized binding informations may result in a second new ligand. By covalently linking the two compounds without altering the protein ligand interactions, a medium-sized molecule could be designed which exhibits a strong TPR2A / TPR2B protein domain affinity. This approach, called fragment based ligand design, was nicely described in literature.[50-52]

4.2.8 Reagents, Methods, and Experiments

For the design of the pharmacophore model and the virtual screening procedure, the UNITY program suite embedded in the SYBYL 7.3 modeling package was utilized (www.tripos.de). During the 3D search, no additional filter options were applied. The screened databases were offered by Asinex

Ltd., Tripos L.P., Ambinter SARL, InterBioScreen Ltd., ChemStar, and ACD Labs Inc. Matching compounds were saved in molecular spreadsheets and visually inspected. For the subset of 36, AutoDock3 docking runs which rely on a grid based algorithm were performed.[53] For all ligand molecules, root atoms and rotatable bonds were defined and the compounds were provided with Gasteiger charges. The 3D structure of the TPR2A binding site was first prepared with the Sybyl 7.3 “Structure Preparation” tool and finally provided with Kollman charges. The utilized grid dimensions were 68 × 64 × 60 points with a spacing of 0.375 Å. Maps were created for the following atom types: C (non-aromatic), C (aromatic), H, O, N, S, P, F, and Cl. Docked structures were clustered with a RMSD value of 2 Å. Docking energies were automatically calculated and are given in kJ / mol.

The ^{15}N labeled and His-tagged Sti1 TPR2B domain was overexpressed in *E. coli* and purified via FPLC (equipped with a nickel column). The protein was dissolved in 50 mM potassium phosphate buffer, including 50 mM KCl and 1 mM tris(2-carboxyethyl)phosphine (TCEP); the pH of the sample solution was 7.5. The six chosen compounds were ordered via eMolecules Inc. (www.emolecules.com) and utilized in the NMR binding studies as delivered (HPLC purified and lyophilized powders). The concentration of the d_6-DMSO stock solution of ligand F was ca. 360 mM.

All NMR measurements were performed on a Bruker 600 MHz spectrometer equipped with a TXI cryogenic probe with z-gradients. A standard Bruker HSQC pulse program using gradient pulses and a common water suppression pulse scheme was applied.[54-57] The temperature for all experiments was 293 K. 1024 data points were recorded in the direct dimension, 90 in the indirect dimension. The spectral widths for proton detection was 13 ppm, for nitrogen detection 32 ppm with a center frequency of 117 ppm.

4.2.9 Declaration

The presented topic was conducted in collaboration with M.Sc. Stephan Lagleder (Technical University Munich, Department Chemistry). S.L. prepared the Sti1 TPR2B domain, prepared all ligand samples, and performed all NMR spectroscopic investigations. The described results have not yet been published. Because of this, no explicit details about the utilized compounds are provided throughout the description of this project.

4.3 Identification of Amyloid Peptide Binders via QSAR Based Methods

Three-dimensional quantitative structure activity relationship (3D-QSAR) is a method for predicting the biological activity of a compound only by help of the knowledge of its structure. For this to work, a general relationship between activity and structure is derived from a compilation of molecules with known activity - the training set. After an evaluation of the universal applicability of the derived relationship by statistical means, the biological activity of novel compounds can be predicted. The most striking advantage over conventional computational activity assays is the possibility to predict ligand activities even if information about the 3D receptor structure lacks. This feature renders QSAR a supplementary method to procedures that are dependent on detailed insight into a receptor conformation (e.g. molecular docking). Apart from that, the *in silico* nature of a QSAR study makes it a fast and cheap method in comparison to experimental screening setups.[58]

4.3.1 The Scientific Question

Alzheimer's disease (AD), a major cause of dementia, afflicts more than thirty million people worldwide.[59] Although enormous efforts have been made to elucidate the molecular cause of the disease, AD still remains incurable and terminal. One of the characteristics of AD is the presence of senile plaques that are associated with neurotoxicity.[59] Whereas several imaging compounds have been proposed for plaque-tracing (for details, see section 4.3.2), more sensitive and selective amyloid peptide binders are needed to improve the early diagnosis of AD. However, the search for improved tracers is very difficult due to little information about the structure of the plaques *in vivo*. The poor water solubility of the plaque-building peptides prevents their adequate conformational analysis by methods like NMR spectroscopy.

Therefore, the aim of this project was to find out whether it is possible to design novel amyloid plaque tracers without having knowledge about the 3D structure of the receptor molecules.

For this purpose, the binding affinity of 65 compounds was elucidated by an experimental setup. The structures and binding affinities of these compounds

was then used to build appropriate training sets for a QSAR studies. After that, general structural properties for high biological activity were proposed, the affinity of novel compounds was predicted, and optimal ligand scaffolds were predicted.

4.3.2 Biological Background of Alzheimer's Disease

Nowadays, a combination of psychiatric evaluation and morphological imaging of the central nervous system atrophy by magnetic resonance imaging is the established diagnosis of Alzheimer's disease. But only neuropathological examinations of the brain from AD patients *post mortem* can give a definitive evidence of AD by demonstrating the presence of senile plaques containing β-amyloid (Aβ) aggregates and neurofibrillary tangles.[60] Aβ-deposits consist of peptides of 40 and 42 amino acids (Aβ40 and Aβ42), forming amyloid fibrils with a characteristic β-plated sheet structure.[61] This sheet structure represents the target for histopathological staining as well as tracers for non-invasive imaging. Non-invasive imaging agents could prove the hypothesis of generation of β-amyloid as a causative event in the development of AD, as well as assess the Aβ-load at a pre-symptomatic stage to provide a differential diagnosis of AD over other neurodegenerative disorders.[60] The regional distribution and concentration of Aβ-plaques can be measured by positron emission tomography (PET). For this, imaging agents are labeled with positron emitting radionuclides.

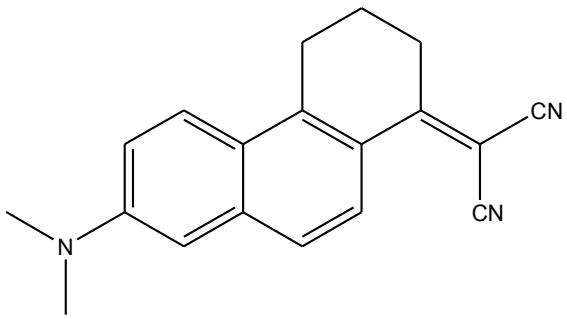


Figure 4.17: Illustration of a member of the FDDNP type ligand series which has been already used in clinical PET studies for tracing amyloid plaques.

The plaque binding radio-labeled Crysamine-G and Congo-Red agents were used as templates for the design of blood-brain barrier (BBB) crossing compounds for *in vivo* imaging.[62-67] Until now, five PET-ligands have been yielded for clinical studies: the thioflavin-T-derivative N-[^{11}C]methyl-6-OH-BTA-1 termed Pittsburgh Compound-B ([^{11}C]PIB), the ^{18}F-labeled benzothiazole

(BTA)-derivative 3′-[^{18}F]FPIB, an ^{11}C and an ^{18}F labeled version of the Congo-Red derivative SB-13, and the amino-naphthyl derivative [^{18}F]FDDNP (see also figure 4.17 that shows the constitution of a member of the FDDNP type ligand series).[59,60]

4.3.3 Design of Various QSAR Models

65 compounds were synthesized and tested for their binding to Aβ40 and Aβ42 plaques *in vitro*. The activities of the compounds are given in % inhibition - the percentage by which the PET signal of the reference structure called BTA1 (see Fig. 4.18) - decreases after the addition of another compound. All used molecules posses the same or a similar scaffold as BTA1.[68]

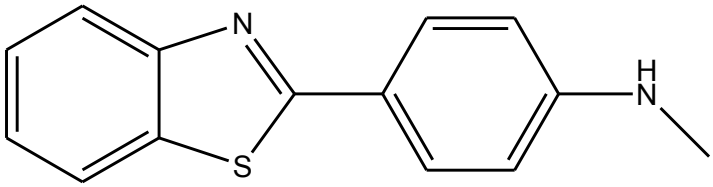


Figure 4.18: Illustration of BTA1, the reference compound used for the activity assay and for the structural QSAR alignment.

As mentioned in chapter 1.2.3, the composition of a training set is crucial for the performance of subsequent predictions by the QSAR model. Therefore, training set compounds were carefully selected from the pool of 65 compounds. As the atomistic structures of the receptor molecules (Aβ plaques) are not known, “active” three-dimensional conformations of the training set compounds could not be determined (as it could be done e.g. via molecular docking). Hence, the basic scaffolds of all compounds were used as fully planar structures and - if possible - all substituents were oriented in a uniform way. Subsequently, the ligand molecules were provided with partial charges (Gasteiger-Hückel) and spatially aligned via a least-square-fit method.[69]

In order to create a meaningful QSAR model, two completely different approaches for assembling the training set compounds were tested. First, only those molecules (out of the 65 available ones) were selected for one training set if they can contribute to optimize the cross-validated q^2 and r^2 statistical values (denoted as training set A). Secondly, a further training set (named as B) was designed since it is known that the mentioned statistical values are not always an adequately measure for the external predictability of a QSAR model (see also e.g. section 1.2.3).[70] For this purpose, compounds were selected in

a more “reasonable” way, thereby not considering the effect of the chosen set on the q^2 and r^2 statistical values. In detail, the final training set B consists of an equal number of compounds that exhibit both very low, very high, and moderate receptor affinity values. The standard error of prediction (SEP) and the standard error of estimate (SEE) are statistical values listing the RMS difference between actual (experimental) and either predicted (cross-validated) or fitted (non-cross-validated) y (here: % inhibition) values. Both are calculated analog to the RMSD computations (see also section 4.3.8). A tabular overview of the created training sets is given in figure 4.19.

Statistics	Set A	Set B
Available Compounds	65	
Selected Molecules	46	33
CoMSIA Fields	s / e / h / a	s / e / h / a
Regression Type	PLS	PLS
Aβ40 LOO Cross-Validation (q^2) *	0.655	0.511
Aβ42 LOO Cross-Validation (q^2) *	0.735	0.732
Aβ40 Coefficient of Determination (r^2) *	0.880	0.845
Aβ42 Coefficient of Determination (r^2) *	0.932	0.922
Aβ40 LOO SEP *	15	21
Aβ42 LOO SEP *	15	20
Aβ40 PLS SEE *	9	12
Aβ42 PLS SEE *	8	11

Figure 4.19: Table of the most important statistics of the training sets that were established for the QSAR study. The meaning of all values is explained in the main text.

* All statistical data are given in inhibition percentage points.

CoMSIA field description: s: steric; e: electrostatic; h: hydrophobic; a: acceptor.

Statistics: PLS: partial-least-square fit; LOO: leave-one-out based cross-validation; SEP: standard error of prediction; SEE: standard error of estimate.

In both training sets, the q^2 and r^2 statistical values for the Aβ42 predictability exceed the respective values for the Aβ40 predictability; this tendency is accompanied by slightly higher errors (SEP and SEE) for the Aβ40 predictions. The described deviation between Aβ40 and Aβ42 can be attributed to the training set composition. Obviously, the used molecules cover a wider range of % inhibition values for Aβ42 plaques than for Aβ40 peptides. The q^2 and r^2 values of the two sets are indicative of statistical relevance as they all exceed the QSAR typical significance threshold of 0.5. This means that the linear dependence of the variance of the predicted % inhibition on the descriptors per molecule is at least 50 %. Except for Aβ40 inhibition, all QSAR models yield excellent q^2 and r^2 values of > 0.7 and > 0.8, respectively.

If one compares both training sets, statistics indicate that set A should result in a better predictability due to higher q^2 and r^2 and lower SEP and SEE values than set B. However, the ability to predict novel compounds (external predictability) has to be evaluated by using a test set of "unknown" ligands (not present in the training set but with known biological activity).[70]

4.3.4 Examination of the Predictability Power via Test Set Ligands

The most reliable way to test the quality of a QSAR model is to predict the receptor affinity of both training set (internal predictability) and novel compounds (external predictability).[70] For this purpose, uniform test sets containing 46 (set A) or 33 (set B) training set compounds plus each time five "unknown" molecules were used. For both models, the % inhibition values (for Aβ40 and Aβ42 plaques) of the chosen compounds was predicted (see Fig. 4.20).

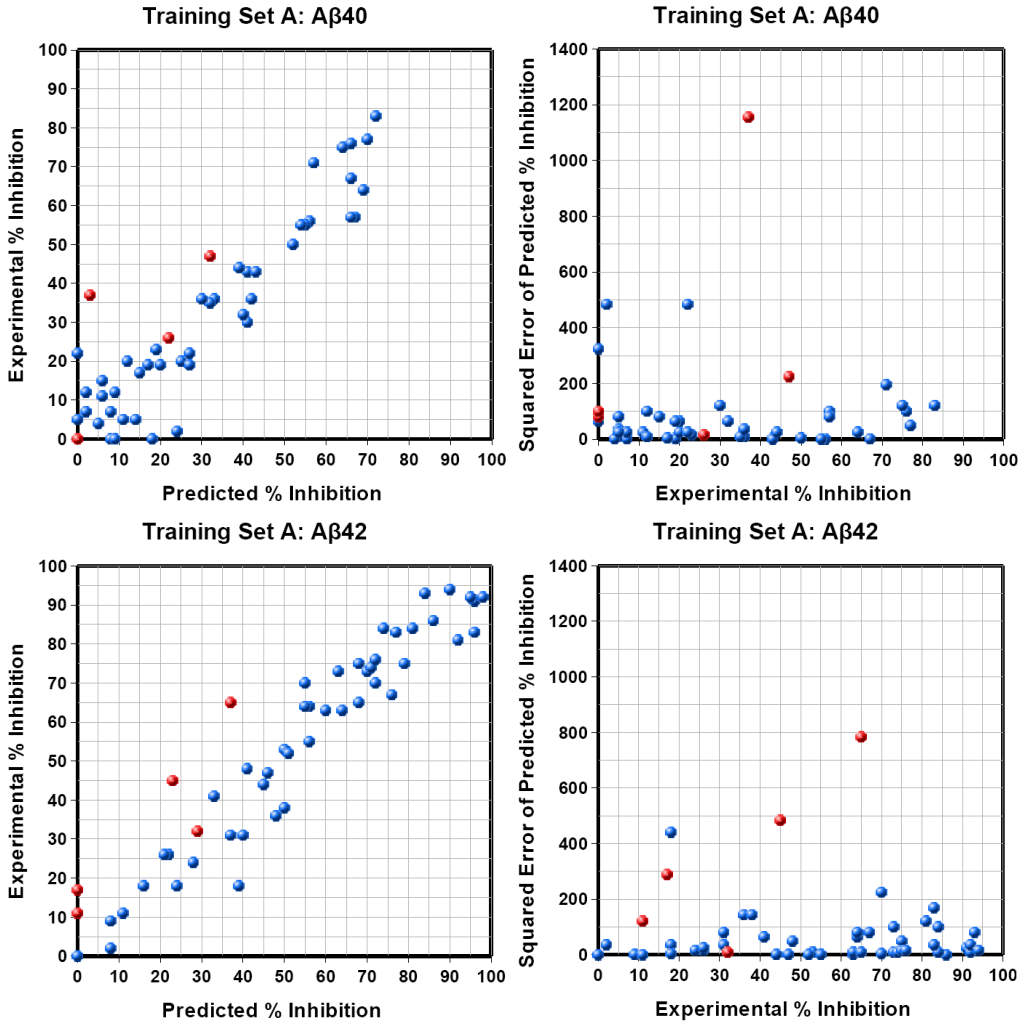
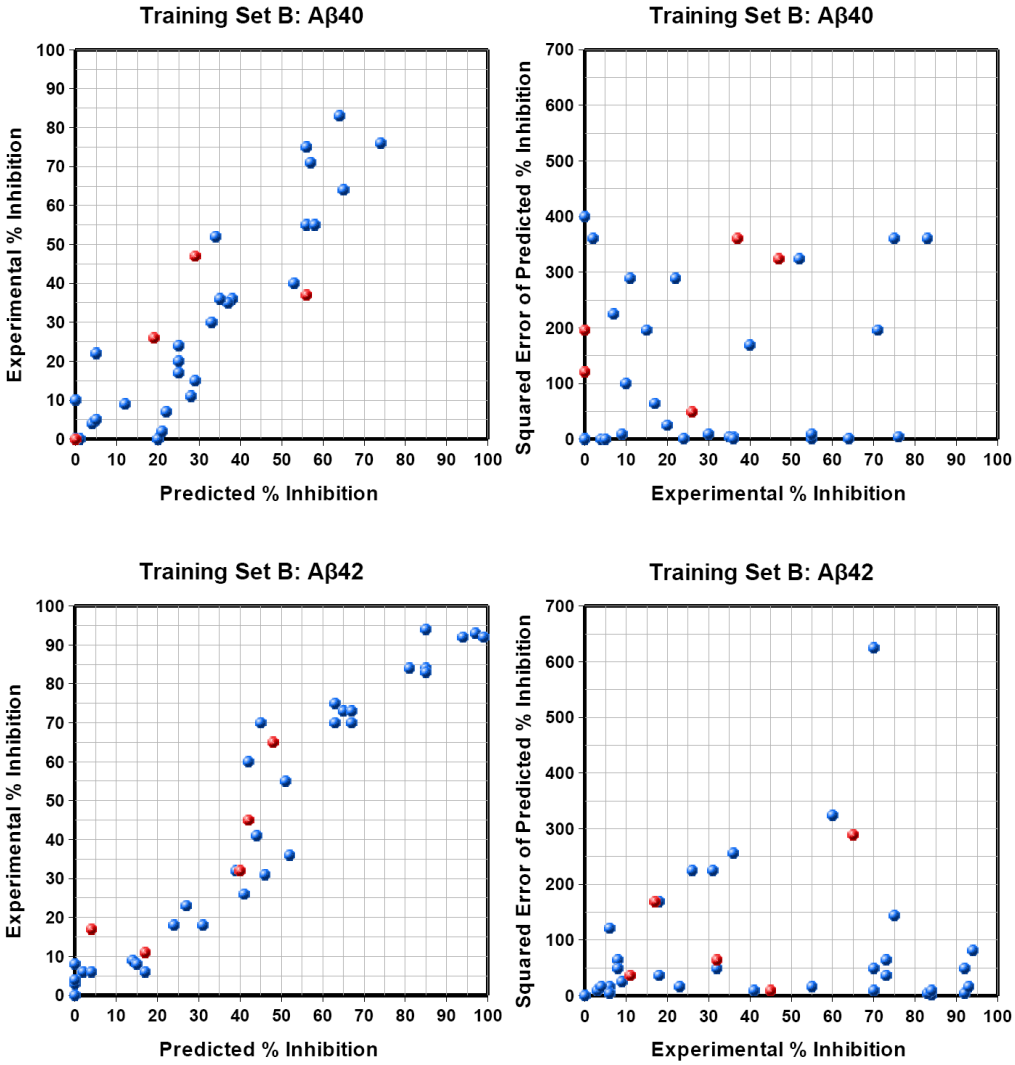


Figure 4.20: Illustration of plots showing correlations between experimental and predicted % amyloid peptide inhibition values for the two training sets A and B, thereby regarding the predictability of Aβ40 and Aβ42 peptide plaque affinities. The squared errors of the shown correlations are also illustrated. The analysis of the plots is given in the main text.

Blue colored data points represent molecules which are present in the training sets; data points colored in red belong to „unknown" compounds.



Continuation of Fig. 4.20

Considering the graphs shown in figure 4.20 (blue colored data points represent molecules from the training set, red colored ones belong to "unknown" compounds), it becomes obvious that most of the data points approximately lie on the diagonals (perfect correlation) of the quadratic plots; this observation is especially true for the prediction of training set compounds which indicates an appropriate internal predictive power of the two QSAR models A and B. As expected, training set A shows better overall predictabilities for already known compounds with the prediction of Aβ42 affinities being slightly more precise (see e.g. the "squared error" plots in figure 4.20). In contrast, training set B is superior with respect to the external predictability which is the more important parameter for evaluating the quality of designed QSAR models. Again, the prediction of Aβ42 affinities is slightly better than the Aβ40 related predictions of % inhibition values (see again the "squared error" plots in figure 4.20).

By calculating the errors of prediction in terms of RMSD values, the internal and external predictabilities of the two different CoMSIA based QSAR models can be evaluated.[70] In this scope, it is important to compare the overall RMSD values of prediction of training set compounds with those of the compounds that are not in the training set: the considered training sets represent a robust QSAR model if the error for the "unknown" compounds is roughly the same as for the training set compounds. In this case, the respective QSAR model provides the possibility to precisely predict novel compounds.[70]

Statistics	RMSD (Known)		RMSD (Unknown)	
	Aβ40	Aβ42	Aβ40	Aβ42
Set A	8	7	18	18
Set B	10	9	14	11

Figure 4.21: Table of RMSD values of the predictabilites of the training sets A and B.

* All values are given in % inhibition.

Known: RMSD values for all training set compounds; Unknown: RMSD values for compounds that are not in the training set.

As can be seen in figure 4.21, the RMSD values for predicting training set compounds and unknown compounds extensively differ (> 10 inhibition % points) for training set A. Hence, the overall robustness of this QSAR model can be supposed to be poor; this becomes also obvious when analyzing the graphical plots shown for model A (see e.g. Fig. 4.20; the data points for the unknown compounds shown in red exhibit larger distances from the diagonal than the data points of the training set compounds shown in blue). In contrast, the RMSD difference between internal and external predictabilities for set B is low (< 5 inhibition % points), again e.g. visualized by similar distances of "known" and "unknown" test set molecule data points from the diagonal (see again Fig. 4.20). Hence, training set B - although having inferior q^2 and r^2 statistical values to training set A and a worse internal predictability - is the more robust QSAR model and allows a more precise prediction of new amyloid peptide plaque binders. However, the quality of the model is not sufficient to precisely predict the Aβ40/42 selectivity of novel compounds due to too high RMSD % inhibition values (see Fig. 4.21).

4.3.5 Propositions for More Effective Amyloid Tracer Molecules

By means of the designed and further on evaluated training set B, physical and chemical features of novel compounds for being highly affine amyloid peptide plaque binders can be predicted with an adequate precision.

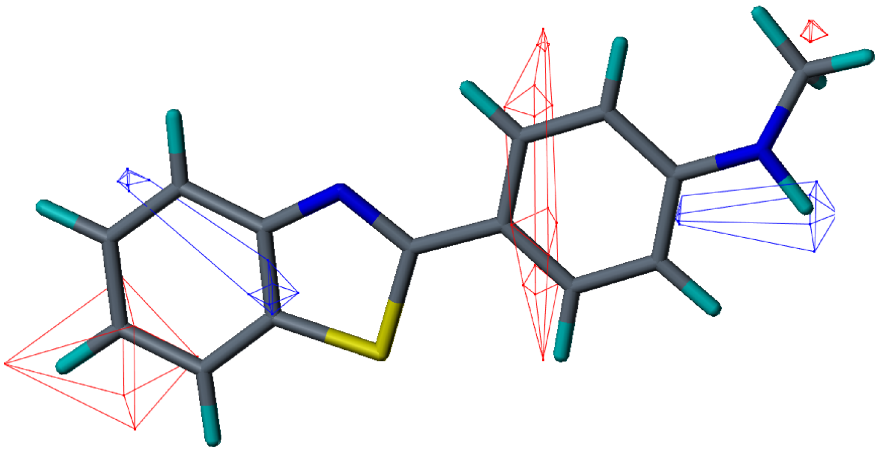


Figure 4.22: Illustration of training set B based QSAR fields projected onto the structure of the reference compound BTA1. As an example, favored areas for positive (red) or negative (blue) partial charges are shown.

For this purpose, QSAR fields are usually projected onto the scaffolds of the utilized training set compounds (see Fig. 4.22 and its caption for color coding informations). Analyzing the created fields as shown above, the computational propositions reveal that the amyloid peptide plaque affinity of the reference molecule BTA1 can be triggered by modifying both its basic scaffold and the its substituents with respect to electrostatic features. For instance, electron-poor groups should be added as substituent to or incorporated into the phenyl ring of the benzothiazole moiety. In addition, the proton of the secondary amino group should be replaced by a more electron-rich chemical group. In order to verify these QSAR suggestions, it is necessary to synthesize and biologically test respective compounds.

4.3.6 Conclusion

Summing up, two QSAR models for predicting the affinity of new amyloid peptide binders were created based on different strategies for establishing compound training sets. In general, both models show very good statistics with respect to high q^2 and r^2 values. However, predictions of QSAR evaluating test sets showed that both statistical values do not guarantee a proper external predictability of a model; for example, model “A” (based on training set A) exhibits unacceptable differences between the RMSD values of the whole test set and the RMSD values of “unknown” compounds. In contrast, the second model (derived from training set B) performed very well in predicting novel test set compounds, proven by a minimal difference between the two respective RMSD values. These results make clear that application of a practical test set including “unknown“ compounds is indispensable for an

evaluation of the predictive power of a QSAR model. Model “B” could be used to predict the affinity of novel compounds that possess a similar basic scaffold as the training set compounds. However, the error of prediction is still too large to differentiate between Aβ40 and Aβ42 selective plaques binders.

4.3.7 Reagents, Methods, and Experiments

All 65 compounds used for QSAR studies were synthesized by the help of commercially available, standard chemical building blocks. Biological testing for amyloid peptide plaque binding will be soon described elsewhere.

All computational operations were carried out on a Suse Linux operating system, using the software suite Sybyl8 (www.tripos.de). QSAR models were derived from CoMSIA fields (used fields: sterics, electrostatics, hydrophobics, H-bond donors, and H-bond acceptor) applying default QSAR settings if not stated otherwise (probe atom with radius: 1 Å, charge: +1, hydrophobicity: +1, hydrogen bond donating: +1, hydrogen bond accepting: +1; grid spacing: 2.0 Å; attenuation factor: 0.3). All molecules were created using the “Sketch Tool” of Sybyl8. The compounds were aligned to the reference molecule BTA1 (see Fig. 4.18) using an 3D alignment tool from the GROMOS96 MD simulation package. In order to achieve a reproducible alignment, residues of the planar molecules were uniformly positioned by modifying the torsional angles. For deriving the final structure activity relationship (SAR), a non-validated partial-least-squares (PLS) regression was applied, finally yielding r^2 statistical values. Cross-validated q^2 values were obtained through leave-one-out (LOO) cross-validation. With the calculated cross-validated coefficient q^2 in hand, the optimum number of QSAR components was obtained. Finally, the components were used for the subsequent partial-least-square regression (no validation). The column filtering box was kept unchecked during all operations. RMSD values were calculated as follows:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{n} (\%actual_i - \%predicted_i)^2}{n}}$$

with n being the number of compounds used for creating the models, %actual being the experimental % amyloid plaque inhibition values, and %predicted being the QSAR based predicted of % inhibition values.

4.3.8 Declaration

The presented topic was conducted in collaboration with B.Sc. Matthias Friedrich (Technical University Munich, Department Chemistry), Dr. Behrooz Yousefi, and Prof. Gjermund Henriksen (Technical University Munich, Klinikum rechts der Isar). B.Y. synthesized all used compounds and was responsible for biological testing. G.H. helped in obtaining precise % plaque inhibition values. M.F. contributed to the design of appropriate training sets, to test new ligand molecules, and to calculate respective QSAR statistics.

The described results are in preparation for publication. Because of this, no details about the utilized compounds and the exact localization of the created QSAR fields are provided throughout the description of this project.

4.4 References

[1] J. Klages, M. Coles, H. Kessler, *Analyst* **2007**, *132*, 693-705.

[2] S. B. Shuker, P. J. Hajduk, R. P. Meadows, S. W. Fesik, *Science* **1996**, *274*, 1531-1534.

[3] C. Dalvit, P. E. Fagerness, D. T. Hadden, R. W. Sarver, B. J. Stockman, *J. Am. Chem. Soc.* **2003**, *125*, 7696-7703.

[4] B. Meyer, T. Peters, *Angew. Chem. Int. Ed.* **2003**, *42*, 864-890.

[5] M. Mayer, B. Meyer, *J. Am. Chem. Soc.* **2001**, *123*, 6108-6117.

[6] K. E. Kövér, G. Batta, *J. Magn. Reson.* **2004**, *170*, 184-190.

[7] C. Dalvit, M. Flocco, M. Veronesi, B. J. Stockman, *Comb. Chem. High Throughput Screen.* **2002**, *5*, 605-611.

[8] T. Tengel, T. Fex, H. Emtenas, F. Almqvist, I. Sethson et al., *Org. Biomol. Chem.* **2004**, *2*, 725-731.

[9] S. Vanwetswinkel, R. J. Heetebrij, J. van Duynhoven, J. G. Hollander, D. V. Filippov et al., *Chem. Biol.* **2005**, *12*, 207-216.

[10] W. Jahnke, *J. Biomol. NMR* **2007**, *39*, 87-90.

[11] M. Coles, M. Heller, H. Kessler, *Drug Discov. Today* **2003**, *8*, 803-810.

[12] B. Luy, A. Frank, H. Kessler, *Conformational analysis of drugs by NMR: Drug Properties: Measurement and Computation*, Wiley-VCH Weinheim, **2008**.

[13] F. Manzenrieder, A. O. Frank, T. Huber, C. Dorner-Ciossek, H. Kessler, *Bioorg. Med. Chem.* **2007**, *15*, 4136-4143.

[14] B. Turk, *Nat. Rev. Drug Discov.* **2006**, *5*, 785-799.

[15] P. A. Bartlett, M. A. Giangiordano, *J. Org. Chem.* **1996**, *61*, 3433-3438.

[16] M. Collinsova, J. Jiracek, *Curr. Med. Chem.* **2000**, *7*, 629-647.

[17] G. R. Stark, P. A. Bartlett, *Pharmacol. Ther.* **1983**, *23*, 45-78.

[18] M. Cushman, J. T. Mihalic, K. Kis, A. Bacher, *J. Org. Chem.* **1999**, *64*, 3838-3845.

[19] T. Yokota, K. Konno, S. Shigeta, A. Holy, J. Balzarini et al., *Antivir. Chem. Chemother.* **1994**, *5*, 57-63.

[20] L. M. Elphick, S. E. Lee, V. Gouverneur, D. J. Mann, *ACS Chem. Biol.* **2007**, *2*, 299-314.

[21] D. Lesuisse, G. Lange, P. Deprez, D. Benard, B. Schoot et al., *J. Med. Chem.* **2002**, *45*, 2379-2387.

[22] P. Gettins, *J. Biol. Chem.* **1988**, *263*, 10208-10211.

[23] H. Umezawa, *Meth. Enzymol.* **1976**, *45*, 678-695.

[24] B. W. Matthews, *Acc. Chem. Res.* **1988**, *21*, 333-340.

[25] P. A. Bartlett, C. K. Marlowe, *Biochemistry* **1987**, *26*, 8553-8561.

[26] D. A. Erlanson, R. S. McDowell, T. O'Brien, *J. Med. Chem.* **2004**, *47*, 3463-3482.

[27] R. A. E. Carr, M. Congreve, C. W. Murray, D. C. Rees, *Drug Discov. Today* **2005**, *10*, 987-992.

[28] H. Kessler, C. Griesinger, J. Zarbock, H. R. Loosli, *J. Magn. Reson.* **1984**, *57*, 331-336.

[29] F. Manzenrieder, A. O. Frank, H. Kessler, *Angew. Chem. Int. Ed.* **2008**, *47*, 2608-2611.

[30] P. Schneider, Y. Tanrikulu, G. Schneider, *Curr. Med. Chem.* **2009**, *16*, 258-266.

[31] M. C. Hutter, *Curr. Med. Chem.* **2009**, *16*, 189-202.

[32] H. M. Sun, *Curr. Med. Chem.* **2008**, *15*, 1018-1024.

[33] G. Klebe, *Drug Discov. Today* **2006**, *11*, 580-594.

[34] W. L. Jorgensen, *Science* **2004**, *303*, 1813-1818.

[35] C. Scheufler, A. Brinker, G. Bourenkov, S. Pegoraro, L. Moroder et al., *Cell* **2000**, *101*, 199-210.

[36] S. Y. Chen, V. Prapapanich, R. A. Rimerman, B. Honore, D. F. Smith, *Mol. Endocrinol.* **1996**, *10*, 682-693.

[37] M. Lassle, G. L. Blatch, V. Kundra, T. Takatori, B. R. Zetter, *J. Biol. Chem.* **1997**, *272*, 1876-1884.

[38] J. Demand, J. Luders, J. Hohfeld, *Mol. Cell. Biol.* **1998**, *18*, 2023-2028.

[39] J. C. Young, W. M. J. Obermann, F. U. Hartl, *J. Biol. Chem.* **1998**, *273*, 18007-18010.

[40] A. Carrello, E. Ingley, R. F. Minchin, S. Tsai, T. Ratajczak, *J. Biol. Chem.* **1999**, *274*, 2682-2689.

[41] B. Honore, H. Leffers, P. Madsen, H. H. Rasmussen, J. Vandekerckhove et al., *J. Biol. Chem.* **1992**, *267*, 8485-8491.

[42] D. F. Smith, W. P. Sullivan, T. N. Marion, K. Zaitsu, B. Madden et al., *Mol. Cell. Biol.* **1993**, *13*, 869-876.

[43] A. K. Das, P. T. W. Cohen, D. Barford, *EMBO J.* **1998**, *17*, 1192-1199.

[44] J. R. Lamb, S. Tugendreich, P. Hieter, *Trends Biochem. Sci.* **1995**, *20*, 257-259.

[45] S. C. Onuoha, E. T. Couistock, J. G. Grossmann, S. E. Jackson, *J. Mol. Biol.* **2008**, *379*, 732-744.

[46] S. Y. Chen, D. F. Smith, *J. Biol. Chem.* **1998**, *273*, 35194-35200.

[47] A. Brinker, C. Scheufler, F. von der Mulbe, B. Fleckenstein, C. Herrmann et al., *J. Biol. Chem.* **2002**, *277*, 19265-19275.

[48] J. Cavanagh, W. J. Fairbrother, A. G. I. Palmer, N. J. Skelton, *Protein NMR Spectroscopy: Principles and practice*, Elsevier Amsterdam, **1996**.

[49] F. Yi, L. Regan, *ACS Chem. Biol.* **2008**, *3*, 645-654.

[50] H. Jhoti, *FASEB J.* **2007**, *21*, A209-A209.

[51] G. Siegal, E. Ab, J. Schultz, *Drug Discov. Today* **2007**, *12*, 1032-1039.

[52] M. Congreve, G. Chessari, D. Tisi, A. J. Woodhead, *J. Med. Chem.* **2008**, *51*, 3661-3680.

[53] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart et al., *J. Comput. Chem.* **1998**, *19*, 1639-1662.

[54] M. Piotto, V. Saudek, V. Sklenar, *J. Biomol. NMR* **1992**, *2*, 661-665.

[55] V. Sklenar, M. Piotto, R. Leppik, V. Saudek, *J. Magn. Reson. A* **1993**, *102*, 241-245.

[56] G. Bodenhausen, D. J. Ruben, *Chem. Phys. Lett.* **1980**, *69*, 185-189.

[57] S. Mori, C. Abeygunawardana, M. O. Johnson, P. vanZijl, *J. Magn. Reson. B* **1996**, *110*, 321-321.

[58] P. Gedeck, R. A. Lewis, *Curr. Opin. Drug Discov. Devel.* **2008**, *11*, 569-575.

[59] L. S. Cai, R. B. Innis, V. W. Pike, *Curr. Med. Chem.* **2007**, *14*, 19-52.

[60] G. Henriksen, B. H. Yousefi, A. Drzezga, H. J. Wester, *Eur. J. Nucl. Med. Mol. Imaging* **2008**, *35,* 75-81.

[61] R. N. Martins, P. J. Robinson, J. O. Chleboun, K. Beyreuther, C. L. Masters, *Mol. Neurobiol.* **1991**, *5*, 389-398.

[62] W. E. Klunk, M. L. Debnath, J. W. Pettegrew, *Neurobiol. Aging* **1994**, *15*, 691-698.

[63] W. E. Klunk, M. L. Debnath, J. W. Pettegrew, *Biol. Psychiatry* **1994**, *35*, 627-627.

[64] W. E. Klunk, M. L. Debnath, J. W. Pettegrew, *Neurobiol. Aging* **1995**, *16*, 541-548.

[65] W. E. Klunk, B. J. Bacskai, C. A. Mathis, S. T. Kajdasz, M. E. McLellan et al., *J. Neuropathol. Exp. Neurol.* **2002**, *61*, 797-805.

[66] N. A. Dezutter, R. J. Dom, T. J. de Groot, G. M. Bormans, A. M. Verbruggen, *Eur. J. Nucl. Med.* **1999**, *26*, 1392-1399.

[67] S. C. Styren, R. L. Hamilton, G. C. Styren, W. K. Klunk, *J. Histochem. Cytochem.* **2000**, *48*, 1223-1232.

[68] H. Toyama, D. Ye, M. Ichise, J. S. Liow, L. S. Cai et al., *Eur. J. Nucl. Med. Mol. Imaging* **2005**, *32*, 593-600.

[69] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, *J. Chem. Theory Comput.* **2008**, *4*, 435-447.

[70] A. M. Doweyko, *J. Comput. Aided Mol. Des.* **2008**, *22*, 81-89.

5 Summing Up and Looking Ahead

Conclusions and Perspectives

Before the results of the present work are summarized, the fruitful interplay of NMR spectroscopy and molecular modeling and its perspectives are shortly discussed. Most of scientific questions accompanying the distinct projects could be answered only by a combined application of NMR techniques and virtual approaches like MD simulations, virtual screening, or molecular docking. In the majority of presented topics, computational tools were utilized to explain NMR spectroscopic outcomes (e.g. RDC prediction) or to convert experimental data into descriptive models (e.g. conformational analysis). In contrast, NMR spectroscopy was also applied to prove the results of molecular modeling approaches (Hop protein ligands). With the aid of both techniques, a plethora of results could be yielded that may preciously contribute to biochemical and pharmacological research.

However, some of the presented findings revealed that the interplay of both methodologies is not perfect. Regarding NMR spectroscopy, especially the application of RDCs for flexible molecules and the estimation of preferred orientations of small and medium-sized molecules in alignment media need to be optimized. Whereas the latter topic was directly treated in chapter 3.2, huge efforts have already been made to accurately fit experimentally gained RDCs onto smaller dynamic molecules like peptides or short nucleic acid chains.[1,2] The advancement of molecular modeling techniques is also an active field of research. For example, many approaches like replica exchange, conformational flooding, local elevation, or coarse grained calculations have recently been developed (but still have to be optimized) to enhance the MD based conformational sampling.[3] The need for such new algorithms became obvious e.g. in chapters 3.1 and 3.3 where even very long MD simulations did not result in fully converged molecular ensembles of Cilengitide or cyclic penta-alanines. Another topic concerns molecular docking tools. As stated in chapter 4.2, virtual screening and docking usually elucidate protein ligands with at best millimolar affinity. The main reason for this drawback is that receptor dynamics are completely neglected in the available tools. Thus, very recently advanced algorithms have been presented which allow at least side chain rotations of receptor amino acids.[4]

In the following section, the results of the present thesis are concluded, thereby considering their impact on the distinct research fields.

1. The effect of proline and *N*-methylated amino acids in conformational design

The main interest in this chapter concerned the question if prolines can be replaced by *N*-methylated amino acids in a cyclic pentapeptide without altering its backbone conformation. Structural analyses via NMR spectroscopy, distance geometry, and MD simulations of various template peptides revealed that the cyclic backbone structure is not significantly modified when one proline in a cyclic pentapeptide is exchanged by one *N*-methylated residue. This outcome allows new strategies in conformational design of cyclic pentapeptides: whereas the utilization of proline, an excellent inducer of stable turn structures, results in the loss of one position where a pharmacophoric side chain group could be introduced; the application of an *N*-methylated amino acid does not lead to this drawback. Thus, it could be stated that *N*-methyl group carrying amino acids are “better” prolines. In contrast, two prolines can not be replaced by *N*-methylated residues without affecting a peptide conformation.

2. Configurational and conformational analysis by an RDC-only driven MD approach

The scientific question to be answered in this chapter was to evaluate the robustness and precision of RDC-only driven MD simulations applied for elucidating molecular structures. By the help of a new alignment medium, an accurate measurement of multiple residual dipolar couplings for the natural product staurosporine was possible. The RDCs were fitted onto different configurational and conformational arrangements via both free and restrained molecular dynamics simulations and an established, static approach (PALES). It could be shown in detail that the right three-dimensional structure of staurosporine was unambiguously identified by both, RDC-coupled and only force field driven MD simulations, thereby providing results of equal quality as the PALES method. Therefore, small and medium-sized organic compounds with unknown configuration and / or conformation can be studied in future with the aid of MD simulations in combination with residual dipolar couplings.

3. NMR based determination of the constitutions of complex bioinorganic compounds

In this section, the expected constitutions of three peptide analogs - parts of potential new biosensors - containing different metal ions were proved by NMR spectroscopic techniques. Two of the molecules comprised, in addition to a ferrocene group, more than ten linearly arranged methylene groups which had to be unequivocally assigned. The challenge in the study of the third compound was to substantiate the coordination of four ligand groups to a central ruthenium ion. Whereas the primary sequence of the ferrocene containing molecules was straightforward via routinely used homo- and heteronuclear experiments recorded at very high magnetic fields, the constitutional analysis of the ruthenium containing compound turned out to be more difficult. By means of proton-nitrogen coupled experiments and an assignment strategy based on exclusion of alternative metal coordination modes, the constitution of the ruthenium complex could be proved beyond doubt.

4. Conformation and dynamics of Cilengitide in different environments

The cyclic peptide Cilengitide *c(RGDf(N^{Me})V-)* is a super-active $\alpha_v\beta_3$ integrin inhibitor. Usually, very high receptor affinities are only obtained when the conformation of the ligand freely tumbling in solution excellently matches its structure in the bound state. Thus, the conformation and dynamics of Cilengitide in four polar and apolar solvents were studied. It could be shown via NMR spectroscopy and MD simulations that the integrin ligand exhibits in all cases almost the same, strongly preferred backbone structure. The investigation of dynamics revealed that the backbone arrangement is highly stable; the strongest flexibilities are found for three peptide bonds which can rotate by ca. 180°, thereby not affecting the overall conformation. A comparison of these findings with the Cilengitide integrin complex showed that the conformation of the ligand in the free state is very similar to the structure when being bound and that the observed dynamics do not interfere with the binding mode.

5. Prediction of residual dipolar couplings via molecular dynamics simulations

An *ab initio* estimation of residual dipolar couplings is of great scientific interest; first, the prediction of RDCs via realistic models would give detailed insight in the interactions of alignment media and solute molecules; additionally, differing RDC signal sets of enantiomeric mixtures arising in chiral alignment media could be stereospecifically assigned to the chiral molecules which might allow an NMR based discrimination of enantiomers. For this purpose, a very simple model of the alignment medium polystyrene/$CDCl_3$ - a short atactic and z-oriented polystyrene chain with fixed backbone end atoms - was applied in stochastic dynamics simulations. The model enabled the computation of converged RDCs for the natural product strychnine which are in much better agreement with experimentally derived data than values obtained by an established steric prediction tool. Furthermore, it could be shown that the alignment of strychnine particularly results from interactions with the polymer backbone.

6. Correlating structural dynamics of peptides with their cellular permeabilities

In contrast to the rapidly increasing number of peptides with interesting biological properties, their application as promising drug candidates in pharmacological research is still limited as peptides are usually not orally available. However, it has recently been demonstrated that $_N$-methylation could trigger the paracellular uptake of cyclic peptides, and that template penta-alanines containing different numbers and patterns of $_N$-methylations exhibit varying cell passing rates. By the help of NMR spectroscopic and MD simulations, a possible correlation of structural features and molecular flexibilities with the extend of absorption could be identified; one backbone amide proton must be highly exposed to the solvent and does not significantly change its orientation over a certain period of time. This observation might be explained by the necessity that stable hydrogen bonds are formed between a peptide and tight-junction flanking proteins which regulate the paracellular transport.

7. Development and testing of a new ligand based NMR screening technology

NMR based ligand screening is a powerful technique for elucidating new drug candidates. However, current methods often suffer from signal overlap in the spectra (e.g. proton-detection) or need additional labeling of ligand molecules (e.g. fluorine screening). In order to overcome these obstacles, a new screening method was developed that is based on phosphorus detection. Since ^{31}P containing molecules usually result in only one NMR resonance, even huge compound libraries could be investigated. Moreover, phosphorus is often an intrinsic element of enzyme inhibitors mimicking the tetrahedral intermediate of a peptide bond hydrolysis; thus, interactions with proteases can be detected without auxiliary ligand labeling. The broad applicability of the new method was not only proved by basic 1D experiments but also via valuable extensions of the approach like signal recovery measurements, reporter based screening, and heteronuclear 2D NMR experiments.

8. Elucidation of a new Sti1 protein ligand via virtual screening and docking

Hop, an adaptor protein that belongs to the group of co-chaperones, mediates the association of Hsp70 and Hsp90. As Hsp90 is a key protein in cancer formation, inhibition of Hsp90-Hop complex formation may block the passage of Hsp90 client molecules and might therefore be a route to novel anti-cancer agents. By means of a pharmacophore model that was designed in accordance with the Hsp90-Hop binding mode, virtual compound libraries were screened for potential Hop binders. The number of found candidates was reduced by a virtual docking approach and visual inspection. Finally, the Hop affinities of several hits were evaluated by target based NMR screening experiments, thereby using the Hop related Sti1 protein as receptor. For the compound which shows the most promising results, a k_D value of ca. 400 µM was experimentally measured what is - compared to the outcomes of other virtual screening studies (usually, “millimolar binders” are found) - an outstanding result.

9. Creation of QSAR models for predicting new amyloid peptide plaque tracers

One of the characteristics of Alzheimer’s disease, a major cause of dementia, is the presence of senile plaques that are associated with neurotoxicity. In order to improve the early diagnosis of the disease, sensitive and selective amyloid peptide binders are needed for PET plaque-tracing. However, the search for improved tracers is very difficult due to little information about the structure of the plaques *in vivo*. Thus, different QSAR models based on a medium-sized compound library containing planar, aromatic molecules were designed. Since it could be clearly shown that the usually calculated statistical parameters are only a measure for the internal predictability of a model, an optimized QSAR training set was created by an individual selection strategy which also allows an excellent external predictability. With the aid of the computed, corresponding QSAR fields, the structural requirements of a compound for exhibiting high amyloid Aβ40 and Aβ42 plaque affinity could now be elucidated.

Final Overview

In order to give a shortly summarized overview of the scientific contributions yielded by this work, the most important and relevant results of the nine presented topics are encapsulated in one sentence as shown in the table below.

Project	Main Outcome
Conformational Design of Cyclic Pentapeptides	It could be shown that one proline can be replaced by one *N*-methylated amino acid in a cyclic pentapeptide without perturbing the overall conformation.
RDC-Only Based Structure Determination	RDC-only driven MD structure determination is well suited to elucidate both the configuration and the conformation of small or medium-sized organic compounds.
Constitutional Analysis of Bioinorganic Molecules	The molecular constitution of three metal-containing, bioinorganic compounds could be unambiguously determined by the help of homo- and heteronuclear NMR experiments.
Conformational Dynamics of an Integrin Inhibitor	Whereas the conformation of the integrin ligand Cilengitide is highly stable in polar and apolar solvents, its flexibility does not negatively affect the binding mode to the receptor.
Prediction of Residual Dipolar Couplings	With the aid of a simple model, RDCs for a natural product could be precisely predicted via MD simulations, thereby being superior to a well established static approach.
Cellular Permeabilities of Cyclic Template Peptides	Solvent accessibility and an adequate orientational stability of at least one backbone amide proton seems to be important for a peptide to display good cellular absorption.
Phosphorus NMR Ligand Screening	A new NMR based ligand screening technique especially qualified for finding protease inhibitors was introduced by using phosphorus nuclei detection.
Virtual Search for Hop / Sti1 Protein Binders	By means of virtual screening and docking, an organic compound that binds with a k_D of ca. 400 µM to the Hop related Sti1 protein was identified.
New Amyloid Plaque Tracer Molecules	A QSAR model of good quality was constructed that allows the prediction of structural requirements of compounds for exhibiting Aβ40 and Aβ42 amyloid plaque affinity.

References

[1] D. MacDonald, P. Lu, *Curr. Opin. Struct. Biol.* **2002**, *12*, 337-343.

[2] G. Kummerlöwe, B. Luy, *Trends Analyt. Chem.* **2009**, *28*, 483-493.

[3] W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen et al., *Angew. Chem. Int. Ed.* **2006**, *45*, 4064-4092.

[4] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew et al., *J. Comput. Chem.* **2009**, *in press.*

Appendix

Abbreviations & Experimental Data

Abbreviations

1D / 2D / 3D	one- / two- / three-dimensional
AD	Alzheimer's disease
AMP-PNP	adenylyl-imidodiphosphate
Aβ	amyloid beta (peptides)
BBB	blood-brain-barrier
C	carbon atom
c	cyclo
calc	calculated
Cbz	benzyloxycarbonyl
CC	computational chemistry
CCR	cross-correlated relaxation
CG	coarse-grained
CLIP-HSQC	clean-inphase-HSQC
COLOC	correlation through long-range coupling
CoMFA	comparative molecular field analysis
CoMSIA	comparative molecular similarity analysis
COSY	correlated spectroscopy
CPMG	Carr-Purcell-Meiboom-Gill (relaxation filter)
CS	chemical shift
CSA	chemical shift anisotropy
CT-COSY	constant-time COSY
d	douplet
dd	double douplet
DDD	distance-driven dynamics
DEPT	distorsion-less enhancement by polarization transfer (spectroscopy)
DG	distance geometry (calculations)
DMSO	dimethylsulfoxide
dPS	deuterated polystyrene
E.COSY	exclusive correlation spectroscopy
edHSQC	multiplicity edited-HSQC
EM	energy minimization
ESI-MS	electrospray-ionization mass spectrometry
et	exchange-transferred
exp	experimental

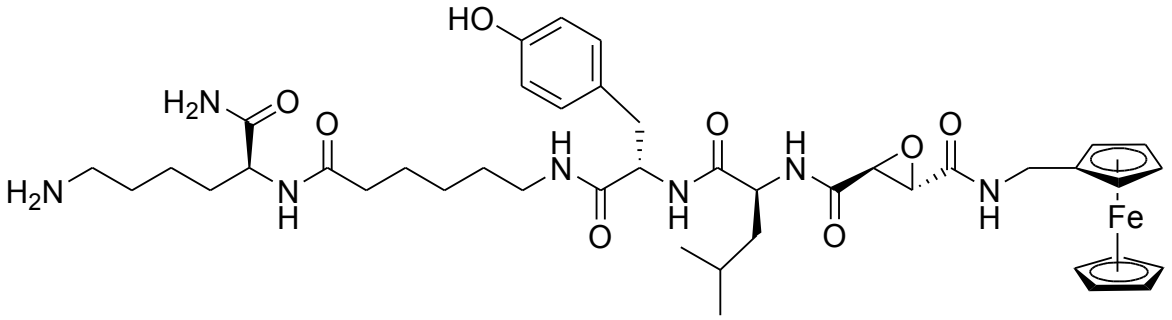
F	fluorine atom
F1 / F2	indirect / direct NMR dimension
FBDD	fragment based drug design
FCN	Glubb
FDA	food and drug administration
FDDNP	2-(1-{6-[(2-fluoroethyl(methyl)amino]-2-naphthyl}ethylidene)malononitrile
FF	force field
FID	free induction decay
fMD	free MD
FPLC	fast performance liquid chromatography
H	hydrogen atom
Het / het	heteronuclear
HETLOC	heteronuclear long range coupling
HMBC	heteronuclear multiple bond correlation
HMQC	heteronuclear multiple quantum coherence
HPLC	high performance liquid chromatography
HSQC	heteronuclear single quantum correlation
INADEQUATE	incredible natural abundance double quantum transfer experiment
ITC	isothermal calorimetry
IUPAC	International Union of Pure and Applied Chemistry
J	scalar (NMR) coupling
JAM	junction adhesion molecule
kD	dissociation constant
kDa	kiloDalton
LD	Langevin MD (simulation)
LE	local elevation
LOO	leave-one-out
Mann	mannitol
MD	molecular dynamics (simulation)
Me	methyl
MeOH	methanol
MIDAS	metal ion dependent adhesion site
MS	mass spectrometry
MW	molecular weight
N	nitrogen atom
NMR	nuclear magnetic resonance (spectroscopy)
NOE	nuclear Overhauser effect (peak volume)
NOESY	nuclear Overhauser effect spectroscopy
NpT	system with constant particle number, pressure, and temperature
NVT	system with constant particle number, volume, and temperature
O	oxygen atom
P	phosphorus atom
P.E.HSQC	primitive exclusive HSQC
PBC	periodic boundary conditions
PBLG	poly-γ-benzyl-L-glutamate
PCA	phosphorus containing analog
PDB	protein data bank
PDMS	polydimetyl
PET	positron emission tomography

PLS	partial least square
PME	particle mesh Ewald
ppm	parts per million
PS	polystyrene / phosphorylated substrate
q	quartet
QM	quantum-mechanical
QNP	quadruple probe head with direct detection
QSAR	quantitative structure acitivity relationship
QXI	quadruple probe head with inverse detection
R1 / R2	longitudinal / transversal relaxation rates
R1ρ	relaxation dispersion (measurement)
RDC	residual dipolar coupling
RDF	radial distribution function
RE	replica exchange
Rex	chemical exchange rate
RF	reaction-field
rMD	restrained MD
RMS	root mean square
RMSD / RMSF	root mean square deviation / fluctuation
ROE	rotating-frame Overhauser effect (peak volume)
ROESY	rotating-frame Overhauser effect spectroscopy
RT	room temperature
Ru	ruthenium atom (ion)
s	singulet
SA	simulated annealing
SAR	structure-activity-relationship
SD	stochastic MD (simulation)
SEE	standard error of estimation
SEP	standard error of prediction
SINA	the most beautiful girl ever seen
STD	saturation transfer difference (NMR spectroscopy)
SVD	singular value decomposition
t	triplet
T1 / T2	longitudinal / transversal relaxation times
TCEP	tris(2-carboxyethyl)phosphine
Test	testosterone
TINS	target immobilized NMR screening
TJ	tight-junction
TOCSY	total correlation spectroscopy
TPR	tetratricopeptide repeat
tr	transient
TXI	triple probe head with inverse detection
UV / VIS	ultraviolet / visible (spectroscopy)
vdW	van der Waals
WaterLOGSY	water ligand optimized gradient spectroscopy

Experimental Data

Below, additional experimental informations are given for those topics of this work which are not yet published. For chapters 4.2 and 4.3, detailed explanations are not presented due to non-disclosure responsibilities.

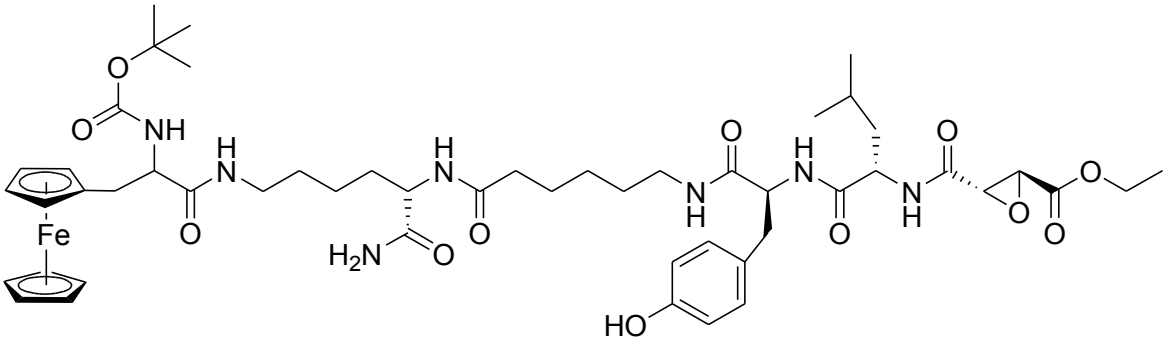
a) Chapter 2.3: NMR chemical shift assignment for the three presented bioinorganic molecules



^{1}H-NMR (900.13 MHz, d_6-DMSO): d = 9.15 (s, 1H, OH_{tyr}), 8.57 (bs, 1H, $NHCH_2Fc$), 8.47 (d, 1H, $^{3}J_{H\text{-}H}$ = 7.9 Hz, NH_{leu}), 8.03 (d, 1H, $^{3}J_{H\text{-}H}$ = 8.2 Hz, NH_{tyr}), 7.80 (d, 1H, $^{3}J_{H\text{-}H}$ = 8.2 Hz, NH_{lys}), 7.76 (m, 1H, NH_{ahx}), 7.68 (bs, 2H, $NH_{2,lys}$), 7.30 (bs, 1H, $CONH_2$), 6.98 (d, 2H, $^{3}J_{H\text{-}H}$ = 8.3 Hz, $CH_{tyr\text{-}bzl}$), 6.97 (bs, 1H, $CONH_2$), 6.64 (d, 2H, $^{3}J_{H\text{-}H}$ = 8.3 Hz, $CH_{Tyr\text{-}Bzl}$), 4.34 (m, 1H, αH_{tyr}), 4.32 (m, 1H, αH_{leu}), 4.21 (m, 2H, CH_{cp}), 4.19 (s, 5H, CH_{cp}), 4.18 (m, 1H, αH_{lys}) 4.11 (m, 2H, CH_{cp}), 4.07 (m, 1H, CH_2Fc), 4.04 (s, 1H, CH_2Fc), 3.62 (m, 1H, CH_{epx}), 3.54 (m, 1H, CH_{epx}), 3.03 (m, 1H, αH_{ahx}), 2.96 (m, 1H, αH_{ahx}), 2.83 (dd, 1H, $^{3}J_{H\text{-}H}$ = 14.3 Hz, 5.8 Hz, βH_{tyr}), 2.77 (t, 2H, $^{3}J_{H\text{-}H}$ = 6.3 Hz, εH_{lys}), 2.69 (dd, 1H, $^{3}J_{H\text{-}H}$ = 14.3 Hz, 8.9 Hz, βH_{tyr}), 2.12 (t, 2H, $^{3}J_{H\text{-}H}$ = 7.3 Hz, εH_{ahx}), 1.66 (m, 1H, βH_{lys}), 1.52 (m, 2H, δH_{lys}), 1.51 (m, 1H, γH_{leu}), 1.51 (m, 1H, βH_{lys}), 1.47 (m, 2H, δH_{ahx}), 1.42 (m, 1H, βH_{leu}), 1.38 (m, 1H, βH_{leu}), 1.33 (m, 2H, βH_{ahx}), 1.32 (m, 1H, γH_{lys}), 1.28 (m, 1H, γH_{lys}), 1.18 (m, 2H, γH_{ahx}), 0.87 (d, 1H, $^{3}J_{H\text{-}H}$ = 6.6 Hz, δH_{leu}), 0.83 (d, 1H, $^{3}J_{H\text{-}H}$ = 6.6 Hz, δH_{leu}). Chemical shifts are given in ppm.

^{13}C-NMR (226.34 MHz, d_6-DMSO): d = 174.3 (CO_{lys}), 172.7 (CO_{ahx}), 171.5 (CO_{leu}), 171.0 (CO_{tyr}), 166.1 (CO_{epx}), 165.7 (CO_{epx}), 156.2 ($C_{tyr\text{-}bzl}$-OH), 130.5 ($C_{tyr\text{-}bzl}$), 128.2 ($C_{tyr\text{-}bzl}$), 115.3 ($C_{tyr\text{-}bzl}$), 85.7 (C_{cp}-CH_2), 68.9 (C_{cp}), 68.5 (C_{cp}), 68.0 (C_{cp}), 54.8 (αC_{tyr}), 53.4 (C_{epx}), 53.1 (C_{epx}), 52.4 (αC_{lys}), 51.7 (αC_{leu}), 41.2 (βC_{leu}), 39.2 (εC_{lys}), 38.9 (αC_{ahx}), 38.3 (CH_2Fc), 37.4 (βC_{tyr}), 35.6 (εC_{ahx}), 31.9 (βC_{lys}), 29.2 (βC_{ahx}), 27.1 (δC_{lys}), 26.5 (γC_{ahx}), 25.4 (δC_{ahx}) 24.6 (γC_{leu}), 23.5 (δC_{leu}), 22.8 (γC_{lys}), 22.0 (δC_{leu}). Chemical shifts are given in ppm.

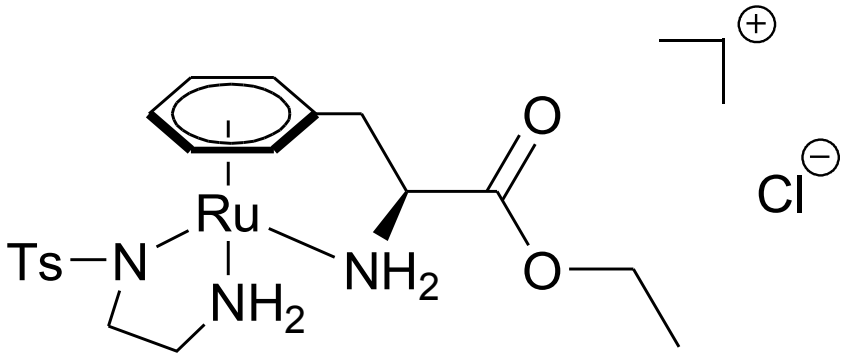
^{15}N-NMR (91.21 MHz, d_6-DMSO): d = 122.9 (N_{leu}), 122.7 ($NHCH_2Fc$), 121.7 (N_{lys}), 117.8 (N_{tyr}), 116.1 (N_{ahx}), 105.0 ($CONH_2$), not detected $NH_{2(lys)}$. Chemical shifts are given in ppm.



^{1}H-NMR (900.13 MHz, d_6-DMSO): d = 9.13 (s, 1H, OH_{tyr}), 8.51 (d, 1H, $^{3}J_{H-H}$ = 8.2 Hz, 25 NH_{leu}), 8.07 (d, 1H, $^{3}J_{H-H}$ = 8.3 Hz, NH_{tyr}), 7.75 (m, 1H, NH_{lys}), 7.74 (m, 1H, NH_{ahx}), 7.72 (bs, 1H, $NHl_{ys,s.c.}$), 7.26 (bs, 1H, $CONH_2$), 6.98 (d, 2H, $^{3}J_{H-H}$ = 8.3 Hz, $CH_{tyr-bzl}$), 6.92 (bs, 1H, $CONH_2$), 6.67 (bs, 1H, NHBoc), 6.63 (d, 2H, $^{3}J_{H-H}$ = 8.3 Hz, $CH_{Tyr-Bzl}$), 4.35 (m, 1H, αH_{tyr}), 4.34 (m, 1H, αH_{leu}), 4.22 (m, 1H, $COOCH_2CH_3$), 4.19 (m, 1H, $COOCH_2CH_3$), 4.16 (m, 2H, CH_{cp}), 4.15 (m, 1H, αH_{lys}), 4.12 (s, 5H, CH_{cp}), 4.04 (m, 2H, CH_{cp}), 3.93 (m, 1H, $FcCH_2CH$), 3.72 (d, 1H, $^{3}J_{H-H}$ = 1.4 Hz, CH_{epx}), 3.60 (d, 1H, $^{3}J_{H-H}$ = 1.4 Hz, CH_{epx}), 3.03 (m, 1H, αH_{ahx}), 3.01 (m, 2H, εH_{lys}), 2.97 (m, 1H, αH_{ahx}), 2.84 (dd, 1H, $^{3}J_{H-H}$ = 14.2 Hz, 5.6 Hz, βH_{tyr}), 2.70 (dd, 1H, $^{3}J_{H-H}$ = 14.2 Hz, 8.8 Hz, βH_{tyr}), 2.69 (m, 1H, $FcCH_2CH$), 2.54 (m, 1H, $FcCH_2CH$), 2.12 (t, 2H, $^{3}J_{H-H}$ = 7.5 Hz, εH_{ahx}), 1.63 (m, 1H, βH_{lys}), 1.52 (m, 1H, γH_{leu}), 1.49 (m, 1H, βH_{lys}), 1.48 (m, 2H, δH_{ahx}), 1.43 (m, 1H, βH_{leu}), 1.39 (m, 1H, βH_{leu}), 1.37 (m, 2H, δH_{lys}), 1.36 (s, 9H, Boc), 1.34 (m, 2H, βH_{ahx}), 1.25 (m, 1H, γH_{lys}), 1.24 (m, 3H, $COOCH_2CH_3$), 1.23 (m, 1H, γHlys), 1.20 (m, 2H, γH_{ahx}), 0.87 (d, 1H, $^{3}J_{H-H}$ = 6.6 Hz, δH_{leu}), 0.83 (d, 1H, $^{3}J_{H-H}$ = 6.6 Hz, δH_{leu}). Chemical shifts are given in ppm.

^{13}C-NMR (226.34 MHz, d_6-DMSO): d = 174.5 (CO_{lys}), 172.5 (CO_{ahx}), 171.8 ($FcCH_2CH(NHBoc)CO$), 171.4 (CO_{leu}), 171.0 (CO_{tyr}), 167.7 (CO_{epx}), 165.2 (CO_{epx}), 156.3 ($C_{tyr-bzl}$-OH), 155.5 (COBoc), 130.5 ($C_{tyr-bzl}$), 128.2 ($C_{tyr-bzl}$), 115.3 ($C_{tyr-bzl}$), 84.9 (C_{cp}-CH_2), 78.7 ($C(CH_3)_3$), 69.5 (C_{cp}), 68.8 (C_{cp}), 67.5 (C_{cp}), 62.0 ($COOCH_2CH_3$), 56.4 ($FcCH_2CH(NHBoc)$), 54.8 (αC_{tyr}), 53.3 (C_{epx}), 52.6 (αC_{lys}), 51.8 (αC_{leu}), 51.6 (C_{epx}), 41.3 (βC_{leu}), 38.9 (εC_{lys}), 38.9 (αC_{ahx}), 37.4 (βC_{tyr}), 35.6 (εC_{ahx}), 33.0 ($FcCH_2$), 32.1 (βC_{lys}), 29.2 (βC_{ahx}), 28.7 (δC_{lys}), 28.7 ($C(CH_3)_3$), 26.5 (γC_{ahx}), 25.5 (δC_{ahx}) 24.6 (γC_{leu}), 23.4 (δC_{leu}), 23.3 (γC_{lys}), 22.1 (δC_{leu}), 14.4 ($COOCH_2CH_3$). Chemical shifts are given in ppm.

^{15}N NMR (91.21 MHz, d_6-DMSO): d = 123.9 (N_{leu}), 121.9 (N_{lys}), 118.1 (N_{tyr}), 115.8 (N_{ahx}), 114.7 ($N_{lys,s.c.}$), 104.8 ($CONH_2$), 90.7 (NHBoc). Chemical shifts are given in ppm.



^{1}H-NMR (400.13 MHz, d_6-DMSO): 7.74 (d, 4 H, $^{3}J_{HH}$ = 7.9 Hz, $C_{Ts}H$ / $\tilde{C}_{Ts}\tilde{H}$), 7.28 (d, 2 H, $^{3}J_{HH}$ = 7.8 Hz $C_{Ts}H$ / $\tilde{C}_{Ts}\tilde{H}$), 7.21 (m, 1 H, CH_2-NH_aH_b), 6.77 (m, 1 H, $\tilde{C}\tilde{H}_2$-$\tilde{N}\tilde{H}_a\tilde{H}_b$), 6.66 (m, 1 H, $C_{meta}H_a$), 6.56 (m, 1 H, $\tilde{C}_{meta}\tilde{H}_a$), 6.44 (m, 1 H, α-CH-$NH_aH_b$), 6.03 (t, 1 H, $^{3}J_{HH}$ = 5.5 Hz, $C_{meta}H_b$), 5.88 (t, 1 H, $^{3}J_{HH}$ = 5.4 Hz, $\tilde{C}_{meta}\tilde{H}_b$), 5.77 (m, 1 H, $\tilde{C}\tilde{H}_2$-$\tilde{N}\tilde{H}_a\tilde{H}_b$), 5.73 (m, 1 H, α-$\tilde{C}\tilde{H}$-$\tilde{N}\tilde{H}_a\tilde{H}_b$), 5.60 (d, 1 H, $^{3}J_{HH}$ = 5.4 Hz, $\tilde{C}_{ortho}\tilde{H}_b$), 5.58 (m, 1 H, $C_{ortho}H_a$), 5.54 (d, 1 H, $\tilde{C}_{ortho}\tilde{H}_a$), 5.48 (d, 1 H, $^{3}J_{HH}$ = 5.7 Hz, $C_{ortho}H_b$), 5.41 (m, 1 H, CH_2-NH_aH_b), 5.20 (dd, 1 H, $^{3}J_{HH}$ = 4.4 Hz, $^{3}J_{HH}$ = 4.4 Hz, $\tilde{C}_{para}\tilde{H}$), 5.11 (dd, 1 H, $^{3}J_{HH}$ = 4.5 Hz, $^{3}J_{HH}$ = 4.5 Hz, $C_{para}H$), 4.84 (m, 2 H, α-CH / α-$\tilde{C}\tilde{H}$), 4.25 (m, 5 H, CH_2CH_3 / $\tilde{C}\tilde{H}_2\tilde{C}\tilde{H}_3$ / α-$\tilde{C}\tilde{H}$-$\tilde{N}\tilde{H}_a\tilde{H}_b$), 3.61 (m, 1 H, α-CH-$NH_aH_b$), 3.28 (dd, 1 H, $^{3}J_{HH}$ = 5.5 Hz, $^{3}J_{HH}$ = 14.5 Hz, β-CH_aH_b), 3.25 (dd, 1 H, $^{3}J_{HH}$ = 5.6 Hz, $^{3}J_{HH}$ = 14.2 Hz, β-$\tilde{C}\tilde{H}_a\tilde{H}_b$), 2.93 (dd, 1 H, $^{3}J_{HH}$ = 13.1 Hz, $^{3}J_{HH}$ = 13.1 Hz, β-$\tilde{C}\tilde{H}_a\tilde{H}_b$), 2.74-2.53 (m, 8 H, CH_2CH_2 / $\tilde{C}\tilde{H}_2\tilde{C}\tilde{H}_2$), 2.67 (m, 1 H, β-$CH_aH_b$), 2.43 (s, 6 H, $C_{Ts}CH_3$), 1.34 (t, 3 H, $^{3}J_{HH}$ = 7.0 Hz, CH_2CH_3), 1.33 (t, 3 H, $^{3}J_{HH}$ = 6.9 Hz, $\tilde{C}\tilde{H}_2\tilde{C}\tilde{H}_3$). Chemical shifts are given in ppm.

^{13}C-NMR (100.51 MHz, d_6-DMSO): 170.5 / 170.4 (COO / $\tilde{C}\tilde{O}\tilde{O}$), 141.3 / 141.1 / 140.9 / 140.7 ($C_{Ts}SO_2$ / $\tilde{C}_{Ts}\tilde{S}\tilde{O}_2$), 129.5 / 129.3 ($C_{Ts}H$ / $\tilde{C}_{Ts}\tilde{H}$), 126.3 / 126.0 ($C_{Ts}H$ / $\tilde{C}_{Ts}\tilde{H}$), 105.1 / 104.5 ($C_{arene}CH_2$ / $\tilde{C}_{arene}\tilde{C}\tilde{H}_2$), 94.7 / 93.0 / 90.8 / 90.2 / 80.0 / 77.7 / 77.3 / 74.5 / 70.2 / 69.5 / 68.1 / 67.1 ($C_{arene}H$ / $\tilde{C}_{arene}\tilde{H}$ / α-CH / α-$\tilde{C}\tilde{H}$), 61.4 (CH_2CH_3 / $\tilde{C}\tilde{H}_2\tilde{C}\tilde{H}_3$), 50.2 / 46.2 / 46.1 ($NCH_2$ / $\tilde{N}\tilde{C}\tilde{H}_2$), 36.9 / 36.8 (β-$CH_2$ / β-$\tilde{C}\tilde{H}_2$), 20.9 / 20.8 ($C_{Ts}CH_3$ / $\tilde{C}_{Ts}\tilde{C}\tilde{H}_3$), 14.0 ($CH_2$-$CH_3$ / $\tilde{C}\tilde{H}_2$-$\tilde{C}\tilde{H}_3$). Chemical shifts are given in ppm.

^{1}H,^{15}N coupled NMR (d_1-$CDCl_3$): {6.9, -1.0} / {5.8, -1.0} / {7.2, -0.1} / {5.4, -0.1} (CH_2NH_aHb / CH_2NH_aHb / $\tilde{C}\tilde{H}2\tilde{N}\tilde{H}_a\tilde{H}_b$ / $\tilde{C}\tilde{H}_2\tilde{N}\tilde{H}_a\tilde{H}_b$), {5.7, 13.2} / {4.3, 13.2} / {6.4, 15.3} / {3.6, 15.3} ($CHNH_aHb$ / $CHNH_aHb$ / $\tilde{C}\tilde{H}\tilde{N}\tilde{H}aH_b$ / $\tilde{C}\tilde{H}\tilde{N}\tilde{H}aH\tilde{b}$). Chemical shifts are given in ppm. The assignment of distinct atoms is shown in chapter 2.3.

b) Chapter 3.1: ROE distance restraints for Cilengitide in four solvents

Cilengitide in H_2O - Distance restraints and violations:

1	ARG	HN	GLY	HN	0.374	0.402	0.028
2	GLY	HA1	ASP	HN	0.305	0.223	-0.082
3	ASP	HN	ASP	HB2	0.355	0.245	-0.110
4	GLY	HA2	ASP	HN	0.296	0.325	0.029
5	ASP	HN	ASP	HB1	0.338	0.285	-0.053
6	ASP	HN	GLY	HN	0.414	0.417	0.003
7	ASP	HN	DPHE	HN	0.368	0.422	0.054
8	ASP	HA	ASP	HN	0.290	0.279	-0.011
9	MEVAL	HA	MEVAL	HB	0.335	0.279	-0.056
10	ASP	HA	ASP	HB2	0.306	0.260	-0.046
11	ASP	HA	ASP	HB1	0.306	0.264	-0.042
12	GLY	HA1	GLY	HN	0.308	0.274	-0.034
13	ARG	HA	GLY	HN	0.294	0.222	-0.072
14	GLY	HA2	GLY	HN	0.342	0.228	-0.114
15	MEVAL	HA	DPHE	HN	0.519	0.487	-0.032
16	DPHE	HA	DPHE	HN	0.350	0.284	-0.066
17	MEVAL	HA	ARG	HN	0.265	0.357	0.092
18	ARG	HN	MEVAL	HB	0.354	0.239	-0.115
19	ASP	HA	DPHE	HN	0.274	0.218	-0.056
20	GLY	HA1	DPHE	HN	0.569	0.550	-0.019
21	DPHE	HN	ASP	HB2	0.524	0.434	-0.090
22	DPHE	HN	ASP	HB1	0.600	0.394	-0.206
23	ARG	HA	ARG	HN	0.278	0.278	0.000
24	ARG	HE	ARG	HG@	0.424	0.262	-0.162
25	GLY	HN	ARG	HB@	0.438	0.382	-0.056
26	ARG	HE	ARG	HB@	0.440	0.354	-0.086
27	ARG	HG@	GLY	HN	0.613	0.425	-0.188
28	DPHE	HN	DPHE	HB@	0.387	0.261	-0.126
29	ARG	HN	ARG	HG@	0.436	0.334	-0.102
30	ARG	HN	ARG	HB@	0.443	0.267	-0.176
31	DPHE	HA	DPHE	HB@	0.365	0.253	-0.112
32	ARG	HE	ARG	HD@	0.373	0.185	-0.188
33	ARG	HA	ARG	HG@	0.377	0.266	-0.111
34	ARG	HA	ARG	HD@	0.511	0.355	-0.156

35	ARG	HA	ARG	HB@	0.368	0.254	-0.114
36	ARG	HN	MEVAL	HN@	0.555	0.304	-0.251
37	ARG	HN	MEVAL	HG1@	0.565	0.352	-0.213
38	DPHE	HN	MEVAL	HN@	0.636	0.458	-0.178
39	MEVAL	HA	MEVAL	HN@	0.597	0.361	-0.236
40	MEVAL	HA	MEVAL	HG2@	0.453	0.274	-0.179
41	DPHE	HA	MEVAL	HG2@	0.714	0.501	-0.213
42	MEVAL	HB	MEVAL	HG2@	0.421	0.241	-0.180
43	MEVAL	HB	MEVAL	HG1@	0.415	0.238	-0.177
44	MEVAL	HA	MEVAL	HG1@	0.447	0.297	-0.150
45	MEVAL	HN@	MEVAL	HB	0.393	0.279	-0.114
46	DPHE	HA	MEVAL	HN@	0.378	0.242	-0.136
47	MEVAL	HG1@	ARG	HE	0.780	0.590	-0.190
48	MEVAL	HN@	GLY	HN	0.766	0.553	-0.213
49	MEVAL	HN@	ASP	HN	0.678	0.533	-0.145
50	ARG	HD@	ARG	HG@	0.424	0.234	-0.190
51	ARG	HD@	ARG	HB@	0.457	0.276	-0.181
52	MEVAL	HG1@	ARG	HG@	0.584	0.457	-0.127
53	MEVAL	HG1@	ARG	HB@	0.644	0.535	-0.109
54	MEVAL	HN@	MEVAL	HG2@	0.680	0.395	-0.285
55	MEVAL	HN@	MEVAL	HG1@	0.925	0.497	-0.428
56	MEVAL	HG1@	MEVAL	HG2@	0.651	0.300	-0.351

All restraint data are $<r^3>^{-1/3}$ averaged and given in nm. @ denotes a pseudo-atom.

Cilengitide in DMSO - Distance restraints and violations:

1	GLY	HN	ARG	HN	0.336	0.409	0.073
2	ARG	HN	ARG	HA	0.252	0.266	0.014
3	GLY	HN	ARG	HA	0.273	0.217	-0.056
4	ASP	HN	GLY	HN	0.378	0.412	0.034
5	GLY	HN	GLY	HA1	0.284	0.280	-0.004
6	GLY	HA1	GLY	HN	0.289	0.280	-0.009
7	GLY	HA2	GLY	HA1	0.199	0.163	-0.036
8	ASP	HN	GLY	HA1	0.309	0.228	-0.081
9	GLY	HN	GLY	HA2	0.332	0.226	-0.106
10	ASP	HN	GLY	HA2	0.262	0.328	0.066
11	DPHE	HN	ASP	HA	0.252	0.229	-0.023

12	ASP	HN	ASP	HB1	0.303	0.284	-0.019
13	ASP	HA	ASP	HB1	0.289	0.266	-0.023
14	ASP	HB2	ASP	HB1	0.210	0.163	-0.047
15	ASP	HN	ASP	HB2	0.310	0.266	-0.044
16	ASP	HB2	ASP	HA	0.296	0.254	-0.042
17	ASP	HN	DPHE	HN	0.351	0.429	0.078
18	DPHE	HN	DPHE	HB2	0.338	0.277	-0.061
19	DPHE	HN	DPHE	HA	0.319	0.276	-0.043
20	DPHE	HA	DPHE	HB1	0.275	0.253	-0.022
21	ASP	HN	ASP	HA	0.282	0.277	-0.005
22	DPHE	HB1	DPHE	HB2	0.202	0.163	-0.039
23	DPHE	HN	DPHE	HB1	0.316	0.242	-0.074
24	DPHE	HA	DPHE	HB1	0.287	0.253	-0.034
25	ARG	HN	MEVAL	HA	0.245	0.322	0.077
26	ARG	HN	MEVAL	HB	0.328	0.251	-0.077
27	MEVAL	HA	MEVAL	HB	0.335	0.279	-0.056
28	ARG	HB@	GLY	HN	0.453	0.396	-0.057
29	ARG	HA	ARG	HD@	0.440	0.365	-0.075
30	ARG	HE	ARG	HD@	0.361	0.185	-0.176
31	ARG	HN	ARG	HG@	0.423	0.336	-0.087
32	ARG	HA	ARG	HG@	0.365	0.264	-0.101
33	ARG	HE	ARG	HG@	0.398	0.263	-0.135
34	ARG	HB@	ARG	HN	0.420	0.271	-0.149
35	ARG	HA	ARG	HB@	0.353	0.255	-0.098
36	ARG	HB@	ARG	HE	0.429	0.350	-0.079
37	ARG	HN	MEVAL	HN@	0.559	0.309	-0.250
38	DPHE	HA	MEVAL	HN@	0.437	0.241	-0.196
39	MEVAL	HN@	MEVAL	HB	0.458	0.277	-0.181
40	MEVAL	HA	MEVAL	HG2@	0.515	0.274	-0.241
41	MEVAL	HB	MEVAL	HG2@	0.489	0.241	-0.248
42	MEVAL	HA	MEVAL	HN@	0.675	0.360	-0.315
43	MEVAL	HB	MEVAL	HG1@	0.479	0.238	-0.241
44	MEVAL	HA	MEVAL	HG1@	0.499	0.300	-0.199
45	ARG	HN	MEVAL	HG1@	0.637	0.357	-0.280
46	ARG	HD@	ARG	HB@	0.453	0.275	-0.178
47	ARG	HB@	ARG	HG@	0.457	0.234	-0.223
48	ARG	HD@	ARG	HG@	0.413	0.234	-0.179
49	ARG	HG@	MEVAL	HG1@	0.700	0.446	-0.254

50	MEVAL	HN@	MEVAL	HG2@	0.934	0.395	-0.539
51	MEVAL	HN@	MEVAL	HG1@	1.368	0.494	-0.874
52	MEVAL	HG1@	MEVAL	HG2@	0.867	0.299	-0.568

All restraint data are $<r^3>^{-1/3}$ averaged and given in nm. @ denotes a pseudo-atom.

Cilengitide in MeOH - Distance restraints and violations:

1	ARG	HA	ARG	HN	0.278	0.219	-0.059
2	ARG	HA	ARG	HE	0.563	0.458	-0.105
3	ARG	HA	GLY	HN	0.296	0.222	-0.074
4	DPHE	HN	DPHE	HB1	0.338	0.263	-0.075
5	DPHE	HN	DPHE	HB2	0.329	0.265	-0.064
6	MEVAL	HA	ARG	HN	0.268	0.229	-0.039
7	MEVAL	HA	GLY	HN	0.528	0.489	-0.039
8	MEVAL	HA	DPHE	HN	0.492	0.490	-0.002
9	ARG	HN	MEVAL	HB	0.344	0.332	-0.012
10	MEVAL	HA	MEVAL	HB	0.335	0.280	-0.055
11	ASP	HA	ASP	HB1	0.302	0.253	-0.049
12	ASP	HB1	DPHE	HN	0.608	0.394	-0.214
13	ASP	HN	ASP	HB2	0.341	0.243	-0.098
14	ASP	HA	ASP	HB2	0.323	0.273	-0.050
15	ASP	HB2	ASP	HB1	0.231	0.163	-0.068
16	ASP	HN	DPHE	HN	0.397	0.406	0.009
17	ARG	HN	GLY	HN	0.396	0.351	-0.045
18	ASP	HN	GLY	HN	0.450	0.419	-0.031
19	GLY	HA2	GLY	HN	0.304	0.230	-0.074
20	DPHE	HA	DPHE	HB1	0.285	0.246	-0.039
21	DPHE	HA	DPHE	HB2	0.306	0.278	-0.028
22	GLY	HA2	ASP	HN	0.333	0.316	-0.017
23	GLY	HA1	GLY	HN	0.344	0.276	-0.068
24	GLY	HA1	ASP	HN	0.284	0.225	-0.059
25	ASP	HA	ASP	HN	0.306	0.273	-0.033
26	ASP	HA	DPHE	HN	0.267	0.220	-0.047
27	ASP	HN	ASP	HB1	0.330	0.311	-0.019
28	DPHE	HA	DPHE	HN	0.344	0.282	-0.062
29	GLY	HN	ARG	HB@	0.449	0.391	-0.058

30	ARG	HA	ARG	HD@	0.456	0.370	-0.086
31	ARG	HE	ARG	HD@	0.383	0.185	-0.198
32	ARG	HN	ARG	HB@	0.432	0.315	-0.117
33	ARG	HA	ARG	HB@	0.368	0.253	-0.115
34	ARG	HE	ARG	HB@	0.442	0.343	-0.099
35	ARG	HN	ARG	HG@	0.454	0.344	-0.110
36	ARG	HA	ARG	HG@	0.387	0.262	-0.125
37	ARG	HE	ARG	HG@	0.418	0.266	-0.152
38	MEVAL	HA	MEVAL	HG1@	0.447	0.299	-0.148
39	MEVAL	HG1@	MEVAL	HB	0.424	0.238	-0.186
40	ARG	HN	MEVAL	HN@	0.525	0.405	-0.120
41	MEVAL	HN@	GLY	HN	0.766	0.609	-0.157
42	MEVAL	HN@	ASP	HN	0.682	0.571	-0.111
43	MEVAL	HN@	DPHE	HN	0.631	0.465	-0.166
44	ARG	HN	MEVAL	HG1@	0.559	0.358	-0.201
45	MEVAL	HG1@	ARG	HE	0.867	0.592	-0.275
46	DPHE	HA	MEVAL	HN@	0.380	0.241	-0.139
47	MEVAL	HN@	MEVAL	HB	0.402	0.277	-0.125
48	MEVAL	HG2@	DPHE	HB2	0.815	0.538	-0.277
49	MEVAL	HA	MEVAL	HG2@	0.458	0.276	-0.182
50	MEVAL	HG2@	MEVAL	HB	0.425	0.241	-0.184
51	ARG	HD@	ARG	HB@	0.462	0.277	-0.185
52	ARG	HB@	ARG	HG@	0.487	0.234	-0.253
53	ARG	HD@	ARG	HG@	0.429	0.233	-0.196
54	MEVAL	HG1@	ARG	HD@	0.788	0.567	-0.221
55	MEVAL	HG1@	ARG	HG@	0.628	0.464	-0.164
56	MEVAL	HG2@	MEVAL	HG1@	0.655	0.300	-0.355
57	MEVAL	HN@	MEVAL	HG2@	0.680	0.397	-0.283
58	MEVAL	HN@	MEVAL	HG1@	0.971	0.494	-0.477

All restraint data are $<r^3>^{-1/3}$ averaged and given in nm. @ denotes a pseudo-atom.

Cilengitide in $CHCl_3$ - Distance restraints and violations:

1	ARGN	HN	ARGN	HA	0.332	0.278	-0.054
2	GLY	HN	ARGN	HA	0.329	0.221	-0.108

3	GLY	HA1	GLY	HN	0.337	0.280	-0.057
4	ASPH	HN	GLY	HA1	0.368	0.228	-0.140
5	GLY	HA2	GLY	HN	0.427	0.226	-0.201
6	ASPH	HN	GLY	HA2	0.378	0.337	-0.041
7	ASPH	HN	ASPH	HA	0.323	0.283	-0.040
8	DPHE	HN	ASPH	HA	0.289	0.220	-0.069
9	ASPH	HN	ASPH	HB1	0.370	0.301	-0.069
10	ASPH	HB1	ASPH	HA	0.373	0.256	-0.117
11	ASPH	HN	ASPH	HB2	0.403	0.261	-0.142
12	DPHE	HN	ASPH	HB2	0.439	0.433	-0.006
13	ASPH	HN	DPHE	HN	0.456	0.404	-0.052
14	DPHE	HN	ASPH	HB1	0.471	0.397	-0.074
15	DPHE	HN	DPHE	HB2	0.351	0.281	-0.070
16	DPHE	HN	DPHE	HA	0.331	0.285	-0.046
17	DPHE	HN	DPHE	HB1	0.346	0.268	-0.078
18	DPHE	HA	DPHE	HB1	0.297	0.252	-0.045
19	ARGN	HN	MEVAL	HA	0.295	0.358	0.063
20	ARGN	HN	MEVAL	HB	0.336	0.242	-0.094
21	MEVAL	HA	MEVAL	HB	0.335	0.278	-0.057
22	ARGN	HE	ARGN	HD@	0.404	0.185	-0.219
23	ARGN	HA	ARGN	HB@	0.413	0.255	-0.158
24	ARGN	HE	ARGN	HB@	0.473	0.337	-0.136
25	GLY	HN	ARGN	HB@	0.527	0.391	-0.136
26	ARGN	HD@	ARGN	HA	0.499	0.345	-0.154
27	ARGN	HA	ARGN	HG@	0.420	0.272	-0.148
28	ARGN	HE	ARGN	HG@	0.444	0.271	-0.173
29	ARGN	HD@	ARGN	HG@	0.455	0.230	-0.225
30	ARGN	HD@	ARGN	HB@	0.513	0.293	-0.220
31	ARGN	HG@	MEVAL	HG1@	0.606	0.462	-0.144
32	ARGN	HG@	MEVAL	HN@	0.783	0.647	-0.136
33	MEVAL	HN@	MEVAL	HB	0.376	0.278	-0.098
34	MEVAL	HA	MEVAL	HG2@	0.417	0.274	-0.143
35	MEVAL	HB	MEVAL	HG2@	0.395	0.241	-0.154
36	ARGN	HN	MEVAL	HG1@	0.591	0.354	-0.237
37	MEVAL	HA	MEVAL	HG1@	0.418	0.297	-0.121
38	MEVAL	HB	MEVAL	HG1@	0.391	0.238	-0.153

39	ARGN	HN	MEVAL	HN@	0.559	0.305	-0.254
40	DPHE	HA	MEVAL	HN@	0.360	0.242	-0.118
41	MEVAL	HN@	DPHE	HB2	0.660	0.463	-0.197
42	MEVAL	HA	MEVAL	HN@	0.579	0.361	-0.218
43	DPHE	HA	MEVAL	HG2@	0.718	0.505	-0.213
44	GLY	HN	MEVAL	HN@	0.741	0.583	-0.158
45	MEVAL	HN@	MEVAL	HG1@	1.104	0.496	-0.608
46	MEVAL	HG1@	MEVAL	HG2@	0.578	0.300	-0.278
47	MEVAL	HN@	MEVAL	HG2@	0.630	0.394	-0.236

All restraint data are $<r^3>^{-1/3}$ averaged and given in nm. @ denotes a pseudo-atom.

c) Chapter 3.2: Measured and PALES-SVD / MD simulations derived RDCs for Strychnine

Strychnine Vector	Measured RDC	PALES-SVD	MD Simulation
C1-H1	-9.3	0.0	-9.2
C2-H2	0.8	-4.7	0.9
C4-H4	-10.5	-0.1	-9.5
C8-H8	3.6	0.7	4.9
C12-H121	-4.5	5.2	-1.2
C12-H122	7.5	1.2	8.8
C13-H13	11.4	5.5	12.1
C14-H14	-3.6	-1.6	-2.8
C15-H15	12.0	4.6	12.2
C16-H161	-9.6	-3.1	-9.6
C16-H162	-2.1	-0.2	-2.3
C17-H17	1.8	-3.2	1.8
C19-H191	0.2	-2.8	0.0
C19-H192	-4.8	3.7	-1.0
C21-H211	6.9	5.7	6.4
C21-H212	-5.4	0.5	-2.3
C23-H23	-6.6	0.7	-7.3
C24-H241	13.5	8.8	14.1
C24-H242	2.2	-2.8	0.1

All values are given in Hz.

d) Chapter 3.3: Caco-2 test results for *N*-methylated penta-alanines

Peptide	Sequence	Conformation	Population	Caco-2 test
P01	c(^{Me}aAAAA-)	homogeneous	100 %	2.33e-07
P02	c(^{Me}a^{Me}AAAA-)	homogeneous	100 %	2.14e-06
P03	c(^{Me}aAAA^{Me}A-)	homogeneous	98 / 2 %	4.40e-06
P04	c(aAA^{Me}A^{Me}A-)	homogeneous	98 / 2 %	3.00e-06
P05	c(^{Me}a^{Me}AAA^{Me}A-)	homogeneous	98 / 2 %	1.05e-06
P06	c(^{Me}aAA^{Me}A^{Me}A-)	homogeneous	98 / 2 %	6.95e-06
P07	c(^{Me}aA^{Me}A^{Me}A^{Me}A-)	homogeneous	98 / 2 %	7.89e-07
S08	c(aAAA^{Me}A-)	inhomogeneous	84 / 13 / 3 %	4.60e-07
S09	c(a^{Me}AAAA-)	inhomogeneous	85 / 11 / 4 %	2.35e-07
S10	c(a^{Me}AA^{Me}AA-)	inhomogeneous	84 / 11 / 5 %	6.20e-07
S11	c(^{Me}a^{Me}A^{Me}AAA-)	inhomogeneous	82 / 11 / 7 %	6.20e-07
S12	c(a^{Me}AA^{Me}A^{Me}A-)	inhomogeneous	84 / 16 %	1.14e-06
S13	c(aA^{Me}A^{Me}A^{Me}A-)	inhomogeneous	85 / 10 / 5 %	3.23e-06

All Caco-2 test results are given in cm / sec permeability. “Me” stands for an N-methyl group, an “A” for an L-alanine, “a” for a D-alanine, and “c” for cyclo.