Application of Computational Methods to the Structural and Functional Properties of Flexible Chiral Molecules

Anwendung computerbasierter Methoden zur Berechnung der strukturellen und funktionellen Eigenschaften flexibler chiraler Moleküle

Der Naturwissenschaftlichen Fakultät

der Friedrich-Alexander Universität Erlangen-Nürnberg

zur

Erlangung des Doktorgrades Dr. rer. nat.

Vorgelegt von

Zlatko Brkljača

aus Pula, Kroatien

Als Dissertation genehmigt
von der Naturwissenschaftlichen Fakultät
der Friedrich-Alexander-Universität Erlangen-Nürnberg

Tag der mündlichen Prüfung: 08. April 2016

Vorsitzender des Promotionsorgans:	Prof. Dr. Jörn Wilms
Gutachter/in:	Prof. Dr. Ana-Sunčana Smith
	Prof. Dr. Timothy Clark

To my family

Zussamenfassung

Die Konzeption von *Flexibilität* und *Chiralität* stellt zwei wesentliche Aspekte modernen chemischen Verständnisses dar, die einer Vielzahl biochemischer und biologischer Prozesse zugrunde liegen. Molekulare Flexibilität und ihre Auswirkungen – einschließlich ihrer fundamentalen Rolle in der Wechselwirkung zwischen Enzymen und Substrat, beziehungsweise zwischen Rezeptoren und Liganden im Allgemeinen – sind Gegenstand der Forschung seit über 70 Jahren. Die Flexibilität zeichnet sich dabei als Eckpfeiler generalisierter katalytischer Mechanismen aus. Auch die Vorstellung von Chiralität ist bereits seit über hundert Jahren etabliert, dennoch werden etliche Auswirkungen dieses Phänomens erst seit kurzer Zeit eingehender untersucht. Neuesten experimenteller und theoretischer Bemühungen zum Trotz fehlt es aktuell noch an einer vollständigen Beschreibung dieses Phänomen. Obwohl gerade das Zusammenspiel von Flexibilität und Chiralität eine entscheidende Rolle auf dem Gebiet der Biochemie und molekularen Medizun spielt – beispielsweise bei der Entwicklung neuer Systeme zur gezielten Pharmakotherapie – ist dieses bis dato jedoch noch weitaus weniger erforscht als die beiden zugrundeliegenden Konzepte.

Ziel dieser Abhandlung ist die Untersuchung des Zusammenwirkens von Flexibilität und Chiralität unter Verwendung aktueller Simulationsverfahren, die von der approximativen klassischen Behandlung bis hin zu hochpräzisen quantenmechanischen Methoden reichen. Zu diesem Zweck wurde der Einfluss der beiden Konzepte in zwei Szenarien untersucht, die sich mit dem Verhalten flexibler und chiraler Moleküle einersteits in Lösung, andererseits in Kontakt mit inorganischen Grenzfälchen befassen.

Der erste Teil der Arbeit befasst sich mit der Untersuchung flexibler chiraler Moleküle in Lösung mittels Zirkulardichroismusspektroskopie (CD). CD stellt eine der zentralen experimentellen Methoden zur Charakterisierung optisch aktiver, chiraler Moleküle dar und findet breite Anwendung in der Untersuchung biologisch relevanter Systeme wie beispielsweise von Peptiden, Proteinen oder Bestandteilen der DNS. Brauchbare Informationen aus experimentellen CD-Spektren zu ziehen ist jedoch mitunter ein sehr komplexes Unterfangen, gerade im Falle hochflexibler Moleküle. Dies macht die Zuhilfenahme theoretischer Verfahren erforderlich um ein detailiertes Verständnis zu erlangen. Vor diesem Hintergrund haben wir durch Kombination von klassischen und quantenmechanischen Simulationen eine Methode zur Berechnung der CD-Spektren flexibler Moleküle entwickelt. In dieser wird zunächst durch Replica-Exchange Molekulardynamik (REMD) ein konvergierter Konformationsphasenraum generiert und anschließend mittels Clusteranalyse ein Satz molekularer Strukturen identifiziert, die diesen repräsentieren. In Übereinstimmung mit früheren Studien konnte ein signifikanter Einfluss des Lösungsmittels auf die resultierenden CD-Spektren nachgewießen

werden. Dieser wurde durch die Berechnung eines gemittelten elektrischen Feldes berücksichtigt, welches sich aus einer Punktladungsverteilung ergibt, die auf einer Reihe von Lösungsmittel-Molekülkonfigurationen im Umfeld aller Peptidkonformationen basiert. Anschließend konnte das CD-Spektrum jeder Konformation durch einen kombiniert quantenmechanischen und molekularmechnischen Ansatz (QM/MM) mit zeitabhängiger Dichtefunktionaltheorie (TD-DFT) berechnet werden. Die durchschnittlichen CD-Spektren erhält man daraufhin als gewichtetes Mittel aller Einzelspektren. Die Gewichtungsfaktoren entsprechen dabei der relativen Häufigkeit der jeweiligen Konformationen aus der Clusteranalyse. Qualitativ entspricht dieses Vorgehen einer sequentiellen Mittelung über die Hauptketten (wodurch die Cluster zusammegefasst werden), die Seitenketten (innerhalb der durch die Hauptkette definierten Cluster) und das Lösungsmittel (für jede repräsentative Konformation). Die Verifizierung dieser Methode erfolgte durch zwei konkrete Fallstudien, welche in den folgenden Absätzen näher erläutert werden.

In unserer ersten Fallstudie wurde die neueentwickelte Methode zur Berechnung der CD-Spektren zweier neuartiger *Rhodomyrtal*-Verbindungen angewandt, die eine stark-antibakterielle Wirkung haben. Durch die aus Kernmagnetresonanztechniken (NMR) bekannten Relativkonfigurationen lag das Interesse in der Bestimmung der Absolutkonfigurationen beider Verbindungen. Dabei gelang mit unserer Methode unter Verwendung von TD-DFT auf B3LYP/6-311G(d) Theorielevel eine hervorragende Übereinstimmung mit experimentellen Spektren. Die drei signifikantesten Merkmale der Spektren konnten nicht nur qualitativ in den berechneten CD-Spektren wiedergefunden sondern auch quantitativ bestätigt werden. Aufgrund dieser Übereinstimmung konnten im Umkehrschluss die Absolutkonfigurationen beider *Rhodomyrtal*-Verbindungen eindeutig bestimmt werden.

Die zweite Fallstudie befasst sich mit der Berechnung der CD-Spektren zweier hochflexibler Opioid-Pentapeptide, Met- und Ada-Enkephalin (Tyr-Gly-Gly-Phe-Met and Tyr-Aaa-Gly-Phe-Met, wobei Aaa eine nicht-essenzielle, adamantylhaltige Amainosäure repräsentiert). Dafür wurden sowohl die zwitterionische als auch die neutrale Form von Met-Enkephalin und den R- und S-Epimeren von Ada-Enkephalin modeliert, da jeweils beide Formen der Peptide im verwendeten Lösungsmittel Trifluoroethanol koexistieren. Experimentelle Studien an Ada-Enkephalin wurden ausschließlich mit einer Epimermischung durchgeführt, weshalb wir – unter Berücksichtigung des Verhältnisses von zwitterionischer und neutraler Form der Peptide, sowie beider Ada-Enkephalin Epimere – kombinierte Spektren berechnet haben. Trotz der damit verbundenen Schwierigkeiten konnte eine gute qualitative Übereinstimmung mit experimentellen Spektren im Bereich unter 230 nm erreicht und deren zentrale Merkmale in den Simulationen reproduziert werden. Allerdings weisen die simulierten Spektren zwischen 230 nm und 240 nm ein Minimum auf, welches sich nicht in den experimentellen Daten wiederfindet und das wir zunächst der ungeeingeten Berücksichtigung eines bestimmten

Anregungszustandes in der TD-DFT zuordneten. Da aktuelle Studien auf Schwierigkeiten bei der Modellierung von Charge-Transfer-Anregungen hindeuten haben wir die Leistungsfähigkeit gängiger Hybridfunktionale, B3LYP und PBE0, langreichweitig korrigierter Funktionale, CAM-B3LYP und ωB97X-D und eines Meta-Hybridfunktionals, M06-2X, im Hinblick auf die ausgewählten Peptide, mit komplexen ab initio RICC2 Rechnungen sowie den experimentell verfügbaren Daten verglichen. Unsere Untersuchungen zeigen relativ gute Übereinstimmungen der korrigierten Funktionale mit RICC2. Gleiches gilt für das Meta-Hybridfunktional M06-2X, wohingegen die anderen Funktionale die zuvor beschriebenen Artefakte in Bezug auf Charge-Transfer-Anregung aufweisen. PBE0, insbesondere aber M06-2X und B3LYP hingegen liefern eine bessere Übereinstimmung mit experimentellen Spektren. Diese Diskrepanz lässt sich dadurch erklären, dass überzählige Charge-Transfer Anregungen, wie sie in B3LYP und PBE0 auftreten, nach der Kombination aller auftretenden Strukturen offenbar vernachlässigbaren Einfluss auf die finalen Spektren haben. Daraus lässt sich folgern, dass die strukturelle Mittellung die Probleme der Funktionale bei Berechnung der CD-Spektren einzelner Strukturen erheblich reduziert. Aus unseren Studien geht daher hervor, dass sich TD-DFT besser zur Untersuchung großer, flexibler Moleküle eignet als das komplexere RICC2 Verfahren. Dies begründet sich zum einen auf der besseren experimentellen Übereinstimmung in Bezug auf bestimmte Anregungsformen, zum anderen darauf, dass die TD-DFT sich auch auf komplexere Molekülstrukturen verhältnismäßig leicht anwenden lässt und die große Anzahl an Konformationen berechnen kann, die notwendig sind um konvergierte CD-Spektren flexibler Moleküle zu erhalten.

Im zweiten Teil dieser Arbeit haben wir uns auf das Verhalten flexibler und chiraler organischer Moleküle an der Grenzfläche zwischen Wasser und Biomineralen konzentriert um den Einfluss beider Eigenschaften auf Biomineralisation beziehungsweise biomineralisationsbasierte Wirkstoffentwicklung unter die Lupe zu nehmen. Dazu wurden die Wechselwirkungen, die innerhalb einer potenziellen Biomineral-Biomolekülverbindung auftreten mittels Molekulardynamik-simulationen die atomare Auflösung erlauben untersucht. Genauer gesagt haben wir das Absorptionsverhalten zweier epimerer Peptide, *R*- und *S*-Sal (N-Sal-Gly-*S*-Asp-*R*-Asp-*S*-Asp und N-Sal-Gly-*S*-Asp-*S*-Asp-*S*-Asp, wobei N-Sal den N-Terminus, der ein Salicylsäurederivat darstellt, bezeichnet), an der stabilen (104) und an der wachsenden (001) Kalzitgrenzfläche untersucht. Zwei gute Gründe sprechen für die Wahl dieses Systems. Zum einen repräsentiert es ein prototypisches Beispiel für biologisch abbaubare Wirkstofftransportsysteme, bei dem Kalzit die Matrix/den Träger darstellt, Salicylsäure die Medikamentenrolle einnimmt und der Rest des Peptids als Linker die Verbindung zum inorganischen Träger herstellt. Zum anderen konnten experimentelle Studien belegen, dass die untersuchten Epimere das Wachstum von Kalzit hemmen und beeinflussen können, also eine entscheidende Rolle in dessen Biomineralisation spielen.

Unsere Resultate legen den Schluss nahe, dass keines der Epimere an die stabile Kalzitoberfläche binden, wohingegen sowohl *R*- als auch *S*-Epimere gut an der (001) Grenzfläche adsorbieren. Dabei geht *R*-Sal mit einem Unterschied von etwa 85 $kJ \cdot mol^{-1}$ in der freien Adsorptionsenergie die stärkere Bindung ein. In beiden Fällen binden die Epimere über drei oder vier negativ geladene Carboxylgruppen an die positiv geladene Hälfte der wachsenden (001) Kalzitoberfläche. Wir erklären die Differenz in der freien Adsorptionsenergie beider Epimere durch die stärkere elektrostatische Wechselwirkung zwischen (001) Kalzit und *R*-Epimer unter der bevorzugt alle vier Carboxylgruppen mit der Ca^{2+}-reichen Oberfläche interagieren 15 % der *S*-Sal Strukturen hingegen adsorbieren lediglich mit drei Carboxylgruppen wodurch sich der enthalpische Beitrag zur freien Energie verringert. Diese Ergebnisse stehen im Einklang mit experimentellen Messungen bei denen *R*-Epimere im Vergleich zu *S*-Epimeren das Wachstum von Kalzit stärker hemmen/beinflussen konnten.

Bei unserer Arbeit handelt sich hingegen um die ersten theoretischen Untersuchungen, die den Einfluss von Chiralität und Flexibilität auf Biomineralisation bzw. biomineralisationsbasierte Wirkstoffentwicklung erklärt. Die Resultate zeigen, dass Chiralität einerseits genutzt werden kann um das Wachstum von Biomineralen feiner zu modifizieren, andererseits dazu die Wirkstoffrate von Medikamenten durch die Stärke der Linker-Biomineral-Wechselwirkung zu verändern und dadurch Einfluss auf die Pharmakodynamik und die therapeutische Wirkung einer Arznei zu nehmen.

Die gezeigten Ergebnisse verdeutlichen das Potential aktueller theoretischer Methoden die erfolgreich bei der Untersuchung des Einflusses von Chiralität und Flexibilität, sowohl in Lösung, als auch an inorganischen/mineralischen Oberflächen, zum Einsatz kamen. Durch den Brückenschlag zwischen Molekülstrukturen einerseits und den Spektren andererseits sind wir in der Lage das Zusammenspiel beider Konzepte erstmals zu erleuchten. Dieses Zusammenspiel wurde auch im Kontext der Biomineralisation untersucht, wo mittels klassicher Molekulardynamiksimulation dessen Rolle in der Wechselwikung zwischen Biomolekül und biomineralischer Oberfläche durch freie Energieprofile der Adsorption erläutert werden konnte. Die Resultate zeigen, dass die Kombination aus Theorie und Experiment nicht nur hilfreich sondern praktisch Voraussetzung ist, um Messungen korrekt interpretieren zu können, welche im Gegenzug dabei helfen bestehende Theorien und Modelle zu validieren und zukünftig zu erweitern.

Abstract

Flexibility and *chirality* of molecules are two cornerstones of modern chemistry, underlying vast array of biochemical and biological processes that take place in living organisms. The former concept and its role in this respect has been investigated and documented for over 70 years, including its critical role in the interactions occuring between enzymes and their substrates and, more generally, between receptors and ligands. Flexibility thus represents one of the mainstays of a generalized catalytic mechanism. While the concept of chirality has been introduced more than a century ago, certain manifestations of this phenomenon have been properly investigated for a relatively short period of time and, despite recent experimental and theoretical efforts, the complete understanding of this phenomenon is yet to be accomplished. Interestingly, although it is precisely the interplay between the two that plays a crucial role in the fields of biochemistry and molecular medicine/pharmacology, e.g. being necessary for the development of conceptually novel systems for targeted drug delivery, it is also significantly more scarcely investigated and understood than either of the two underlying concepts.

The aim of this thesis is to investigate the interaction between flexibility and chirality using *state-of-the-art* computational methods, ranging from highly accurate quantum mechanical treatments to more approximate classical methods. We have thereby investigated the interplay of the two concepts in two different scenarios, dealing with the behavior of flexible and chiral molecular species in solution (in the bulk of the solvent) and on the inorganic surface, respectively.

In the first part of this thesis we focused on the flexible, chiral molecules in solution, which we decided to inspect using the most powerful manifestation of chirality, namely using circular dichroism (CD) spectroscopy. CD is one of the key experimental methods employed in the structural characterization of optically active chiral molecules and, as such, it is widely used in studies of biologically important systems, such as peptides, proteins and DNA molecules. However, extracting useful information from experimentally obtained CD spectra can be a complex task, especially in the case of highly flexible molecules. Thus, to obtain a detailed understanding of the experiments, theoretical methods and resources need to be employed. In this respect, we developed a general methodology for calculating the CD spectra of flexible molecules by combining two "worlds", namely classical and quantum mechanics. Our method is based on the generation of a converged conformational phase space, obtained from advanced classical molecular dynamics simulations, namely replica exchange molecular dynamics, which is followed by finding a set of structures representing the entire phase space using clustering analysis. In agreement with previous studies, we have observed strong solvent effects on the spectra. These were taken into account by calculating an

average field consisting of point charges originating from a number of solvent configurations around each peptide conformation in the subset. Subsequently, the CD spectrum of each solvated conformation was obtained using the quantum mechanics/molecular mechanics (QM/MM) method incorporating a time-dependent density functional theory (TD-DFT) calculation. The average CD spectra were obtained by taking the mean of all weighted single CD spectra, where the weighting factor of an individual spectrum is given by the cluster-generated population fraction represented by the respective conformation. Qualitatively, our procedure can be considered as the sequential averaging over the solvent (for each representative conformation), the sidechains (inside the backbone-based clusters), and the backbone (combining the principal clusters). The verification of our methodology came in terms of two separate case studies, which are elaborated in more details in the following paragraphs.

In our first case study we applied the developed methodology to calculate the CD spectra of two novel *Rhodomyrtal* compounds which possess potent antibacterial activity. With the knowledge of the relative configurations, known from NMR techniques, the task at hand was to determine the absolute configuration of the two compounds. We were thereby able to sucessfully reproduce the experimental CD spectra using our strategy and utilizing TD-DFT method using B3LYP /6-311G(d) level of theory, obtaining an excellent agreement between experimentally measured and calculated CD spectra. More precisely, the three most dominant spectral features exhibited experimentally by these species are both qulitatively and quantitatively reproduced by the theoretically obtained CD spectra. In tur, this enabled us to unambigously assign the absolute configurations of both *Rhodomyrtal* compounds.

We followed with our second case study, in which we used the developed methodology in an attempt to calculate CD spectra of the two highly flexible opioid pentapeptides, which are found to exhibit a pronounced anti-tumor activity, namely Met- and Ada-enkephalin (Tyr-Gly-Gly-Phe-Met and Tyr-Aaa-Gly-Phe-Met, respectively, where Aaa denotes an unnatural adamantyl containing amino acid). We therefore modeled both the zwitterionic and neutral forms of the enkephalins and additionally both the *R*- and the *S*-epimers of Ada-enkephalin, which was necessary due to the fact that zwitterionic and neutral forms of the peptides coexist in trifluoroethanol, also taking into account the fact that the available experiments were conducted on an epimeric mixture of Ada-enkephalin. Thus, to make an appropriate comparison with experiment, we have produced composite spectra that account for the appropriate contributions of the zwitterionic and neutral forms of the peptides, as well as the expected epimeric ratio in the case of Ada-enkephalin. Despite the complexity of the task, qualitative agreement with experiment has been obtained below 230 nm and the main features of the measured spectra have been reproduced by our calculations. However, the simulated spectra show a persistent minimum between about 230-240 nm that is not present in the experimental measurements and which

we initially assumed to arise from the poor treatment of a particular excited state by TD-DFT. Motivated both by this finding and by recent studies that have pointed to problems with modeling charge transfer excitations, we benchmarked the performance of the selected global hybrid functionals (GHFs), namely B3LYP and PBE0 functionals, long-range corrected functionals (LCFs), more precisely CAM-B3LYP and ωB97X-D functionals, and one hybrid meta functional (HMF), namely M06-2X, against high level *ab initio* RICC2 calculations for selected peptide structures. Furthermore, we compared the performance of the functionals with the experimentally available data. Our results show that long-range corrected functionals correlate relatively well with RICC2 calculations, as does the meta-hybrid M06-2X, while the global hybrid functionals exhibit the aforementioned charge-transfer artifacts. Moreover, PBE0 and even more so M06-2X and B3LYP produce spectra in better agreement with the experimental data. We have clarified this apparent discrepancy by finding that the surplus charge-transfer excitations, exhibited by B3LYP and PBE0, seem to have a negligible contribution to the final spectra, once appropriate structural averaging is performed. This implies that the averaging procedure substantially reduces the effects of the problems exhibited by the GHFs in the CD spectra of the individual model structures. In this respect, our study strongly suggests that TD-DFT is likely to be preferred compared to the stronger RICC2, not only because of its better agreement with experiment for certain transition classes but also because it can be conveniently used, even for relatively large molecules, to evaluate the substantial number of conformations required to achieve converged CD spectra for flexible molecules.

In the second part of the thesis we concentrated on the behavior of flexible chiral organic molecules at the water – biomineral interface, in an attempt to uncover the roles of flexibility and chirality in biomineralization and biomineralization-inspired drug design. We thus decided to explore the interactions occurring within a potential biomineral – biomolecule composite using advanced molecular dynamics simulations, which enabled us to obtain atomistic details at the interface between biomolecules and mineral surfaces. More precisely, using advanced simulation techniques we investigated the adsorption behavior of two epimeric peptides, namely *R*- and *S*-Sal (N-Sal-Gly-*S*-Asp-*R*-Asp-*S*-Asp and N-Sal-Gly-*S*-Asp-*S*-Asp-*S*-Asp respectively, where N-Sal denotes the N-terminal residue which is a salicylic acid derivative), on both the stabile (104) and growing (001) surfaces of calcite. The reasoning behind this particular biomolecule – biomineral combination is twofold; on one hand, it represents a prototypical example of the biodegradable drug delivery system, with calcite serving as the matrix/carrier and salicylic acid playing the role of an archetypical drug, while the reminder of the peptide acts as the linker through which the drug attaches to the inorganic carrier. On the other hand, the studied epimers were experimentally shown to inhibit and modify the growth of calcite, showcasing that they play a significant role in the biomineralization of calcite.

Our research suggests that neither epimer binds to the stabile surface of calcite, while both the *R*- and *S*-epimers adsorb very strongly to the calcite (001) surface. *R*-Sal was found to bind more favorably, with the difference in the calculated adsorption free energies between the two forms amounting to approximately 85 $kJ \cdot mol^{-1}$. In both cases the epimers adsorb to the positively charged half of the growing (001) calcite surface through either three or four available negatively charged carboxyl groups. We explained the observed difference in free energies of adsorption between two epimers by noting that the *R*-epimer has a stronger electrostatic interaction with the (001) calcite surface, exhibiting a predominant adsorption mode in which all four carboxyl groups interact with Ca^{2+} rich surface. More precisely, approximately 15% of *S*-Sal structures are found to be adsorbed via three carboxyl groups, thus lowering the enthalpic contribution to the free energy. This finding agrees very well with the experimental measurements, where it was found that *R*-epimer inhibits/modifies the growth of the calcite mineral more strongly compared to its epimeric counterpart. Our work is the first theoretical study to shed light on the role of the chirality and flexibility on biomineralization and biomineralization-inspired drug design. Our results suggest that chirality can be used, on one hand, as a fine-tuning tool via which one can monitor and modify the growth of the biomineral, while, on the other hand, it may also serve to change and tweak the drug delivery rates by varying the strength of the linker – biomineral interactions, in turn changing the pharmacodynamical response and therapeutic efficacy of the targeted drug delivery systems.

The results of this thesis emphasize the power of *state-of-the-art* theoretical methods, which we successefully used to illuminate the roles of *chirality* and *flexibility*, both for the case of the flexible, chiral molecules in the solution and at inorganic/mineral interfaces. We thereby elucidated the interaction between these two crucial concepts by establishing the bridge between the molecular structure and spectra of highly flexible species for the first time. The interplay between the two has also been investigated in the contexts of biomineralization and biomineralization-inspired drug design. In that context, we were thereby able to clarify the role of flexibility and chirality in the interactions between biomolecules and biomineral surfaces by calculating free energy profiles of adsorption using advanced classical molecular dynamics simulations. Importantly, we find that the interplay between the theory and the experiment is not only helpful, but rather necessary when trying to properly interpret the experimental measurements, which, in turn, enable us to validate and further develop existing theories and models.

Acknowledgments

To begin with, I would like to express my sincere gratitude and appreciation to my supervisor Prof. Dr. Ana-Sunčana Smith and to my co-supervisor Prof. Dr. David M. Smith, who are not only my scientific mentors, but also my mentors in life. Words cannot describe how much I have learned and, perhaps even more importantly, how much sheer fun I had discussing science and the world around us with you. Ana, thank you for your unbelievably profound and deep scientific insights, for your decisiveness which I so often lack, and which I admire you for. Last but not least, I thank you from the bottom of my heart for your constant belief in me, which at times I certainly have not deserved. David, as the movie title correctly states, there are only a few good man, and you are one of them. I feel at the same moment honored, unbelievably lucky, and humbled for having you as my scientific father—the one every PhD student could only wish for. Perhaps most importantly, thank you both for understanding me when I felt nobody else did; you intrinsically motivated me more than any other people I have ever met. In my mind, you will always be my mentors, in my heart, my family.

I am grateful to the Elite Network of Bavaria, the Cluster of Excellence: Engineering of Advanced Materials (EAM), the Unity through Knowledge Fund (UKF) and the European Social Fund (ESF; project MIPoMat) for their financial support.

Sincere gratitude goes to the entire staff of the Cluster of Excellence: Engineering of Advanced Materials and to the Friedrich Alexander University. You were all so very polite and prompt when it came to resolving my never-ending requests. Special thanks go to Waltrud Meinecke – the number of times you have helped me is so large that it almost became uncountable, I can only try to repay you for that with your favorite Kinder Bueno packages.

The work in this thesis was mainly obtained through the use of the computing facilities of the Regionales Rechenzentrum Erlangen (RRZE). I hereby thank the entire staff at the RRZE, and especially Dr. Thomas Zeiser, who was always more than willing to resolve all the issues I experienced, and did so in the shortest possible time.

During my PhD research training I had a great pleasure to work in two different places, namely at the Cluster of Excellence: Engineering of Advanced Materials in Erlangen and at the Institute Ruđer Bošković, situated in Zagreb. Here I would firstly like to express my everlasting gratitude to the German part of my life, namely Sara Kaliman, Mislav Cvitković, Zoran Miličević, Dr. Karmen Čondić-Jurkić, Dr. Timo Bihr, Daniel Schmidt, Robert Stepić, Damir Vurnek and his wife Dragana, Michael Klimczak, and Jayant Pande.

Sara, I start (and always end up) with you. My life and time spent in Erlangen would never be the same without you, you read through me better than any other female I have ever met, which is most probably because you yourself are semi-male. I miss you so much, and always will (mainly because nobody wants to play table tennis with me anymore).

Dear Miclaf, you are my favorite taxi driver in the entire galaxy (and you better be, as you are the star of the BlaBla car, let us not forget that). Jokes aside, you made my life richer and fuller, both in Erlangen and in Zagreb, our discussions made my day even when everything seemed to pale. As a famous Greek adage says: "Će kandžalošemo i opijamooo se viiineeeeem."

Zoran, perhaps the greatest legend of them all. Simply undefeatable, and if you had that chance to spend any time with him, unforgettable. Especially when he combines subjects such as basketball, whiskey, precious metals, and water models in the span of 10 seconds. He will always be remembered as the MPA, i.e., Mister Positive Attitude. Go there and beyond...

My dear Karmen, you brought more sadness and more joy into my life in the year of our Lord 2012 than any other person, male or female. If only I was Australian, sky would have been my limit... but honestly, spending time with you was simply a blast, you are just amazin'.

Dear Timothy Dalton and Daniel (don't get immediately mad for putting the two of you together in the acknowledgments, after all, you are Timo & Mimo), thank you both for teaching me that Mislav is properly pronounced as Miclaf. Thank you Timo for our unbelievable strolls to McDonald's (or as they say it, there and back). Daniel, if for nothing else, thank you for showing me that there exists a human more precise than the atomic clock. Both of you, live long and prosper.

Herr Stepić, what can I tell you that you do not already know... you are the blond embodiment of me, perhaps a bit more redneck, but I guess that will be taken as an additional compliment in your case.

Dear Vurnekci, your arrival in Erlangen brought a special joy in my life, as for the first time I was able to experience sheer joy accompanied by the worst cringeworthy humor imaginable, you are really a fantastic combo, now being furthered by the precious little Alien. Sorry, Elin. Thank you for your existence, you are indeed unique.

Dear Micha (Michael), I am really grateful for the existence of ionic liquids, otherwise I would have never met the guy who listens to a death metal in the BMW Z3 convertible, and prepares some of the best sushi west of China, simply unbelievable combination of attributes. On a more serious note, you are a great person and a good friend, my last years in Erlangen would have been much less enjoyable if it was not for you.

Jayant Pande is not a giant, but he could easily be a panda, as he adores everything bamboo-like. A legend in his own right, I had more fun spending time solving math/logic puzzles with him than watching Die Hard, which says a lot. Of course, we had ample time for solving these puzzles, which we eventually did, all thanks to Jayant's amazing eating proficiency (our lunches lasted at least 2 hours). Thank you Jayant, thank you for everything, it was an amazing journey, never to be forgotten...

I am also sincerely grateful to all members of both the Cluster of Excellence: Engineering of Advanced Materials and to my colleagues at the Institute for Theoretical Physics I, you made me feel at home.

Let me now jump a couple of hundred kilometers to the southeast, whereby I wish to acknowledge my friends and colleagues from the Institute Ruđer Bošković, namely Marko Hanževački, Dr. Jurica Novak, Josip Vlajčević, Jakov Lovrić, Nataša Vučemilović-Alagić and Dr. Momir Mališ.

Markec, also known as the best organic chemist in the world, you make any office experience more enjoyable and more laughable. I cannot imagine the Institute without you and without the amazing energy you bring to work.

Dear Jurek, you are one of the coolest people out there, and I always regret not spending more time with you and Tomislav Jednačak a.k.a. TJ, the parties you two "crash" are the stuff of legends.

Dear Josip and Jakov, the so-called east side gang of the office, you make my every day somewhat sweeter, most probably because the coffee I drink with you has always too much sugar in it. The discussions I have with the two of you are often the most involving and fun part of my day, always reinvigorating me.

Dear Nataša, even though I know you only for half a year, you have become such an essential part of my life that I can barely imagine a day in office without you. Thank you for bringing joy and smile to my face when I felt a weights of a thousand tons upon my shoulders, you truly are a small wonder from Split.

Dear Momir, you are, and always will be, my best friend, and a person who helped me in times when I felt the entire world turned its back on me. You are my greatest supporter, and an unselfish ray of sun which brightened even my darkest days, for which I will always stay in debt to you, my brother in heart.

I would also like to acknowledge all of the competent Institute Ruđer Bošković staff, and especially the members of the group for quantum organic chemistry which I have not already explicitly mentioned, you made my stay at the Institute more enjoyable and more interesting.

This PhD thesis would have never been finished without a constant support from my friends, who were always there for me, in good and bad. I hereby wish to thank Vladimir Damjanović, Nikola Bregović, Anđela Šarić, Matija Juršić, Marko and Matej Hrelja, Marko Mutapčić, Igor Malić, Vedran Guštin, Franko Šarić and Domagoj Mišković for being the best pals a man can wish for, you gave me strength when I thought I have none.

Special thanks goes to Michael Klimczak and my new colleague Dr. Christian Wick for proofreading portions of this thesis, Marko Hanževački for teaching me how to prepare stylish and beautiful figures, Zoran Miličević for his constant support and unselfish help, Jurica Novak for his great insights into the quantum side of the reality, Mislav Cvitković for immensely helpful and exhilarating discussions which inspired me when I found myself stuck in the "mud", and Sara Kaliman, for being Sara. At this point I have to mention three kings and dearest of my friends, namely Jayant Pande, Momir Mališ and Ivan Jelenić, that I am eternally grateful to, and without whom the journey I undertook would be much more bitter and certainly googolplex times less enjoying. Once again, thank you all.

Finally, at the end of this journey through the acknowledgments, I have to thank a small, yet to me most precious group of people, my family. Greatest of all thanks goes to them, they were and are always with me, no matter the distance separating us.

Table of Contents

1. Introduction

The *chirality* of a molecule and its *flexibility* represent two pillars of modern chemistry, underlying a multitude of chemical and biological processes that occur on the planet Earth, and more specifically, in the human body. The former phenomenon, namely chirality (from the Greek word *kheir*, hand), denotes a lack of symmetry at the molecular scale. In this respect, a chiral molecule *a priori* exists in two asymmetrical forms, which we informally call the "right" and "left" forms. These forms are non-superimposable, mirror images of each other, thus resembling our hands (hence the term chirality). Due to the fact that the most notable biological molecules are chiral, including proteins, enzymes, amino acids, carbohydrates, nucleosides and a number of alkaloids and hormones,[1] both chirality and its manifestations are of preeminent importance in chemistry, biochemistry and biology.[2] For example, chirality plays a central role in pharmacology, where it governs both the bioavailability and the activity of the drug at hand. This is due to the fact that the human body serves as an outstanding chiral processing unit, converting each chiral and/or racemic drug in its own way. By utilizing distinct metabolic pathways, different pharmacological activity and different responses at the level of the entire organism, are achieved.[1] The importance of chirality can be best illustrated on the examples of amino acids and sugars, representing the building blocks of proteins and RNA/DNA, respectively. In this respect, the distinguishing feature of life is its exclusive dependence on the *S*-forms of amino acids and the *R*-forms of sugars, whereby their enantiomeric counterparts remain largely uninvolved in biochemical processes.

The concept of chirality, through its realization in chiroptical phenomena, also plays a crucial role in the elucidation of the structures and absolute configurations of chiral molecules/moieties. In particular, Circular Dichroism (CD) spectroscopy, together with X-ray and nuclear magnetic resonance (NMR) techniques, has become a key experimental method in the structural characterization of optically active/chiral molecules. It is predominantly used to determine and monitor the changes (for example, unfolding of proteins) in the secondary structures or conformations of (macro)molecules, particularly polypeptides,[3] proteins[4] and DNA,[5] as their secondary structure is sensitive to the present temperature, environment and/or pH.[6-8] However, although CD spectroscopy represents a powerful means to determine the structure of bio(macro)molecules, we are still far from reaching a full understanding of the extent and scope of the problems to which this method can be successfully applied. This limitation becomes especially evident when attempting to use CD for the characterization of (highly) flexible molecular species.

The second underlying concept of this thesis, namely molecular flexibility, is well known to have a fundamental role in a wide variety of chemical and biochemical processes. For example, in the case of polymers molecular flexibility of the main chain can strongly affect their physical properties,[9] which can ultimately give rise to rubber-like elastic properties of these chemical species.[10,11] Recently, molecular flexibility was found to have important implications in other material science applications, such as functional dyes,[12] liquid crystals,[13] and in the rapidly growing fields of flexible optics[14] and flexible electronics.[15] Perhaps even more dominant is the role of molecular flexibility in biochemistry,[16,17] playing there a governing part in the enzyme-substrate and/or receptor-ligand interactions, where the versatility of conformational states influences the adaptability of these pairs of molecules. Moreover, in pharmacology, the availability of flexible substituents near pharmacophores influences the possibility of additional binding outside an active site or receptor, which can lead to inhibition or antagonism. Important examples of these are the inhibition of HIV-1 replication,[18] antimicrobial defense,[19] as well as cold active enzymes[20] and small heat shock proteins.[21] Although the importance of molecular flexibility has been contested in certain applications,[22] it is generally accepted that it plays a fundamental role in processes such as molecular binding,[23] molecular recognition,[24] and enzymatic catalysis, thereby being one of the three cornerstones of a generalized catalytic mechanism.[25]

Although molecular flexibility is involved in and influences an enormous range of scientific studies, it is rather complicated to investigate it via experimental techniques. This includes the aforementioned, chiroptical methods, such as CD spectroscopy, which can be used to probe the structure of a chiral compound. Despite the fact that this method represents one of the best experimental solutions of the described problem, it turns out that its usefulness is determined largely by the properties of the specific system we are interested in. This is related to the so-called "experimental problem",[26] which arises due to the fact that a large number of experimental observables (among which are also CD spectra) result from averaging over the entire molecular ensemble. Thus, when a molecule exhibits only a small degree of conformational freedom, the advanced experimental techniques can indeed be successfully used to elucidate the structural characteristics of these systems.[4,27,28] Very small molecules (up to 20-30 atoms) represent one such case, while, perhaps surprisingly, large biomolecules (more than 500 atoms), which "live" on the other side of the size spectrum, represent the other. This behavior stems from the fact that, due to their size, large biomolecules commonly form stable tertiary structures, where one structure usually dominates (called the native structure for the case of protein/DNA in a living system).[29] The remainder of the size spectrum is populated with often extremely flexible species, where the concept of a predominant state loses meaning, as now an entire plethora of different molecular conformations exists, each of them contributing to the experimentally measured observable. Additional complexity comes in the form of the observation that similar average values may result from widely disparate distributions (no bijective

property exists), hence it becomes difficult to directly extract the information about the flexibility from only the knowledge of the measured observable, even for previously mentioned "easy" cases. Therefore, even in the light of the development of advanced experimental techniques, computational simulation plays a vital role in obtaining direct information about flexibility and dynamics.[26] The theoretical approach is thus crucial in both understanding the experimental results and in expanding the underlying conceptual framework.

Theoretical methods are becoming increasingly important in all aspects of natural sciences, proving their increasing significance in areas as distinct as surface science,[30,31] catalysis,[25,32] spectroscopy[33] and environmental science.[34] As these methods range from extremely accurate to very approximate, one can use them to investigate an immensely wide scope of natural phenomena. For example, if we are interested in obtaining accurate predictions of molecular properties of small systems, we can employ the so-called *ab initio* methods, which are completely based on quantum mechanics and on the underlying physical constants. As we move along the size spectrum, we are driven toward more approximate methods, termed *semi-empirical* or *empirical*, which, through the application of the empirical parameters, opt to significantly decrease the computational demand without drastically impacting the accuracy of the results. It is important to note that both *ab initio* and *semi-empirical* approaches include approximations, such as the fundamental approximations to the underlying equations of quantum mechanics (e.g. Born-Oppenheimer approximation) or the ones employed to limit the size of the system (e.g. periodic boundary conditions). Only relatively small portion of systems can be treated directly with quantum mechanical methods, as the size of the investigated systems usually does not allow the use of computationally very demanding QM calculations. Thus, to investigate large systems we are forced to abandon the treatment of molecular interactions via quantum mechanics, employing the classical formalism instead. The concept behind this is known as *molecular mechanics* (MM), whereby one evaluates the potential surface by taking into account approximations arising from classical mechanics, in turn significantly decreasing the computational demand compared to the more involving QM calculations. MM methods are usually employed when the long-time phase space behavior of the investigated system is of a predominant significance, i.e when the exact information about electron motions in the system is of a secondary importance. This is the case in the studies involving protein/peptide – inorganic surface adsorption thermodynamics, as well as in the conformational investigations of proteins/peptides. In this respect, *molecular dynamics* simulations, due to the immense increase in the computational power in recent years, are becoming able to access experimentally relevant timescales, on one hand being used to interpret experiments, while on the other hand using experiments to validate and improve already existing models. This interplay between experiments and simulations serves as a new paradigm of 21st century science.

In this thesis we use theoretical techniques coming from both the classical and quantum „worlds“, in an attempt to properly describe features of highly flexible and chiral molecular systems. We have thereby meticulously researched the interplay of the two concepts in two important scenarios, dealing with flexible and chiral molecules in solution and at the interface between water and inorganic interfaces.

In the first scenario we consider the phenomenon of circular dichroism (CD) spectroscopy, thereby attempting to establish the link between *molecular structure* and the *CD spectrum* for highly flexible molecules. This represents a rather daunting task, as it only very recently became feasible for theoretically developed methods to properly describe the two underlying concepts. More precisely, only recently did advancements in classical methods enable us to investigate the so-called conformational phase space of a flexible molecule in solution with a desired accuracy, thus enabling a proper description of the molecular ensemble, which represents a crucial ingredient when attempting to resolve the „experimental problem“.[26] On the other hand, only very recently the quantum methods became sufficiently precise, reliable, and finally, but not less importantly, fast enough to enable the calculation of the CD spectra of complex (systems possessing a wide variety of different excitations) and flexible molecules. The emergent conclusion of our work is that, despite the significant complications that arise, our developed methodology, employing *state-of-the-art* computational methods and statistical tools, indeed enabled us to, for the first time, establish the aforementioned *structure – spectrum* link.

In the second scenario, we tackle the phenomenon of biomineralization and biomineralization-inspired drug design using advanced molecular dynamics simulations. Some of the most interesting systems in this respect, attracting interest from both the pharmaceutical and food industry, are the bioinorganic composites of calcite (calcium carbonate, $CaCO_3$) functionalized by small, biologically active molecules, with the aim of controlled drug delivery.[35,36] In this respect, the advantages of these systems when compared with the micellar carriers are their potentially higher activities and selectivities. However, our understanding of this process at the molecular scale is quite limited, in turn restricting our ability to manipulate with the morphological, chemical and finally, pharmaceutical properties of these systems. Moreover, although peptides represent the most potent biological systems used to functionalize calcite, the role of their flexibility and chirality on the biomineralization and on the properties of biomineralization-inspired drug delivery process remains even more elusive. We thus decided to investigate the interactions of calcite with two highly active peptides/epimers, which are experimentally found to strongly interact with the biomineral. The chosen biomolecule – biomineral composites represent a prototypical drug delivery systems, where the N-termini of the chosen peptides (salicylic acid) plays the role of a drug, while their remaining parts perform the role of a binder/linker via which they attach to calcite carrier. We characterized these composite systems in depth by

employing advanced classical molecular dynamics simulations with enhanced sampling. On one hand, this approach allowed us to analyze the conformational behavior of the adsorbed peptides in detail, while, on the other hand, it permitted us to investigate the underlying thermodynamics of the process by calculating free energy profiles of adsorption. Taken together, our findings enabled us to shed light on the role of flexibility and chirality on both biomineralization and biomineralization-inspired drug design.

We start our journey by elaborating the basic theoretical concepts behind this work in the following chapter, which is followed by the exposition of the results of our research in their full detail.

(1) Nguyen, L. A.; He, H.; Pham-Huy, C. *Int. J. Biomed. Sci.* **2006**, *2*, 85.

(2) Autschbach, J. *Chirality* **2009**, *21*, 116.

(3) Gopal, R.; Park, J. S.; Seo, C. H.; Park Y. *Int. J. Mol. Sci.* **2012**, *13*, 3229.

(4) Greenfield, N. J. *Nat. Protoc.* **2006**, *1*, 2876.

(5) Kypr, J.;Kejnovská, I.; Renčiuk, D.; Vorličková, M. *Nucleic Acids Res.* **2009**, *37*, 1713.

(6) Hirst J. D.; Bhattacharjee S.; Onufriev A. V. *Faraday Discuss.* **2002**, *122*, 253.

(7) Kelly, S. M.; Price, N. C. *Biochim. Biophys. Acta* **1997**, 1*338*, 161.

(8) Pace, C. N. *Methods Enzymol.* **1986**, *131*, 266.

(9) Simha, R. *J. Chem. Phys.* **1945**, *13*, 188.

(10) Bunn, C. W. *Proc. R. Soc. Lond. A* **1942**, *180*, 82.

(11) Bunn, C. W. *Proc. R. Soc. Lond. A* **1942**, *180*, 67.

(12) Marina, M.; Safont-Sempere, M.; Osswald, P.; Stolte, M.; Grüne, M.; Renz, M.; Kaupp, M.; Radacki, K.; Braunschweig, H.; Würthner, F. *J. Am. Chem. Soc.* **2011**, *133*, 9580.

(13) Sebastián, N.; López, D. O.; Diez-Berart, S.; de La Fuente, M. R.; Salud, J.; Pérez-Jubindo, M. A.; Ros, M. B. *Materials* **2011**, *4*, 1632.

(14) John, G.; Jadhav, S. R.; Menon, V. M.; John, V. T. *Angew. Chem. Int. Ed.* **2012**, *51*, 1760.

(15) Park, S.; Wang, G.; Cho, B.; Kim, Y.; Song, S.; Ji, Y.; Yoon, M.-H.; Lee, T. *Nature Nanotech.* **2012**, *7*, 438.

(16) Qi, S.; Krogsgaard, M.; Davis, M. M.; Chakraborty, A. K. *Proc. Natl. Acad. Sci. U. S. A* **2006**, *103*, 4416.

(17) Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. *J. Med. Chem.* **2002**, *45*, 2615.

(18) Teixeira, C.; Serradji, N.; Amroune, S.; Storck, K.; Rogez-Kreuz, C.; Clayette, P.; Barbault, F.; Maurel, F. *J. Mol. Graphics Modell.* **2013**, *44*, 91.

(19) Ohkura, K.; Shinohara, Y.; Hori, H. *Anticancer Res.* **2011**, *31*, 2561.

(20) Papaleo, E.; Tiberti, M.; Invernizzi, G.; Pasi, M.; Ranzani, V. *Curr. Protein Pept. Sci.* **2011**, *12*, 657.

(21) Jaya, N.; Garcia, V.; Vierling, E. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 15604.

(22) Hou, T.; Wang, J.; Zhang, W.; Xu, X. *J Chem. Inf. Model*. **2007**, *47*, 460.

(23) Forrey, C.; Douglas, J. F.; Gilson, M. K. *Soft Matter* **2012**, *8*, 6385.

(24) Janin, J.; Sternberg, M. J. E. *F1000 Biol. Reports* **2013**, *5*, 2.

(25) Hammes, G. G.; Benkovic, S. J.; Hammes-Schiffer, S. *Biochemistry* **2011**, *50*, 10422.

(26) van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; Kastenholz, M. A.;Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F. A.; Yu, H. B. *Angew. Chem. Int. Ed.* **2006**, *45*, 4064.

(27) Leschziner, A. E.; Nogales, E. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 43.

(28) Schuler, B.; Hofmann, H. *Curr. Opin. Struct. Biol.* **2013**, *23*, 36.

(29) Stryer, L.; Berg, J. M.; Tymoczko, J. L. (2002), *Biochemistry* (5th ed.), New York: W. H. Freeman.

(30) Michalka, J. R.; McIntyre, P. W.; Daniel Gezelter, J. *J. Phys. Chem. C* **2013**, *117*, 14579.

(31) Kamali, R.; Kharazmi, A. *Int. J. Ther. Sci.* **2011**, *50*, 226.

(32) Mavrikakis, M. *Nature Mater.* **2006**, *5*, 847.

(33) Jacob, C. R. (2011) *Computational Strategies for Spectroscopy: from Small Molecules to Nano Systems* (1st ed.), edited by V. Barone, Wiley, New York.

(34) Ling, L.; Fan, M.; Wang, B.; Zhang, R. *Energy Environ. Sci.* **2015**, Advance Article.(DOI?)

(35) Fujiwara, M.; Shiokawa, K.; Morigaki, K.; Zhu, Y; Nakahara, Y. *Chem. Eng. J.* **2008**, *137*, 14.

(36) Wang, C.; He, C.; Tong, Z.; Liu, X.; Ren, B.; Zeng, F. *Int. J. Pharm.* **2006**, *308*, 160.

2. Theoretical Background

2.1. Quantum Calculations

Quantum mechanics is the science of small, enabling us to explain the behavior of matter and its interactions with energy on the atomic and sub-atomic scales. Thus, a proper investigation of the processes that include the electron motion and changes in the electron density which readily occur on molecular/atomic scale, such as "breaking" and "making" of chemical bonds, we have to employ quantum mechanics. In the following sections we are going to introduce basic principles of quantum mechanics, thereby also introducing a number of computational techniques arising from them.

2.1.1. Schrödinger Equation in a Nutshell

Most of the modern computational chemistry heavily relies on one of the most important scientific concepts developed in the 20th century, namely the Schrödinger wave function, usually denoted by Greek letter Ψ. The wave function is a complex single-valued function which contains the complete information about the system. Due to its nature it does not possess any physical meaning by itself, however, the scalar product of the wave function with its complex conjugate, $\Psi(q,t)\Psi^*(q,t)$, bears the notion of the probability of finding a particle at a particular position q and time t, the concept known as Born interpretation. The time-dependent Schrödinger's equation, completely determining the future of the system, is given by:

$$i\hbar\frac{\partial}{\partial t}\Psi(q,t) = \boldsymbol{H}\Psi(q,t). \tag{2.1}$$

Here $\boldsymbol{H}$ represents the famous Hamiltonian operator, which describes the total energy of the system at a given time t. When Hamiltonian operator $\boldsymbol{H}$ does not depend on time, it can be shown that the "full" wave function $\Psi(q,t)$ can be written as a product of two separate functions $\Psi(q,t) = \psi(q)\varphi(t)$, where $\psi(q)$ and $\varphi(t)$ depend only on the spatial coordinates and time, respectively. We are then in the position to write time-independent Schrödinger equation as:

$$\boldsymbol{H}\psi(q) = E\psi(q), \tag{2.2}$$

$$i\hbar\frac{\partial\varphi(t)}{\partial t} = E\varphi(t), \tag{2.3}$$

where $\psi(q)$ represent the eigenfunctions of the time-independent Hamiltonian operator, with the corresponding eigenvalues E, which are interpreted as the energy. Eq. 2.2 is easily solved, yielding

$$\varphi(t) = e^{-iEt}\,. \tag{2.4}$$

Therefore, the overall solution of the time-dependent Schrödinger equation for this special case becomes

$$\Psi(q,t) = \psi(q)e^{-iEt}. \tag{2.5}$$

The Hamiltonian operator $\boldsymbol{H}$ for charged particles in atomic units is given by the following expression:

$$\boldsymbol{H} = -\sum_{i=1}^{N}\frac{1}{2}\boldsymbol{\nabla}_i^2 - \sum_{A=1}^{M}\frac{1}{2M_A}\boldsymbol{\nabla}_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{\mathbf{q}_{iA}} + \sum_{i}^{N-1}\sum_{j>i}^{N}\frac{1}{\mathbf{q}_{ij}} + \sum_{B>A}^{M}\frac{Z_A Z_B}{\mathbf{Q}_{AB}}, \tag{2.6}$$

where $\boldsymbol{\nabla}_i^2$ is the Laplace operator acting upon particle i. The first term in the expression describes the kinetic energy of the electrons within the wave mechanics formulation, the second one describes the kinetic energy of the nuclei, the third term (or sum) represents the energy arising from the attraction between the electrons and the nuclei, while the last two terms give the contributions due to the repulsive Coulombic forces between nuclei and electrons among themselves, respectively. The indices i, j represent electrons, indices A, B denote nuclei, while N and M represent the number of electrons and nuclei, respectively. The above is the time-independent non-relativistic formulation of the Schrödinger equation. Depending on the boundary conditions imposed onto the system of interest, the Schrödinger equation is often solvable for particular values of E and it cannot be solved analytically for most of the systems, exceptions being the one-electron systems. However, because the (atomic) nuclei are considerably heavier than electrons, their velocities are much lower in comparison. We can thus approximate the behavior of the systems by assuming that the electrons move in a field of fixed nuclei, which is the so-called Born-Oppenheimer approximation. This approximation performs so well that the electronic ground state properties remain almost completely unaffected, and is thus used by default in quantum chemistry. The overall wave function Ψ_{tot} describing the state of the molecules can now be written as

$$\Psi_{\mathrm{tot}}(q_{el}, q_N) = \Psi_{el}(q_{el}, q_N)\Psi_N(q_N), \tag{2.7}$$

$$(H_{el} + V_N) = U\Psi_{el}. \tag{2.8}$$

The Hamiltonian H_{el} that arises is called the electronic Hamiltonian, whereas the solutions to the resulting Schrödinger equation are called the electronic wave function and electronic energy. The solutions to the obtained electronic Schrödinger equation are eigenfunctions $\psi(q)$, describing different stationary states of the system. Each of the obtained eigenstates comes with a corresponding eigenvalue E, representing the energy of its respective eigenfunction. The state with the lowest energy/eigenvalue (E_0) is called the electronic ground-state (ψ_0). The ground-state is most commonly described by a product of one-electron molecular orbitals ($\Phi_\mu(q)$), which account for the spatial

distribution of a single electron. Any proper description of an electron would be incomplete without introducing the spin of the electron into the formalism, which is usually done using two orthonormal spin functions, represented as $\alpha(s)$ and $\beta(s)$, denoting the spin up and spin down eigenfunctions, respectively. A complete wave function of an electron is now constructed by taking the product of a spin function and the molecular orbital, giving rise to a so-called spin orbital $\chi(q,s)$, which fully describes the behavior of a single electron. At this point one can attempt to construct a wave function characterizing the system consisting of *n* electrons. This can be accomplished by making a product of *n* spin orbitals $\chi(q,s)$, which in turn represents the most basic form of the complete wave function. However, the total wave function for any fermion (electron in this particular case) must be antisymmetric with respect to an interchange in coordinates of any two of them. There exists a certain mathematical form of the many electron wave function that fulfils this requirement, namely a Slater determinant, which can be written as

$$\psi(q) = \frac{1}{\sqrt{n!}} \begin{vmatrix} \chi_1(1) & \chi_2(1) & \cdots & \chi_n(1) \\ \chi_1(2) & \chi_2(2) & \cdots & \chi_n(2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(n) & \chi_2(n) & \cdots & \chi_n(n) \end{vmatrix}. \tag{2.9}$$

Here the interchange of two electrons corresponds to an exchange of two of the rows in the determinant, and thus a change in the sign of the overall wave function, thereby guaranteeing that the antisymmetry condition is appropriately treated. Additionally, this determinant obeys the famous Pauli exclusion principle, which claims that two electrons cannot possess the same set of quantum numbers, from which it follows that a specific molecular orbital cannot be simultaneously occupied with electrons of the equal spin.

Molecular orbitals $\varphi_i(q)$ are most commonly constructed from a set of known (basis) functions. In this respect, any adequately defined basis functions can be used to construct molecular orbitals, including polynomials, exponentials, Gaussians, etc. The chosen basis functions have to fulfill two main criteria; the underlying physics has to be satisfied, i.e., the functions have to approach zero when the electron – nucleus distance becomes large enough, and secondly, the chosen functions should allow a relatively simple evaluation of the arising integrals. Combined, these two criteria imply that the most favorable way to construct MOs is to employ atomic orbitals, usually denoted by $\rho_\mu(q)$, which are centered on the atoms of the investigated molecule. In that case we obtain the expression known as the linear combination of atomic orbitals (LCAO):

$$\varphi_i(q) = \sum_\mu c_{\mu i}\, \rho_\mu(q). \tag{2.10}$$

Overall, to solve the Schrödinger equation we only have to determine the molecular-orbital expansion coefficients $c_{\mu i}$ in the Eq. 2.10, thereby fully specifying the ground-state electronic wave function. The coefficients $c_{\mu i}$ are arrived at using the variational principle, which guarantees that any approximated wave function will, for its eigenvalue, produce an energy greater than the real energy, which would represent the exact solution of the Schrödinger equation. Thus, the variational method represents a powerful mathematical apparatus which is able to provide us with an upper limit for the exact energy of the ground state, whereby one, by minimizing the energy with respect to the coefficients $c_{\mu i}$, obtains the wave function optimally describing the system. Additionally, the variational principle guarantees that the results improve, i.e., the obtained ground state energy becomes lower when more basis set functions are employed. We thus inspect the convergence and quality of our results by performing calculations with larger basis sets.[1]

2.1.2. Hartree-Fock (HF) Theory

In this section we focus on Hartree-Fock (HF) theory,[2,3] which is a cornerstone of quantum chemistry, and which laid the groundwork for a number of modern *ab initio* quantum mechanical methods (usually termed *post-Hartree-Fock* methods). As explained in the previous section, the coordinates of the electrons in the investigated system have to be made separable in order to solve the Schrödinger equation. This can be achieved by constructing an approximate Hamiltonian, whereby the electron interactions are not accounted for directly, but rather in an average fashion. More precisely, each electron experiences repulsion due to an average position of the remaining electrons in the system. By taking this into account and solving the emerging Schrödinger equation one can calculate the exact one-electron eigenfunctions, representing the optimal one-electron wave functions. Finally, the variational method can be applied to minimize the total energy, thereby producing the following set of equations, usually called Hartree-Fock equations:

$$\boldsymbol{F}\varphi_i = \varepsilon_i \varphi_i \,. \tag{2.11}$$

Here $\boldsymbol{F}$ represents the so-called Fock operator, while the energy corresponding to the one-electron molecular orbital φ_i is denoted with ε_i. The Fock operator is a sum of a one-electron Hamiltonian and the Hartree-Fock potential which is the average potential felt by electron i due to electron-electron repulsion. Solving Hartree-Fock equations involves the expansion of the molecular orbitals $\varphi_i$using a set of one-electron basis set functions, giving rise to Eq. 2.10. The coefficients $c_{\mu i}$ and orbital energies ε_i in the Equations 2.10 and 2.11 respectively can be obtained via the Roothan-Hall equations[4,5]

$$\sum_{\mu=1}^{N} c_{\mu i} \left(F_{\mu\nu} - \varepsilon_i S_{\mu\nu}\right) = 0. \tag{2.12}$$

When the system contains N basis functions/atomic orbitals, both $F_{\mu\nu}$ and $S_{\mu\nu}$ are the elements of $N \times N$ Hermitian matrices, former being known as the Fock matrix ($\boldsymbol{F}$), while the latter the so-called overlap matrix ($\boldsymbol{S}$). The above equation can be written as a generalized eigenvalue problem

$$\boldsymbol{FC} = \boldsymbol{SC\varepsilon}. \tag{2.13}$$

Here $\boldsymbol{C}$ is the N-order matrix containing the coefficients $c_{\mu i}$ as its elements, with $\boldsymbol{\varepsilon}$ represented by the diagonal matrix having ε_i for its elements. The Hartree-Fock differential equation has now been converted to a matrix equation which can be solved by standard matrix techniques. The problem of calculating the molecular orbitals φ_i reduces to the problem of calculating the set of unknown expansion coefficients $c_{\mu i}$. However, the explicit dependence of the Fock operator on the $c_{\mu i}$ coefficient values makes the Roothaan-Hall equations non-linear. Thus, as mentioned earlier, these equations have to be solved in an iterative fashion. This is performed by making a guess for the orbitals, where the behavior of each electron is described as if it interacts only with the nucleus, experiencing all the other electrons through the average field created by them. The quality of the orbitals is improved iteratively, and this process is discontinued when the coefficients $c_{\mu i}$ remain unchanged in two successive iterations (the described procedure is also known as the *self-consistent field* (SCF) method).

2.1.3. Basis Set

The wave function is usually expanded in the form of Eq. 2.10, with ρ_μ being the basis functions and $c_{\mu i}$ their respective expansion coefficients. Using more basis functions enables us to lower the energy obtained from Hartree-Fock procedure, with this trend progressing until the basis set approaches completeness (point at which Hartree-Fock limit is reached). However, exact expansion would necessitate an infinite number of basis functions, which is impossible to implement in practice, leading to the usage of finite basis sets, from which stems the so-called basis set truncation error.[6] For the electronic structure calculations one typically uses two types of basis functions, namely the Slater Type Orbitals (STO) and Gaussian Type Orbitals (GTO). Although the STOs ($\rho_\mu \sim e^{-\zeta r}$) describe the atomic orbitals as appearing on atoms in molecules more precisely, the problem appears when a molecule contains more than 2 atoms, as then the two-electron integrals arise, which cannot be evaluated analytically but only numerically, representing a computationally demanding task. On the other hand, GTOs ($\rho \sim e^{-\zeta r^2}$) are significantly more computationally feasible, since the integrals appearing during HF procedure can be solved analytically. It became a usual practice to prepare linear combinations of GTOs (*primitives*, PGTO) in order to approximate more accurate STOs, thereby obtaining contracted Gaussian type orbitals (*contracted*, CGTO).[7]

This thesis contains primarily the applications of the Pople basis set 6-31G(d), which belongs to the so-called split valence basis sets. In this basis set a contraction of six PGTOs gives rise to core orbitals, while the inner and outer parts of the valence orbitals are represented by a contraction of three PGTOs and one PGTO, respectively. Finally, in this particular case *p*-type polarization functions have been added to hydrogen atoms, while *d*-type polarization functions have been added to the atoms belonging to the second row of the periodic table, respectively. We also utilized the basis sets of Dunning and co-workers, namely TZVP,[8] while also using a larger Pople basis set, namely 6-311++G(2df,2p), for testing purposes.

2.1.4. Drawbacks of Hartree-Fock Method

HF method has two major drawbacks. Firstly, each energy eigenfunction is assumed to be describable by a single Slater determinant. Secondly and more importantly, HF lacks the proper description of the electron correlation, namely, while the method gives rise only to the average effect of inter-electronic repulsion, the instantaneous interaction among the electrons is neglected (although it fully accounts for the electron exchange). Despite the aforementioned flaws, HF method can, for example, be successfully applied to obtain optimized geometries and to calculate electrostatic potential around the molecule of interest, for which we employed it in this thesis (see Section 2.2.4.1 for details). However, the highlighted aspects of HF are especially unfavorable in the calculation of the electronic excited states, where a time-dependent (TD) extension of HF would have to be applied. Due to the described problems a number of different methods made to improve HF have been developed. Examples of such methods are the Møller-Plesset perturbation theory (MP)[9], and the configuration interaction (CI)[10] and coupled-cluster (CC)[11,12] approaches. On the other hand, density functional theory (DFT)[13] provides a conceptually different theoretical framework to incorporate the effects of electron correlation, depending on the electron density as the underlying quantity. Of the aforementioned methods, we have employed both the TD-DFT methods and coupled-cluster approach to describe the excited state properties of molecules. The theories behind these two approaches are explained in the following sections.

2.1.5. Density Functional Theory (DFT)

DFT[13] represents a formally exact approach to solve a static many-electron problem, by completely bypassing the formalism of the many-electron wave function and replacing it with the electron density, which now serves as a fundamental quantity. The main advantage of this theory compared to the traditional wave function – based methods lies in the fact that it substitutes the many-body problem involving n electrons ($3n$ spatial coordinates) with a problem of finding electron density, which depends on three spatial coordinates. The modern DFT is based on the work of Hohnberg and Kohn[14] and is governed by three main principles. The first of these (Hohnberg-Kohn existence theorem) states

that for any system of interacting particles, the electron density uniquely determines the external potential, thereby implying that the ground state energy of the system is a unique functional of the electron density. The second Hohnberg-Kohn theorem[7] states that the ground state energy can be obtained using the variational principle, where the electron density minimizing the ground state energy represents the exact ground state density. One can thus apply the equivalent mathematical apparatus used in the Hartree-Fock methodology to find the best possible electron density, which in turn corresponds to the lowest energy of the system. In short, the Hohnberg-Kohn theorems state that it is possible to uniquely determine all ground state properties of the system without an inherent knowledge of the wave function. However, the proofs of the two theorems are non-constructive in their nature, i.e., they do not tell us how to obtain ground state energy from the electron density nor do they suggest a way of obtaining the electron density. This problem was resolved by Kohn and Sham by introducing an auxiliary fictitious system of non-interacting electrons experiencing an effective one-electron potential (known as Kohn-Sham potential),[15] which are postulated to have the same overall ground state density as the real system of interacting particles. The ground state energy of the real system can then be expressed as a functional of the electron density, and is usually written in the following form:

$$\begin{aligned} E(\rho) = \;& E^T(\rho) + E^V(\rho) + E^J(\rho) + E^{XC}(\rho) \\ = \;& E^T(\rho) + \int d^3q\rho(\mathbf{q})v_0(\mathbf{q}) + \frac{1}{2}\int d^3q \int d^3q' \frac{\rho(\mathbf{q})\rho(\mathbf{q}')}{|\mathbf{q}-\mathbf{q}'|} + E^{XC}(\rho). \end{aligned} \tag{2.14}$$

Here $E^T(\rho)$ corresponds to the kinetic energy of the fictitious system of non-interacting electrons, $E^V(\rho)$ denotes the electron-nuclei attraction, and $E^J(\rho)$ describes the Coulombic repulsion between the electrons. The final term, namely $E^{XC}(\rho)$, represents the exchange-correlation energy which contains all the corrections to the first three terms and contributions not taken into account by them, including also the corrections to the kinetic energy arising from the electron-electron interactions.

The electronic density of the ground state of the interacting system can be written in terms of auxiliary one-electron orbitals $\varphi_i(\mathbf{q})$,

$$\rho_0(\mathbf{q}) = \sum_{i=1}^{N} |\varphi_i(\mathbf{q})|^2. \tag{2.15}$$

These one-electron orbitals are known as Kohn-Sham orbitals, and are obtained iteratively, as mentioned above, following exactly the methodology developed for the Hartree-Fock approach, however this time by solving the one-electron Kohn-Sham equations:

$$\left(-\frac{\nabla^2}{2} + v_s[\rho](\mathbf{q})\right)\varphi_i(\mathbf{q}) = \varepsilon_i\varphi_i(\mathbf{q}). \tag{2.16}$$

One can generalize Eq. 2.16 by introducing the spin of the electrons, namely, for the system consisting of *N* electrons, where $N = N_{\uparrow} + N_{\downarrow}$, the ground state electron density can be written as the sum of the spin "up" and "down" electron densities:

$$\rho_0(\mathbf{q}) = \rho_{0\uparrow}(\mathbf{q}) + \rho_{0\downarrow}(\mathbf{q}) = \sum_{\sigma=\uparrow,\downarrow} \sum_{i=1}^{N_\sigma} |\varphi_{i\sigma}(\mathbf{q})|^2. \tag{2.17}$$

Although it can be shown that there exists an unique functional describing exchange-correlation energy $E^{XC}(\rho)$, which holds for all systems and all cases, the exact form of this functional is unknown, except for certain special cases. Thus, in order to use DFT one needs to approximate the exchange-correlation functional, thereby three main approaches have been developed – the local density approximation (LDA), the generalized gradient approximation (GGA) and the hybrid method. In all three aforementioned approaches one writes the exchange-correlation functional $E^{XC}(\rho)$ as the sum of two terms, namely the exchange term $E^{X}(\rho)$ and the correlation term $E^{C}(\rho)$, representing the simplest possible approximation,

$$E^{XC}(\rho) = E^{X}(\rho) + E^{C}(\rho). \tag{2.18}$$

In the most naive of the three approximations, the LDA, one assumes that the electron density behaves as a uniform electron gas, with the value of the potential at some point depending only on the value of the electron density evaluated at that point. The generalization of the LDA that takes into account the spin of the electrons is known as the local spin density approximation (LSDA), in which the electron density is given as a sum of α and β densities. However, due to the generally inhomogeneous distribution of the electron density in molecular species, the LDA approximation cannot be used to reliably and accurately model chemical processes. Improvement can be achieved by employing GGA, where one includes the variations in the density by considering the gradient of the density, yielding an overall substantial increase in the accuracy of DFT calculations. Finally, the members of the third group, namely the hybrid functionals, provide further improvements by mixing certain portions of LDA/GGA based components with a fraction of the HF exchange energy – as HF predicts the exact exchange energy for non-interacting electrons. Perhaps the most well-known example of hybrid functionals is the three-parameter model developed by Becke (B3).[16] Combining this hybrid functional with Lee-Yang-Parr's Correlation Functional (LYP) produces B3LYP,[17] representing one the most widely used functionals in computational chemistry.

2.1.6. Time-Dependent Density Functional Theory

In this section (parts of the text were directly taken from [18]) we consider the time-dependent density functional method (TD-DFT),[19] representing a direct extension of the DFT. This method enables one

to characterize a system of interest in the presence of external force fields, namely magnetic and electric fields, enabling the calculation of the excited state properties. Hohenberg-Kohn-Sham DFT is based on the Rayleigh-Ritz variation principle. In the case of a time-dependent external potential, however, no minimum principle exists. Instead, there is a stationary-action principle. The starting point of studying time-dependent systems is the electronic Schrödinger equation

$$\mathrm{i}\frac{\partial}{\partial t}\Psi(\mathbf{x}_1,\ldots,\mathbf{x}_N,t) = \boldsymbol{H}(t)\Psi(\mathbf{x}_1,\ldots,\mathbf{x}_N,t), \tag{2.19}$$

with the Hamiltonian of the system being

$$\boldsymbol{H}(t) = \boldsymbol{T} + \boldsymbol{V}(t) + \boldsymbol{W} \tag{2.20}$$

$$\boldsymbol{H} = -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 + \sum_{i=1}^{N} v(\mathbf{q}_j,t) + \sum_{j>i}^{N}\frac{1}{\mathbf{q}_{ij}}. \tag{2.21}$$

Here the term $\boldsymbol{W}$ vanishes if one deals with a system of non-interacting electrons. According to the Runge-Gross theorem,[19] for a given initial wave function, there exists a unique mapping between the time-dependent external potential of a system and its time-dependent density. Overall, this implies that the wave function at any time t is known up to the time-dependent phase, and can be represented as a functional of the density and the initial state Ψ_0 only, namely as

$$\Psi(t) = e^{\mathrm{i}\alpha(t)}\Psi[\rho,\Psi_0](t). \tag{2.22}$$

Analogously to the case of the ground state, the time-dependent density can also be calculated by considering an auxiliary system of non-interacting electrons in which the electrons move in an effective time-dependent potential. This potential is unique by the virtue of Runge-Gross theorem applied to Kohn-Sham system. To derive time-dependent Kohn-Sham equations we have to assume the existence of the time-dependent reference system with the non-interacting particles (electrons) with the external potential $V_s(\mathbf{q},t)$[20,21] and orbitals $\varphi_i(\mathbf{q},t)$ from which the electronic density $\rho_s(\mathbf{q},t)$ follows, which is by definition equal to the exact electronic density $\rho(\mathbf{q},t)$ of the real interacting system

$$\rho(\mathbf{q},t) = \rho_s(\mathbf{q},t) = \sum_{i=1}^{N}|\varphi_i(\mathbf{q},t)|^2. \tag{2.23}$$

In analogy with the time-independent case, the time-dependent Kohn-Sham orbitals are now obtained by solving the one-electron Schrödinger equation (under the condition that the external potential exists)

$$\mathrm{i}\frac{\partial}{\partial t}\varphi_i(\mathbf{q},t) = \left(-\frac{1}{2}\nabla_i^2 + V_s(\mathbf{q},t)\right)\varphi_i(\mathbf{q},t), \tag{2.24}$$

$$\mathrm{i}\frac{\partial}{\partial t}\varphi_i(\mathbf{q},t) = \boldsymbol{F}^{KS}\varphi_i(\mathbf{q},t), \tag{2.25}$$

where $\boldsymbol{F}^{KS}$ is the Fock operator, while the $V_s(\mathbf{q},t)$ represents the time-dependent external potential of the non-interacting system, and is given by

$$V_s(\mathbf{q},t) = V(\mathbf{q},t) + \int d^3q' \frac{\rho'(\mathbf{q}',t)}{|\mathbf{q}-\mathbf{q}'|} + A^{XC}(\rho), \tag{2.26}$$

with $V(\mathbf{q},t)$ being the external potential of the interacting system. All of the exchange-correlation effects are captured in the time-dependent exchange correlation functional $A^{XC}(\rho)$, playing the role of the exchange-correlation functional $E^{XC}(\rho)$ from the time-independent case, and, analogously to $E^{XC}(\rho)$, its exact form is unknown.[1] Approximation which enables the application of the time-independent exchange-correlation functionals in the time-dependent problems is the so-called adiabatic local density approximation (ALDA), reminiscent of the LDA, which assumes the slow change of the electron density with time, thereby implying

$$A^{XC}(\rho) \cong E^{XC}(\rho) \tag{2.27}$$

2.1.6.1. Linear Response TD-DFT (LR-TD-DFT)

In the cases in which the electron density of the excited state does not differ significantly from the density of the ground state, e.g. in the case of a small perturbation, instead of solving a full time-dependent Schrödinger equation it is more practical to use the linear response theory,[22-26] especially in the calculations of spectroscopic observables, such as excitation energies, oscillatory strengths and rotatory strengths. The response of any quantum-mechanical observable $\hat{\alpha}$ on the perturbation is given by the difference of the time-dependent expectation value of the observable ($\alpha(t)$) and the inital value (α_0), and can be represented by the following expansion

$$\alpha(t) - \alpha_0 = \alpha_1(t) + \alpha_2(t) + \alpha_3(t) + \cdots, \tag{2.28}$$

where $\alpha_1(t)$ represents the linear respose.

The linear response can be written as

$$\alpha_1(t) = \int_{-\infty}^{+\infty} dt' \chi_{\alpha\beta}(t-t')F(t'), \tag{2.29}$$

where $F(t)$ is the external field coupled with the observable β and turned on at the moment t_0 on, while $\chi_{\alpha\beta}$ is the retarded response function. To apply the linear response theory to TD-DFT, we first define the time-dependent external potential

$$v(\mathbf{q},t) = v_0(\mathbf{q}) + v_1(\mathbf{q},t)\Theta(t-t_0), \tag{2.30}$$

where $v_1(\mathbf{r},t)$ denotes a small perturbation which is turned at time t_0, and $\Theta(t-t_0)$ represents the step function. Now let the system be in its ground state during the time $t < t_0$. Then the linear response of the electron density equals

$$\rho_1(\mathbf{q},t) = \int dt' \int d^3q' \chi_{\rho\rho}(\mathbf{q},t,\mathbf{q}',t')v_1(\mathbf{q}',t'), \tag{2.31}$$

where the so-called density-density response function $\chi_{\rho\rho}$ is given by

$$\chi_{\rho\rho}(\mathbf{q},\mathbf{q}',t-t') = -\mathrm{i}\Theta(t-t')\langle\Psi_0|[\hat{\rho}(\mathbf{q},t-t'),\hat{\rho}(\mathbf{q}')]|\Psi_0\rangle. \tag{2.32}$$

By applying the Fourier transformation to the Eq. 2.32 we can obtain the frequency-dependent response of the electron density

$$\rho_1(\mathbf{q},\omega) = \int d^3q' \chi_{\rho\rho}(\mathbf{q},\mathbf{q}',\omega)v_1(\mathbf{q}',\omega). \tag{2.33}$$

Here $\chi_{\rho\rho}(\mathbf{q},\mathbf{q}',\omega)$ is the Lehmanns' representation of the density-density response function and is given by

$$\chi_{\rho\rho}(\mathbf{q},\mathbf{q}',\omega) = \sum_{n=1}^{\infty}\left\{\frac{\langle\Psi_0|\hat{\rho}(\mathbf{q})|\Psi_n\rangle\langle\Psi_n|\hat{\rho}(\mathbf{q}')|\Psi_0\rangle}{\omega-\Omega_n+\mathrm{i}\eta} - \frac{\langle\Psi_0|\hat{\rho}(\mathbf{q}')|\Psi_n\rangle\langle\Psi_n|\hat{\rho}(\mathbf{q})|\Psi_0\rangle}{\omega+\Omega_n+\mathrm{i}\eta}\right\}, \tag{2.34}$$

where $\{\Psi_n\}$ denotes the set of eigenfunctions of the Hamiltonian $\boldsymbol{H}_0$, with the ground state wave function Ψ_0 of the energy E_0, while $\Omega_n = E_n - E_0$ represents the energy of the n-th excitation. However, this interacting response function $\chi_{\rho\rho}(\mathbf{q},\mathbf{q}',\omega)$ is immensly hard to calculate, thus we apply the non-interacting Kohn-Sham system to resolve this problem. More precisely, by equating the response of the density of the multi-particle interacting system with the density response of the non-interacting one we obtain the Dyson equation, the central equation of TD-DFT, which connects the

interacting ($\chi_{\rho\rho}(\mathbf{q},t,\mathbf{q}',t')$) and non-interacting ($\chi_{\rho\rho S}(\mathbf{q},t,\mathbf{q}',t')$) response functions and which can be written as

$$\begin{aligned}\chi_{\rho\rho}(\mathbf{q},t,\mathbf{q}',t') = \quad & \chi_{\rho\rho S}(\mathbf{q},t,\mathbf{q}',t') + \int d\tau \int d^3x \int d\tau' \int d^3x' \\ & \cdot \chi_{\rho\rho S}(\mathbf{q},t,\mathbf{x},\tau)\left\{\frac{\delta(\tau-\tau')}{|\mathbf{x}-\mathbf{x}'|} + f_{xc}(\mathbf{x},\tau,\mathbf{x}',\tau')\right\}\chi_{\rho\rho}(\mathbf{x}',\tau',\mathbf{q}',t'),\end{aligned} \tag{2.35}$$

where $f_{xc}(\mathbf{x},\tau,\mathbf{x}',\tau')$ represents the time-dependent exchange-correlation kernel

$$f_{xc}(\mathbf{q},t,\mathbf{q}',t') = \left.\frac{\delta v_{xc}[\rho](\mathbf{q},t)}{\delta\rho(\mathbf{q}',t')}\right|_{\rho_0(\mathbf{q})}. \tag{2.36}$$

By applying the Fourier transformation to the LR-TD-DFT equation we obtain the response of the density in the frequency domain, representing an analog to Eq. 2.34, namely

$$\rho_1(\mathbf{q},\omega) = \int d^3q' \chi_{\rho\rho S}(\mathbf{q},\mathbf{q}',\omega)\left[v_1(\mathbf{q}',\omega)\int d^3x\left\{\frac{1}{\mathbf{q}'-\mathbf{x}} + f_{xc}(\mathbf{q}',\mathbf{x},\omega)\right\}\rho_1(\mathbf{x},\omega)\right]. \tag{2.37}$$

Here $\chi_{\rho\rho S}(\mathbf{q},\mathbf{q}',\omega)$ represents the Fourier transform of the time-dependent response function in the Kohn-Sham formalism with the non-interacting particles

$$\chi_{\rho\rho S}(\mathbf{q},\mathbf{q}',\omega) = \sum_{j,k=1}^{\infty}(f_k - f_j)\frac{\varphi_j^0(\mathbf{q})\varphi_k^{0*}(\mathbf{q})\varphi_j^{0*}(\mathbf{q}')\varphi_k^0(\mathbf{q}')}{\omega - \omega_{jk} + \mathrm{i}\eta}, \tag{2.38}$$

with f_j and f_k representing the occupation numbers of the Kohn-Sham orbitals, possessing value 1 and 0 for the occupied and virtual orbitals, respectively, with $\omega_{jk} = \varepsilon_j - \varepsilon_k$ being the difference between the eigenvalues (energies) of the Kohn-Sham orbital. Moreover, only the terms in which one index denotes the occupied and the other one the virtual orbitals contribute to the sum in Eq. 2.38. Most importantly, the response function $\chi_{\rho\rho S}(\mathbf{q},\mathbf{q}',\omega)$ of the non-interacting Kohn-Sham system has poles where the excitation energies of the Kohn-Sham system appear. It is thus exactly the Dyson equation that enables us to determine the excitation energies of the system of interest.

The most important application of TD-DFT in chemistry is the matrix formulation of the described LR-TD-DFT developed by Casida,[25] enabling the calculations of excitation energies, and both the oscillator and rotatory strengths, from which one builds the ultraviolet (UV) and circular dichroism (CD) spectra, respectively. By considering the response of the Kohn-Sham density matrix, he derived the commonly known "Casida" equations, which are similar in structure with time-dependent Hartree-Fock, and are what is coded in most of the quantum packages today.

TD-DFT calculations, in particular when using hybrid functionals, like already mentioned B3LYP, and sufficiently large basis set, can provide very accurate results for excitation energies, electric and magnetic dipole moments for well localized states involving mainly single excitations. This guarantees a wide application of TD-DFT for calculation of chiroptical properties of a number of different chiral compounds,[27] ranging from metal complexes[28,29] to biologically important systems.[30-32] Nevertheless, implementations of TD-DFT may be associated with somewhat unpredictable precision,[33] particularly for larger systems.[30] In the context of electronic spectra calculations, TD-DFT is known to underestimate the excitation energies of transitions with a significant charge transfer (CT) character,[34-36] including those in peptide systems.[37] This problem stimulated the development of long-range corrected functionals (LCFs), including the relatively successful Coulomb-attenuated functional CAM-B3LYP[38] and the dispersion-encompassing ωB97X-D.[39] The hybrid meta GGA functional (HMF) M06-2X,[40] on the other hand, has been shown to predict CT transitions, with intermediate spatial overlap, with accuracy similar to that of CAM-B3LYP.[36] Moreover, it was very recently shown that M06-2X provides reasonable excitation energies for a large set of conjugated molecules.[41] However, even the more traditional global hybrid functionals (GHFs), such as B3LYP[17] and to a lesser extent PBE0,[42] have proven quite accurate in some cases.[29-43] Indeed, recent extensive benchmarking studies found that the B3LYP functional often exhibits lower deviations from experiment than, for example, CAM-B3LYP.[33,44] These conclusions have been recently generalized in an review article.[45]

2.1.7. Ab Initio Quantum Mechanics – Coupled Cluster Methods

This section provides a short description of the coupled-cluster theory,[12] focusing on the CC2 methodology, which we used in this study to calculate excited state properties of small peptide fragments. The methodology follows the general overview of the coupled cluster theory presented in [1]. The exponential ansatz for coupled cluster theory is given by:

$$\Psi_{CC} = e^{\boldsymbol{T}}\Psi_{HF} \tag{2.39}$$

where Ψ_{HF} denotes the Hartree-Fock wave function which serves as a reference, while $e^{\boldsymbol{T}}$ represents an operator defined through a Taylor expansion in $\boldsymbol{T}$:

$$e^{\boldsymbol{T}} = 1 + \boldsymbol{T} + \frac{\boldsymbol{T}^2}{2!} + \frac{\boldsymbol{T}^3}{3!} + \cdots = \sum_{i=0}^{\infty} \frac{\boldsymbol{T}^i}{i!} \tag{2.40}$$

The operator $\boldsymbol{T}$ is known as the cluster operator, and is represented by the sum of the i-th particle operators $\boldsymbol{T_i}$ $(i = 1,2,3,\ldots,n)$, n denoting the number of electrons in the system:

$$\boldsymbol{T} = \boldsymbol{T}_1 + \boldsymbol{T}_2 + \boldsymbol{T}_3 + \cdots + \boldsymbol{T}_n. \tag{2.41}$$

When operators $\boldsymbol{T_i}$ act on the HF reference function they produce all the *i*-th excited Slater determinants, i.e., $\boldsymbol{T}_1$ includes only the excitations of single electrons, while $\boldsymbol{T}_2$ includes the ones involving two electrons, etc.[1] Each generated determinant comes with a coefficient which is called the cluster amplitude. The operator $e^{\boldsymbol{T}}$ expresses the coupled cluster wave function Ψ_{CC} as a linear combination of the Slater determinants that include both the reference Hartree-Fock wave function Ψ_{HF} and, additionally, all transitions of the electrons from the occupied orbitals into the unoccupied/virtual ones. The arising total wave function incorporates these excited states, thereby addressing the electron correlation effects. Importantly, it is enough to know the singles and doubles cluster amplitudes, and two-electron molecular orbital integrals, to determine the coupled cluster correlation energy.[1]

It is a usual to truncate the obtained cluster operator $\boldsymbol{T}$ at the certain point in the expansion (see Eq. 2.41), which in turn defines a particular method. For example, the *coupled-cluster doubles* (CCD) method, which represents the lowest level approximation, is obtained when only $\boldsymbol{T}_2$ is included, i.e., $\boldsymbol{T} = \boldsymbol{T}_2$, whereby only the double excitations are included.[1] *Coupled-cluster singles and doubles* (CCSD) marks the improvement which includes the first two operators, i.e., $\boldsymbol{T} = \boldsymbol{T}_1 + \boldsymbol{T}_2$, thereby incorporating both the single and double excitations. The next method in line is termed *coupled-clusters singles, doubles and triples* (CCSDT), including additional triple excitations, and already providing very accurate results for correlation energies; however, it can usually be employed only in very small systems due to excessive computational costs. For this reason more approximate methods have been developed, whereby excitations of higher order are introduced through the perturbative framework, lowering the computational demand without serious performance losses.[1] In this respect, we highlight two such approaches for calculations of excited state properties, namely the CC2[46,47] and the CC3[48], of which we will characterize only the CC2 methodology, as we chose to use it in this thesis. The CC2 method is obtained from CCSD by including only the doubles contributions stemming from the lowest order in perturbation theory – the perturbation being defined as in MP theory.[1] In effect, the amplitude equations for the ground-state doubles amplitudes reduce to an MP2-like expression, except that the two-electron integrals are now being transformed by the $\boldsymbol{T}_1$ operator. With regard to the excited state calculations, the second order corrections are employed when treating singles dominated excitations, while only the zeroth order correction is utilized for the actual double excitations.

As we discussed in [18], coupled cluster method, in particular CC2, has been shown to be quite affordable, as it enables the possibility to study relatively large systems on a high level of accuracy. Moreover, of all the wave functions/*ab initio* based methods for computing excitation energies, it seems to have the best cost/performance ratio. It should be noted that, despite its sophistication, the CC2 method is not infallible in terms of its predictions of excitation energies,[49,50]

and in some cases, exhibits agreement with experiment on a level similar to some TD-DFT functionals.[33,51] Nevertheless, the CC2 methodology can be expected to result in predictions free from DFT-related artifacts and to thus provide benchmark quality results in this respect.

2.2. Classical Mechanics

Although QM methods represent the most accurate way to treat any system of interest, it being either biological or chemical, the investigated systems usually consist of thousands of atoms, making the application of QM methods unfeasible. One is then driven to employ classical mechanics, in which the potential energy of the system is given as a function of nuclear coordinates only. Electron motion is then implicitly include, and the propagation of system/particles is performed classically, i.e., via Newton's laws. To gain an insight into the behavior of the system at hand and possibly estimate important physical quantities describing the system one needs to apply statistical mechanics and thermodynamics, the basics of which will be presented in the following sections.

2.2.1. Statistical Mechanics

Macroscopic systems generally contain a vast number of particles (N) whose evolution in time is governed by the Schrödinger equation and Newton's laws. Molecular dynamics simulations represent a powerful theoretical tool which enables us to obtain a detailed insight into the system of interest at the atomic scale, thereby providing us with the position and momentum of each particle present in the simulation. The relation between the simulations and observable properties of large systems is established through the theoretical groundwork known as statistical mechanics, providing a rigorous mathematical apparatus arising from the interplay of classical mechanics and probability theory.

The thermodynamic state of the system is usually identified with a rather small number of suitable parameters (usually known as state variables), e.g. temperature T, pressure p and the number of particles N. The values of remaining thermodynamic/macroscopic properties can be elucidated by employing the equation of state or other basic thermodynamics equations. On the other hand, the microscopic state of the system represents a particular microscopic configuration of the thermodynamic system. It is completely determined by specifying the positions of the atoms $\boldsymbol{q}$ and their respective momenta $\boldsymbol{p}$. Each microstate can be represented as a unique point in the multidimensional space containing all possible microstates of the given system (commonly termed phase space):

$$(p^N, q^N) \equiv (\boldsymbol{p}_1, \dots, \boldsymbol{p}_N; \boldsymbol{q}_1, \dots, \boldsymbol{q}_N). \tag{2.42}$$

Each point in this $6N$-dimensional phase space describes the entire state of the system. This brings us to the notion of the ensemble, which is defined as all the possible microstates that our system can

occupy under certain constraints that characterize the system macroscopically, such as constant temperature and constant volume. Importantly, fundamental postulate of statistical mechanics tells us that each microstate is *a priori* equiprobable if the system resides in the state of thermodynamic equilibrium.

In statistical mechanics, the observed values of certain observables are defined as the averages of the entire ensemble. The average of an ensemble is given by

$$\langle Q \rangle_{ensemble} = \iint Q\,(\boldsymbol{p},\boldsymbol{r})\rho(\boldsymbol{p},\boldsymbol{r})\mathrm{d}\boldsymbol{p}\mathrm{d}\boldsymbol{r}, \tag{2.43}$$

where $Q(\boldsymbol{p},\boldsymbol{r})$ denotes the observable that possesses a continuous spectrum of values, which is now of course the function of both the momentum $\boldsymbol{p}$ and the position $\boldsymbol{r}$. $\rho(\boldsymbol{p},\boldsymbol{r})$ denotes the probability density of the ensemble, and possesses the following form:

$$\rho(\boldsymbol{p},\boldsymbol{r}) = \frac{1}{Z}\exp[-H(\boldsymbol{p},\boldsymbol{r})/\mathrm{k_B}T], \tag{2.44}$$

where H represents the Hamiltonian of the investigated system, $\mathrm{k_B}$ is the Boltzmann constant, while Z denotes the partition function:

$$Z = \iint \exp[-H(\boldsymbol{p},\boldsymbol{r})/\mathrm{k_B}T]\,\mathrm{d}\boldsymbol{p}\mathrm{d}\boldsymbol{r}. \tag{2.45}$$

We can thus observe that, once we possess the partition function Z, we are able to calculate all of the thermodynamic properties of interest. Moreover, the partition function is related to the most important state functions, such as the internal energy, Helmholz free energy, entropy, etc.[52] However, this particular integral is extremely hard to calculate, more precisely, it is usually analytically unsolvable, as all possible states of the system have to be taken into account. This is circumvented by using numerical methods, such as molecular dynamics or Monte Carlo simulations. For example, in MD simulations each point in the phase space is obtained in discrete time steps. Thus, to properly calculate necessary ensemble averages, the MD simulation has to visit all of the microstates allowed by the imposed macroscopic constraints. More precisely, the simulated trajectory allows us to obtain the desired observable Q as a time average, which is given by

$$\langle Q \rangle_{MD} = \lim_{\tau\to\infty}\frac{1}{\tau}\int_0^{\tau} Q(\boldsymbol{p}(t),\boldsymbol{r}(t))\,\mathrm{d}t \approx \frac{1}{M}\sum_{i=1}^{M} Q_i(\boldsymbol{p},\boldsymbol{r}). \tag{2.46}$$

In the Eq. 2.46, the time of the simulation is represented by τ, M is the number of steps involved, while $Q_i(\boldsymbol{p},\boldsymbol{r})$ denotes the value of the observable at the i-th step of the simulation. Thus, although the

time averages of the ensemble can be obtained using numerical techniques, the experimentally observable quantities are represented by ensemble averages. This conundrum is resolved by one of the most profound axioms of statistical mechanics, namely the ergodic hypothesis, which postulates that

$$\langle Q \rangle_{ensemble} = \langle Q \rangle_{MD}. \tag{2.47}$$

This fundamental hypothesis states that, if the system is propagated endlessly in time, it will eventually visit all possible states. Thus, if a simulation is ergodic, and is propagated for a sufficient amount of time to visit all relevant microstates of the system, one can derive meaningful thermodynamic, structural and dynamic information from it. Although it is believed that majority of systems obey ergodicity, proper sampling using MD simulations can be rather problematic, as often only portions of phase space are properly sampled, while other parts remain unvisited, which is usually a product of high energy barriers between the states in the system. As it will be presented in the following sections (section 2.2.5), a number of advanced MD techniques have been developed to circumvent these problems.

2.2.2. Ensembles

As previously mentioned, by fixing a small amount of macroscopic variables, such as temperature, volume, pressure, etc., one can define a macroscopic state. Depending on the choice of fixed variables, different ensembles are produced. Below we comment the most commonly encountered ensembles:

Microcanonical ensemble (*NVE*) – all states where the number of atoms N, the volume of the system V and the total energy E of the system are constant. This ensemble corresponds to the isolated system, with no exchange of the energy with the surroundings, i.e., it gives rise to the adiabatic system.

Canonical ensemble (*NVT*) – all states where the number of atoms N, the volume of the system V and the temeperature T are kept fixed. In this case the system is able to exchange the energy with the environment.

Isobaric-isothermal ensemble (*NPT*) – all states where the number of atoms N, the volume of the system V and the temperature T are preserved. This perhaps represents the most natural ensemble to use from the chemistry point of view, as these represent the usual conditions under which experiments are conducted.

Grand canonical ensemble (μVT) – all states where chemical potential μ, the volume of the system V and the temeperature T are mainteined fixed. To be able to keep the chemical potential μ fixed the system is allowed to exchange the matter with the surroundings.

2.2.3. Molecular Dynamics (MD)

With the proliferation of computational power in the recent decades, computational simulations became a necessary tool in solving a wide spectrum of problems arising in physics, chemistry, and science of materials. While the theoretical framework used to study complex systems, namely statistical mechanics, was forged relatively long ago, certain problems are yet intractable using existing experimental and purely theoretical/analytical approaches, in spite of the fact that recent improvements in experimental techniques can provide detailed information of the system of interest at the microscopic level. In this respect, computer simulations allow one to obtain a detailed insight into the atomic and/or molecular level behavior of the system, in turn acting as a bridge between theory and experiment. The main caveat lies in the fact that the computer simulations require a number of input parameters to appropriately describe the system, i.e., the input parameters and the emerging Hamiltonian of the system have to mirror the reality in order to produce useful results. Thus, the basic step in any computer simulation is a proper definition of the system Hamiltonian, which is used to obtain the statistical ensemble of the system by employing different ensemble generating algorithms, that in turn enables one to successfully determine the properties of interest. Traditional simulation methods are generally divided into two classes, namely stochastic and deterministic. The most famous representative of the former class is the Monte Carlo method, while molecular dynamics based approaches represent a cornerstone of the latter.

Molecular dynamics simulations use classical mechanics, namely Newton's equations of motion, which are integrated to evolve the system in time, resulting in the change of the particle positions and in the assignment of velocities corresponding to new positions at every step of the simulation. In this way the entire phase space can be obtained, from which detailed dynamical information about the system can be extracted. One can essentially consider molecular dynamics as the combination of laboratory experiments and theory, with the method itself often being viewed as a "virtual" experiment. Molecular dynamics simulations enable exploration of the complex, dynamical processes which occur in various biologically relevant systems. More precisely, they are readily used to investigate protein stability, conformational changes in molecules, adsorption processes, protein and peptide folding, ion transport through the biological membranes etc. However, to properly describe these systems, one has to construct models representing the reality, i.e., models that agree with experimental measurements. In this respect, the quality of MD simulations can be usually probed by testing them against the experimental observables, including adsorption energies, UV/CD spectra, NMR results, etc. Moreover, it is necessary for MD simulations to satisfy fundamental physical laws, for example energy conservation. Thus, to properly set up and perform molecular dynamics simulation, a number of steps have to be undertaken.

Firstly and most importantly, one has to choose a force field/parameterization which describes the behavior of the investigated system well. Basically, the force field gives rise to the Hamiltonian of the system of interest, whereby all of the interactions occurring between the particles in the system, i.e., bonded and non-bonded ones, are described by analytical expressions which classically approximate quantum-mechanical molecular systems. It is common practice to assume only pairwise interaction between the particles, which significantly lowers the computational cost. The force field is discussed in depth in section 2.2.4.

Secondly, as it is analytically impossible to resolve the equations of motion for all but the simplest systems, they have to be resolved numerically. Due to this, the MD technique is based on the so-called integration algorithms, which are used to propagate both the particles' positions and velocities from time t to $t + \Delta t$. The chosen algorithm has to satisfy a number of criteria – namely, it has to maintain the total energy and total momentum in the system constant, it has to be efficient, and it has to allow for a large enough of a time step. Importantly, the chosen time step has to account for the fastest motion occurring in the simulation, i.e., the time step has to be smaller than the fastest motion. This motion usually belongs to the vibration of the bonds containing hydrogen atoms.[1] However, in the modern MD simulations these bonds are usually constrained using the so-called *SHAKE* algorithm, which readily enables time steps in the 2-10 fs range. The most popular integration algorithms are the Verlet algorithm, Leap-frog, Velocity Verlet and Beeman's algorithm. All of these algorithms start from a premise that the positions, the velocities and the accelerations can be approximated with the Taylor series:

$$\begin{aligned} \vec{r}(t+\delta t) &= \vec{r}(t) + \vec{v}(t)\delta t + \frac{1}{2}\vec{a}(t)\delta t^2 + \dots, \\ \vec{v}(t+\delta t) &= \vec{v}(t) + \vec{a}(t)\delta t + \frac{1}{2}\vec{b}(t)\delta t^2 + \dots, \\ \vec{a}(t+\delta t) &= \vec{a}(t) + \vec{b}(t)\delta t + \dots. \end{aligned} \tag{2.48}$$

As an example of the integration algorithms we present the Velocity-Verlet algorithm:

$$\begin{aligned} \vec{r}(t+\delta t) &= \vec{r}(t) + \vec{v}(t)\delta t + \frac{1}{2}\vec{a}(t)\delta t^2, \\ \vec{v}(t+\delta t) &= \vec{v}(t) + \frac{1}{2}[\vec{a}(t) + a(t+\delta t)]\delta t. \end{aligned} \tag{2.49}$$

This algorithm enables one to obtain the positions, velocities and accelerations at a time t, and it represents one of the most commonly used integrators.

Finally, we have to choose the statistical ensemble and methods via which main thermodynamic quantities are controlled. In the case when no external potentials are present the Hamiltonian of the system remains conserved, making microcanonical (NVE) the preferred ensemble. However, other ensembles, such as NPT and NVT, can also be used. For example, for the purposes of

chemically relevant systems, which are usually held under a constant atmospheric pressure, with the enthalpy being the state function of choice, the NPT ensemble becomes a more natural ensemble to use than NVE. Thus, in order to calculate crystal lattice, bulk properties of minerals, infrared (IR) spectra or thermal expansion coefficients one should choose the NPT ensemble. On the other hand, if one wants to calculate the surface energy or is interested in the process of binding, i.e., wants to calculate adsorption energy, then the NVT ensemble is the preferred one. Moreover, due to the high stability of the NVT ensemble algorithms (NPT can be somewhat erratic in its behavior, especially when dealing with large systems), it is very common to prepare a system using the NPT ensemble, thereby obtaining the desired density of the system, and then propagate the production run (usually a 10-100 ns long run from which the ensemble averages are gathered and which is used for the analysis) using the NVT ensemble. The control of the temperature during simulations is achieved through the use of thermostats (usually Andersen, Berendsen, Nosé-Hoover and Langevin) while the pressure is most commonly controlled by using the Berendsen weak-coupling scheme or the Parrinello-Rahman barostat. In this work we use stochastic Langevin dynamics for this purpose, which approximates the canonical ensemble, and which is discussed in the following section.

2.2.3.1. Langevin Dynamics (LD)

The Langevin equation is a stochastic differential equation which amends Newton's equation by adding two additional terms used to incorporate neglected degrees of freedom. The first and second terms in Langevin's equation represent a frictional force and a random force $\vec{R}$, respectively. By employing this model, the effect of the solvent, which is otherwise not explicitly treated, is taken into account (although not to the full extent) in terms of frictional drags on the solute, and also via random collisions describing the jostling nature of the solvent molecules due to their thermal motion. Langevin's equation for the i-th particle is given by

$$\vec{F}_i - \gamma_i \vec{v}_i + \vec{R}_i(t) = m_i \vec{a}_i. \tag{2.50}$$

where γ_i represents the friction coefficient acting on the i-th particle, while $\vec{F}_i$ represents the overall force on the i-th atom. The connection between the friction coefficient γ_i and the random force $\vec{R}_i$ comes from the fluctuation-dissipation theorem, whereby it is presumed that the time average of the random force is zero. Additionally, if one assumes that the random force remains uncorrelated at different points in time, the following relation is obtained:

$$\langle \vec{R}_i(t) \cdot \vec{R}_i(t') \rangle = 2k_B T \gamma_i \delta(t - t'). \tag{2.51}$$

The temperature of the system is maintained using exactly this expression. Overall, the quasi-solvent introduced to the system through LD allows improved conformational sampling compared to its Newtonian counterpart. However, one should not consider LD as essentially representing molecular

dynamics with the implicit solvent, though it can be used for that purpose; while LD simulates certain aspects of the implicit solvent models, both the electrostatic (coming from the solvent; intrinsic ones are always taken into account) and the hydrophobic effects are not taken into account. Langevin dynamics is also used in the molecular dynamics with an explicit solvent where it plays the role of the thermostat, as mentioned above, and is used for this purpose throughout this thesis.

2.2.4. Force Fields

The main approximation on which the entire molecular modelling is based is the Born-Oppenheimer approximation, with which it is possible to view the energy of a molecule in a specific electronic state just as a function of the coordinates. The methods of molecular mechanics neglect the motion of electrons and calculate the energy of the system as a function of nuclear coordinates only.[7,53] At the heart of every MD simulation lies the force field, which describes all of the interactions present in the system of interest, thereby defining the Hamiltonian in these "classically" described systems. Naturally, the Hamiltonian in these cases is constructed by employing expressions that describe bonding and non-bonding interactions between atoms and molecules in a simplified manner. Here one can, using the classical terms parameterized to reproduce QM calculations and/or fitted to the experimental data, obtain a classical description of the potential energy of the system, which is known as a force field/parameterization. The modern force fields contain at least four terms that are used to describe both the intra- and intermolecular interactions. One of the most commonly used force fields is the AMBER force field,[54] in which the function of the potential energy is defined as:

$$\begin{aligned} V(r) \quad &= \sum_{\text{bonds}} k_b\,(b-b_0)^2 + \sum_{\text{angles}} k_\theta(\theta-\theta_0)^2 \\ &+ \sum_{\text{torsions}} k_\phi[\cos(n\phi+\delta)+1] + \sum_{\substack{\text{nonbonded}\\ \text{pairs}}} \left[\frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}}\right]. \end{aligned} \tag{2.52}$$

The first three functional terms describe the interactions between bonded atoms, namely the bonds (1-2 interactions), the angles (1-3 interactions) and torsional angles (1-4 interactions), respectively. The first two of these are most commonly approximated with the harmonic potential (in certain force fields a Morse potential is used instead of a harmonic oscillator for the description of 1-2 interactions). The last term describes all of the non-bonded interactions (non-bonded interactions between the first two neighbors are neglected, while the interactions between the third neighbors are usually scaled down), where the electrostatic interactions are described with the Coulomb law, while the van der Waals forces are approximated using the Lennard-Jones potential. The latter one can also be described by using the Buckingham potential, as it is well established that the repulsion arises exclusively due to the overlapping electronic wave-functions, and that the electronic density decreases almost exponentially

with the distance from the nucleus.[1] Together this implies that one could use the exponential function to describe the repulsive part of the van der Waals potential, thus obtaining:

$$E_{\mathrm{vdW}}(r) = Ae^{-Br} - \frac{C}{r^6}. \tag{2.53}$$

Today a number of different force fields and software packages for modelling various systems are readily available, among which the most popular are AMBER, GROMACS and OPLS. These force fields are developed specifically to describe biologically important systems, such as peptides, proteins, DNA, lipids etc. The AMBER package[55] contains the so-called general AMBER force field (*gaff*[56]) which can be used to parameterize a general small organic molecule – the obtained parameters can be mixed with other existing and more specialized AMBER force fields. In this thesis, we used AMBER force fields (*ff99SB*[57], *ff03*[58], *ff12SB*[59], *gaff*[56]) to parameterize the investigated peptide systems and small organic molecules. In the simulations where peptides interact with the calcite surface (Chapter 4) we utilized the highly specialized force field by Pavese *et al.*[60] to describe the calcite crystal. Coupling of the AMBER and Pavese force field, as well as the details of the two are described in the section 4.2.

As mentioned earlier, force field parameters are usually obtained by either fitting to the experimental measurements or data calculated using high level of theory. While all of the aforementioned force field terms play a role in the simulations of small organic molecules and peptides in particular, the most important term is the electrostatic one, as even small errors in these parameters can induce rather large changes in both the conformational behavior of a molecule and its absorptive potential, with both of these properties being crucial in our work. Thus, as the partial charges of atoms play a particularly important role in the classical descriptions of our systems, we inspect the theoretical basis for determining partial atomic charges in depth in the following subsection.

2.2.4.1. Restrained Electrostatic Potential (RESP)

The concept of partial atomic charges underlies the entire chemistry, where even a qualitative understanding of their nature can be a powerful tool when investigating chemical reactivity and physical properties of a chemical system. To calculate the partial atomic charges for the purpose of molecular dynamics simulations one can use the *Electrostatic Potential* (ESP) method, which is a method based on the assignment of charges to the atomic centers using the least square fitting method. This procedure ensures that the obtained charges reproduce electrostatic potential generated using quantum mechanical methods (often HF/6-31G* level of QM theory is used for the calculation of the potential). The major disadvantage of this method is the ill prediction of the so-called „buried“ charges, namely the charges that lie far away from the solvent molecules and which are found at a

rather large distance from the points on the QM-obtained ESP grid.[61] The described problems arise mainly from the statistical nature of the mentioned process, as during the fitting stage, exactly the „buried“ charges can fluctuate the most resulting in a minimal improvement of the fit.

To resolve these issues the method called *Restrained Electrostatic Potential* (RESP)[61] was introduced, using which the problems arising during the above described fitting procedure can be greatly diminished, thereby only slightly decreasing the quality of the produced charges. This is accomplished by introducing a penalty function which restraints the partial atomic charges, in turn setting them on a certain target charge. More precisely, an RESP fit is usually performed in two stages, where in the first stage partial charges of all atoms are left free to vary. On the other hand, in the second stage of the fit partial charges of all atoms are restrained, except for degenerate hydrogen atoms which are refit having their charges set to the same value. The effect of this procedure is twofold. Firstly, the RESP determined charges of „buried“ atoms are of a smaller magnitude compared to the ones produced by the usual ESP procedure, thus being closer to the expected values. Secondly, and perhaps even more importantly, this procedure reduces the polarity of the otherwise overly polarized C-H bonds, resulting in a more accurate parameterization of organic molecules. Therefore, when using the target charges (set to be zero to begin with) together with the hyperbolic penalty function, one can successfully obtain partial atomic charges, which also satisfy the intramolecular electrostatics. In practice, it was shown that the charges which are well behaved in the ESP method are virtually unchanged after RESP treatment, while the charges of all of the atoms belonging to the nonpolar groups (especially the „buried“ nonpolar groups) change significantly and obtain more intuitive and reasonable values.[62]

2.2.5. Free Energy/MD Methods for Improved Sampling

Free energy represents one of the most significant concepts in chemistry, governing the chemo-physical behavior of biologically relevant systems, such as the spontaneity of a certain process. In this respect, some of the most important chemical and physical processes are viewed in terms of a free energy landscape, whereby the free energy is given as a function of a reaction coordinate. The processes in which this concept plays the essential role include folding of molecules, in particular peptides and proteins, their aggregation, their adsorption to inorganic/organic interfaces, etc. A free energy landscape, similarly to a geographic one, contains valleys and hills, where valleys correspond to stable (and potentially meta-stable) states, while the hills represent the barriers between the states. The height of the hill determines the transition rates between the states, and the higher this barrier is, the slower this transition becomes. Similarly, the wider and deeper the valley in the free energy landscape, the more populated it is. Biomolecular systems usually “live” in a complex free energy landscape, with numerous states and barriers separating them. Naturally, a complete knowledge of a

free energy landscape would provide full information about the behavior of the system at hand, enabling one, for example, to find free energy profiles along any reaction coordinate of interest.

One can attempt to construct the free energy as a function of a reaction coordinate ξ using a brute force approach, namely via the unbiased MD simulations, where the free energy is expressed as

$$F(\xi) = -k_{\mathrm{B}}T\, ln(P(\xi)). \tag{2.54}$$

Here $P(\xi)$ denotes the probability of finding the system at a certain value of the reaction coordinate ξ. However, adequate sampling represents a daunting task for "normal" MD simulations when dealing with complex systems, as they possess rugged free energy landscapes, leading to the quasi non-ergodic behavior. More precisely, the molecule/system of interest can easily become stuck/kinetically trapped in the local minimum of the free energy landscape, and if the temperature/energy of the system is insufficiently large for the molecule to cross the high free energy barrier between this and some other minimum in the free energy landscape, the MD simulation will effectively sample only a small region of the entire phase space, in turn giving a strongly biased insight into the behavior of the system. Thus, a number of methods have been developed with the aim of circumventing this problem, in turn enabling proper sampling of the region of interest and the extraction of the desired free energy information, such as thermodynamic integration, free energy perturbation, enhanced sampling, etc.

An important subset of these "enhanced sampling" methods is based on the introduction of a biasing potential to the Hamiltonian of the system, thereby improving the sampling of the problematic portions of the free energy landscape. Amongst the most popular such techniques are umbrella sampling, metadynamics and steered molecular dynamics. The difference in the aforementioned techniques lies in the way the biasing potential is implemented, namely, the first approach (umbrella sampling) employs a fixed biasing potential; in the metadynamics, the biasing potential is added in an iterative fashion, while in the steered molecular dynamics a time-dependent biasing potential is imposed – forcing otherwise unfavorable events to happen, i.e., enabling crossing of the high-energy barriers in the free energy landscape. However, in all aforementioned approaches one has to be able to adequately choose the reaction coordinate, often representing a problematic issue, which, if not properly treated, can lead to unwanted and potentially disastrous results. This can be avoided by using an unbiased sampling technique, namely replica exchange molecular dynamics (REMD).[63] In the first part of our work we decided to apply REMD simulations (Chapter 3), while in the second part we utilized the umbrella sampling technique (Chapter 4), as it was possible to unambiguously determine the reaction coordinate of interest in the latter case. These chosen methods are discussed in more details in the following sections.

2.2.5.1. Replica Exchange Molecular Dynamics (REMD)

In this section we introduce the concepts behind the replica exchange molecular dynamics (REMD)[63] following the presentation of the method in [64]. Historically, replica exchange (RE) method was invented to improve the dynamical properties of Monte Carlo simulations, however, it was soon afterwards also adopted for the MD simulations, giving rise to replica exchange molecular dynamics method. Classical molecular dynamics consists of a single stochastic process, i.e., using a specific setup of the system of interest, one attempts to investigate the entirety of the phase space. On the other hand, in the REMD method, replicas/copies of the investigated system, identical in all respects except for the temperature, are given a chance to exchange the temperatures at which they are simulated (after a certain time interval) thereby circumventing the previously mentioned kinetic traps. However, the REMD method evades the problem of kinetic traps using exactly the temperature exchange between the copies of the system; with each exchange of the copies the system is „stirred", thus, if it ever becomes kinetically trapped during the duration of the simulation, it has a significantly higher chance of escaping it compared to the "normal" MD simulations.

The REMD method involves simulating a number of copies of the investigated system, with each copy being independently simulated. As previously mentioned, the copies/replicas of the system differ only in temperature at which they are propagated. The simulations are performed by choosing an initial set of target temperatures $(T_1, \dots, T_m, T_n, \dots, T_N,)$ for N copies/replicas, with coordinates $\boldsymbol{q}_1, \dots, \boldsymbol{q}_m, \boldsymbol{q}_n, \dots, \boldsymbol{q}_N$, where $\boldsymbol{q}_i$ represents the coordinates of the i-th replica. To begin with, the i-th replica starts at the i-th temperature. Replicas are then let to propagate and achieve the target temperatures. After a predefined time intervals each replica attempts to exchange its temperature with the neighboring ones using the Monte Carlo criterion:

$$P_{\text{exchange}} = \min\left[1, \exp\left[\left(\frac{1}{kT_i} - \frac{1}{kT_j}\right)(E(\mathbf{q}_m) - E(\boldsymbol{q}_n))\right]\right]. \tag{2.55}$$

The exchange between replicas occurs only when Eq. 2.55 is satisfied, so that $T(\mathbf{q}_i) = T_n$ and $T(\mathbf{q}_j) = T_m$. After the exchange, the propagation of replicas continues for a certain predefined amount of time, when a new exchange is attempted. From Eq. 2.55, it is possible to observe that only the neighboring replicas possess energies close enough to have sufficiently large exchange probabilities, thus the exchanges are attempted only in those cases.

When performing the REMD, a number of factors have to be taken into account to provide a successful outcome of the simulation. Firstly, one has to ensure that replicas can visit the entire temperature space available. Secondly, the temperature of the lowest replica has to be sufficiently small to allow a proper sampling of low-energy regions of the free energy landscape. Thirdly, the highest temperature has to allow for energetically high barriers to be crossed over. Finally, the

exchange rates between the neighboring replicas should be high enough (15-25% is found to be an optimal range of acceptance probabilities), with an additional requirement, namely, the spacing between replicas (in the temperature space) has to be such that it ensures approximately equal exchange rates between any two replicas – to guarantee proper sampling.

If successfully performed, the REMD provides us with a detailed view of the free energy landscape of a system of interest, enabling us not only to extract relevant thermodynamic data at the temperature of interest, but to follow temperature dependency of the free energy surface throughout the utilized temperature range. The major limitation of this approach lies in the fact that the exchange rate between the neighboring copies of the system is ought to fall in the 15-25% range, as mentioned previously, implying that, for systems containing tens to hundreds of thousands of atoms, the number of replicas needed to ensure the aforementioned ratio would need to increase rather rapidly, making the method inapplicable when dealing with large biomolecular systems. Despite these disadvantages, the REMD has been successfully applied in protein folding,[65,66] showcasing its particular usefulness in the study of smaller peptides.[64] A number of various REMD algorithms have been developed in recent years, whereby, instead of temperature, the replicas differ in other properties, for example in their Hamiltonian (energy). Moreover, REMD simulations allow us to calculate the free energy of a relevant/chosen reaction coordinate by employing Eq. 2.54.

2.2.5.2 Umbrella Sampling

The free energy difference when going from state *A* to state *B* is given by

$$\Delta F = F_B - F_A = k_B T \ln \frac{P_A}{P_B}. \tag{2.56}$$

Here P_A and P_B denote the probabilities to find the system in a state A or B, respectively. This difference can in principle be obtained from the MD simulation by evaluating the probability distribution along the reaction coordinate which connects the states of interest; in an ideal scenario, this can be performed by directly counting the number of appearances of the system in the corresponding states found in the MD trajectory. This technique, however, is only rarely applicable; more precisely, it can only be employed when both the free energy difference itself and the energy barrier between the two states of interest is small. If one deals with a complex molecular species living in a rugged free energy landscape, the states A and B will very often be separated by a high energy barrier, i.e., if the "normal" MD simulation starts in any of the two states, for example A, it will most likely sample only the portion of the phase space belonging to the state A, while the state B will remain poorly sampled, or even worse, not sampled at all. Umbrella sampling[67] is a method that modifies the Hamiltonian of the system by adding a suitably chosen biasing potential (umbrella potential), thereby flattening the free energy landscape between the regions in which the ergodicity is

hindered, in turn drastically enhancing sampling of the problematic parts of the configuration space. The added biasing potential, under the assumption that it depends only on the chosen reaction coordinate, can be subtracted, yielding the unbiased free energy profile of the reaction coordinate.

The introduced umbrella potential would, in an ideal scenario, be the exact inverse/negative of the free energy, which would allow for equal sampling of the states of interest. However, this represents exactly the physical quantity one is attempting to calculate, thus this "approach" cannot be used in practice. Instead, the most common strategy employed is to divide the reaction coordinate in a number of small sampling windows, or intervals, with each interval having a different umbrella potential (appearing most commonly in the shape of a harmonic potential) that improves sampling around a predefined value of the reaction coordinate $\xi(r)$. The exact strength of the biasing potential depends on the position of its respective interval in the reaction coordinate, and is usually found via trial and error. The result of an umbrella sampling is a series of histograms containing the biased distribution of the reaction coordinate from each sampled interval. The weighted histogram analysis method (WHAM)[68] can then be applied to produce unbiased histograms and to combine them, thus generating the free energy profile along the chosen reaction path.

2.3. References

(1) Jensen, F. *Introduction To Computational Chemistry*, J. Wiley & Sons Inc., New York, 2007.

(2) Hartree, D. R. *Proc. Cam. Phil. Soc.* **1928**, *24*, 246.

(3) Fock, V. Z. *Phys.* **1930**, *61*, 126.

(4) Roothan, C. C. *Rev. Mod. Phys.* **1951**, *23*, 69.

(5) Hall, G. C. *Proc. Roy. Soc. (London)* **1951**, *A205*, 541.

(6) Atkins, P.; Friedman, R. *Molecular Quantum Mechanics*, Oxford University Press, Oxford, 2005.

(7) Cramer, C. J. *Essentials of Computational Chemistry – Theories and Models*, Second Edition, J.Wiley & Sons Inc., New York, 2004.

(8) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.

(9) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(10) Carsky, P. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Clark, N. L., Gasteiger, J., H. F. S., III, Schreiner, P. R., Eds.; Wiley: Chichester, U.K., p 485, 1998.

(11) Gauss, J. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Clark, N. L., Gasteiger, J., H. F. S., III, Schreiner, P. R., Eds.; Wiley: Chichester, U.K., p 615, 1998.

(12) Frey, P. A. *Annu. Rev. Biochem.* **2001**, *70*, 121.

(13) Parr, R. G.; Yang, W., *Density Functional Theory of Atoms and Molecules*; Oxford Univ. Press, New York, 1989.

(14) Kohn, W.; Hohenberg, P. *Phys. Rev. B* **1964**, *136*, 864.

(15) Kohn W.; Sham L. J. *Phys. Rev.* **1965**, *140*, 1133.

(16) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.

(17) Kohn, H.; Becke, A. D.; Parr, R. G. *J. Phys. Chem.* **1996**, *100*, 12974.

(18) Brkljača, Z.; Mališ, M.; Smith, D. M.; Smith, A.-S. *J. Chem. Theory Comput.*, **2014**, *10* , 3270.

(19) Runge E.; Gross, E. K. U. *Phys. Rev. Lett.* **1984**, *59*, 997.

(20) Mearns, D.; Kohn, W. *Phys. Rev. A* **1987**, *35*, 4796.

(21) van Leeuwen, R. *Phys. Rev. Lett.* **1999**, *82*, 3863.

(22) Dreuw, A.; Head-Gordon, M. *Chem. Rev.* **2005**, *105*, 4009.

(23) Engel, E., Dreizler, R. M. *Density Functional Theory*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

(24) Ullrich, C. A. *Time-dependent Density-functional Theory - Concepts and Applications*, Oxford University Press, 2012.

(25) Casida, M. E. in the *Recent Advances in Density Functional Methods*, editor D.P. Chong, World Scientific, Singapure, p 155, 1995.

(26) Casida, M. E.; Huix-Rotllant, M. *Annu. Rev. Phys. Chem.* **2012**, *63*, 287.

(27) Warnke, I.; Furche, F. *WIREs Comput. Mol. Sci.* **2012**, *2*, 150.

(28) Autschbach, J.; Jorge, F. E.; Ziegler, T. *Inorg. Chem.* **2003**, *42*, 2867.

(29) Rudolph, M; Autschbach, J. *J. Phys. Chem. A* **2011**, *115*, 14677.

(30) Kaminský, J.; Kubelka, J.; Bouř, P. *J. Phys. Chem. A* **2011**, *115*, 1734.

(31) Aquino, A. J. A.; Nachtigallova, D.; Hobza P.; Truhlar, D. G.; Hättig, C.; Lischka, H. *J. Comput. Chem.* **2011**, *32*, 1217.

(32) Wakai, A.; Fukasawa, H.; Yang C.; Mori, T.; Inoue Y. *J. Am. Chem. Soc.* **2012**, *134*, 4990.

(33) Jacquemin, D.; Wathelet, V.; Perpète, E. A.; Adamo, C. *J. Chem. Theory Comput.* **2009**, *5*, 2420.

(34) Kuritz, N.; Stein, T.; Baer, R.; Kronik, L. *J. Chem. Theory Comput* **2011**, *7*, 2408.

(35) Peach, M. J. G.; Benfield, P.; Helgaker, T.; Tozer, D. J. *J. Chem. Phys.* **2008**, *128*, 044118.

(36) Li, R.; Zheng, J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2010**, 1*2*, 12697.

(37) Tozer, D. J.; Amos, R. D.; Handy, N. C.; Roos, B. O.; Serrano-Andrés, L. *Mol. Phys.* **1999**, *97*, 859.

(38) Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51.

(39) Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615.

(40) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.

(41) Charaf-Eddin, A.; Planchat, A; Mennucci, B.; Adamo, C.; Jacquemin, D. *J. Chem. Theory Comput.* **2013**, *9*, 2749.

(42) Ernzerhof, M.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 5029.

(43) Jamorski, C.; Foresman, J. B.; Thilgen, C.; Lüthi, H.-P. *J. Chem. Phys.* **2002**, *116*, 8761.

(44) Guido, C. A.; Knecht, S.; Kongsted, J; Mennucci, B. *J. Chem. Theory Comput.* **2013**, *9*, 2209.

(45) Laurent, A. D.; Jacquemin, D. *Int. J. Quant. Chem.* **2013**, *113*, 2019.

(46) Christiansen, O.; Koch, H.; Jørgensen, P. *Chem. Phys. Lett.* **1995**, *243*, 409.

(47) Hald, K.; Hättig, C.; Yeager, D. L.; Jørgensen, P. *Chem. Phys. Lett.* **2000**, *328*, 291.

(48) Christiansen, O.; Koch, H.; Jørgensen, P. *J. Chem. Phys.* **1995**, *103*, 7429.

(49) Silva-Junior, M. R.; Schreiber, M.; Sauer, S. P. A.; Thiel, W. *J. Chem. Phys.* **2008**, *129*, 104103.

(50) Schreiber, M.; Silva-Junior, M. R.; Sauer, S. P. A.; Thiel, W. *J. Chem. Phys.* **2008**, *128*, 134110.

(51) Crawford, T. D.; Stephens, P. J. *J. Phys. Chem. A* **2008**, *112*, 1339.

(52) Chandler, D. *Introduction to Statistical Mechanics*, Oxford University Press, New York, 1987.

(53) Leach, A. R. *Molecular Modelling*, Prentice Hall, 2001.

(54) Ponder, J. W.; Case D. A. *Adv. Protein Chem.* **2003**, *66*, 27.

(55) Case, D. A.; Darden, T. A.; Cheatham, T.E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman ,P. A. *AMBER 10*, University of California, San Francisco 2008.

(56) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157.
(57) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712.
(58) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J. W.; Wang, J.; Kollman, P. A. *J. Comput. Chem.* **2003**, *23*, 1999.
(59) Case, D. A.; Darden, T. A.; Cheatham, T.E. III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsich, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman , P. A. The FF12SB force field, AmberTools 13 Reference Manual (2013) 27–29.
(60) Pavese, A.; Catti, M.; Parker, S. C.; Wall, A. *Phys. Chem. Miner.* **1996**, *23*, 89.
(61) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.
(62) Wang, Y.; Cieplak, P.; Kollman, P. A. *J. Comp. Chem.* **2000**, *21*, 1049.
(63) Sugita Y.; Okamoto. Y. *Chem. Phys. Lett.* **1999**, *314*, 141.
(64) Sabonmatsu, K. Y.; Garcia, A. E. *Proteins***, 2002**, *43*, 225.
(65) Andrec, M.; Felts, A. K.; Gallicchio, E.; Levy, R. M. *Proc. Natl. Acad. Sci. U.* **2005**, *19*, 6801.
(66) Zhou, R. *Methods Mol. Biol.* **2007**, *350*, 205.
(67) Kästner, J. *Comput. Mol. Sci.* **2011**, *1*, 932.
(68) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comp. Chem.* **1992**, *13* , 1011.

3. Calculation of the CD Spectra of Flexible Molecules from Their Conformational Phase Space

3.1. Introduction

Dichroism (greek *dikhroos*, two colors) is an optical phenomenon which denotes the dependence of the absorption of the electromagnetic radiation of specific frequency on its polarization during its passing through or its reflectance from matter.[1] Depending whether we are interested in the frequency dependence of the absorption of the linearly or circularly polarized light we differ linear (LID) and circular dichroism (CD). Although both phenomena arise due to specific structural symmetries of investigated systems/molecules, linear dichroism can only be manifested in systems where certain symmetry is expressed on the macroscopic scale. For this reason only the substances showing long-range ordering (most commonly crystals) can exhibit linear dichroism. On the other hand, circular dichroism is the direct consequence of intrinsic symmetry of molecules, thus it can be manifested in all three aggregate states. More precisely, to exhibit circular dichroism the molecule has to be chiral (greek *kheir*, hand, shown in Figure 3-1), which means that it does not possess symmetry elements such as mirror planes (σ), inversion centers (i) or rotation-reflection axes (Sn). The two possible enantiomers/mirror images are usually referred to as the *R*- and *S*-enantiomers (*R* standing for *rexus*, *S* for *sinister*, Figure 3-1). Moreover, the active molecules in the investigated system should not just be of the enantiomorphic kind, one of the enantiomers has to be present in excess, otherwise the effects mutually annul. For example, racemic mixtures, in which two enantiomers are found in the same ratio, do not exhibit circular dichroism. Moreover, if two chemically equivalent substances differ in the chirality of one or more chiral centers, but not in all of them, i.e. they are not enantiomers, they constitute epimers.

Due to the fact that an overwhelming majority of biologically important molecules are chiral, chirality, and its manifestation in chiroptical effects such as electronic and vibrational circular dichroism, is of paramount importance in chemistry, biochemistry, and neighboring disciplines.[3] In

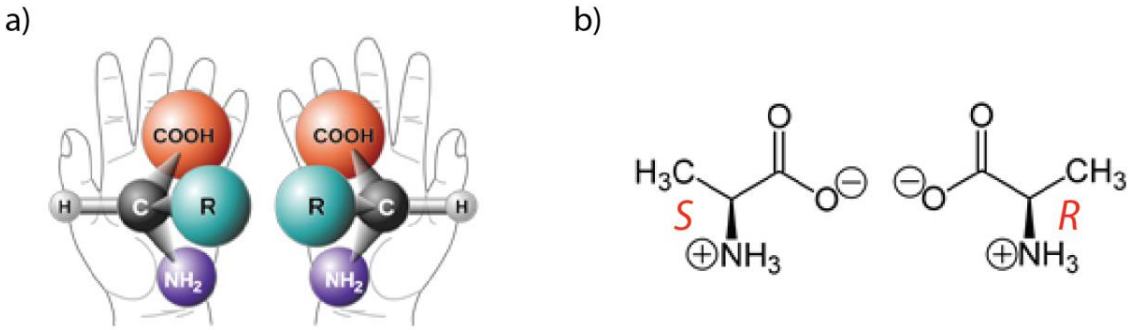


Figure 3-1. Chirality of a molecule shown on the example of a) a generic amino acid and b) zwitterionic form of Alanine, the smallest chiral amino acid.[2]

this respect, the electronic CD spectroscopy, together with X-ray and NMR techniques, represents the key experimental method used in the structural characterization of optically active chiral molecules. This method is widely used in studies of biologically important systems, such as proteins, DNA and more generally, chiral organic compounds. Focusing on the case of proteins, secondary structural motifs, including the most predominant ones such as α-helices, β-sheets, β- and γ-turns and random coil structures, produce spectral features of particular shapes in the far-ultraviolet part of the CD spectrum, differing both in the position and in the magnitude of the exhibited spectral elements.[4] As an example, in an α-helix, excitations of the $\pi \rightarrow \pi*$ type of the peptide bond result in a strong positive band at about 190 nm and in a negative feature appearing at around 208 nm. Moreover, electronic transition from a lone pair orbital on oxygen atom, n_O, to the $\pi*$ orbital induces a negative band that is usually situated at about 222 nm. In similar fashion, other motifs give rise to different and often specific spectral features, in turn enabling us to distinguish between them using CD spectroscopy. This method is particularly useful for monitoring structural changes resulting from changes in environmental conditions such as temperature, ionic strength, and pH, whereby one can monitor protein/DNA folding events.[5] However, even though CD spectroscopy is very sensitive to the conformation of the studied molecule its interpretation is usually based on the comparison of the CD spectrum of a molecule of interest to a reference set consisting of CD spectra of compounds with known X-ray or NMR structures.[6] The quality of such an empirical approach strongly depends on the structural similarity between the compound of interest and the reference set. Approximate theoretical approaches have been successfully used to complement the experimental measurements and to correlate the molecular properties and spectral features, examples of which are methods for proteins with well-defined secondary structures.[7-9]

As we elaborated in [10] and [11], CD spectroscopy has also been extensively applied to the structural characterization of smaller organic molecules, such as peptides, even though these species generally possess less well defined secondary structures.[12,13] Thus, in this situation, theoretical methods also provide an attractive means to support experiment and assign certain spectral features to specific structural motifs. However, for these, usually very flexible systems, obtaining a reliable theoretical description has proven to be a challenging task,[14] to differ from the aforementioned case of molecules with well-defined secondary structure, where the semi-empirical methods for the calculation of CD spectra enabled successful application of the theory.[7-9,15-18] On one hand these species typically have an ill-defined secondary structure yielding a very wide phase space. Since the CD spectrum emerges as an ensemble average, the capacity to sample and manipulate this wide phase space significantly affects the result. On the other hand, the entire molecule of interest is typically exposed to the solvent. Even if the individual solvent molecules are not chiral, they will surround the chiral peptides in an asymmetric manner and thus contribute to the CD spectra.[19] This subtle but strong effect appears very challenging for semi-empirical methods.[8] A natural compromise between accuracy

and cost for the calculation of peptide spectra TD-DFT. This approach has performed reasonably well in a number of applications,[1] ranging from metal complexes[20,21] to biologically important systems.[22-24] Hence, TD-DFT calculations become a method of choice for the CD spectra of small molecules. Indeed, experimental CD spectra of short constrained α-helices were reproduced successfully with these more sophisticated methods, after taking the solvent effects into account.[25]

We decided to tackle the above described problem by combining a number of *state-of-the-art* computational methods. By combining the advanced classical molecular dynamics (MD) simulations and the quantum (TD-DFT) calculations we developed a procedure which enables us to calculate CD spectra of a molecule of interest. Our results show that even though the errors of the calculations are not negligible, the main features of the experimental CD spectra are well reproduced in our calculations. Consequently, the link between the CD spectra and the conformational phase space of flexible molecules can be established for the first time.

3.2. Theory of Circular Dichroism

Chirality of molecules is spectroscopically observed in the terms of different refractive indices for left and right circularly polarized light, implying that the extinction coefficients for left and right circularly polarized light are different. This phenomenon is called CD, and the CD spectrum is experimentally obtained when the differential absorbance/dichroism of left and right circularly polarized light ($\Delta\varepsilon$) is measured as a function of wavelength λ,[26] namely

$$\Delta\varepsilon = \varepsilon_L - \varepsilon_R. \tag{3.1}$$

On the other hand, the quantity that describes circular dichroism theoretically is referred to as the rotatory strength R, which relates to the experimentally measured differential absorbance $\Delta\varepsilon$ through the following expression:[27]

$$R = \frac{3(2303)\hbar c}{16\pi^2 N_A}\int \frac{\Delta\varepsilon(\lambda)}{\beta\lambda}\mathrm{d}\lambda = 2.297\cdot 10^{-39}\int \frac{\Delta\varepsilon(\lambda)}{\beta\lambda}d\lambda. \tag{3.2}$$

The parameter β in the Eq. 3.2 is the so-called Lorentz correction, accounting for the average effect of the local field (e.g. solvent) on the chromophore.

We can theoretically obtain the rotatory strength using the Rosenfeld equation given by

$$R_{0k} = Im\{\langle\psi_0|\boldsymbol{\mu}|\psi_n\rangle\cdot\langle\psi_n|\boldsymbol{m}|\psi_0\rangle\}, \tag{3.3}$$

where R_{0k} denotes the rotatory strength of a transition from the ground state to an electronically excited state n, $\boldsymbol{\mu}$ and $\boldsymbol{m}$ represent the electric transition and magnetic transition dipole moments, respectively, while ψ_0 and ψ_n denote the wave functions of the ground state and the nth excited state. As can be seen from Eq. 3.3, the rotatory strength of a particular transition is calculated as the imaginary part of the scalar product of two integrals, namely $\langle\psi_0|\boldsymbol{\mu}|\psi_n\rangle$ and $\langle\psi_n|\boldsymbol{m}|\psi_0\rangle$. The former integral gives the probability that an electron transition from the ground state to the nth excited state occurs, overall representing a light induced oscillating dipole,[28] with electric transition moment $\boldsymbol{\mu}$

serving as a linear displacement of charge. The latter integral, namely $\langle\psi_k|\boldsymbol{m}|\psi_0\rangle$, can be thought of as a current loop induced by light, with magnetic transition dipole moment $\boldsymbol{m}$ marking the angular displacement of charge, in turn causing the charge circulation. It is thus precisely the interplay of electric transition dipole moment $\boldsymbol{\mu}$ and magnetic transition dipole moment $\boldsymbol{m}$ that produces a helical displacement of charge, in turn causing different absorption of left and right circularly polarized light by chiral molecules.

Using the TD-DFT and coupled cluster implementations described in Theoretical Background (Sections 2.1.6 and 2.1.7, respectively) it is now possible to calculate both the excitation energies and rotatory strengths for the electronic transitions of the investigated molecules. The CD spectra can be constructed as a convolution of overlapping Gaussian functions, each representing an electronic transition:

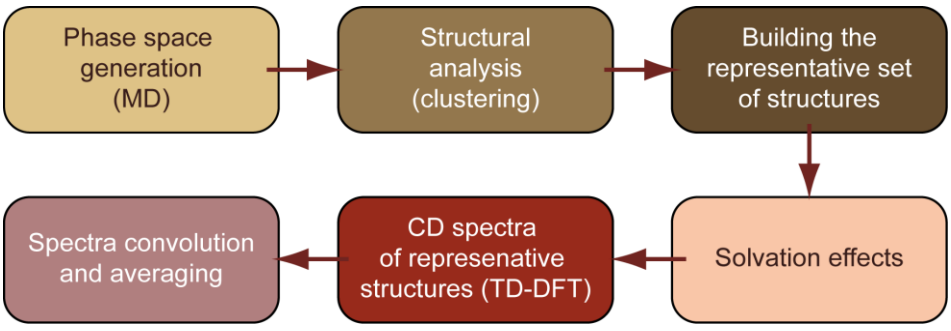
$$\Delta\varepsilon(E) = \frac{1}{2.297\times10^{-39}}\frac{1}{\sqrt{2\pi\sigma^2}}\sum_{i}^{A}\Delta E_i R_i e^{\frac{-(E-\Delta E_i)^2}{2\sigma^2}} \qquad (3.4)$$

Here, A represents the total number of calculated excited states and σ is the width of the Gaussian functions. ΔE_i and R_i are the excitation energies and rotatory strengths for the transition i, respectively.

The subsequent sections of this chapter are mainly based on our three publications (references 10, 11, and 29). Portions of the text and figures used are reprinted with permission from *J. Chem. Theory Comput.* **2012**, *8* (5), pp 1694-1705 (Copyright © 2012 American Chemical Society), *J. Chem. Theory Comput.* **2014**, *10* (8), pp 3270-3279 (Copyright © 2014 American Chemical Society) and *RSC Adv.* **2014**, *4*, pp 13514-13517 (Copyright © Royal Society of Chemistry 2014).

3.3. Methodology

We introduced a general methodology for calculating the CD spectra of flexible molecules in a recent study.[10] In short (Scheme 3-1), the method is based on the generation of a converged conformational phase space, obtained from advanced classical molecular dynamics simulations, namely REMD, followed by finding a set of structures representing the entire phase space using clustering analysis.



Scheme 3-1. Generalized procedure for calculating CD spectra of a flexible molecule.[11]

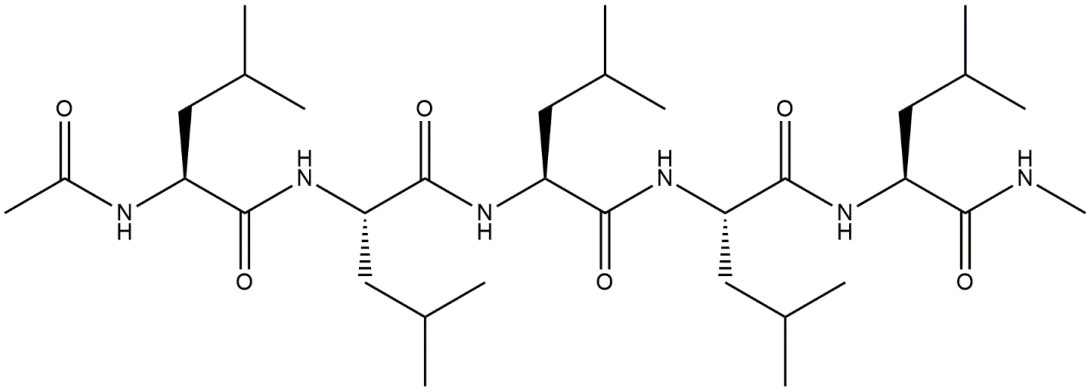


Figure 3-2. Schematic representation of a protected form of Pentaleucine (Leu_5).

The effects of the solvent are introduced through a modified average Coulomb field[30,31] for each conformation.[10] In the next step, an electronic structure method is chosen to calculate the individual CD spectra, which are subsequently combined using appropriate weights to produce the final spectrum. Our methodology will now be presented in a step-wise manner, using a model pentapeptide system, namely the protected from of *S*-Pentaleucine ($AceLeu_5Nme$, Figure 3-2), for testing purposes.

3.3.1. Parameterization

First and necessary step in any study involving molecular dynamics simulations is a choice of a suitable force field or, in other words, of a suitable parameterization. In this respect we deal with two separate cases, namely with the case of a generic organic molecule, for which no set of specialized force field parameters exist and with the case of peptides, for which a highly specialized parameterizations/force fields have been developed.

General organic molecule - parameterize the chosen organic molecule using the general AMBER force field (*gaff*).[32] The missing parameters, namely the atomic charges, are then to be obtained from the one-conformer RESP fit. If the molecule of interest exhibits high flexibility, a multi-conformer approach is recommended. The potential itself is obtained from the HF/6-31G(d) quantum mechanical method, which is consistent with *gaff*[32] methodology. Prior to calculating the electrostatic potential the structure of the molecule of choice is a subject to the optimization also using the HF/6-31G(d) level of theory.

Peptides - parameterize the peptides using, for example, specialized Duan et al. *ff03*,[33] while the missing parameters, necessary to simulate special amino acid residues, such as an unnatural amino acid residues or neutral terminal amino acids, should be obtained in a manner consistent with the *ff03*[33] force field. Specifically, the charges of these residues are to be obtained from a restrained (RESP) two-conformer (extended and alpha-helix) fit to the electrostatic potential. Thereby, the potential is

obtained using the B3LYP/cc-pVTZ//HF/6-31G(d,p) quantum mechanical method with IEFPCM (ε = 4.335) model representing a (low-polarity) polarizable organic continuum.[33] Prior to calculating the electrostatic potential, all structures have to be subjected to constrained optimizations using the HF/6-31G(d,p) level of theory. The dihedral angles (Φ,Ψ) have to be fixed at (-60°,-40°) and (-120°,140°), for the alpha-helix and the extended conformations, respectively.

Solvation - solvate a molecule of interest with *n* molecules of a desired slovent. For the purposes of solvating model Pentaleucine system, a cubic box containing 512 molecules of trifluoroethanol (TFE) was built using Packmol.[34] Parameters for the TFE molecule were obtained from the R.E.DD.B. (RESP and ESP database, parameter code name W-15).[35] The initial size of the box was chosen such that the density was 1.393 g/cm^3, which is typical for TFE at T=300K. The resulting box was equilibrated under NPT conditions and used as a template to solvate the molecules in regular truncated octahedron unit cells. For example, in our model case exactly 350 molecules of TFE (box size of 43.0 Å) was used to solvate the pentapeptide (Figure 3-2).

3.3.2. MD Equilibration

To prepare the system for the production molecular dynamics simulations, one has to go through multiple steps, which include the minimization of the system, thereby removing possible close contacts between the solvent and the solute molecules. Afterwards a set of short molecular dynamics simulations is performed to obtain the proper starting conditions of the system for the production run. Namely, The minimization of this complete periodic system consists of several steps: (*i*) minimization

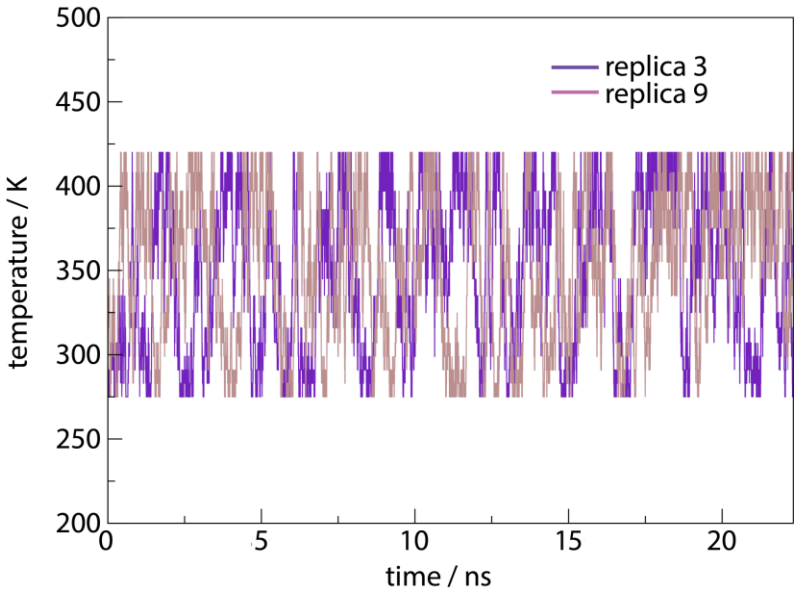


Figure 3-3. Replicas 3 and 9 (corresponding to initial temperatures of 291 K and 345 K, respectively) of *S*-Pentaleucine were taken from 22 ns long REMD run. First 2.5 ns were omitted (equilibration time). It is possible to observe that the replicas sampled the entire temperature space.

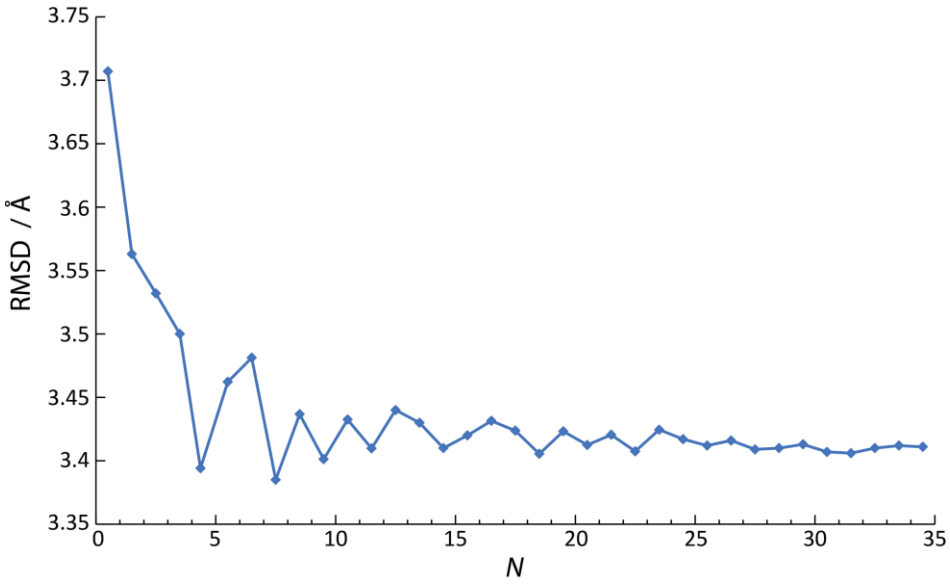


Figure 3-4. The convergence of REMD was tested using an adaptation of the procedure outlined in ref. 36. The trajectory of *S*-Pentaleucine corresponding to T = 299 K was randomized. Subsequently, the average structures of the first 1000×N structures from the randomized trajectory (N goes from 1 to 34) were found in a cumulative fashion. Finally, RMSDs of the average structures from the reference structure were evaluated (reference structure was the starting structure for the MD simulation; however, the convergence of the results is not sensitive to choice of the reference structure).

with the fully restrained solvent, (*ii*) 50 ps of Langevin dynamics, in the NVT ensemble at 300 K, with the collision frequency of 1 ps^{-1} and a time step of 2 fs , restraining both the molecule and the solvent, (*iii*) minimization with reduced positional restraints on the solvent and the molecule, (*iv*) minimization with no positional restraints, (*v*) a further 50 ps of Langevin dynamics at 300 K (NVT ensemble) with no positional restraints, (*vi*) to equilibrate the density to its stable value a final 50 ps Langevin dynamics simulation was performed at 300 K with a constant pressure of 1 atm (NPT).

3.3.3. Phase Space Generation

To generate the conformational phase space we use the advanced replica exchange molecular dynamics (REMD), where 16 replicas of the previously equilibrated system were set on different target temperatures, usually ranging between 270 and 420 K (in the case of *S*-pentaleucine the target temperatures were set at T = 275, 283, 291, 299, 307, 316, 326, 335, 345, 355, 365, 375, 386, 397, 408, 420 K). The temperatures were chosen such that the exchange probability between the replicas falls between 15 and 20% (consult the Theoretical Background, Section 2.2.5.1, for the detailed description of how to obtain the target temperatures for a specific system). The exchanges were attempted every 500 steps (each ps). In between, all replicas were subject to Langevin dynamics in the NVT ensemble with 2 fs time step and a collision frequency of 1 ps^{-1} (dynamics was performed with periodic boundary conditions, whereas the long-range electrostatic interactions beyond an 8 Å cutoff

were taken into account using the particle mesh Ewald method). The 16 replicas of *S*-Pentaleucine were then propagated for a total of 352 ns (22 ns per replica). The coordinates of the entire system were saved every 0.5 ps, which, upon omitting first 2.5 ns from the subsequent analysis (REMD equilibration time), gave rise to 34000 conformations at T = 300 K comprising the conformational space of *S*-Pentaleucine. All molecular dynamics simulations were performed using AMBER 10.[37] Aspects of the convergence of the REMD simulations, including the structural ensembles and the temperature sampling are shown in Figures 3-3 and 3-4, respectively.

3.3.4. Structural Analysis

In this section we consider two powerful mathematical tools commonly used in the statistical analysis of MD data and give a brief elaboration of these concepts following the recent review [39]. Principal component analysis (PCA) is a statistical procedure used to describe the structure of high-dimensional data by reducing its dimensionality. The most common usage of this statistical tool is in exploratory data analysis, enabling one to expose the internal structure of the data and to detect correlations in large data sets.[39] Metaphorically, PCA can be thought of as a lower dimensional "shadow"/projection of a high-dimensional object, which is formed when the object is observed from the most informative viewpoint. This method was initially developed in 1901 by Pearson,[40] as an analogue of the principal axis theorem in mechanics, and it was independently reinvented and named 29 years later by Hotelling.[41] To this day PCA remains one of the most popular and most widely used dimension reduction techniques.[39,42]

PCA has become an invaluable tool in the analysis of MD trajectories,[43,44] where it can be, for example, used to characterize the conformational phase space of a molecule of interest and to monitor the conformational changes in the underlying structures. For example, in the case of peptides, PCA can be used to visualize conformational patterns of all amino acid residues of the investigated peptide simultaneously.

The principal components are obtained by diagonalization of the covariance matrix $\mathbf{C}$ of the data at hand (the data in the case of MD trajectories are simply the atomic positions of the atoms):

$$\mathbf{C} = \mathbf{W}\boldsymbol{\Lambda}\mathbf{W}^{T}$$

As an outcome of this procedure we obtain the matrix $\mathbf{W}$ which contains the eigenvectors (covariance matrices are symmetric, therefore the calculated eigenvectors exhibit orthogonality), and the diagonal matrix $\boldsymbol{\Lambda}$, containing the corresponding eigenvalues as diagonal entries. The eigenvector accounting for the highest variance in the data is called the first principal component (first PC), the second one accounts for the second highest variance in the data (second PC), etc. In favorable cases, only a small set consisting of the first few principal components can be used in analysis, while the remaining

components can be ignored without significantly affecting the results. The PC analysis assumes linearity, thus it is limited to expressing the data in the form of a linear combinations of basis vectors.[39]

Clustering analysis is a mathematical technique which enables one to divide a set of objects into subgroups (i.e. clusters) of similar objects, whereby the precise notion of similarity depends on the problem at hand. This implies that clustering algorithms in general need a measure of similarity of objects ("distance" in the similarity space), which is then used to partition the data set by maximizing distances between the members of different clusters and by minimizing the distances between the members of the same cluster. There exist two main clustering approaches, namely partitional and hierarchical algorithms.[45] Hierarchical clustering creates a decomposition of the data set and results in a hierarchical tree, thereby allowing for nested clusters. On the other hand, partitional clustering works by constructing different partitions of the data set, and can be, for example, used to group similar conformations found using MD simulation into non-overlapping clusters.

Partitional techniques attempt to optimize a certain pre-defined criterion function to partition a data set into a pre-defined number of clusters, which is usually achieved by minimizing the square root error function, representing the most common criterion.[39,46] The *K*-means algorithm represents one of the simplest and oldest such algorithms, and is one of the most widely used clustering approaches. The first step of the *K*-means algorithm is to randomly place a pre-defined number of centroids (*K*, one for each cluster), each denoting the average position of its respective cluster; when clustering coordinates from the MD simulations, the centroid simply represents the conformation that best describes all of the conformations within that cluster. Secondly, initial *K* clusters are formed by assigning each point in the data set to the nearest centroid. Lastly, new positions of *K* centroids are now found by calculating the mean of the initially formed clusters. This process is iterated until the clusters remain stable, i.e. centroids do not change their position from the previous step. Although it is possible to show that this procedure always converges, the obtained clusters do not necessarily provide the optimal configuration. This problem stems from the non-deterministic nature of this algorithm, i.e. the outcome depends on the initial random distribution of centroids. It is advisable in this respect to repeat the clustering of the data a larger number of times when using the *K*-means algorithm, which can significantly diminish otherwise exhibited problems. This algorithm is known to perform rather poorly when the "expected" clusters are non-spherical in nature and when the clusters strongly vary in size or densities.[45] When applied to MD simulation data this clustering usually produces homogeneously sized clusters.[39,47] The recent study showcased *K*-means, together with average-linkage clustering algorithm, as the best performing clustering algorithm for analysis of MD simulation data, prompting us to use it in the analysis of our simulation data.[47]

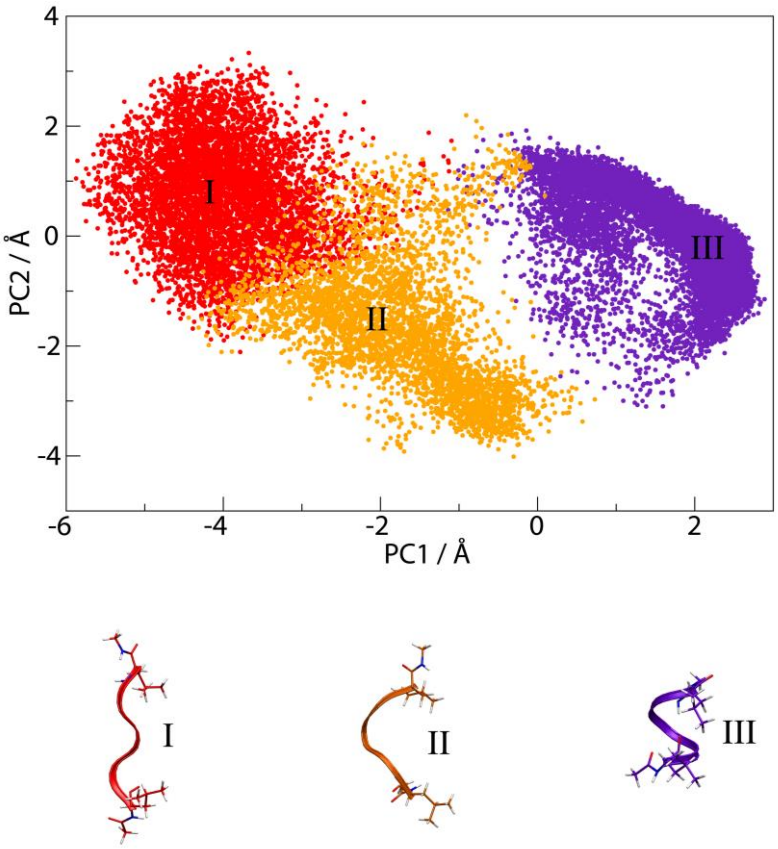


Figure 3-5. Top: Conformational phase space of *S*-Pentaleucine at T = 299 K, after clustering. Each cluster is represented by a Roman numeral. Bottom: Ribbon renderings of three representative zwitterionic Met-enkephalin conformations (conformations most similar to the centroids of their respective clusters) are shown. Conformations were plotted using the VMD visualization software.[38]

Both the PCA and the clustering using the *K*-means algorithm were performed using AmberTools.[48] In the analysis of the conformational phase space, in the case of peptides the α-carbon atoms were used (for a general organic molecule a careful choice of atoms has to be made so that the important motions of a molecule are properly collected), whereas for CD spectra calculations, further sub-clustering was performed on the basis of the positions of all atoms. Finally, the representative structures for each cluster were determined as the structure with the lowest RMS deviation from the relevant centroid. To showcase this procedure, all 34000 structures of *S*-Pentaleucine were used to prepare principal components which were then projected to the first two principal components, which upon visual inspection could be associated with the end-to-end distance and the twist of the backbone, respectively. In this way, 2-D plots, where each conformation is represented by a single data point, were constructed for a molecule of interest (Figure 3-5). The *K*-means clustering was performed afterwards, producing in the case of our model *S*-Pentaleucine system three well defined clusters (Figure 3-5), whose representative structures are also shown in Figure 3-5.

3.3.5. CD spectra Calculations

In this thesis, calculation of the CD spectra is performed using a QM/MM approach, in which the entire peptide was treated quantum mechanically using TD-DFT, whereas the effect of the solvent was included through the introduction of partial charges at the positions of the solvent nuclei, as obtained from the molecular dynamics simulations. The spectra were usually obtained with the B3LYP functional, which unless otherwise specified was used in conjunction with the 6-31G(d) basis set (for the solvent and convergence testing presented in the following two sections (3.3.5.1 and 3.3.5.2) we used smaller STO-3G basis set, enabling a significant speed up of the necessary calculations), which was benchmarked against larger 6-311++G(d,p), showing it to be sufficient to obtain converged spectra (Section 3.3.5.5). On the other hand, the performance of the B3LYP (and four additional TD-DFT functionals) was extensively investigated, benchmarking it both against the *ab initio* RICC2 method and against the experiment (Section 3.5.7).

3.3.5.1. Solvent effects

The effect of differing solvent configurations on the CD spectra is clearly demonstrated in Figure 3-6. Therein, each spectrum corresponds to the same conformation of *S*-Pentaleucine, each with a unique solvent configuration. This result confirms the marked sensitivity of both the ΔE_i and R_i to the presence of an asymmetric solvent structure.[49] Clearly, in order to obtain a reliable and reproducible CD spectrum for a single peptide conformation, one needs to average over many configurations of the solvent.[50] In the current application, we generated the required configurations in the following manner. First, we extracted and minimized a selected snapshot from the REMD trajectory. The resulting configuration was then subjected to constant temperature molecular dynamics (NVT) with strong

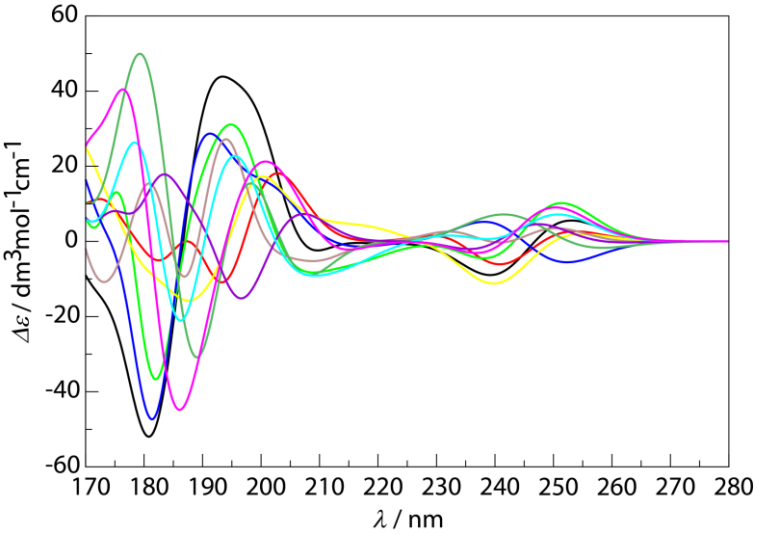


Figure 3-6. CD spectra of a single *S*-Pentaleucine conformation with different configurations of TFE. The spectra were calculated using B3LYP/STO-3G level of quantum theory, each convoluted by setting $A = 72$ and $\sigma = 0.15$ eV (see Section 3.2).

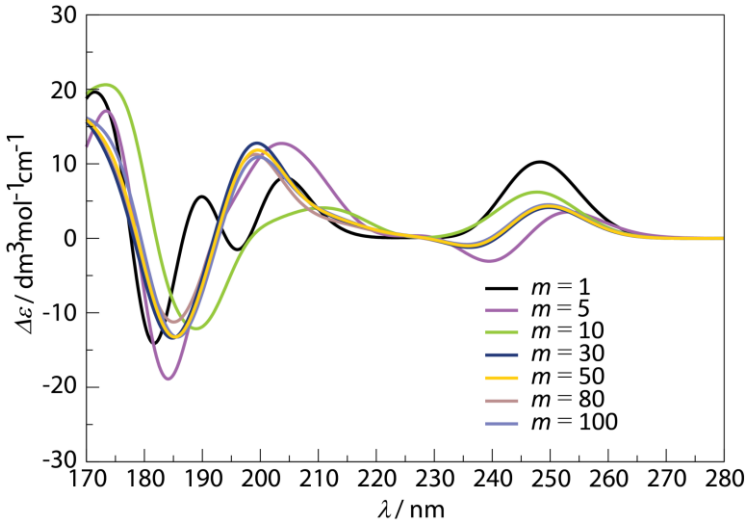


Figure 3-7. Averaged CD spectra after finding a mean of *m* spectra examples of which are shown in Figure 3-6. The CD spectra were obtained using the B3LYP/STO-3G level of theory.

positional restraints on the peptide. After an equilibration period, the solvent coordinates were extracted every 10 ps and, after alignment with respect to the minimized peptide conformation, used as a set of independent solvent configurations.

The CD spectrum of the selected peptide conformation was then calculated under the influence of each of the obtained solvent configurations (Figure 3-6), using the B3LYP/STO-3G level of theory for this particular task, setting A = 72 and σ = 0.15 eV (see Theory of CD, Section 3.2). The resulting spectra were combined to give the solvent-averaged spectrum for the relevant peptide conformation. The convergence of such an average is shown in Figure 3-7 using m = 5, 10, 30, 50, 80 and 100 solvent configurations. It can be seen that while reasonable convergence can already be achieved with 30 solvent configurations, the use of 50 is preferable. Such a procedure is, however, prohibitively expensive for general use, since many peptide conformations would need to be treated in this way in order to obtain a CD spectrum representative of the overall structural ensemble.

To render an adequate treatment of the solvent more feasible, we have employed an approach in which we first combined the distinct solvent configurations and then evaluated their collective effect on the CD spectrum. Specifically, by superposing n solvent configurations (obtained as per the description above), each with its partial charges scaled by n, we generated a single averaged solvent configuration. The CD spectrum for the relevant peptide configuration was then calculated under the influence of this single configuration. Such a procedure can be considered as a simplified version of the Averaged Solvent Electrostatic Potential (ASEP) developed by Aguilar and coworkers[51] which closely resembles the Average Solvent Electrostatic Configuration (ASEC) approach suggested by Coutinho et al.[52] The resulting dependence of the spectrum on the number of solvent configurations used to construct the average configuration is shown in Figure 3-8. The comparison with the stepwise

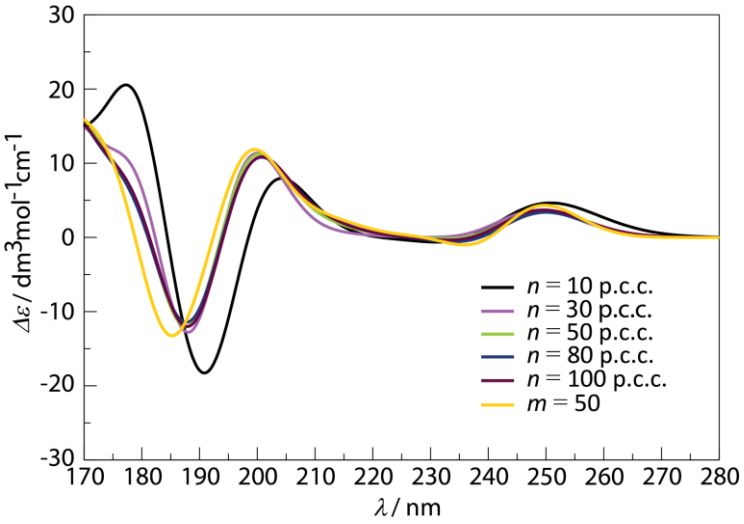


Figure 3-8. The same *S*-Pentaleucine conformation from Figure 3-6 and 3-7 is used to test the convergence of CD spectra when the ASEC approach is applied. The convergence of the CD spectrum again occurs when 50 point-charge configurations (p.c.c.) are employed in order to build the ASEC (calculated using B3LYP/STO-3G level of theory, with A = 100 and σ = 0.19 eV). For comparison, the spectrum obtained by averaging over m = 50 simulations (Figure 3-7, yellow curve) is also plotted (yellow curve).

calculation (averaging over 50 spectra as in Figure 3-7) shows that excellent agreement is obtained after including the contribution of 50 solvent configurations to the superposed average.

3.3.5.2 Calculating CD spectra of a single cluster

Under the assumption that the conformational phase space of each particular molecule is well represented by the previously described results of the REMD simulations, we proceed with calculating the CD spectra of a single cluster. This will be exemplified for the well-defined cluster III of *S*-Pentaleucine (Figure 3-5). Despite the fact that the individual structures in this cluster have similar backbone conformations, their (solvent averaged) CD spectra (calculated as discussed in the Section 3.3.5.1) differ appreciably (Figure 3-9a). This suggests that the side-chains have considerable effects on the individual CD spectra. Consequently, it is clear that a single structure cannot represent the whole cluster but a set of the representative conformations, which will be the basis for the calculation of the CD spectra, needs to be constructed. We obtained this set through sub-clustering of the initial backbone-based cluster by applying the *K*-means clustering algorithm, taking into account the positions of all the atoms in the molecule of interest.

The size of the representative set is determined by monitoring the convergence of the spectra after averaging over a number of representative structures in the set (Figure 3-9b). In the specific case of the cluster III of *S*-Pentaleucine, sub-clustering was performed to give sets with 3, 9, 18, 27 and 36 and 45 structures, respectively. The average CD spectrum of the set was constructed by weighted

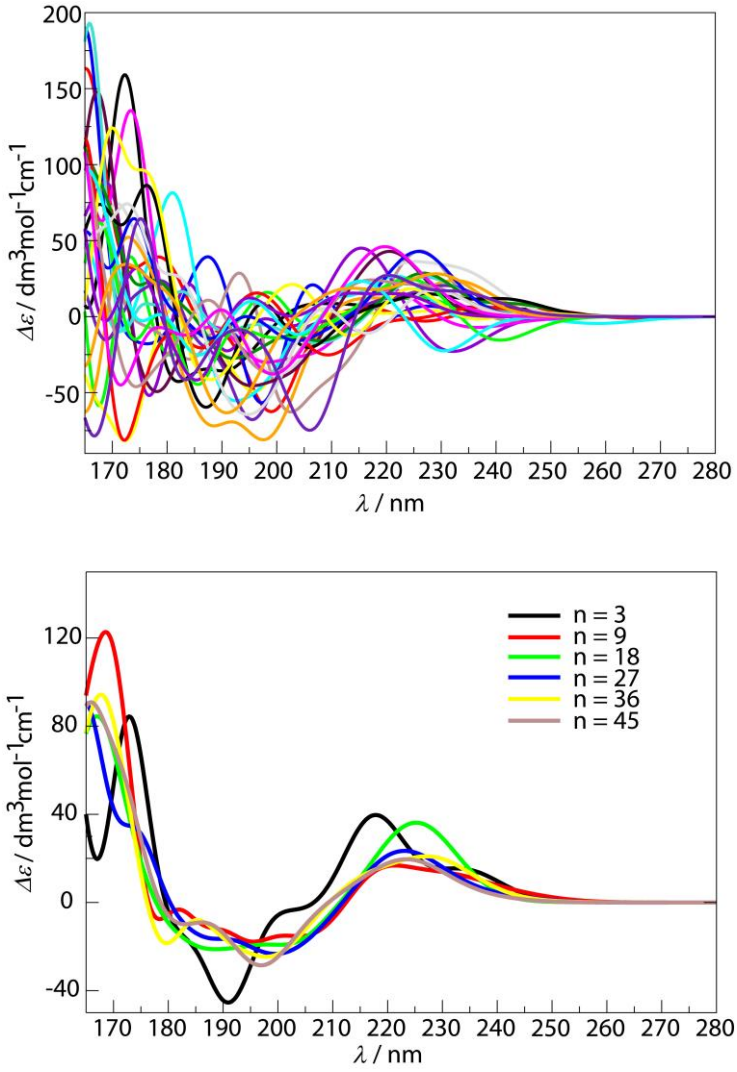


Figure 3-9. Top: single CD spectra of 27 representative conformations of *S*-Pentaleucine obtained after sub-clustering of cluster III into 27 sub-clusters (obtained using our implementation of the ASEC approach, with B3LYP/STO-3G level of theory, where we set A = 100 and σ = 0.19 eV, see Section 3.3.5.1 and 3.2). Bottom: convergence seen when 27 representative conformations are used in order to obtain the average CD spectrum of cluster III (n represents the number of sub-clusters in which the cluster III is divided).

averaging. Thereby, the weight of each individual spectrum is given by the population fraction of the respective sub-cluster in the context of the whole cluster. As can be seen from Figure 3-9b, the spectrum begins to converge already for a set of 18 representative conformations, and very good results are obtained for a set with 36 characteristic conformations, each representing a sub-cluster with, on average, 680 conformations. While further fragmentation of the phase space is possible, this tends to produce a large number of sub-clusters containing only 4-5 conformations, with badly defined centroids.

3.3.5.3. Calculating CD spectra of the entire phase space

In order to build the CD spectra emerging from the entire phase space, we have opted to retain the above sampling frequency to arrive at an average of 680 conformations within a given sub-cluster. In the context of a phase space built from 34000 structures, this amounts to a total of 50 sub-clusters and a corresponding set of 50 characteristic structures. The final combination of the corresponding spectra can be considered as averaging of the different backbone structures of the principal clusters (Figures 3-5 and 3-10) to produce the overall spectrum of a given molecule/peptide. Under such circumstances, we used A=100 and σ = 0.135 eV (B3LYP/6-31G(d) level of theory). To test the consistency of this methodology in its entirety, we calculated the overall CD spectrum of *R*-Pentaleucine, which is the enantiomer of *S*-Pentalucine. The phase space and the CD spectrum for *R*-Pentaleucine were generated according to the procedure outlined above. In the ideal case, due to the mirror symmetry, the resulting

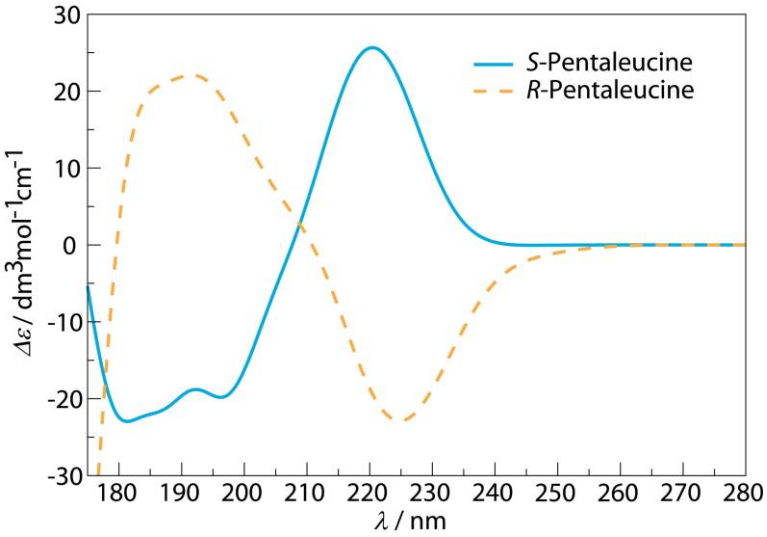


Figure 3-10. Comparison of *R*- and *S*-Pentaleucine spectra.

spectrum should be identical but opposite to that of *S*-Pentaleucine. However, because of errors which accumulate in each step, from generating the phase space with a classical force field, over the sampling and the building of a subset of solvated representative structures, to the TD-DFT calculations, the sum of the two spectra is not exactly zero, and deviations of up to 10-20% have been observed (Figure 3-10). Nevertheless, the main features of both the *R*- and *S*-Pentaleucine spectra are identical and opposite, with main peaks appearing at exactly the same locations. Notably, these spectra emerged from two completely independent calculations performed on two strongly related but distinct peptides. Since the flexible Pentaleucine possesses a relatively wide conformational phase space, which is usually most prone to errors, this good correspondence supports the applicability of our method.

3.3.5.4. Influence of different force fields on the CD spectra

To investigate the effect of different force fields on the CD spectrum of a molecule of interest we again employed our model *S*-Pentaleucine system. We thereby above presented methodology to prepare the CD spectra of *S*-Pentaleucine by generating conformational phase space employing three different and highly specialized force fields, namely *ff03*,[33] *ff99SB*[53] and *ff12SB*[54] from the Amber suite of force fields. Then, using our established strategy, we gathered 50 representative structures from each of these phase spaces. Interestingly, though different parameterizations give rise to somewhat different conformational phase spaces, the CD spectra seem rather well preserved in all cases, giving rise to very similar CD spectra (Figure 3-11). This being so we decided to employ the force field *ff03*,[33] parameterized specifically for peptide/protein simulations, for the remainder of the peptides studied in this thesis.

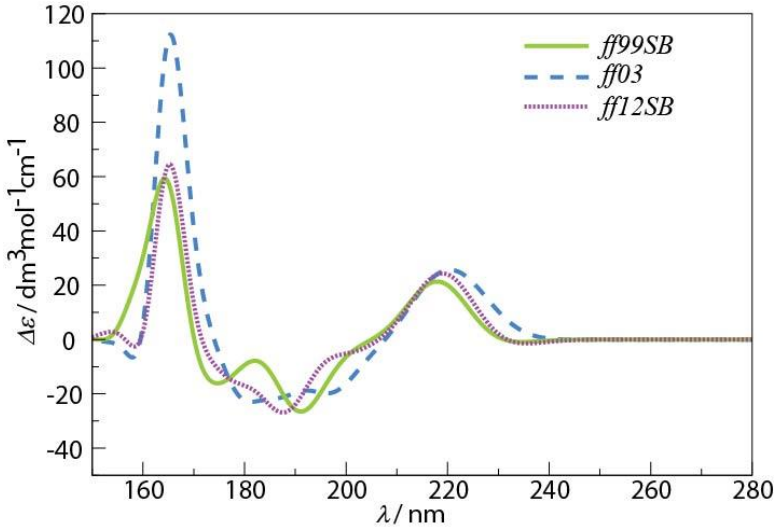


Figure 3-11. CD spectra of *S*-Pentaleucine obtained using *ff99SB*, *ff03* and *ff12SB* force fields.

3.3.5.5. Influence of the basis set on the CD spectra

We have explored the effect of using a larger basis set on the calculated CD spectra. For this purpose we employed B3LYP functional together with the Pople 6-31G(d) basis set, which is the basis set of choice in this work, and also the significantly larger 6-311++G(d,p). Inspection of the result immediately reveals the high degree of similarity between the 6-31G(d) and 6-311++G(d,p) spectra, with both spectra exhibiting same spectral features (Figure 3-12). It thus seems that the results obtained with the 6-31G(d) set represent a good approximation to those emerging from the use of the much larger 6-311++G(d,p). This lends confidence to the continued use of 6-31G(d), as we have done throughout this work.

At this point we have presented our methodology and, as we have seen, it indeed seems capable of tackling the problem at hand. However, the final stage of our methodology, namely the test

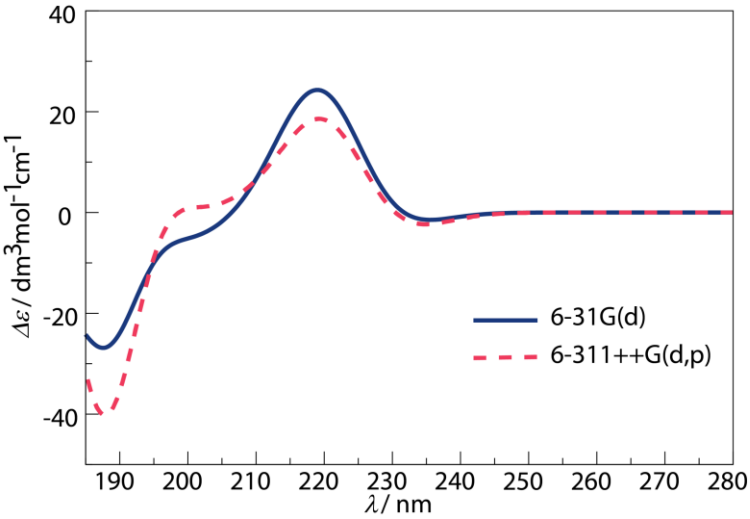


Figure 3-12. CD spectra of *S*-Pentaleucine obtained using 6-31G(d) and 6-311++G(d,p) basis sets.

against the experiment, is yet to be performed. With this in mind we present two different cases, whereby we investigate the applicability of our methodology in the case of a) small organic molecules and b) highly flexible peptide species.

3.4. Case 1: CD Spectra of Rhodomyrtals

3.4.1. Introduction

The genus *Rhodomyrtus* (Myrtaceae) consists of 24 species, which are native to Asia, Malesia, Melanesia, and Australia. During the past few years, a number of unusual phloroglucinol derivatives such as rhodomyrtone, rhodomyrtosones, and tomentosones, have been isolated from *Rhodomyrtus tomentosa*.[29] Among them, rhodomyrtone displayed potent anti-bacterial activity, especially against methicillin-resistant *Staphylococcus aureus*. Chemical investigation of a relative to *R. tomentosa*, namely *R. psidioides* plant, which is commonly called native guava, led to the isolation of two novel compounds that possess a new phloroglucinol-coupled eudesman skeleton and one known compound, namely eucalyptin A.[29] While the relative configurations of these compounds were sucessefully determined using experimental techniques, in particular using COSY and NOESY NMR measurements, the absolute configuration could not have been established, as the aforementioned techniques remain "blind" to this molecular property. Thus, the CD spectra of both of these compounds were experimentally obtained, baring in mind that the only way to establish the absolute configuration in this case is by a direct comparison of the experimental results with the theoretically obtained spectra.[29] We thus decided to theoretically obtain the ECD spectra of compounds **1** and **2** and of their enantiomeric counterparts (Figure 3-13) by employing our methodology (Section 3.3), where the CD spectrum arises from a set of representative conformations of the molecule of interest. As the identical procedure was used in the case of both compounds, we present it only for compound **1**.

3.4.2. Simulation Details

Using our established methodology we first modeled compounds **1** and **2** using general AMBER force field (*gaff*).[32] The missing parameters, namely the atomic charges, were obtained from the one-conformer RESP fit. Upon solvating compound **1** with 758 molecules of methanol we proceeded with a combination of minimizations and dynamics simulations in explicit solvent (for details see Methodology, Section 3.3.2), which resulted in the equilibrated system at 300 K. The conformational

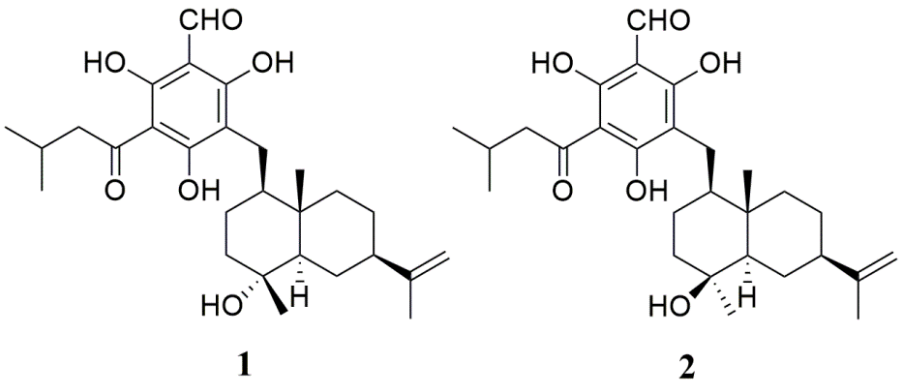


Figure 3-13. *Rhodomyrtal* compounds **1** and **2**.[29]

phase space was generated using REMD, where 16 replicas of the previously equilibrated system were set on different target temperatures (T = 275, 282, 289, 296.5, 304, 312, 320, 328,336.5, 345, 354, 363, 372, 381.5, 391 and 402 K). The exchanges were attempted every 500 steps. In between, all replicas were subject to Langevin dynamics in the NVT ensemble with 2 fs time step and a collision frequency of 1 ps^{-1} (dynamics was performed with periodic boundary conditions, whereas the long-range electrostatic interactions beyond an 8 Å cutoff were taken into account using the particle mesh Ewald method). The 16 replicas were propagated for a total of 256 ns (16 ns per replica). The coordinates of the entire system were saved every 0.5 ps, which, upon omitting first 2.5 ns from the subsequent analysis (REMD equilibration time), gave rise to 27000 conformations at T = 296.5 K comprising the conformational spaces of compound **1** and **2**, respectively.

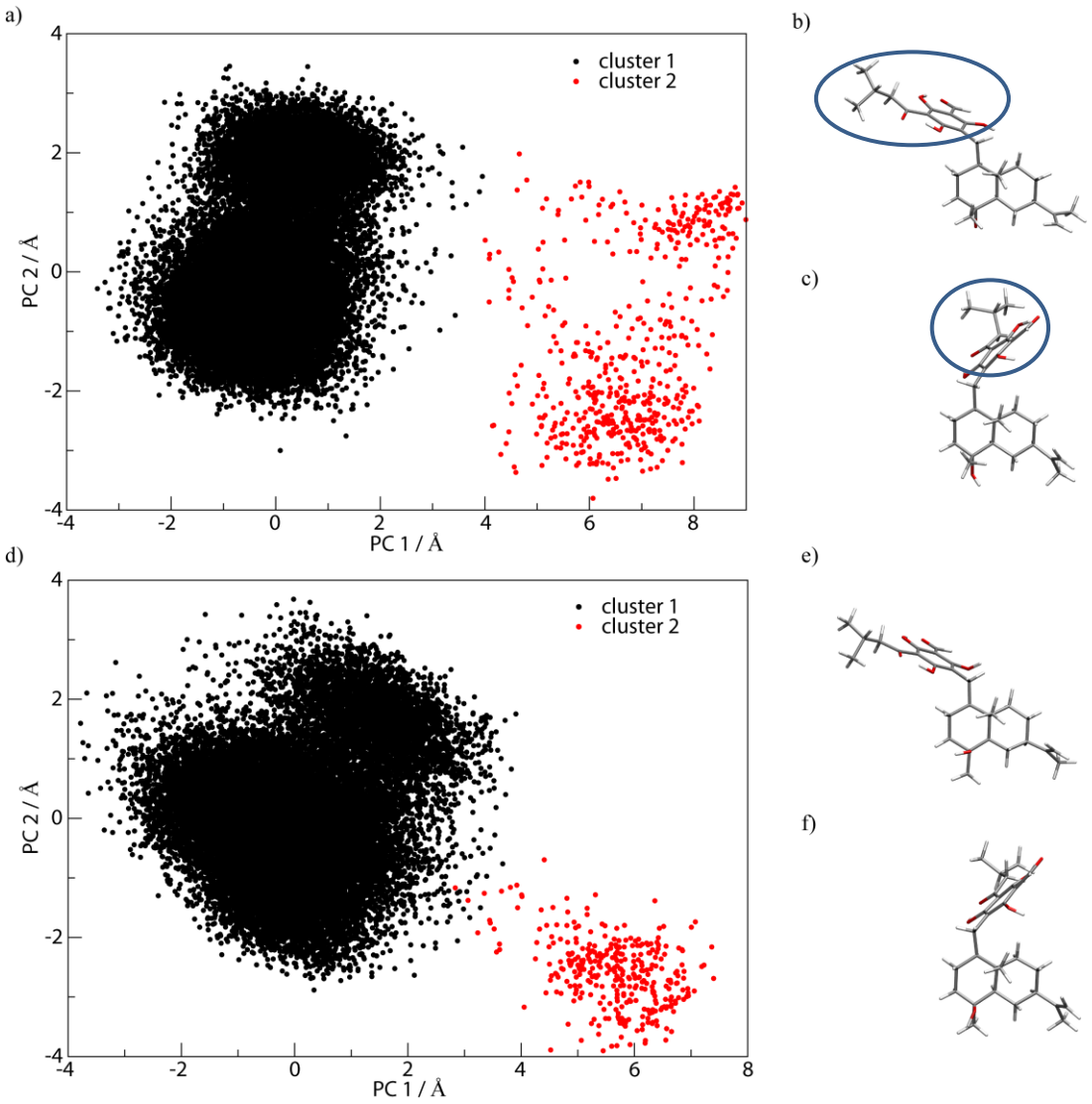


Figure 3-14. Conformational phase space of compounds a) **1** and d) **2**. The representative conformations of b) and e) cluster 1, and c) and f) cluster 2, for compounds **1** and **2**, respectively.[29]

3.4.3. Conformational Phase Space of Rhodomyrtals

We firstly analyzed the conformational phase space of compounds **1** and **2**, thereby employing principal component analysis and clustering (*K*-means algorithm), with the latter producing two distinct clusters of conformations (clusters and representative conformations of both compounds are shown in Figure 3-14). Both the PCA and clustering were based on 10 carbon atoms representing the backbone of both compounds (shown for compound **1**, Figure 3-15). We can easily observe that the two compounds live in a rather similar conformational phase spaces, which should come to no surprise, as the only difference between the two is the chirality of a carbon atom in the trans-decalin moiety (Figure 3-13). Additionally, in both cases cluster 1 contains the vast majority of the conformations, namely more than 95% of all conformations belong to this cluster (Figure 3-14a) and d), black). The main difference between the representative conformations of clusters 1 and 2 for both compounds is the relative position of the dominant side chain (enclosed in blue ellipses) to the trans-decalin moiety in compounds **1** and **2**, as can be observed from b) and c), and from e) and f), respectively.

3.4.4. CD Spectra of Rhodomyrtals – Elucidating the Enantiomery

To calculate the CD spectra of the two compounds we firstly subclustered the two clusters found after the initial conformational analysis, which gave rise to 50 conformations representing entire

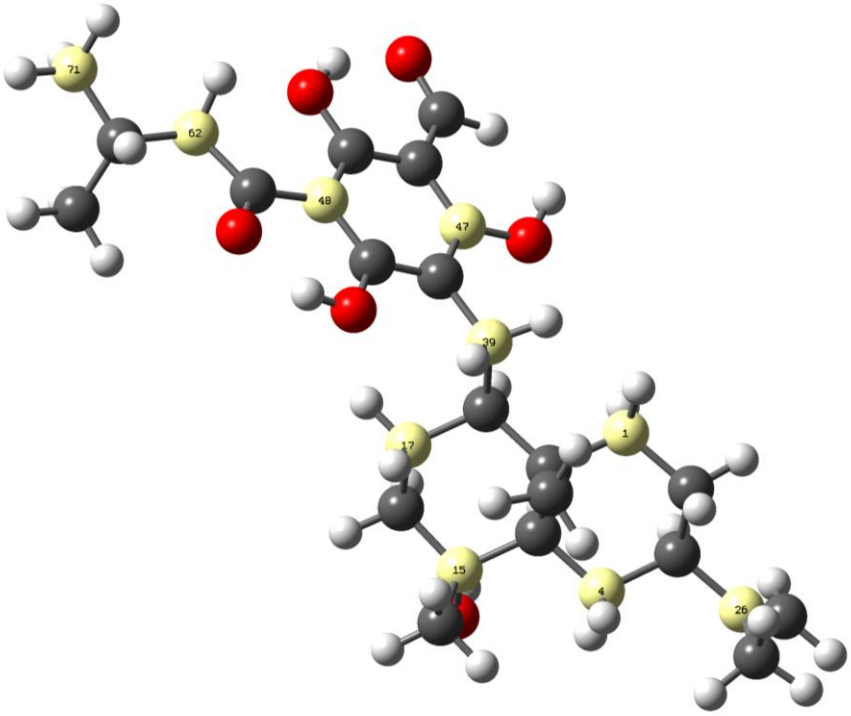


Figure 3-15. Set of ten carbon atoms on which PCA and clustering of the compound **1** were based are shown in yellow. The equivalent set of atoms was chosen for the analysis of the compound **2**.[29]

conformational phase space of compound **1** and **2**. The CD spectra of 50 representative conformations (each classically minimized prior to the spectra calculations) were now calculated using TD-DFT, where B3LYP/6-31+G(d) level of theory was used, taking into account first 40 excited states. The solvent effect was accounted for by using our already described approach (ASEC, see Solvent effects Section of the Methodology, Section 3.3.5). Finally, the average CD spectrum of compounds **1** and **2** was obtained as a weighted average of 50 individual spectra, which were prior to the averaging convoluted with Gaussian functions of width 0.18 eV. Thereby, the weight of each individual spectrum was given by the population fraction of the respective subcluster. The average CD spectra, representing entire conformational phase space of *Rhodomyrtal* compounds **1** and **2** are shown in Figure 3-16, together with the independently calculated CD spectra of their respective enantiomers, again obtain using the described procedure. We find that the theoretically obtained CD spectra enable us to unambiguously determine the absolute configuration of both experimentally obtained compounds. Namely, we can observe from Figure 3-16 that in both cases we can easily conclude what enantiomer is represented in the experiment, as all of the main features of the experimental CD spectra are exhibited with the enantiomers marked red (compounds **1** and **2**, Figure 3-16a and b). More precisely, the dominant spectral features in the CD spectra of compounds **1** and **2**, namely the minima found at around 212 nm, the strong maxima lying at about 263 nm and the wide minima appearing at 295 nm, are all found to lie at almost identical positions in the obtained experimental CD spectra, with the intensities and the qualities of these features following very well their counterparts from the experimental CD spectra. On the other hand, the calculated CD spectra of their respective enantiomers (shown in blue) exhibit identical and opposite spectral features to the ones observed for the experimentally obtained CD spectra of *Rhodomyrtal* compounds.[29]

This study represents the first "benchmark against the experiment" of our developed

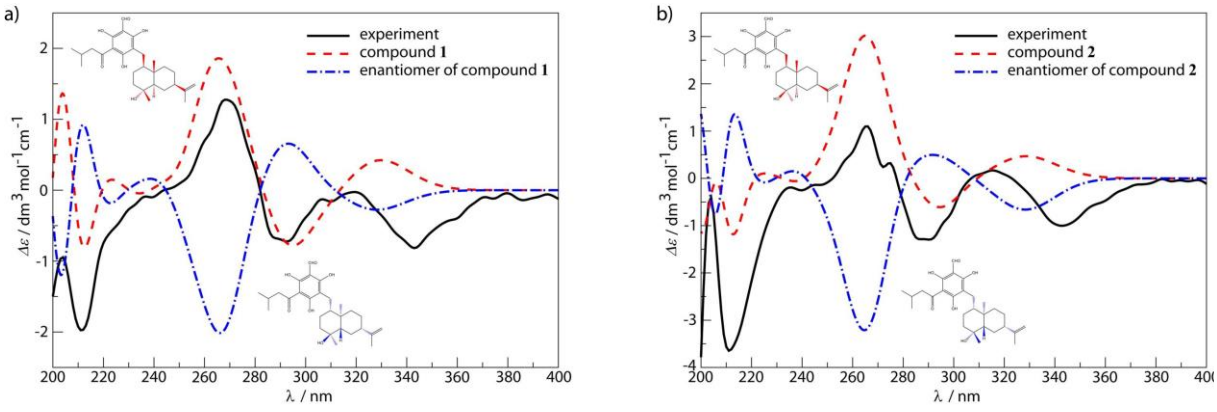


Figure 3-16. Experimental CD spectra of compounds **1** and **2** (black, a) and b) respectively) and the corresponding theoretical CD spectra calculated using B3LYP (red dashed lines, a) and b) respectively). Additionally, the CD spectra of the enantiomers of compounds **1** and **2** are shown (blue dash-dotted line, a) and b), respectively).[29]

methodology, in which we observed an excellent agreement between the theoretically obtained and experimentally measured CD spectra, enabling us in turn to successfully and unambiguously determine the absolute configurations of *Rhodomyrtal* compounds **1** and **2**. Although these species do not possess as high flexibility as the species studied in our following case study (Section 3.5), we still deem it necessary to perform the averaging procedure (see Section 3.3.7), as we observed that a relatively similar structures of these compounds can exhibit rather distinct spectral patterns. With this we turn to somewhat more challenging task of calculating CD spectra of highly flexible opioid peptides.

3.5. Case 2: CD Spectra of Met- and Ada-Enkephalin

3.5.1. Introduction

Opioid peptides belong to a group of biologically relevant signaling molecules that inhibit neurotransmitter release upon activation of the appropriate opioid receptor.[55] These peptides are known to play a significant role in regulating dopamine release,[56] drug-induced reward,[57] as well as in pain mediation, opiate dependence, and euphoria.[57] Possible keys to their activity are the aromatic residues found in positions 1 and 4 in their sequences.[57]

The most ubiquitous opiod peptide is Met-enkephalin (Tyr-Gly-Gly-Phe-Met, Met), also known as the opioid growth factor (OGF). Met-enkephalin plays a major role in cell proliferation and tissue organization during development, as well as in cellular renewal, wound healing, and angiogenesis. However, particularly interesting is its capacity to inhibit tumor-cell division which is done in a receptor-mediated fashion.[58-61] While it is known that Met-enkephalin binds to three classical receptors, namely the μ-, κ- and δ-receptors, it was shown that ligands selective for these three receptors do not affect cell growth either *in vitro* or *in vivo*.[58] Subsequently, it was found that the Met-enkephalin also binds to an OGF-receptor (OGFr) positioned on the nuclear envelope.[62] OGFr, previously known as the ζ-receptor has no significant homologies to classical opioid receptors. Nevertheless, it is precisely the Met-enkephalin-OGFr axis that was found to be responsible for the anti-tumor activity.[60-62] This implies that Met-enkephalin must be transported through the cell membrane in order to interact with the OGFr.

Following these discoveries, various unnatural analogues of Met-enkephalin containing lipophilic moieties were synthesised[58] and tested for improved or novel activity profiles toward malignant diseases. The analogues had either backbone modifications or were covalently conjugated with a lipid or a polymeric tail. However, analogues in which both tyrosine and phenylalanine residues had undergone modifications expressed smaller affinity for the OGFr and in turn had diminished anti-tumor activity.[62] Other types of substitutions promoted the passive or active cellular uptake of peptides without greatly affecting the affinity, thus resulting in increased activity.[58,59]

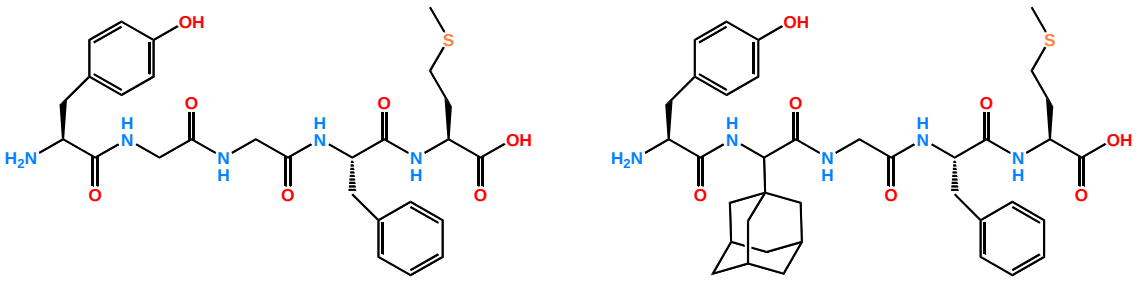


Figure 3-17. Met-enkephalin (Met, left) and Ada-enkephalin (Ada, right).[10]

An interesting group of peptides in this respect are those containing the adamantyl functional group. For example, the incorporation of (*R*,*S*)-(1-adamantyl)glycine (Aaa) in position 2 of the Met sequence results in Ada-enkephalin (Tyr-Aaa-Gly-Phe-Met, Ada, Figure 3-17).[58] This peptide showed a tenfold increase in the anti-tumor activity when compared with Met-enkephalin in the case of HEp-2 cells (human larynx carcinoma) and a twofold increase with SW-620 tumor cells (colon carcinoma).[58] This efficiency may arise from the capability of the hydrophobic Aaa to induce both β-turn and/or γ-turn[63] spatial conformations, which then increase the hydrophobicity of Ada-enkephalin and its membrane permeability.[64] These results indicate that the limiting step in anti-tumor activity may be the diffusion of peptides through the membrane, which in turn may be related to the structure of peptides in nonpolar environments.

The structure of Met-enkephalin has been studied both in water and in nonpolar solvents. In water, both NMR results and molecular dynamics simulations show Met-enkephalin to possess an essentially random distribution of conformers, with no unique native structure.[56] On the other hand, NMR experiments in binary bilayered mixed micelles showed relatively well defined structures.[65] These latter results, however, have not thus far been reproduced theoretically.

As an alternative to NMR, information about the secondary structures of proteins and peptides can be obtained by measuring CD spectra, as explained in the Introduction (Section 3.1). In the case of Met-enkephalin and Ada-enkephalin, such experiments were performed in the low-polarity solvent 2,2,2-trifluoroethanol. The spectrum of Ada-enkephalin displayed a deep negative minimum at 200 nm, which was assigned as representing a high population of ordered conformers, with likely folded backbones. Met-enkephalin was assumed to remain unfolded because its spectrum remained positive for all wavelengths between 190 and 240 nm.[59] Such conclusions support the hypothesis that the antiproliferative activity is connected to the folding ability of investigated peptides.[59]

Ada-enkephalin has two epimers, for which different antitumor capacities have been measured. The *S*-epimer (containing *S*-Aaa) penetrates rapidly to the cellular interior, while the *R*-epimer (containing *R*-Aaa) does not even bind to the outer membrane, and therefore does not enter the cell.[58] This points to possible differences in the physical and structural properties of the epimers. However, the discussed CD measurements of Ada-enkephalin were conducted on a 1:1 mixture of the two epimers. Hence, the origin of the dominant minimum in the spectrum cannot be clearly associated with certain structural preference. Furthermore, the side-chain chromophores present in the two aromatic residues (Tyr1 and Phe4), present in all peptides, may have as strong a contribution to the CD spectra as the backbone structure itself.[66-68] In the context of comparisons with experiments, an additional difficulty is the potential coexistence of the zwitterionic and neutral forms of the peptide. These forms may have very different phase spaces but will both contribute to the measured spectrum with statistical weights given by their concentration ratio, which is often not known. As a

consequence, the experimental CD spectra of highly flexible peptides consisting of both aromatic and non-aromatic constituents have not yet been successfully reproduced by theoretical means.

We use our developed methodology to tackle the described difficulties that arise when trying to calculate CD spectra of highly flexible molecule. In this respect, we first investigate the conformational phase spaces of Met-enkephalin and the Ada-enkephalin epimers in TFE, in both the neutral and zwitterionic forms. On the basis of these phase spaces we calculate the CD spectra by means of TD-DFT calculations. After estimating the concentration ratio of the two forms, we construct the analogue of measured CD spectra of these two peptides. Our results show that even though the errors of the calculations are not negligible, the main features of the experimental CD spectra are well reproduced in our calculations, and that TD-DFT is indeed a method of choice when calculating CD spectra of small and flexible molecules.

3.5.2. Pepetide Preparation

All peptides were modeled using the Duan et al. *ff03*[33] and the general Amber force fields,[32] as part of the Amber 10 simulation package.[37] Specifically, we constructed the zwitterionic forms of: (i) Met_Z - Met-enkephalin (TyrGly-GlyPheMet); (ii) *R*-Ada_Z - the *R*-epimer of Ada-enkephalin (Tyr*R*-AaaGlyPheMet), (iii) *S*-AdaZ - the *S*-epimer (Tyr*S*-Aaa-GlyPheMet). Furthermore, we considered the neutral forms of: (iv) Met_N (Tyr_0GlyGlyPheMet_0); (v) *R*-AdaN (Tyr_0*R*-AaaGlyPheMet_0) and (vi) *S*-Ada_N (Tyr_0*S*-AaaGlyPheMet_0). Here Tyr_0 and Met_0 represent the neutral forms of the N- and C-terminal tyrosine and methionine residues, respectively.

The missing parameters for Tyr_0, Met_0, and Aaa amino acids were obtained in a manner consistent with the ff03 force field.[33] Specifically, the charges were obtained from a restrained (RESP) two-conformer (extended and alpha-helix) fit to the electrostatic potential. The potential was obtained from the B3LYP/cc-pVTZ//HF/6-31G(d,p) quantum mechanical method with IEFPCM ($\varepsilon = 4.335$) model representing a (low-polarity) polarizable organic continuum.[33] Prior to calculating the electrostatic potential, all structures were subject to constrained optimizations using the HF/6-31G(d,p) level of theory. The dihedral angles (Φ,Ψ) were fixed at (-60°,-40°) and (-120°,140°), for the alpha-helix and the extended conformations, respectively. All aforementioned calculations were performed using the Gaussian 09 package.[69]

3.5.3. Simulation Details

Starting from an initially linear conformation, the initial folded structures for the dynamics with explicit solvent were obtained through a combination of minimizations and dynamics simulations using the Generalized Born representation of implicit solvent available in AMBER.[37] All subsequent explicit simulations were performed with periodic boundary conditions whereas the long-range electro-static interactions beyond 8 Å cutoff were calculated using the particle mesh Ewald method.[37]

The previously mentioned cubic box containing 512 molecules of TFE (see Section 3.3.1.) was used to solvate the peptides. The resulting box was equilibrated under NPT conditions and used as a template to solvate the folded peptides in regular truncated octahedron unit cells. All forms of Met and Ada were solvated with 226 and 347 molecules of TFE (box sizes were 37.4 Å and 42.6 Å), respectively. The minimization and preparatory MD simulations of this complete periodic system consisted of several steps, each of them presented in full detail in Methodology (Section 3.3.2).

The configuration space of each peptide was generated by combining two REMD simulations performed in Amber 10.[37] The starting conformations of the second run were chosen to be as different as possible from that of the first run. In each run, sixteen replicas of the fully equilibrated system were set at different temperatures (T = 275, 283, 291, 299, 307, 316, 326, 335, 345, 355, 365, 375, 386, 397, 408, 420 K). The exchanges were attempted every 500 steps. In between, all replicas were subject to Langevin dynamics in the NVT ensemble with a 2 fs time step and a collision frequency of 1 ps^{-1}. The 16 replicas were propagated for a total of 176 ns (11 ns per replica). The coordinates of the entire system were saved every 0.5 ps. The first 2.5 ns were omitted from the analysis. Consequently, 17000 structures from replicas at T=299 K were extracted from each REMD run, giving rise to a total of 34000 conformations comprising the conformational space of each of the six peptides.

3.5.4. Conformational Phase Space of Peptides

In order to explore the conformational phase spaces in a comparative fashion, all structures from all three forms of both peptides, associated with a temperature of 299 K, were combined into a s ingle set of structures. This set was used to determine a common set of principal components, based on the positions of α-carbon atoms (see Section 3.3.4). Each structure was then projected onto the first two principal components, which upon visual inspection, could be associated with the end-to-end distance and the twist of the backbone, respectively. In this way, 2-D plots, where each conformation is represented by a single data point, were constructed for each peptide simulated. The associated free energy surfaces were then calculated from the local density of points in this representation. In addition, the trajectories of each peptide were clustered, also on the basis of their *α*-carbon positions, so as to typically produce five clusters. Exceptionally, if such a procedure would yield substantial overlaps in the space of the first two principal components, only four clusters were constructed.

3.5.4.1. Zwitterionic Forms

The results of the described procedure for the zwitterionic peptides confirm the somewhat intuitive expectation that Met_Z has a less structured phase space than either *R*-Ada_Z or *S*-Ada_Z (Figure 3-18). Indeed, one can observe that an entire section of the phase space is inaccessible to the two Ada_Z variants. This is presumably because the Aaa group cannot adjust favorably to the semi-open conformation of the backbone, typical for cluster III of Met_Z (Figure 3-19). The more surprising

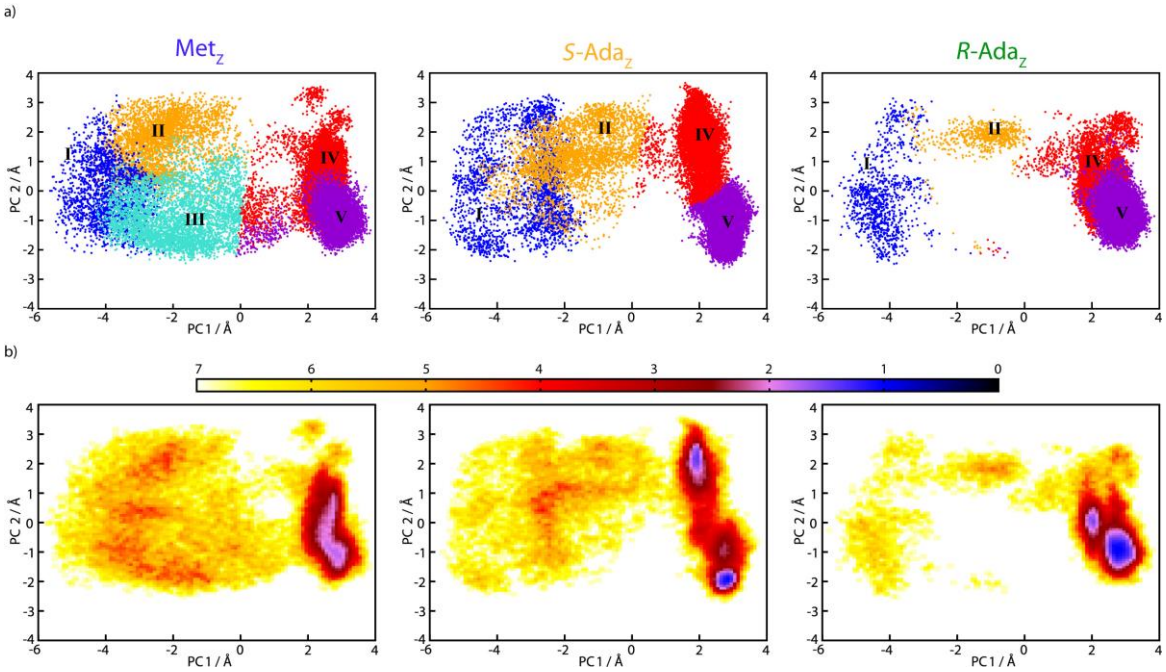


Figure 3-18. a) Conformational phase space of Met$_Z$, *S*-Ada$_Z$, and *R*-Ada$_Z$ respectively at T = 299 K, after clustering. Each cluster is represented by Roman numeral. b) Free energy surfaces for Met$_Z$, *S*-Ada$_Z$, and *R*-Ada$_Z$ respectively. The free energy surface is given by $W/kT = -\ln(N/N_{tot})$ where N_{tot} represents the total number of configurations of each individual peptide (N_{tot} = 34000).[10]

difference, however, is that observed between the phase spaces of *S*-Ada$_Z$, and *R*-Ada$_Z$, with the latter being considerably more constrained. Interestingly, this effect emerges from differences in how the direct Coulomb interaction between the charged ends of the peptides, which is not screened efficiently by the low-polarity solvent, competes with the twisting of the backbone. The former phenomenon seems to have a dominant effect on the conformational phase space of all three zwitterionic forms. Consequently, the completely open structures, typical for clusters I and II, are scarce in comparison to folded (U-shaped) conformations (clusters IV and V) that together comprise about 70%, 80% and 95% of all conformations of Met$_Z$, *S*-Ada$_Z$ and *R*-Ada$_Z$, respectively (see Figure 3-19 for details on cluster populations and representative structures).

From the point of view of the 2-D representations of the free energy, all folded structures appear at values of PC1>1, where the clustering algorithm found two distinct clusters (IV and V). Indeed, two separate deep minima are observed for both epimers of Ada$_Z$, albeit at different positions in the PC2 coordinate (Figure 3-18b). In the case of Met$_Z$, the two minima seem to merge into a single, considerably shallower and wider minimum. Closer inspection of the representative structures of clusters IV and V (Figures 3-20 and 3-21, respectively) shows that structures in cluster IV are dominated by the electrostatically driven formation of a salt bridge between the N- and C-termini of

Peptide	Cluster I	Cluster II	Cluster III	Cluster IV	Cluster V
Met_Z	5.7	9.7	12.5	41.7	30.4
S-Ada_Z	7.5	10.0	-	41.9	38.8
R-Ada_Z	2.9	2.0	-	28.5	66.6

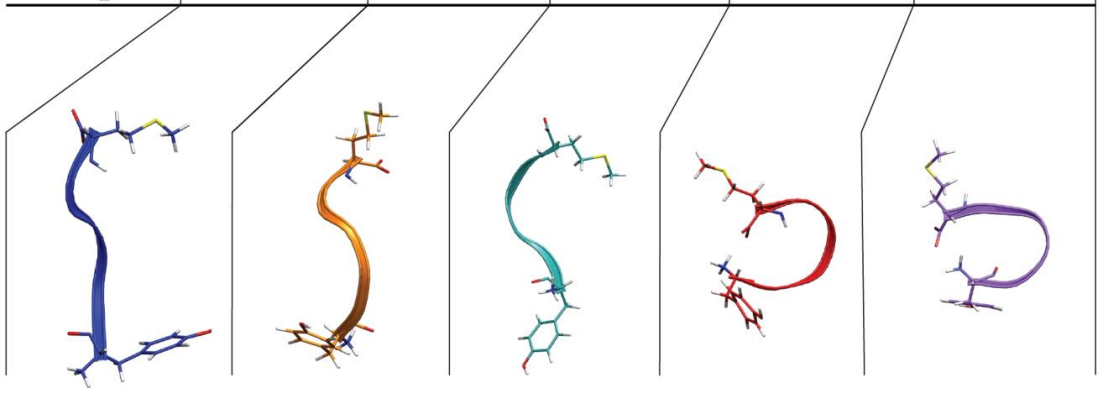


Figure 3-19. Cluster population percentages of Met_Z, *S*-Ada_Z and *R*-Ada_Z. Ribbon renderings of five representative Met_Z conformations (conformations most similar to the centroids of their respective clusters) are shown.[10]

all three peptides. In the case of cluster V, the dominance of this bridge is somewhat reduced and all three representative structures have a β-turn hydrogen bond between Phe4 and Tyr1.

Further analysis of the representative structures from cluster IV (Figure 3-20) shows that here, both Met_Z and *R*-Ada_Z have U shaped backbones, with the upper and lower parts of the backbones well aligned. In the case of Met_Z, the structure is maintained by two hydrogen bonds, out of which one, present in 17% of conformations, is between Met5 and Gly2 and induces β-turns of the type II'. Due to the lack of the backbone twist, the centroids of cluster IV of Met_Z and *R*-Ada_Z have roughly the same projections on PC2. However, as a consequence of the two hydrogen bonds at the C-terminus,

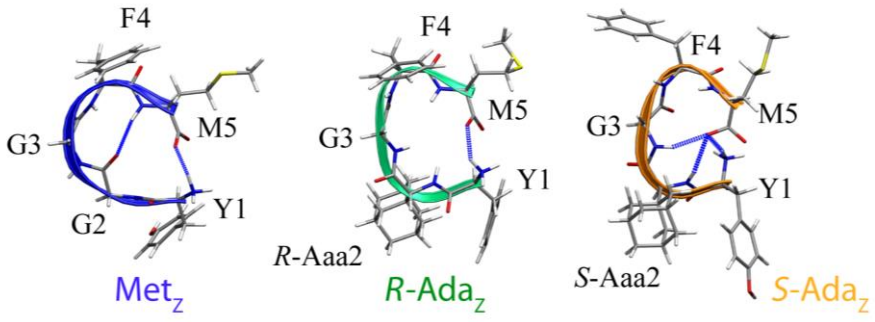


Figure 3-20. Representative structures of cluster IV for all three peptides. Apart from the salt bridge between the N-terminus of Tyr1 and the C-terminus of Met5, characteristic for the dominant structures of all three peptides, Met_Z possesses a β-turn hydrogen bond between Met5 and Gly2. S-AdaZ shows two hydrogen bonds formed by the interaction of C-terminus of Met5 with Gly3 and S-Aaa2.[10]

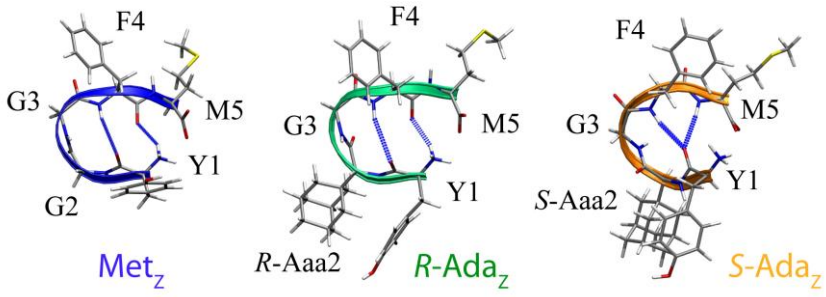


Figure 3-21. Representative structures of cluster V for all three peptides. Dominant structures of Met$_Z$ and *R*-Ada$_Z$ have a strong hydrogen bond between N-terminus of Tyr1 and Phe4 as well as a β-turn hydrogen bond between Phe4 and Tyr1. This same hydrogen bond exists also in the representative structure of *S*-Ada$_Z$ but is bifurcated with a second bond from Met5 to Tyr1.[10]

the representative structure of *S*-Ada$_Z$ has a more twisted backbone. Consistently, the free energy minimum and the position of the centroid in the conformational phase space are shifted to considerably larger PC2 values.

The clusters V can be described, for Met$_Z$ and *R*-Ada$_Z$, as nearly circular islands occupying practically identical regions of the conformational phase space. They are both occupied with β-turns of the type II' with about 50% of both Met$_Z$ and *R*-Ada$_Z$ structures possessing the identical β- turn hydrogen bond between Phe4 and Tyr1, and a very similar backbone appearance (Figure 3-21). On the other hand, *S*-Ada$_Z$ possesses a bifurcated hydrogen bond of the $i + 4 \rightarrow i$ type, characteristic for α-helical structures, placing the free energy minimum lower in terms of its PC2 projection. However, the strong electrostatic interaction between the charged end groups prevents the formation of a pure α-helix structure. As a result, the conformations of *S*-Ada$_Z$ form mainly structures that resemble β-turns of type II.

3.5.4.2. Neutral Forms

The lack of the charged end groups in the neutral forms of the peptides results in more evenly populated conformational spaces. In addition, some new parts of the phase space, that are completely unoccupied by the zwitterionic forms, become apparent. In the case of Met$_N$, all five clusters have similar populations (Table 3-1) and occupy areas of similar sizes in the phase space (Figure 3-22). Consequently, the free energy plot of Met$_N$ contains a number of connected shallow minima. One of these can be associated with the sub-cluster Va in which helical structures of 3_{10} type are most common (Figure 3-23a).

S-Ada$_N$ and *R*-Ada$_N$ are also found to visit virtually the entire phase space. However, in both cases, more pronounced minima in the free energy surface can be observed. Specifically, about 75% of *S*-Ada$_N$ conformations are extended, which provides a free energy centered around cluster I.

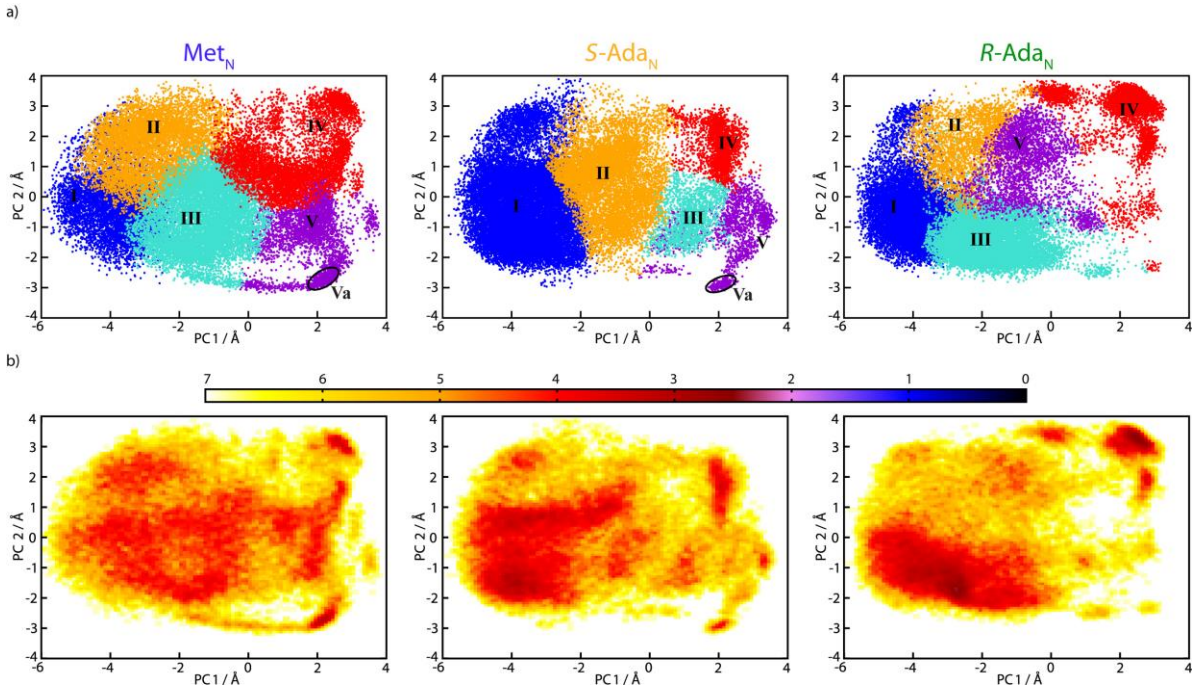


Figure 3-22. a) Clustered conformational phase space spanned by 34000 conformations and b) the respective free energy plots of Met_N (left), *S*-Ada_N (middle), and *R*-Ada_N (right) at $T = 299$ K. Each cluster is represented by a Roman numeral.[10]

Moreover, like in Met_N, a minimum can be associated with the sub-cluster Va for *S*-Ada_N. Due to different hydrogen bonding patterns (Figure 3-23b), true α-helices can be found for the latter peptide. In addition to these features, inspection of the middle panel in Figure 3-22 shows a wide minimum at a PC2 value of about 1, stretching over all negative values of PC1, as well as a minimum with folded structures close to the centroid of cluster IV (Figure 3-23c). Both of these minima are significant for *S*-Ada_N but are occupied to a much lesser extent for Met_N and *R*-Ada_N.

The most dominant feature of the *R*-Ada_N conformational space is a broad minimum at low values of both PC1 and PC2, to which extended and semi-folded structures (typical for clusters I and III, respectively) contribute. Indeed, these two clusters account for nearly 70% of the distribution for *R*-Ada_N. Another interesting part of the phase space is the far upper right-hand corner of the PC plot. For both *R*-Ada_N and Met_N, this region, which belongs to cluster IV, is populated by quasi left helices (Figure 3-23c). In the case of *S*-Ada_N, however, the formation of a proper left-helix is prevented by the

Table 3-1. Population percentages for the clusters of the neutral structures.[10]

Peptide	Cluster I	Cluster II	Cluster III	Cluster IV	Cluster V
Met_N	17.8	22.0	28.7	19.0	12.5
***S*-Ada_N**	57.0	25.9	5.1	6.8	5.4
***R*-Ada_N**	33.9	7.7	34.9	14.9	8.8

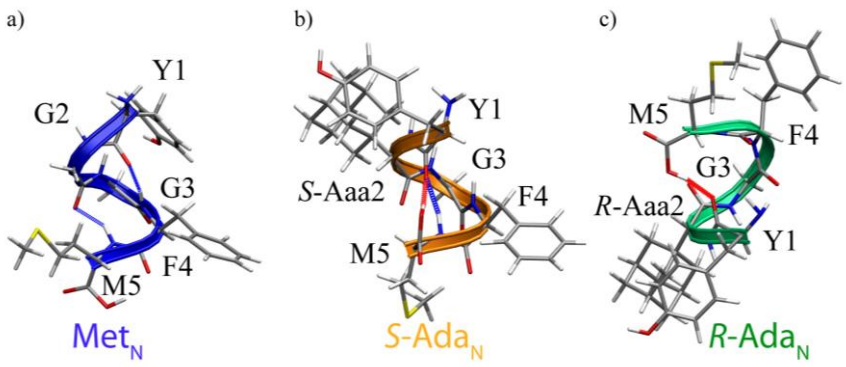


Figure 3-23. a) An example of 3_{10} helical structure of Met_N, b) α-helical structure of *S*-Ada_N, and c) A conformation of cluster IV of *R*-Ada_N that is an example of quasi left helical structures, which can be found in the upper right part of the phase space of *R*-Ada_N, a) and b) are both from the cluster Va. The Met_N conformation possesses a hydrogen bond between Phe4 and Tyr1, and between Met5 and Gly3. The *S*-Ada_N structure possesses a bifurcated hydrogen bond from Tyr1 to Met5, and to the C-terminal. The first bond is an α-helix bond whereas the second further stabilizes the conformation.[10]

S-Aaa residue and this particular section of the phase space is completely inaccessible. Similarly, *R*-Ada_N does not exhibit a significant population in the vicinity of the Va region. Thereby, the *R*-adamantyl residue apparently retards the formation of both α- and 3_{10}-helices.

3.5.4.3. Conformational Analysis Summary

Our simulations show that in TFE, the zwitterionic peptides show a much larger propensity for folded structures than the neutral forms, which, in turn, exhibit a considerably higher proportion of extended structures. This appears to be the result of the dominant electrostatic force acting between the charged end groups in the zwitterionic forms. Since such a force is absent in the neutral forms, they are able to occupy the phase space in a less restricted manner. This result contrasts with the situation in water, where the polar solvent successfully screens the charged termini. As a result, in water, Met_Z is equivalently abundant in all accessible parts of the conformational phase space, in a similar way as Met_N is in TFE.[70] The comparison between the various peptides in TFE shows a trend in the extent of ordering such that Met < *S*-Ada < *R*-Ada, irrespective of the nature of the terminal groups (Figures 3-18 and 3-22). The surprising realization is that the two epimers of Ada actually have very different phase spaces. However, even though *R*-Ada phase spaces are more restricted than those of *S*-Ada (that are superficially more similar to Met), we have shown that R-Ada and Met actually have a much larger overlap of representative structures (Figures 3-20 and 3-21). The representative conformations for *S*-Ada, in both the neutral and zwitterionic forms, consistently exhibit a different hydrogen bonding pattern than those of the other two peptides, which is also evident in the contrasting CD spectra. These differences may be related to the enhanced capacity of *S*-Ada to penetrate cellular membranes and act as an anti-tumor agent.

3.5.5. CD Spectra of Zwitterionic and Neutral Forms

Here we present the CD spectra of all of the forms of Met and Ada-enkephalin studied in this thesis. We thus again employ our strategy, thereby calculating CD spectra of each of the 50 representative structures of each form by taking into account the solvent effect using our implementation of the ASEC procedure (B3LYP/6-31G(d) level of quantum theory, setting A = 100 and σ = 0.135 eV, see Sections 3.2 and Section 3.3.5.5). Then, 50 single CD spectra were combined to give the average CD spectrum of each form by a weighted averaging procedure (see Methodology for deeper elaboration, Section 3.3.5.2 and 3.3.5.3). Figure 3-24a shows that the CD spectra of the zwitterionic peptides all exhibit negative minima. Specifically, Met_Z possesses a deep and wide minimum between 190 nm and 210 nm, followed by a positive peak at about 220 nm. This latter signal is associated with the folded conformations residing in the cluster V, which dominates the spectrum. We find, rather interestingly, that *R*-Ada_Z has a spectrum with very similar peak positions as Met_Z, but of different intensities (Figure 3-24a). However, we also observe that the two peptides exhibit rather similar conformational phase spaces. We thus wanted to test the following hypothesis: do the similar phase spaces imply similar CD spectra? To investigate this hypothesis we identified the cluster V of Met_Z that

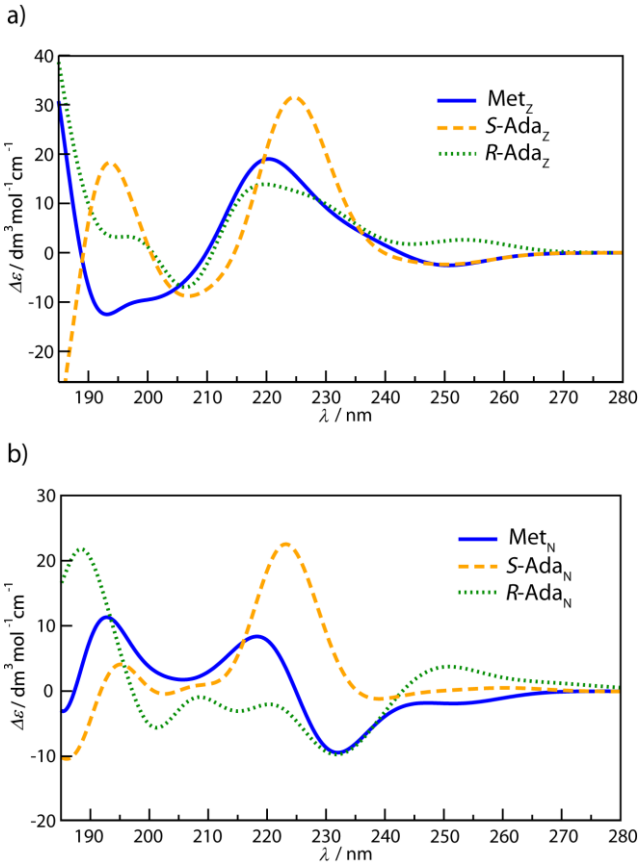


Figure 3-24. CD spectra of zwitterionic and neutral forms of pentapeptides are shown in the upper and lower panel, respectively.[10]

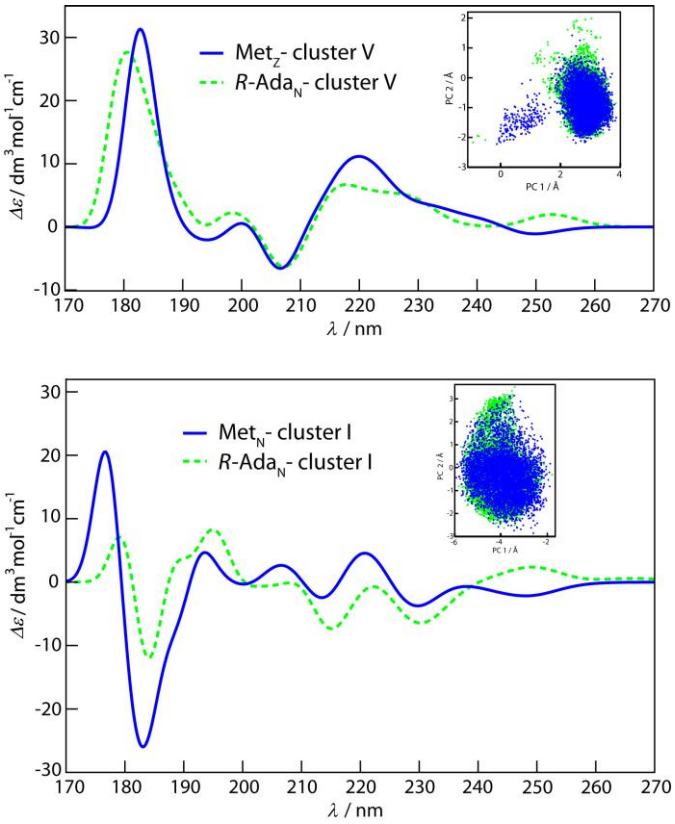


Figure 3-25. Average CD spectra corresponding to clusters V of *R*-Ada$_Z$ and Met$_Z$ and clusters I of Met$_N$ and *R*-Ada$_N$ are shown in the upper and lower panel, respectively. The overlay of the clusters in the phase space is plotted in the insets.[10]

corresponds closely to the cluster V in *R*-Ada$_Z$, both in terms of the positions of the centroids and in terms of very similar representative structures (Figure 3-21). To obtain spectra of comparable intensity, both spectra are normalized by the total number of conformations in the respective clusters. The results of our inspection are shown in the upper panel of Figure 3-25, where the striking similarity between the average CD spectra of the clusters V of Met$_Z$ and *R*-Ada$_Z$ is apparent.

Since the cluster V is very compact and well defined, we have extended our testing to a pair of similar, but more dilute and spread clusters, which could be expected to be considerably more challenging for our method. We thus repeated the same procedure for the clusters I of Met$_N$ and *R*-Ada$_N$ (lower panel in Figure 3-25). Again, a remarkable agreement of the two spectra was obtained.

Apart from confirming the reliability of our approach, these results show that even though structures with very similar backbone shape can have very different spectra due to the side-chain conformations, this effect can be effectively averaged (Figure 3-25) in much the same way as the solvent can (Figures 3-7 and 3-8). Namely, the sub-clustering on the basis of all atoms can be regarded as sampling the side-chain conformations of structures that belong to a cluster with a well-defined backbone structure as determined by initial clustering over only the C_α atoms. This hierarchical sampling is very clear when the phase space is condensed (cluster V). However, in the sparse phase

space, the backbone and side-chain sampling may be somewhat overlapping, without truly affecting the result (cluster I), if a sufficiently large set of representing structures is considered.

We can now turn to the spectra of the neutral species. In this respect, we find that the spectrum of Met_N (Figure 3-24b) is very different to the spectrum of Met_Z. It remains positive for transitions between 185 and 225 nm, with peaks at 193 and 220 nm. The spectrum of *R*-Ada_N becomes mostly negative after the peak at 185 nm. Interestingly, *S*-Ada shows a very pronounced peak at 225 nm, in both its zwitterionic and neutral forms. Also, both the *S*-Ada CD spectra have a peak at 195 nm, although of different intensities. The only difference between the two is the negative minimum in the spectrum of *S*-Ada_Z (208 nm) that is not apparent in the spectrum of *S*-Ada_N.

Because the final spectra are the result of averaging 50 individual spectra, which are quite distinct, it is generally difficult to assign the spectral features in Figure 3-24 to any specific electronic transitions. Nevertheless, we have examined the underlying transitions, and were thereby able to comment, for example, on some aspects of the π-π^* states of the aromatic chromophores. The ^1L_b states of Tyr and Phe are calculated to appear at approximately 250 and 233 nm, respectively. These values are some 25 nm lower than the corresponding experimental values.[71,72] The ^1L_a states of the same sidechains are calculated to lie at approximately 213 and 208 nm, respectively. Both wavelengths are within 10 nm of their corresponding experimental values.[72] Even though these states can carry significant rotatory strengths at the level of an individual structure, we found that the average rotatory strengths across the entire phase space are close to zero. In this respect, such states can be expected to have only a very small net effect on the final CD spectra. This result, which is particularly apparent in the wavelength region around 250 nm, highlights the importance of averaging over a large number of individual structures.

3.5.6. Comparison of Theoretical and Experimental Spectra

In order to compare the calculated CD spectra of the individual peptides to those measured for Met and Ada in TFE,[59] two further steps are required. The first concerns the estimation of the equilibrium ratio of the neutral to zwitterionic forms (N:Z) of the individual peptides in TFE.

We base our calculations of the N:Z ratio in TFE on the measurements of the analogous ratio for Leu-enkephalin in dimethyl sulfoxide (DMSO), which was experimentally determined to be 3:2.[73] Our own simulations of Leu-enkephalin in TFE indicate that its conformational behavior is practically identical to that of Met, for both the neutral and zwitterionic forms (Figure 3-26). On this basis, we believe that 3:2 is also a good approximation of the N:Z ratio of Met in DMSO. Using this as a starting point, we have elected to estimate the N:Z ratio of Met in TFE by means of a free energy cycle describing the transfer from one solvent to the other (Scheme 3-2).

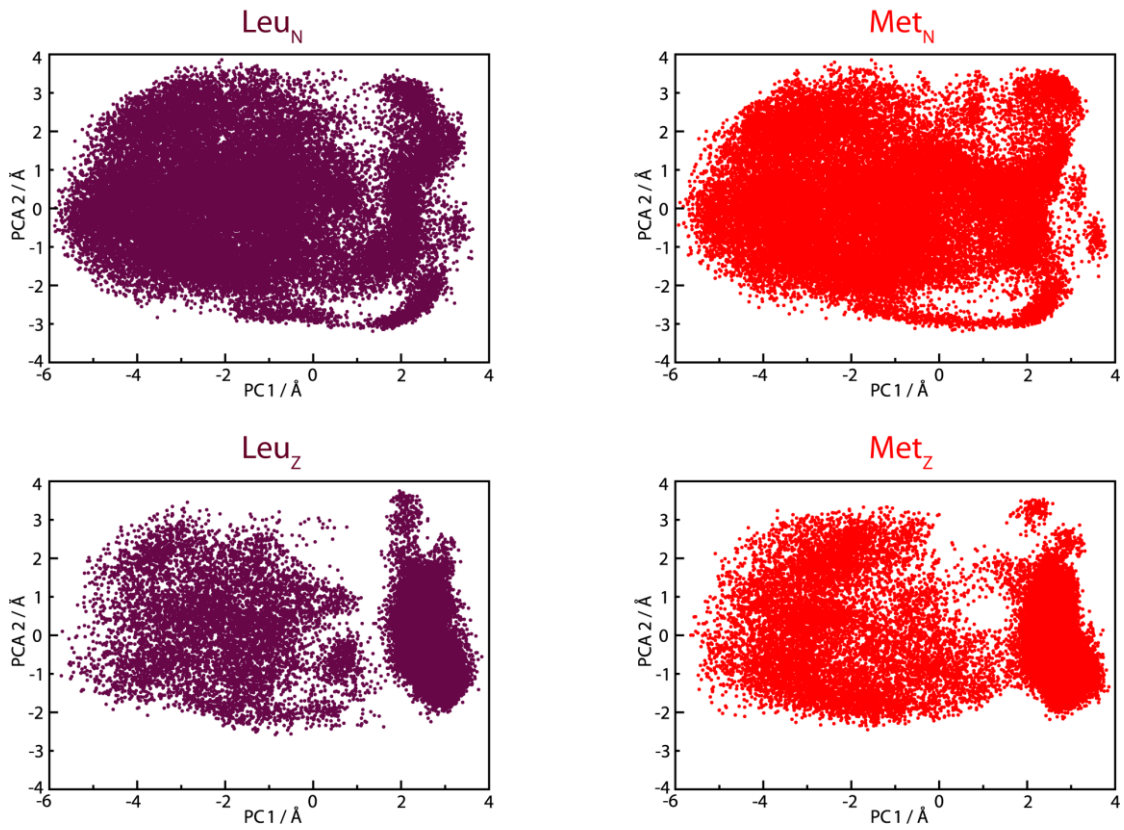
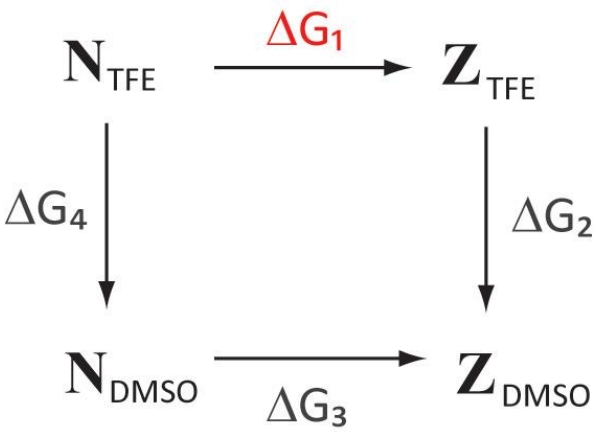


Figure 3-26. Conformational phase spaces of Leu_N ($Tyr_0GlyGlyPheLeu_0$) and Leu_Z (TyrGlyGlyPheLeu) compared with the conformational phase spaces of Met_N and Met_Z, respectively. The parameters of the neutral C-terminal amino acid residue Leu_0 were obtained in a similar fashion to the parameters of Met_0 and Tyr_0 residues (see Methods). Both Leu_N and Leu_Z were solvated with 226 molecules of TFE in a regular truncated octahedron unit cells, analogously to the Met_N and Met_Z in TFE. The procedure described in Section 3.3 (Methodology) was used in order to obtain 34000 conformations of both Leu_N and Leu_Z. To obtain the presented phase spaces, conformations of Leu_N and Leu_Z are projected onto first collective principle component basis vectors used in Figures 3-18 and 3-22.[10]

The desired free energy difference (for $[N:Z]_{TFE}$) emerges as $\Delta G_1 = \Delta G_3 + \Delta G_4 - \Delta G_2$, where ΔG_3 is assigned as 1 kJ mol^{-1}, based on $[N:Z]_{DMSO}$=3:2. We have estimated the (vertical) transfer free energies (ΔG_2 and ΔG_4), via the Zwanzig free energy perturbation scheme 3-2, where the following holds:

$$\begin{aligned}\Delta G\left(Met_{TFE} \rightarrow Met_{DMSO}\right) &= -RT\ln\left\langle \exp\left(\frac{-\left(E_{DMSO}-E_{TFE}\right)}{RT}\right)\right\rangle_{TFE} \\ &= -RT\ln\sum_{i=1}^{50} w_i \exp\left(\frac{-\left(E_{i,DMSO}-E_{i,TFE}\right)}{RT}\right)\end{aligned} \qquad (3.5)$$

More specifically, we have evaluated the requisite differences in the solvation free energies by solving the linearized Poisson-Boltzmann equation within the Adaptive Poisson-Boltzmann Solver (APBS) program package.[74] Therein, the peptides were assigned the same *ff03*-derived parameters as described



Scheme 3-2. The free energy cycle used to estimate the equilibrium ratio between the neutral (N) and zwitterionic (Z) forms in TFE (ΔG_1) based on the experimentally obtained ratio (ΔG_3) for the analogous equilibrium for Leu-enkephalin in DMSO. See text.[10]

in the Methodology (Section 3.3.1). TFE and DMSO were represented as continuous media with dielectric constants of 26.8 and 47.2, respectively, and the non-polar solvation contributions were assumed to be equal in the two solvents. The individual calculations were performed for the same 50 representative structures used to obtain the CD spectra for each peptide. Similarly, we used the previously determined population fractions of the corresponding sub-clusters to represent the statistical weights (w_i) in the above expression.

Even though we expect this perturbation scheme to be most accurate for Met-enkephalin, we also performed the same calculations for both epimers of Ada-enkephalin (with ΔG_3 = 1 kJ mol^{-1}). For all three peptides, we find that the zwitterionic forms are present in fractions of 8-18% in TFE. Considering the uncertainties associated with this calculation, we estimate the fraction for all peptides to be very similar and equal to about 10% ($[N:Z]_{TFE}$=9:1). This is qualitatively consistent with the expectation that the zwitterionic form should be less favored in the less polar solvent. Consequently, we have used this ratio for weighting the contributions of neutral and zwitterionic forms to all composite spectra. One should note that increasing the contribution of the zwitterionic form up to 20% does not significantly affect the final CD spectra (Appendix, Figure A-1).

The second aspect that needs to be taken into account before making the final comparison between the published experimental and theoretical spectra is the fact that both epimers of Ada were present (in an equimolar ratio) under the experimental conditions. Consequently, the theoretical spectra were combined according to the expressions: Met=0.9Met_N+0.1Met_Z, and Ada=0.9(*S*-Ada_N+*R*-Ada_N)/2+ 0.1(*S*-Ada_Z+*R*-Ada_Z)/2, for the purposes of comparison to experiment.

The resulting final spectra are shown in Figure 3-27a and compared to the experimental spectra[59] in Figure 3-27b. The comparison immediately reveals that a number of features of the experimental spectra have been correctly reproduced. In the range of 185 to 230 nm, both the

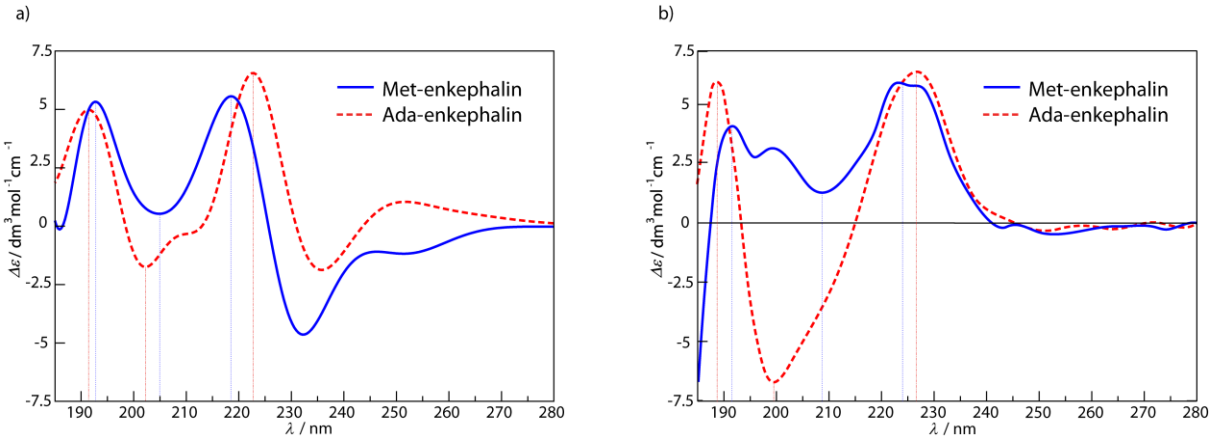


Figure 3-27. a) Calculated CD spectra for Met-enkephalin and Ada-enkephalin. b) Experimentally obtained CD spectra of Met-enkephalin and an epimeric mixture of Ada-enkephalin in TFE.[10]

experimental and the theoretical CD spectra of Ada-enkephalin have a pronounced maximum at ~190 nm, followed by a wide minimum between 200-215 nm. The strong peak at ~224 nm obtained from calculations corresponds to the peak present at 227 nm in the experimental CD spectrum. Importantly, the analysis of the components contributing to the spectra shows that the minimum at 200 nm does not appear because of enhanced contribution of strongly folded structures of Ada, as was suggested previously. Instead, we find that the strongest contribution at these wavelengths comes from *R*-Ada$_N$, which has a relatively disperse phase space.

A qualitatively good agreement was also obtained in the case of Met-enkephalin, where the correspondence between the theoretical and the experimental spectra is found in both peaks at ~192 nm and ~220 nm. Likewise, between the peaks, the spectra remain positive with a minimum appearing at 205 nm theoretically and 208 nm experimentally.

Between 230 and 240 nm, a relatively strong minimum appears in our calculations, even though the experimental spectra have no such particular features in this range. Taking into account the individual contributions of the forms, it is possible to see that it predominantly arises from the neutral forms of Met and *R*-Ada. It was not, however, possible to identify characteristic peptide structures responsible for this signal as it was present in nearly all spectra of the relevant representative structures. A closer investigation of the excited states, however, has allowed us to pinpoint its origin.

We were able to identify an electronic transition of Met$_N$ consistently centered around 230 nm. Unlike the π-π* states discussed above, the rotatory strengths of this state did not average to zero over the conformational ensemble. Rather, the individual rotatory strengths are consistently negative and the result is the anomalous minimum present in the final calculated spectra. It is important to note that this is the only feature of the final spectra that we are able to unambiguously relate to a specific excited state in this way.

The state involves excitations originating from various occupied orbitals into the π* orbital of the terminal COOH group. Among these are local promotions from the non-bonding orbital of the terminal carbonyl oxygen. Such n-π* states tend to appear around 205 nm in simple carboxylic acids.[75] In the present case, the mixing of less local contributions from the occupied π orbitals of adjacent amide bonds and the Phe side-chain, as well as non-bonding orbitals from nearby backbone oxygens, seems to lower the transition energy. In this respect, we hypothesized that the appearance of the erroneous minimum/transition at the terminal COOH group is connected to the use of TD-DFT, as implementations of TD-DFT may be associated with somewhat unpredictable precision,[76] particularly for larger systems.[22] In particular, we conjectured that the use of B3LYP functional was to be blamed, as the observed artifact seemed like a classic case of B3LYP artificially stabilizing so-called charge-transfer excitations, producing the feature that does not appear in the experiment.

3.5.7. To B3LYP or not to B3LYP, the Question Is Now

Due to the fact that both theoretically obtained CD spectra reproduce experimental spectra below 230 nm very well (see Figure 3-27), and as the only obstacle from obtaining an excellent agreement are the features/minima observed at approximately 230 nm, we were strongly motivated to investigate this issue in depth, having in mind that the choice of the electronic structure method is clearly an important component in our methodology, and moreover, that this investigation promises to generally deepen and expand our understanding of the excited states and their underlying role in the CD spectra. We thus decided to investigate the behavior of selected GHFs (B3LYP and PBE0), LCFs (CAM-B3LYP and ωB97X-D) and the HMF M06-2X for calculating CD spectra of peptides (this choice is elaborated in detail in Theoretical Background, Section 2.1.6), with Met-enkephalin serving as our modeling system.[10] To properly test the performance of the aforementioned TD-DFT functionals we benchmarked them both against the experiment[59] and against the stronger and more expensive *ab initio* Resolution-of-Identity Second-Order Approximate Coupled-Cluster (RICC2) method (see Theoretical background for a deeper discussion, Sections 2.1.6 and 2.1.7), which, although not flawless with respect to excited state predictions,[76-79] it should not exhibit DFT-related artifacts, thereby being able to provide benchmark quality results in this respect. For the purpose of the latter we modified Met_N

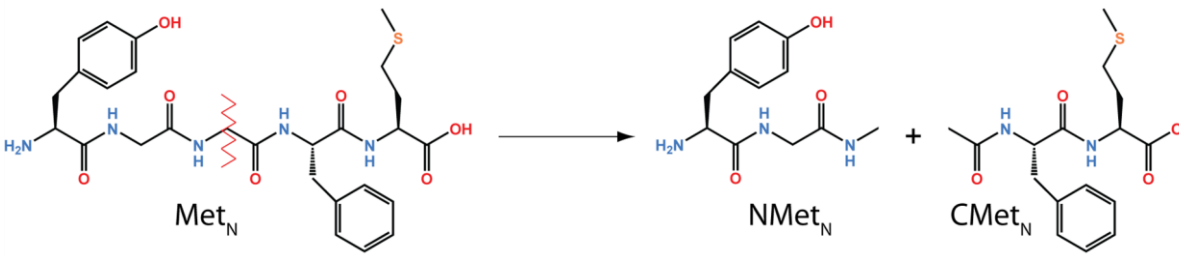


Figure 3-28. Neutral Met-enkephalin (Met_N) is shown on the left. Two neutral model peptides $NMet_N$ and $CMet_N$ are prepared by cutting along either the C_α(Gly3)-C_O(Gly3) or the C_α(Gly3)-N(Gly3) bond.[11]

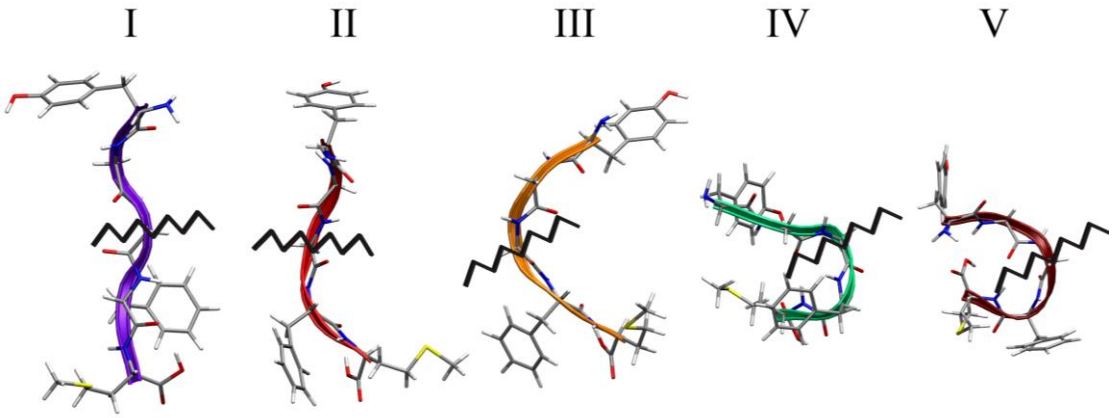


Figure 3-29. Five representative Met_N conformations.[11]

and Met_Z, thereby obtaining smaller model peptide systems for which the RICC2 calculations were feasible.

3.5.7.1. Preparation of the Model Systems

The neutral model structures (Figure 3-29) were prepared from five Met_N conformations extracted from five clusters (I-V) of the conformational phase space (Figure 3-22). Each of the structures (Figure 3-29) was cut along the C_α(Gly3)-C_O(Gly3) bond or the C_α(Gly3)-N(Gly3) bond (shown in Figure 3-27). Capping the vacant positions with hydrogen atoms gives rise to the N-terminal model $NMet_N$ ($Tyr_0GlyNme$), and the C-terminal model $CMet_N$ ($AcePheMet_0$). Zwitterionic models (Figure 3-29) of the C-terminus ($CMet_Z$) were prepared by cutting and capping five zwitterionic Met-enkephalin structures (Met_Z, TyrGlyGlyPheMet), along the C_O(Gly2)-N(Gly3) bond. The original five structures were obtained from the conformational phase space of Met_Z (Figure 3-18). Using parameters from the *ff*03[33] force field, supplemented with those previously developed for Tyr_0 and Met_0,[10] the CH_3 groups of the N-methyl (Nme in $NMet_N$) and acetyl (Ace in $CMet_N$) residues, as well as the NH_3^+ group of $CMet_Z$, were subjected to steepest descent minimization, while leaving the rest of the system frozen (Amber 10).[37] Thereby each model system was solvated with 226 molecules of TFE, using parameters from the RESP and ESP database.[35]

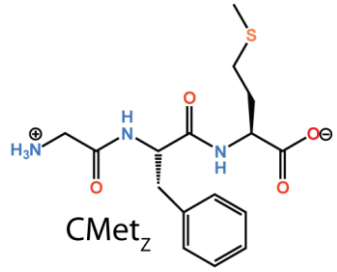


Figure 3-30. Zwitterionic model peptide structure ($CMet_Z$).[11]

3.5.7.2. Excited State Calculations

The model systems were treated quantum mechanically, while the effects of the solvent were introduced through a our implementation of the Average Solvent Electrostatic Configuration

approach,[31] by spatially superimposing scaled charges from 50 uncorrelated solvent configurations (see Figures 3-7 and 3-8). In the present applications, the solvent configurations were extracted every 10 ps from a series of 300 K NVT molecular dynamics simulations, with restrained solutes. In the first instance, excitation energies and rotatory strengths were calculated using the RICC2 implementation in TURBOMOLE V6.3.1,[80] with TZVP[81] for both the atom-centered and the auxiliary basis sets. For comparison, TD-DFT calculations were performed with Gaussian 09,[69] using the CAM-B3LYP, ωB97X-D, M06-2X, B3LYP and PBE0 functionals, in combination with the 6-31G(d) basis set. Using the calculated excitation energies and rotatory strengths, CD spectra were constructed by convoluting a series of state specific Gaussian functions of width σ = 0.135 eV, unless otherwise stated.

To properly asses the performance of the TD-DFT functionals, a state-by-state analysis is required. This analysis is presented in Tables 3-2 and 3-3 for $CMet_N$ and $NMet_N$, wherein the exited states are classified according to the dominant excitations in a given transition, using RICC2 as a reference. The wavelengths (Table 3-2) and rotatory strengths (Table 3-3) are obtained as averages over five $CMet_N$ and five $NMet_N$ structures, respectively. The standard deviations reflect the variation of the excitation energies and rotatory strengths, of a given transition, as a function of geometry. Any surplus states found in the TD-DFT calculations have been omitted from this analysis.

3.5.7.3. RICC2 Calculations

This method, chosen as the benchmark, predicts an average of seven excited states above 185 nm in the case of $CMet_N$, with four such states being found for $NMet_N$. The said states are organized in Tables 3-2 and 3-3, according to their characteristic average wavelength, in decreasing order (depicted in Appendix, Table A-1). As can be seen from the standard deviations, the effects of geometry on the excitation energies are relatively small but may be more significant for the rotatory strengths.

A detailed inspection of the rotatory strengths in Table 3-3 reveals that, for $CMet_N$, there are three important transitions. The same transitions dominate the average CD spectrum, which is shown in Figure 3-31a. The most prominent transition is a C-terminal excitation at 220 nm (n_O(COOH) + π(F-M) → π*(COOH)), which exhibits a strongly negative rotatory strength and thus results in a deep minimum in the CD spectrum. A small maximum, which originates from the n_O(F) → π*(F-M) amide transition, is evident at 210 nm. A second transition involving the C-terminal carboxylic group (188 nm, n_S(M) → π*(COOH)) is also associated with a negative average rotatory strength, which is manifested in the spectrum as a small minimum at low wavelengths. This transition is associated with a non-local (CT) character and is subject to a relatively large variation in wavelength in response to changes in geometry, as evidenced by the corresponding standard deviation (especially with TD-DFT). At even shorter wavelengths, RICC2 identifies additional states with CT character, out of which the π (F) → π*(COOH) transition at 180 nm has a relatively significant, positive average rotatory strength.

Table 3-2. Average wavelengths together with their standard deviations for excited states in $NMet_N$ and $CMet_N$ (lying above 185 nm) are found from the values obtained for five model structures, respectively. Colors are used for clarity. The states are ordered according to the average wavelength obtained by RICC2.[11]

Excited state[a]	Wavelength / nm					
	RICC2	CAM-B3LYP	ωB97X-D	M06-2X	B3LYP	PBE0
Tyrosine ^1L_b	259.2 ± 1.6	242.2 ± 1.4	242.4 ± 1.4	238.9 ± 1.4	250.1 ± 1.5	245.9 ± 1.
Phenylalanine ^1L_b	243.4 ± 0.6	227.8 ± 0.8	227.9 ± 0.7	225.5 ± 0.8	233.5 ± 0.7	229.4 ± 0.
n_O(COOH) + π(F-M) → π*(COOH)	219.1 ± 0.6	223.3 ± 0.9	222.9 ± 0.9	233.0 ± 0.8	228.3 ± 1.6	225.5 ± 0.
n_S(M) → n_S(M)*	213.0 ± 1.9	216.7 ± 2.3	215.6 ± 2.3	224.9 ± 2.3	222.2 ± 3.1	217.9 ± 3.
n_O(Y) → π*(Y-G)	212.5 ± 1.3	214.7 ± 1.3	214.3 ± 1.4	221.6 ± 1.1	220.4 ± 1.2	213.3 ± 4.
n_O(F) → π*(F-M)	210.8 ± 3.7	210.8 ± 3.8	210.7 ± 3.5	217.6 ± 4.0	221.1 ± 3.5	213.3 ± 2.
n_O(G) → π*(G-Nme)	205.8 ± 2.4	209.3 ± 2.3	208.7 ± 2.4	216.2 ± 2.2	213.7 ± 3.3	210.6 ± 2.
n_O(Ace) → π*(Ace-F)	205.4 ± 2.1	208.3 ± 1.9	207.7 ± 1.9	214.8 ± 2.2	212.8 ± 2.0	209.7 ± 2.
Tyrosine ^1L_a	203.1 ± 0.8	206.6 ± 0.8	206.0 ± 0.7	201.7 ± 0.8	212.1 ± 0.5	212.1 ± 2.
Phenylalanine ^1L_a	193.3 ± 0.5	201.2 ± 0.7	200.7 ± 0.7	195.6 ± 0.9	208.8 ± 1.1	204.6 ± 1.
n_S(M) → π*(COOH)	188.8 ± 2.1	191.5 ± 7.6	189.2 ± 7.3	193.0 ± 7.2	259.5 ± 12.4	244.3 ± 10

[a]The single capital letters are used to denote the amino acids. The sequence of $CMet_N$ and $NMet_N$ in this notation would thus be written as Ace-F-M and Y-G-Nme, respectively. Aromatic excitations are denoted using the standard nomenclature. Orbitals of the π type of the amide bond between Phe and Met are denoted as π(F-M). Lone pairs are denoted as n with their elemental symbol subscripted. Thus n_O(Ace) refers to a non-bonding orbital on the carbonyl oxygen of acetyl residue, n_S(M) is a non-bonding orbital on the sulfur atom of the Met side chain, while n_O(COOH) is a non-bonding orbital on the carbonyl oxygen of the C-terminal carboxylic acid group. Antibonding orbitals are denoted with a * in the standard manner (the exception is n_S(M)*, which denotes an unoccupied orbital). The principle transitions are shown graphically in Appendix (Table A-1).

Table 3-3. Rotatory strengths together with their standard deviations for excited states in $NMet_N$ and $CMet_N$ (above 185 nm) are found from the values obtained for five model structures, respectively. Colors are used for clarity. The states are ordered according to the average wavelength obtained by RICC2.[11]

Excited state[a]	Rotatory strength / 10^{-40}erg esu cm Gauss^{-1}					
	RICC2	CAM-B3YLP	ωB97X-D	M06-2X	B3LYP	PBE0
Tyrosine ^1L_b	6.1 ± 9.0	7.9 ± 14.1	7.4 ± 13.1	8.7 ± 14.8	7.6 ± 14.6	7.8 ± 14.
Phenylalanine ^1L_b	0.1 ± 0.3	0.7 ± 0.9	0.6 ± 0.8	1.2 ± 2.0	0.2 ± 1.5	1.2 ± 1.8
n_O(COOH) + π(F-M) → π*(COOH)	-31.0 ± 7.6	-18.9 ± 6.2	-20.2 ± 6.5	-18.2 ± 5.7	-23.1 ± 7.6	-16.6 ± 9.7
n_S(M) → n_S(M)*	0.3 ± 0.9	1.2 ± 2.7	1.5 ± 2.7	1.4 ± 2.5	3.3 ± 6.3	1.7 ± 2.7
n_O(Y) → π*(Y-G)	10.5 ± 8.0	7.4 ± 7.3	7.9 ± 6.9	3.7 ± 5.9	8.3 ± 7.0	10.6 ± 3.4
n_O(F) → π*(F-M)	7.5 ± 9.5	7.3 ± 8.9	4.1 ± 10.2	6.7 ± 9.0	7.9 ± 13.9	8.4 ± 14.
n_O(G) → π*(G-Nme)	-3.2 ± 7.6	-5.8 ± 8.7	-4.7 ± 7.5	-2.9 ± 5.0	1.0 ± 3.7	-6.4 ± 10.
n_O(Ace) → π*(Ace-F)	0.6 ± 4.5	-0.4 ± 2.2	0.6 ± 3.2	0.6 ± 2.9	-1.7 ± 3.8	-2.6 ± 4.1
Tyrosine ^1L_a	1.0 ± 22.1	3.8 ± 25.3	1.4 ± 22.6	3.7 ± 28.9	1.9 ± 12.7	4.4 ± 8.6
Phenylalanine ^1L_a	3.2 ± 8.1	3.2 ± 2.5	3.7 ± 2.8	3.0 ± 4.8	3.5 ± 9.2	6.4 ± 14.
n_S(M) → π*(COOH)	-13.1 ± 20.3	-6.3 ± 8.9	-11.6 ± 14.6	-1.6 ± 4.6	-0.7 ± 1.8	-4.2 ± 6.4

Due to the broadening of the spectral lines, this transition may affect the theoretical spectra at wavelengths that are experimentally accessible (>185nm).

In the case of $NMet_N$, there are two dominant transitions that result in two positive peaks in the CD spectrum (Figure 3-31b). These peaks are associated with an amide transition at 212 nm (n_O(Y) → π*(Y-G)) and the aromatic Tyrosine L_b transition at 259 nm. The latter also exhibits the highest wavelength observed in all the studied model systems.

Experimentally, the Tyrosine L_b transition occurs at around 275 nm,[71,72] meaning that RICC2 underestimates its wavelength by some 15 nm. Similar deviations are observed for the wavelengths of

the Phenylalanine L_b and L_a transitions, which are measured at 260 and 210 nm,[72] respectively. The wavelength for the Tyrosine L_a state (λ_{exp}=230 nm), on the other hand, is underestimated by RICC2 by approximately 25 nm. These deviations point to some systematic discrepancies in the prediction of energies of aromatic transition by RICC2, also noticed previously.[82]

3.5.7.4. Long-Range Corrected Functionals

The behaviors of the two LCFs considered in the present application (CAM-B3LYP and ωB97X-D) are remarkably similar (Figure 3-31). For example, above 185 nm, both LCFs predict an average of between 7 and 8 states for $CMet_N$, and 4 states for $NMet_N$, (Table 3-4), which is also in good agreement with RICC2. Similarly, the LCFs are the only functional class that completely reproduces the ordering of the states predicted by RICC2 (compare columns two to four in Table 3-2). Furthermore, the LCFs exhibit the best overall agreement with the average transition wavelengths of the reference calculations, while the rotatory strengths are reasonably well reproduced. The agreement with RICC2 is particularly good in the case of the C-terminal excitation (~220 nm) that dominates the $CMet_N$ spectrum, as well as with the amide transitions. The latter is reflected in the similarities of the average CD spectra, particularly with respect to amide-type transitions. However, in the comparison with both RICC2 and the experiment, the LCFs perform relatively poorly for the aromatic L_b transitions, shifting them to considerably lower wavelengths. This can be seen, for example, in the misplaced second maximum in the average $NMet_N$ spectrum (Figure 3-31b). Despite these deviations, one can conclude that both LCFs exhibit nearly quantitative agreement with RICC2, a finding further supported by a very recent benchmarking study.[83] This conformity is not only apparent on the level of the average transitions and spectra but also on the level of single structures, as shown in Figure 3-32.

Table 3-4. Number of excited states (N_λ) found above 185 nm for all 15 model systems.[11]

Model Structure	RICC2	CAM-B3LYP	ωB97X-D	M06-2X	B3LYP	PBE
			N_λ			
$CMet_N$ I	8	10	8	9	26	20
$CMet_N$ II	6	7	7	7	25	20
$CMet_N$ III	7	7	7	7	26	20
$CMet_N$ IV	7	8	7	8	25	19
$CMet_N$ V	7	7	7	8	26	22
$NMet_N$ I	4	4	4	4	13	9
$NMet_N$ II	4	4	4	4	15	10
$NMet_N$ III	4	4	4	4	17	13
$NMet_N$ IV	4	4	4	4	16	12
$NMet_N$ V	4	4	4	4	14	9
$CMet_Z$ I	7	7	7	8	35	24
$CMet_Z$ II	8	8	7	9	30	22
$CMet_Z$ III	8	7	7	7	33	24
$CMet_Z$ IV	7	8	7	7	42	32
$CMet_Z$ V	9	8	7	9	36	27

3.5.7.5. The Hybrid Meta Functional

M06-2X performs similarly to the LCFs in terms of its comparison with RICC2. This is particularly true with respect to the number of states (Table 3-4) and their rotatory strengths (Table 3-3). However, larger deviations are obtained for the transition wavelengths and, in some cases, the order of the states is inverted (compare the second and fifth columns in Table 3-2). Most non-aromatic transitions, including those with CT character, are subject to a relatively consistent but not entirely systematic red shift of around 10 nm. This is most visible in the positions of the dominant spectral features above 200 nm in the $CMet_N$ average spectrum, as well as in the position of the nearly lost maximum associated with the amide transition in the $NMet_N$ average spectrum (Figure 3-31). The performance of the HMF for the L_b transitions is similar to that of the LCFs although the former compares better than the latter to RICC2 for the L_a transitions. On the whole, M06-2X functional reproduces the majority of the key elements of the transitions and spectra predicted by RICC2 but, due to the larger deviations in wavelengths, the agreement remains qualitative.

3.5.7.6. Global Hybrid Functionals

The B3LYP and PBE0 functionals typically overestimate the number of excited states for the systems studied herein. For example, in the case of $CMet_N$, B3LYP and PBE0 find about 25 and 20 states, respectively, in the same wavelength region where RICC2 predicts 7 (Table 3-4). This difference between the two GHFs is to be expected and stems from the fact that the higher proportion of exact exchange present in PBE0 leads to less artificially created excited states than for B3LYP.[84] Otherwise, however, the two GHFs exhibit a relatively high degree of mutual agreement, with the caveat that many of the PBE0 transitions tend to be somewhat blue shifted relative to B3LYP. This has the effect of a relatively uniform leftwards displacement of the average PBE0 spectra in Figure 3-31, with respect to their B3LYP counterparts. From all the functionals tested, B3LYP has the smallest deviations from RICC2 in terms of wavelengths for the aromatic L_b transitions (also seen in the

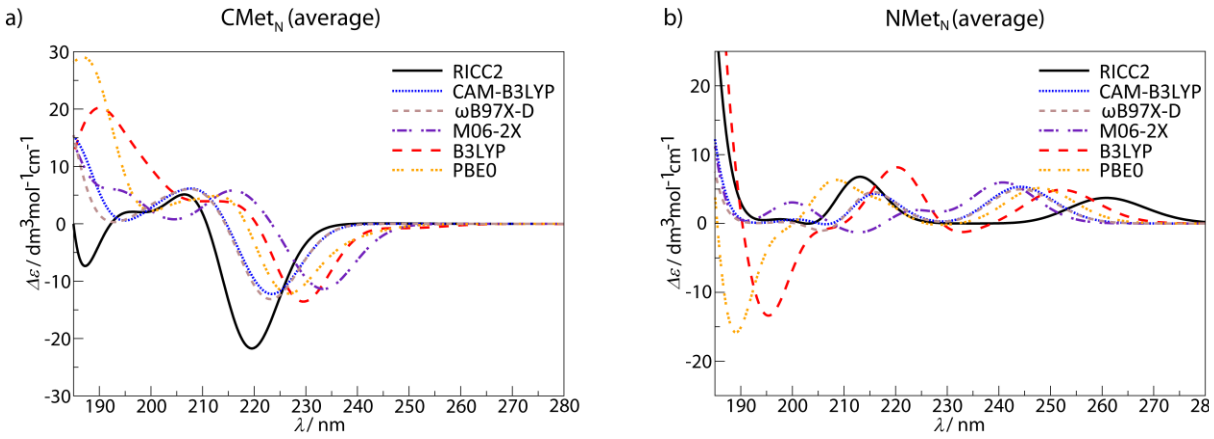


Figure 3-31. CD spectra averaged over five structures of a) $CMet_N$ and b) $NMet_N$.[11]

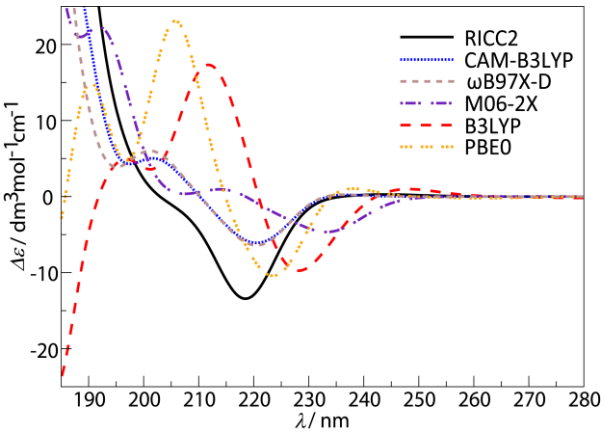


Figure 3-32. Example CD spectra of a single structure of $CMet_N$.[11]

average spectrum of $NMet_N$). At the same time, the GHF L_a transitions are relatively strongly red shifted with respect to RICC2, bringing them closer to the experimental values. The non-aromatic transitions calculated by the GHFs follow most closely the M06-2X results, and therefore appear at longer wavelengths than the RICC2 results. The exception is the CT transition ($n_S(M) \rightarrow \pi^*(COOH)$) that is significantly over-stabilized by the GHFs, but its contribution to the overall rotatory at large wavelengths becomes small after averaging over five structures, especially for B3LYP.

The majority of the surplus states, found by the GHFs but not by RICC2 or the other functionals, appear at wavelengths shorter than 210 nm and are typically associated with a non-local (CT) character. The main consequence of these states is the qualitative disagreement with RICC2 in the spectra of single structures, particularly in the range of wavelengths below 210 nm (see, for example, Figure 3-32). Further inspection of the rotatory strengths of the surplus transitions suggests that their contribution is small and sporadic, although numerous. Hence, it is only natural that their impact on the average spectrum diminishes with increasing the number of considered structures. As a result, the performance of both GHFs improves significantly upon averaging (Figure 3-31).

3.5.7.7. Zwitterionic Forms

The occurrence of CT artifacts in TD-DFT could be expected to be even more prominent in zwitterionic systems. To evaluate these effects in the context of our study, we compare the average spectrum of the $CMet_Z$ structures obtained with RICC2 to the TD-DFT functionals (Figure 3-33). Interestingly, the results are consistent with those obtained for the neutral structures. Both the LCFs and the MHF agree with RICC2 in the number of predicted excitations (Table 3-4). The spectra predicted by LCFs reproduce the RICC2 excitation energies well, although the rotatory strengths appear weaker. M06-2X shows very similar rotatory strengths to LCFs, but the excitation energies are red shifted, albeit to a lesser extent than in the neutral system. B3LYP and, to a lesser extent PBE0, deviates even more in terms of the average number of predicted excitations (Table 3-4) and again,

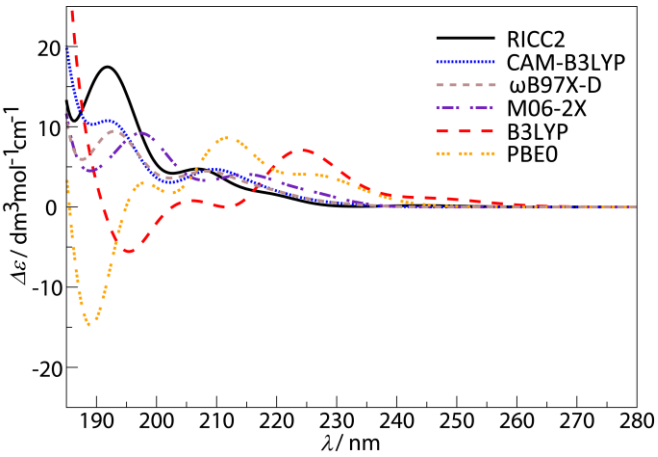


Figure 3-33. Average CD spectra of 5 structures of $CMet_Z$.[11]

larger deviations from the benchmark curve can be seen. However, even in these cases, the performance of the GHFs improves significantly upon averaging.

3.5.7.8. Additivity of $NMet_N$ and $CMet_N$

It is interesting to consider to what extent the sum of the spectra of the two peptide fragments (Figure 3-28) compare to the spectrum of Met_N. We perform this analysis on the example of the averaged $NMet_N$ and $CMet_N$ spectra, calculated with CAM-B3LYP (Figure 3-31), and show (Figure 3-34) that their sum exhibits the same key features as the spectrum arising from averaging over the corresponding five full structures of Met_N (Figure 3-29).

Because of the size of Met_N, it is difficult to perform the analogous comparison with the RICC2 approach. Interestingly, however, there is a reasonable agreement between the averaged

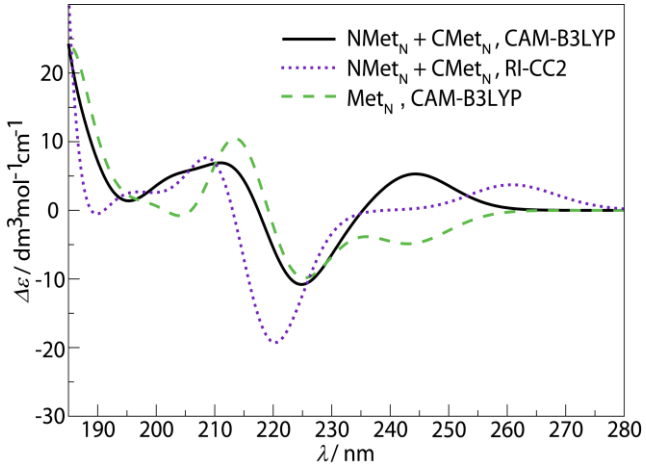


Figure 3-34. Superposition of average $NMet_N$ and $CMet_N$ spectra (CAM-B3LYP - black solid line, RICC2 - violet dotted line) are compared to the average CD spectrum of the corresponding five Met_N structures (CAM-B3LYP - green dashed line).[11]

spectrum of Met_N calculated with CAM-B3LYP and the spectrum that is obtained from the sum of the fragment spectra obtained with RICC2. On this basis, it is plausible to expect that, in this specific case, the RICC2 spectra of Met_N would resemble the CAM-B3LYP spectra of Met_N shown in Figure 3-34. This presumption is supported by the quantitative analysis presented in the previous section but, due to the differing size dependence of the two methods, is not expected to hold either rigorously or generally.

3.5.7.9. Met_N

To further investigate the performance of the TD-DFT functionals, we performed an analysis of the underlying excitations and constructed the average CD spectra of Met_N from 50 structures that were representative of the entire conformational space. We compare these calculations to the experimentally obtained spectrum[59] (Figure 3-35) using the fact that, in TFE, the neutral form of the peptide (Met_N) is expected to dominate. As we have previously witnessed, the zwitterionic form (Met_Z) is estimated to account for 10% of the N:Z equilibrium and therefore has only a minimal impact on the combined spectra (compare Figure 3-27a with the Figures 3-24a and b with respect to Met-enkephalin and its forms).

Consequently, we calculate the CD spectrum of Met_N for all the five functionals and the results are shown, together with the experimental spectrum, in Figure 3-35. Overall, we find very good agreement between B3LYP and M06-2X, as well as a reasonable correspondence of these two functionals with the experimental CD spectrum. The PBE0 spectrum for Met_N closely parallels that of B3LYP but, as was the case for the model structures, it is marginally but relatively uniformly shifted towards lower wavelengths. The spectra obtained with CAM-B3LYP and ωB97X-D once again exhibit practically identical behavior. Although the LCF spectra exhibit the same general features as the other functionals, they are blue shifted by approximately 10 nm, a result which is also consistent with those obtained for the smaller model structures.

The only larger discrepancy between all the calculated spectra and the experiment is the strong minimum between 230 and ~235 nm seen in the M06-2X and B3LYP spectra, just below 230 nm for PBE0 and at ~224 nm for CAM-B3LYP and ωB97X-D (Figure 3-35a). This is exactly the same problematic feature observed in the spectra of Met- and Ada-enkephalin obtained using B3LYP functional (Figure 3-27a). This minimum is, in the case of all functionals, induced by the dominant $n_O(COOH) + \pi(F\text{-}M) \rightarrow \pi^*(COOH)$ excitation, also appearing in $CMet_N$ (Figure 3-31a). The resemblance of the CAM-B3LYP and RICC2 spectra (Figure 3-34), as well as the systematic appearance of this transition with even more negative rotatory strengths within the RICC2 approach (Tables 3-3 and 3-4), suggests that this transition is not unique to TD-DFT. Furthermore, since this transition only arises for neutral peptides, one could imagine that the contribution of the zwitterionic

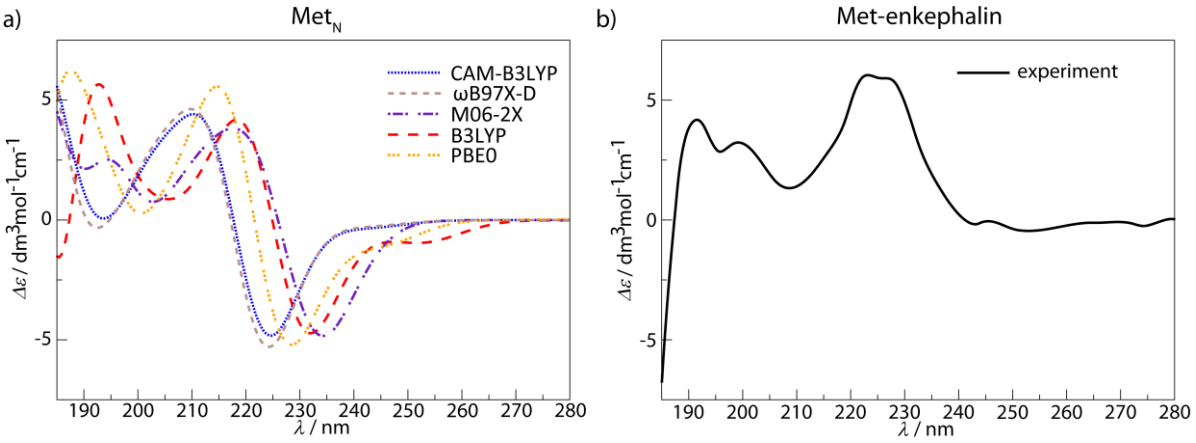


Figure 3-35. a) Theoretically calculated CD spectra of Met_N. b) Experimentally obtained CD spectrum of Met-enkephalin in TFE.[11]

form to the experimental spectra is larger than expected. However, the superposition of the Met_Z and Met_N spectra with zwitterionic fractions larger than 10% does not improve the overall comparison with experiments. This is true for all the functionals, even though at least M06-2X and the LCFs do not appear to suffer from CT artifacts when applied to (C)Met_Z (average CD spectra of Met_Z are shown in Figure 3-36). On the basis of these considerations, we are confident to conclude that the persistent appearance of the n_O(COOH) + π(F-M) → π*(COOH) excitation is not related to the TD-DFT approach and its potential difficulties with charge-transfer transitions. We can at this point tentatively propose two possible sources of this problematic feature, taking into regard the fact that our extensive testing strongly suggests that we can eliminate as the causes of the artifact both the convergence aspects of the CD spectra production (number of structures used to obtain the spectrum) and the size of the basis set used to calculate CD spectra (Figure 3-12). By eliminating these we are left with two most probable causes that can potentially lead to the appearance of this feature, first one being the underlying force field, which possibly generated somewhat biased conformational phase space of peptides, thereby favourizing structures that exhibit very negative rotatory strengths of the problematic excitation, in turn leading to the problematic minimum. For the second possible cause we tentatively propose quantum effects the solvent molecules (TFE) have on the peptides, more precisely, the interaction of the n_O/π^* orbitals of the COOH group with the orbitals arising from TFE, which was not taken into account, could significantly change to the response of this particular excitation, thereby possibly resolving the observed artifact.

The strongest correspondence between the theoretical and the experimental spectra is in the peak measured at ~220 nm. With M06-2X and B3LYP, it also appears at ~220 nm, while PBE0 and the LCFs center it, less accurately, closer to 210 nm. This maximum corresponds to the combination of two amide n → π* excitations, namely n_O(Y) → π*(Y-G) and n_O(F) → π*(F-M), which were also the dominant excitations in $NMet_N$ and $CMet_N$, respectively (Tables 3-3 and 3-4). Between this and the adjacent peak, all five spectra remain positive, with minima appearing at 205, 203, 200, 193 and 192

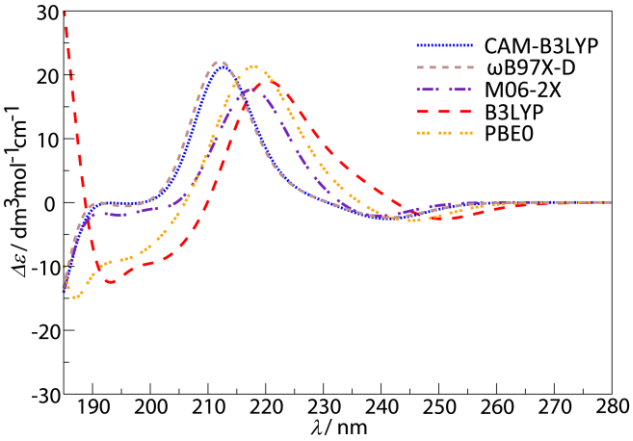


Figure 3-36. Theoretically calculated CD spectra of Met_Z. No strong minimum at around 230 nm is observed, in contrast to all the average CD spectra of the neutral species (Figures 3-27, 3-35). However, even though Met_Z spectra correlate better with the experiment in this respect, they also lack any discernible resemblence to the experimental spectrum below 220 nm (Figure 3-27).[11]

nm, for B3LYP, M06-2X, PBE0, CAM-B3LYP and ωB97X-D, respectively. The corresponding minimum in the overall positive experimental CD spectrum lies at ~208 nm.

The next prominent feature in the experimental spectrum is the positive peak centered at around ~195 nm. Although this wavelength borders on the edge of the window for reliable measurements and may hence be accompanied by a minimally accentuated uncertainty, it is nevertheless instructive to consider the behavior of the TD-DFT functionals in this region. Both the B3LYP and M06-2X spectra exhibit a positive peak at around the correct wavelength, while the analogous PBE0 feature is again blue shifted. A similar but broader maximum is found for the LCFs between 175 and 190 nm, which is indeed below the wavelengths accessed in the discussed experiments. A more detailed analysis of excitations in the CAM-B3LYP and ωB97X-D peaks over all 50 structures reveals that the dominant contributions come from local $\pi(Y) \rightarrow \pi^*(Y)$ transitions (which potentially represent the low frequency tail of the B_a and B_b excitations) and, to a lesser extent, from various $\pi/n_O(amide) \rightarrow \pi^*(amide)$ transitions. Despite the long-range corrections inherent to these functionals, we furthermore find numerous transitions with a strong CT character in this wavelength range. The most common are excitations of the type $\pi/n_O(amide) \rightarrow \pi^*(aromatic)$ and $\pi(aromatic) \rightarrow \pi^*(amide)$, as well as $n_S(M) \rightarrow \pi^*(amide)$ and $n_s(M) \rightarrow \pi^*(F)$.

Guided by the assignments of excitations from the LCF spectra, we analyzed the excitations contributing to the peaks observed in the B3LYP and PBE0 spectra around 195 and 190 nm, respectively. Despite the occurrence of many more CT transitions with the GHFs, the overall conclusions regarding the identification of the key contributors are virtually identical to those outlined above for the LCFs. Importantly, however, the overall spectral feature produced by B3LYP (and to a

lesser extent PBE0) is more consistent with the experiment than that arising from CAM-B3LYP and ωB97X-D. Interestingly, B3LYP is the only functional that exhibits the same behavior as the experimental spectrum in the lowest wavelength region (below 190 nm) following the maximum at 195 nm, although, as mentioned earlier, the uncertainties in the experimental curve are expected to increase at these lower wavelengths.

The M06-2X functional, on the other hand, gives the ratio of the average rotatory strengths for the two dominant peaks in the overall spectrum (Figure 3-35a) most similar to that in the experiment. With this functional, however, the peak at 195 nm is associated with the Phenylalanine L_a excitation (average rotatory strength of $\sim 5 \cdot 10^{-40}$ erg esu cm Gauss^{-1}), coupled with a small contribution from the $n_S(M) \rightarrow \pi^*(COOH)$ excitation, consistent with the result over five structures (Table 3-2). This is despite the fact that the rotatory strengths of the L_b and L_a aromatic transitions, averaged over all 50 structures, are negligible, for all other investigated functionals. For example, B3LYP predicts the Phenylalanine L_a excitation at ~206 nm, and with an average rotatory strength close to zero.

It should be noted that, relative to the HMF and LCFs, the GHFs frequently over-stabilize CT transitions which then appear over the entire wavelength range of interest. A typical example is the $n_S(M) \rightarrow \pi^*(COOH)$ transition, which, with B3LYP, appears with an average wavelength of 253.4 nm and a mean rotatory strength of $-0.25 \cdot 10^{-40}$ erg esu cm Gauss^{-1} (over 50 structures). This is even closer to zero than after averaging over five structures (Table 3-3). Such a trend, where the impact of spurious CT excitations tends to vanish after averaging over a sufficient number of structures, seems to apply quite generally.

3.5.7.10. Excited States Discussion

While modeling the CD spectra of flexible molecules is the only possible way to associate the measured spectra with their conformational phase spaces, our work clearly demonstrates the complexity of this task. The helpful fact here is that, irrespective of the choice of the method, the relevant transitions and their excitation energies depend only weakly on the choice of peptide conformation. On the other hand, the rotational intensity of these excitations may change significantly with the peptide configuration. Since the measured signal contains contributions from all possible conformations, to correctly reproduce the experimental spectrum, appropriate averaging over theoretical spectra from a number of structures must be performed.

The effect of averaging is perhaps best seen in the comparison of the Met_N spectra shown in Figures 3-34 and 3-35a, where the averaging is performed over a set of 5 and 50 structures, respectively. In the former case, one of the dominant features of the spectrum is the Tyrosine L_b peak at longer wavelengths. However, after appropriate averaging, the contribution of this excitation becomes negligible. This is also generally true for the intensity of the two local Phenylalanine

Table 3-5. Mean signed errors (MSE) of TD-DFT transition wavelengths from RICC2 values are calculated as $\frac{1}{n}\sum_i TDDFT_i - RICC2_i$. $TDDFT_i$ and $RICC2_i$ correspond to the values of the wavelength of the particular excitation in the *i*-th peptide model found using the TD-DFT method of interest and RICC2, respectively.[11]

Excitation type	Excited state	Wavelength / nm				
		CAM-B3LYP	ωB97X-D	M06-2X	B3LYP	PBE
Aromatic excitations	Tyrosine ^1L_b	-17.0	-16.8	-20.3	-9.1	-13.3
	Phenylalanine ^1L_b	-15.6	-15.6	-17.9	-9.9	-14.0
	Tyrosine ^1L_a	3.4	2.8	-1.4	8.8	8.8
	Phenylalanine ^1L_a	7.9	7.4	2.4	15.4	11.3
CT	n_S(M) -> π*(COOH)	5.7	3.2	7.1	74.8	59.0
Amide excitations	n_O(Y) -> π*(Y-G)	2.2	1.8	9.2	8.0	1.1
	n_O(F) -> π*(F-M)	0.0	-0.1	6.8	9.3	2.5
	n_O(G) -> π*(G-Nme)	3.5	2.9	10.4	8.0	4.8
	n_O(Ace) -> π*(Ace-F)	2.9	2.3	9.4	7.4	4.3

Table 3-6. Mean signed errors (MSE) of TD-DFT transition rotatory strengths from RICC2 values are calculated as $\frac{1}{n}\sum_i TDDFT_i - RICC2_i$. $TDDFT_i$ and $RICC2_i$ correspond to the values of the rotatory strength of the particular excitation in the *i*-th peptide model found using the TD-DFT method of interest and RICC2, respectively.[11]

Excitation type	Excited state	Rotatory strength / 10^{-40} erg esu cm Gauss^{-1}				
		CAM-B3LYP	ωB97X-D	M06-2X	B3LYP	PBE
Aromatic excitations	Tyrosine ^1L_b	1.8	1.3	2.6	1.6	1.8
	Phenylalanine ^1L_b	0.6	0.5	1.1	0.1	1.1
	Tyrosine ^1L_a	2.8	0.4	2.6	4.9	7.5
	Phenylalanine ^1L_a	0.0	0.5	0.8	-3.5	3.2
CT	n_S(M) -> π*(COOH)	10.0	7.1	12.0	12.9	11.7
Amide excitations	n_O(Y) -> π*(Y-G)	-3.2	-2.6	-6.8	-0.5	5.6
	n_O(F) -> π*(F-M)	-0.2	-3.4	-0.9	1.8	0.9
	n_O(G) -> π*(G-Nme)	-2.7	-1.6	0.3	0.6	-3.2
	n_O(Ace) -> π*(Ace-F)	-1.0	0.0	0.0	-2.3	-3.1

transitions. Interestingly, on the level of a single structure, these excitations are associated with small rotatory strengths while the Tyrosine transitions appear to be very sensitive to the geometry, leading to the notably large standard deviations (see Table 3-3).

In the context of using RICC2 calculations as a benchmark for the TD-DFT approach (Tables 3-5, 3-6), both the Tyrosine and Phenylalanine L_b transitions are blue shifted for all investigated functionals (first two rows in Table 3-5). Nevertheless, the excitation energies predicted by B3LYP for these transitions agree best with RICC2, while M06-2X disagrees most. On the contrary, L_a transitions predicted with TD-DFT are typically red shifted, with B3LYP deviating most from RICC2, while M06-2X shows nearly quantitative agreement. Interestingly, the sensitivity of the rotatory strength to the peptide conformation is also reflected in the comparison between high and low accuracy approaches. Thereby, TD-DFT relatively accurately reproduces the rotatory strengths predicted by RICC2 for Phenylalanine transitions, but the deviations are significantly larger for Tyrosine excitations, particularly with B3LYP (compare the first 4 rows in Table 3-6).

From the perspective of the experimentally available results, which indicate a blue shift for all aromatic transitions predicted by RICC2, one can conclude that TD-DFT is inferior to the higher-level approach when L_b transitions are considered, but that it performs better for L_a transitions. A natural extension of the above conclusion is that B3LYP outperforms the other functionals in modeling aromatic transitions, especially in terms of their excitation energies. This is particularly important when assessing the reliability of the maximum associated with the Phenylalanine L_a transition, predicted only by M06-2X at 195 nm, in the spectrum of Met_N.

The comparison of TD-DFT with RICC2 is especially useful for non-local transitions involving charge transfer, even though the direct comparison with experiments is currently challenging. In the case of the systems studied herein, RICC2 predicts one principle transition of this type ($n_S(M) \rightarrow \pi*(COOH)$) in the wavelength range of interest (> 185 nm). This same transition is also found by TD-DFT. For CAM-B3LYP, ωB97X-D and M06-2X, it appears at somewhat larger, but relatively similar wavelengths (fifth row in Table 3-5). For B3LYP and PBE0, the red shift is considerably more pronounced.

The comparison between TD-DFT and the experiment reveals that the agreement is especially significant for the amide excitations, which are responsible for the dominant maximum in the CD spectrum of Met_N (Figure 3-35). The latter is best reproduced by M06-2X and B3LYP, suggesting that these functionals are particularly suitable for modeling this transition type, a conclusion supported by previous B3LYP calculations of helical Alanine chains.[22] From this point of view, the amide transitions predicted by CAM-B3LYP and ωB97X-D are blue shifted, which is interesting from the perspective that these functionals nearly quantitatively reproduce the RICC2[72] results, particularly for amide excitations (final four rows in Table 3-3). Similar overall conclusions regarding the applicability of TD-DFT functionals were obtained in the case of transition metal complexes.[21]

Thus, we overall find that, although it can be expected for the TD-DFT approaches to exhibit artificial or over-stabilized CT excitations, we find these to be prominent only for the GHF in the considered energy range. These artifacts are obvious on the level of a single peptide structure (Figure 3-32), where the GHF spectra fail to reproduce features systematically predicted by other methods for wavelengths below 210 nm. However, the performance of both B3LYP and PBE0 improves significantly upon averaging over many structures, for all investigated models. The understanding of such a trend emerges from an in-depth analysis of the surplus states, whose rotatory strengths are either insignificant or behave essentially randomly as a function of geometry. Consequently, the averaging procedure, essential to calculations of CD spectra of flexible molecules, becomes beneficial for the accuracy of TD-DFT. Such improvement is apparent for both PBE0 and B3LYP but is perhaps more significant in the latter case, which ultimately provides a spectrum exhibiting substantial

agreement with the main features of its experimentally observed counterpart, in turn becoming the functional of choice with regard to the studied systems, despite the artifacts it exhibits on the level of single structures.

3.6. Conclusion

We have introduced a general methodology for calculating the CD spectra of flexible molecules. Our method is based on the generation of a converged conformational phase space, obtained from advanced classical molecular dynamics simulations, namely REMD, which is followed by finding a set of structures representing the entire phase space using clustering analysis. In agreement with previous studies, we have observed strong solvent effects on the spectra. These were taken into account by calculating an average field consisting of point charges originating from a number of solvent configurations around each peptide conformation in the subset. Subsequently, the CD spectrum of each solvated conformation was obtained using the QM/MM method incorporating TD-DFT calculation. The average CD spectra were obtained by taking the mean of all weighted single CD spectra, where the weighting factor is given by the population fraction of the total phase space represented by the conformation giving the respective single CD spectrum. Qualitatively, our procedure can be considered as the sequential averaging over the solvent (for each representative conformation), the sidechains (inside the backbone-based clusters), and the backbone (combining the principal clusters).

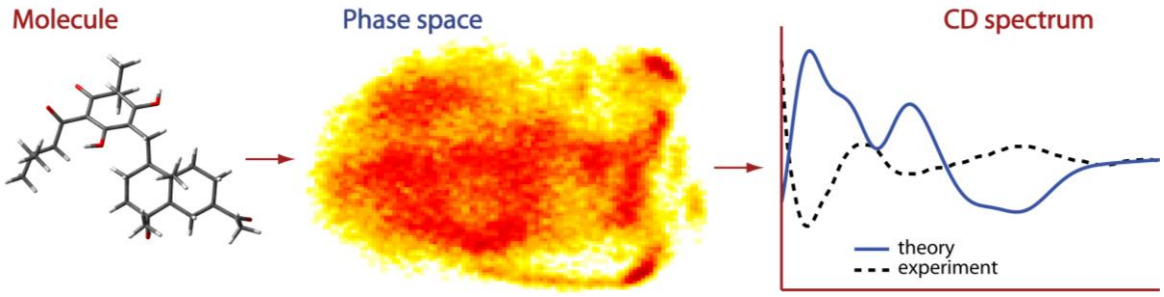


Figure 3-37. Our methodology presented in short. Parameterize the molecule of interest with the suitable force field, generate the conformational phase space using advanced molecular dynamics, and finally, construct the average CD spectrum by weighted averaging of the CD spectra of *n* structures (representing the phase space), where we use TD-DFT as a quantum method of choice for calculation of CD spectra.

In our first case study we applied the developed methodology to calculate the CD spectra of two novel *Rhodomyrtal* compounds (Figure 3-13), in an attempt to determine the absolute configuration of the two compounds. We were able to successfully reproduce the experimental CD spectra using the described procedure and utilizing the B3LYP functional, thereby obtaining an excellent agreement between experimentally measured and calculated CD spectra (Figure 3-16). More precisely, all of the most dominant spectral features are both quantitatively and qualitatively reproduced by the theoretically obtained CD spectra, in turn enabling us to unambiguously assign the absolute configurations of both the *Rhodomyrtal* compounds.[29]

In our second study we applied our methodology in an attempt to calculate CD spectra of the two highly flexible opioid peptides, namely Met-enkephalin and Ada-enkephalin. We thus used REMD to obtain the conformational phase spaces of both the zwitterionic and the neutral forms of Met-enkephalin and of both the *R*- and the *S*-epimers of Ada-enkephalin, which was necessary due to the fact that all these forms coexist in TFE. Despite the numerous complications, qualitative agreement with experiment has been obtained below 230 nm and the main features of the measured spectra have been reproduced by our calculations. However, the simulated spectra show a persistent minimum between about 230-240 nm that is not present in the experimental measurements and which we hypothesized to arise from a poor treatment of a particular excited state by TD-DFT. Motivated by this finding we decided to investigate the performance of TD-DFT functionals in detail, by benchmarking the performance of five functionals (B3LYP, CAM-B3LYP, M062X, PBE0, ωB97X-D), both against the experiment and against a higher level of theory, namely RICC2. We have thereby shown that the TD-DFT approach can indeed be used to calculate the properties necessary to derive the CD spectra of flexible peptides, finding that this approach is not related to the appearance of the aforementioned artifact/minimum. In this respect, we tentatively propose both the underlying force field and the neglected quantum effects of the solvent on the peptides as possible sources of this problematic feature, which, if improved/taken into account, could possibly resolve this problematic issue. On the other hand, the charge transfer artifacts, occurring with some functionals, are significantly diminished by the essential process of averaging the spectrum over the entire conformational space. Interestingly, we found that despite its complexity, RICC2 is not always capable of providing accurate predictions of excitation energies[76-79] and, for certain transition types investigated herein, we also found indirect evidence for non-negligible deviations from experimental values. For the smaller model systems, CAM-B3LYP (along with ωB97X-D) was found to exhibit nearly quantitative agreement with RICC2 for all three transition types and could thus be used as a reasonable indicator of the RICC2 result for the larger system. In that context, the LCFs (CAM-B3LYP and ωB97X-D) showed a less satisfactory agreement with experiment than PBE0, which was in turn slightly inferior to B3LYP and M06-2X. Thus, when considering the reproduction and understanding of experimental CD spectra, the choice of the functional should be adjusted to the types of transitions expected to dominate.[85] Here, in some cases, TD-DFT may actually be preferable to the RICC2 approach. This is not only because of its better agreement with experiment for certain transition classes but also because it can be conveniently used, even for relatively large molecules, to evaluate the substantial number of conformations required to achieve converged CD spectra for flexible molecules.

The method developed herein establishes a direct link between the conformational phase space of a small flexible molecule in solution and its CD spectrum. The conclusion of our work is that the current state-of-the-art theoretical methods show significant promise in reproducing experimentally observed CD spectra of flexible, solvated molecules. In the context of the methodology presented in

this work, further development of the force fields and functionals for TD-DFT calculations can only yield a more quantitative agreement. We show that theoretical calculations become necessary for interpretations of the experimental measurements, which cannot be performed in an unambiguous manner when attempting to determine the correspondence between the structures in the underlying ensemble and the measured CD spectrum.

3.7. References

(1) Warnke, I.; Furche, F. *WIREs Comput. Mol. Sci.* **2012**, *2*, 150.

(2) https://en.wikipedia.org/wiki/Chirality.

(3) Autschbach, J. *Chirality* **2009**, *21*, 116.

(4) Brahms S.; Brahms, J. *J. Mol. Biol.* **1980**, *138*, 149.

(5) Hirst J. D.; Bhattacharjee S.; Onufriev A. V. *Faraday Discuss.* 2002, *122*, 253.

(6) Greenfield, N. J. *Nature Protoc.* **2007**, *1*, 2876.

(7) Woody, R. W. *Monatsh. Chem.* **2005**, *136*, 347.

(8) Jiang, J.; Abramavicius, D.; Bulheller, B. M.; Hirst, J. D.; Mukamel, S. *J. Phys. Chem. B* **2010**, *114*, 8270.

(9) Abramavicius, D.; Jiang, J.; Bulheller, B. M.; Hirst, J. D., Mukamel, S. *J. Am. Chem. Soc.* **2010**, *132*, 7769.

(10) Brkljača, Z.; Čondić-Jurkić, K.; Smith, A-S.; Smith, D. M. *J. Chem. Theory Comput.* **2012**, *8*, 1694.

(11) Brkljača, Z.; Mališ, M.; Smith, D. M.; Smith, A.-S. *J. Chem. Theory Comput.* **2014**, *10* , 3270.

(12) Greenfield, N. J. *Anal. Biochem.* **1996**, *235*, 1.

(13) Gopal, R.; Park, J. S.; Seo, C. H.; Park, Y. *Int. J. Mol. Sci.* **2012**, *13*, 3229.

(14) Glättli, A.; Daura, X.; Seebach, D.; van Gunsteren, W. F. *J. Am. Chem. Soc.* **2002**, *124*, 12972.

(15) Besley, N. A.; Hirst, J. D. *J. Am. Chem. Soc.* **1999**, *121*, 9636.

(16) Hirst, J. D.; Colella, K.; Gilbert, A. T. B. *J. Phys. Chem. B* **2003**, *107*, 11813.

(17) Gilbert, A. T. B.; Hirst, J. D. *J. Mol. Struct.: THEOCHEM* **2004**, *675*, 53.

(18) Rogers, D. M.; Hirst, J. D. *Chirality* **2004**, *16*, 234.

(19) Fidler, J.; Rodger, P. M.; Rodger, A. *J. Chem. Soc., Perkin Trans. 2* **1993**, *2*, 235.

(20) Autschbach, J.; Jorge, F. E.; Ziegler, T. *Inorg. Chem.* **2003**, *42*, 2867.

(21) Rudolph, M; Autschbach, J. *J. Phys. Chem. A* **2011**, *115*, 14677.

(22) Kaminský, J.; Kubelka, J.; Bouř, P. *J. Phys. Chem. A* **2011**, *115*, 1734.

(23) Aquino, A. J. A.; Nachtigallova, D.; Hobza P.; Truhlar, D. G.; Hättig, C.; Lischka, H. *J. Comput. Chem.* **2011**, *32*, 1217.

(24) Wakai, A.; Fukasawa, H.; Yang C.; Mori, T.; Inoue Y. *J. Am. Chem. Soc.* **2012**, *134*, 4990.

(25) Kaminský, J.; Kubelka, J.; Bouř, P. *J. Phys. Chem. A* **2011**, *115*, 1734.

(26) Kelly, S. M; Price, N. C. *Curr. Protein Pept. Sci.* **2000**, *1*, 349.

(27) Schellman, J. A. *Chem. Rev.* **1975**, *75*, 323.

(28) Condon, E. U.; Altar, W.; Eyring, H. *J. Chem. Phys.* **1937**, *5*, 753.

(29) Shou, Q.; Smith, J. E.; Mon, H.; Brkljača, Z., Smith, A.-S.; Smith, D. M.; Griesser, H. J.; Wohlmuth, H. *RSC Adv.* **2014**, *4*, 13514.

(30) Sanchez, M. L.; Aguilar, M. A.; Olivares del Valle, F. J. *J. Comput. Chem.* **1997**, *18*, 313.

(31) Coutinho, K.; Georg, H. C.; Fonseca, T. L.; Ludwig, V.; Canuto, S. *Chem. Phys. L.* **2007**, *437*, 148.

(32) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157.

(33) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J. W.; Wang, J.; Kollman, P. A. *J. Comput. Chem.* **2003**, *23*, 1999.

(34) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. *J. Comp. Chem.* **2009**, *30*, 2157.

(35) Dupradeau, F-Y.; Cézard, C.; Lelong, R.; Stanislawiak, E.; Pecher, J.; Delepine, J. C.; Cieplak, P. *Nucleic Acids Res.* **2008**, *36*, 360.

(36) Garcia, A. E.; Herce, H.; Paschek, D. *Annu. Rep. Comput. Chem.* **2006**, *2*, 83.

(37) Case, D. A.; Darden, T. A.; Cheatham, T.E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman ,P. A. *AMBER 10*, University of California, San Francisco 2008.

(38) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33.

(39) Wolf, A.; Kirschner, K. N. *J. Mol. Model.*, **2013**, *19*, 539.

(40) Pearson, K. *Philos. Mag.* **1901**, *2*, 559.

(41) Hotelling, H. *J. Educ. Psychol.* **1933**, *4*, 417.

(42) Jolliffe, I. T. *Principal Component Analysis.* New York: Springer; 2002.

(43) Amadei, A.; Linssen, A. B.; Berendsen, H. J. *Proteins* **1993**, *17*, 412.

(44) Maass, A.; Tekin, E. D.; Schüller, A.; Palazoglu, A.; Reith, D.; Faller, R. *Biochim. Biophys. Acta*. **2010**, *1804*, 2003.

(45) Tan, P. N.; Steinbach, M.; Kumar, V. *Introduction to Data Mining*, Addison-Wesley, chapter *Cluster Analysis: Basic Concepts and Algorithms*. pp. 487–568, 2006.

(46) Jain, A. K.; Murty, M. N.; Flynn, P. J. *ACM Comput. Surv.* **1999**, *31*, 264.

(47) Shao, J.; Tanner, S. W.; Thompson, N.; Cheatham, T. E. *J. Chem. Theory Comput.* **2007**, *3*, 2312.

(48) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. *J. Comput. Chem.* **2005**, *26*, 1668.

(49) Fidler, J.; Rodger, P. M.; Rodger, A. *J. Chem. Soc., Perkin Trans. 2* **1993**, *2*, 235.

(50) Jensen, L.; Swart, M.; van Duijnen, P. T.; Autschbach, J. *Int. J. Quant. Chem.* **2006**, *106*, 2479.

(51) Sanchez, M. L.; Aguilar, M. A.; Olivares del Valle, F. J. *J. Comput. Chem.* **1997**, *18*, 313.

(52) Coutinho, K.; Georg, H. C.; Fonseca, T. L.; Ludwig, V.; Canuto, S. *Chem. Phys. Lett.* **2007**, *437*, 148.

(53) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712.

(54) Case, D. A.; Darden, T. A.; Cheatham, T.E. III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsich, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman , P. A. The FF12SB force field, AmberTools 13 Reference Manual (2013) 27–29.

(55) Kendrew, J. C. In *The Encyclopedia of Molecular Biology*; Blackwell Science: Oxford, United Kingdom, 1994; pp. 772-781.

(56) Sanbonmatsu, K. Y.; Garcia, A. E. *Proteins* **2002,** *46*, 225.

(57) Terenius, L. *Alcohol* **1996**, *13*, 31.

(58) Horvat, Š; Mlinarić-Majerski, K.; Glavaš-Obrovac, Lj.; Jakas, A.; Veljković, J.; Marczi, S.; Kragol, G.; Roščić, M.; Matković, M.; Milostić-Srb, A. *J. Med. Chem.* **2006**, *49*, 3136.

(59) Grdičak, M.; Supek, F.; Kralj, M.; Majer, Z.; Hollosi, M.; Šmuc, T.; Mlinarić-Majerski, K.; Horvat, Š. *Amino Acids* **2009**, *38*, 1185.

(60) Cheng, F.; McLaughlin, P. J.; Verderame, M. F.; Zagon, I. S. *Mol. Cancer* **2008**, *7*, 5.

(61) Cheng, F.; McLaughlin, P. J.; Verderame, M. F.; Zagon, I. S. *Cancer Res.* **2007**, *67*, 10511.

(62) Zagon, I. S.; Verderame, M. F.; McLaughlin, P. J. *Brain Res. Rev.* **2002**, *38*, 351.

(63) Kuroda, Y.; Ueda, H.; Nozawa, H.; Ogoshi, H. *Tetrahedron Lett.* **1997**, *38*, 7901.

(64) Knipp, G. T.; Vander Velde, D. G.; Siahaan, T. J.; Borchardt, R. T. *Pharm. Res.* **1997**, *14*, 1332.

(65) Marcotte, I.; Separovic, F.; Auger, M.; Gagne, S. M. *Biophys. J.* **2004**, *86*, 1587.

(66) Andrew, C. D.; Bhattacharjee, S.; Kokkoni, N.; Hirst, J. D., Jones, G. R.; Doig, A. J. *J. Am. Chem. Soc.* **2002**, *124*, 12706.

(67) Bhattacharjee, S.; Tóth, G.; Lovas, S.; Hirst, J. D. *J. Phys. Chem. B* **2003**, *107*, 8682.

(68) Tanaka, T.; Kodoma, T. S.; Morita, H. E.; Ohno, T. *Chirality* **2006**, *18*, 652.

(69) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.;

Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg,J.J.;Dapprich,S.;Daniels,A.D.;Farkas,O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, revision A.02*; Gaussian, Inc.: Wallingford, CT, 2009.

(70) Su, L.; Cukier, R. I. *J. Phys. Chem. B* **2007**, *111*, 12310.

(71) Zhang, L.; Peslherbe, G. H.; Muchall, H. M. *Photochem. Photobiol.* **2006**, *82*, 324.

(72) Berova, N.; Nakanishi, K.; Woody, R. W., eds. In *Circular Dichroism : Principles and Applications*, 2nd edition, Wiley-VCH, New York, 2000; p. 75.

(73) Gerothanassis, I. P.; Birlirakis, N.; Karayannis, T.; Tsikaris, V.; Sakarellos-Daitsiotis, M.; Sakarellos, S.; Vitouxb, B.; Marraud, M. *FEBS Lett.* **1992**, *298*, 188.

(74) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037.

(75) Limão-Vieira, P.; Giuliani, A.; Delwiche, J.; Parafita, R.; Mota, R.; Duflot, D.; Flament, J.-.P.; Drage, E.; Cahillane, P.; Mason, N. J.; Hoffmann, S. V.; Hubin-Franskin, M.-J. *Chem. Phys.* **2006**, *324*, 339.

(76) Jacquemin, D.; Wathelet, V.; Perpète, E. A.; Adamo, C. *J. Chem. Theory Comput.* **2009**, *5*, 2420.

(77) Silva-Junior, M. R.; Schreiber, M.; Sauer, S. P. A.; Thiel, W. *J. Chem. Phys.* **2008**, *129*, 104103.

(78) Schreiber, M.; Silva-Junior, M. R.; Sauer, S. P. A.; Thiel, W. *J. Chem. Phys.* **2008**, *128*, 134110.

(79) Crawford, T. D.; Stephens, P. J. *J. Phys. Chem. A* **2008**, *112*, 1339.

(80) TURBOMOLE V6.3 2011, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH.

(81) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.

(82) Parac, M.; Grimme, S. *J. Phys. Chem. A* **2002,** *106*, 6844.

(83) Guido, C. A.; Knecht, S.; Kongsted, J; Mennucci, B. *J. Chem. Theory Comput.* **2013**, *9*, 2209.

(84) Goerigk, L.; Grimme, S. *J. Phys. Chem. A* **2009**, *113*, 767.

(85) Laurent, A. D.; Jacquemin, D. *Int. J. Quant. Chem.* **2013**, *113*, 2019.

4. Biomineralization and Biomineralization-Inspired Drug Design: Mineral - Peptide Interactions

4.1. Introduction

The field of interface chemistry has been heavily focused in recent years on the development of systems that could be used for the controlled introduction and release of active pharmaceutical compounds in the living organisms and tissues.[1] In this respect, knowledge and understanding of basic interactions between the active species (drugs) and the carrier used for their distribution is vital. It was only recently that inorganic carriers were given special attention due to their unique properties which make them useful in regulating drug delivery.[2,3] Perhaps the most promising of these carriers is the mineral calcite ($CaCO_3$), representing one of the most abundant minerals on the Earth's surface. This mineral is one of the most interesting candidates for the use as a mineral carrier of pharmaceutical substances, as it represents a prototypical example of a nontoxic and biocompatible material, and is extraordinary chemically active toward the organic molecules containing polar hydroxyl and carboxyl groups.[1,4-6] The bioactivity of calcite is perhaps most easily observed, for example, in marine organisms, such as mollusks and crustaceans, where it forms hierarchical structures under the influence of the biopolymers.[7] For these reasons calcite represents an ideal biomineral for the investigation of intricate biomineral - active organic compound interactions occurring on the surfaces of crystals, enabling us to investigate both the possible growth mechanisms of biominerals and the interactions that have a predominant effect on the properties of biodegradable drug delivery systems.

The most common experimental strategies used to investigate the biomineralization and biomineralization-inspired drug design of $CaCO_3$ are heavily focused on the usage of additives, which are of a rather wide chemical scope, ranging from various ionic species, such as copper,[8] manganese,[9] iron and other divalent cations,[10] phosphate species[11,12] to low molecular-weight molecules,[13,14] synthetic polymers,[15] and finally all the way up to bioorganic (macro)molecules.[16-18] The latter ones are especially interesting, as it was found that, differently from their non-biogenic counterparts, they exhibit a number of important advantages.[19] As an example, it is well known that biomolecules can act as catalysts for crystal formation by lowering the energy of activation,[20,21] in turn drastically altering the physical conditions necessary for this process to take place.[19] Moreover, it was found that the mechanical performance of these materials often excels over the one observed in the case of their synthetic counterparts.[22] In this respect, it was found that proteins, due to their electrostatic properties, hydrophobicity, as well as their shape and size, can interact differently, and bind to the surfaces of calcite – producing diverse effects on the growth of a calcite biomineral, from inhibition to morphological growth modulation.[23-27] Interestingly, it was found that acidic proteins and peptides most readily interact with calcite.[28,29] This is famously showcased in nature on the example of the shell of the red abalone, where green organic sheets intersperse through the nacre. It was found that a

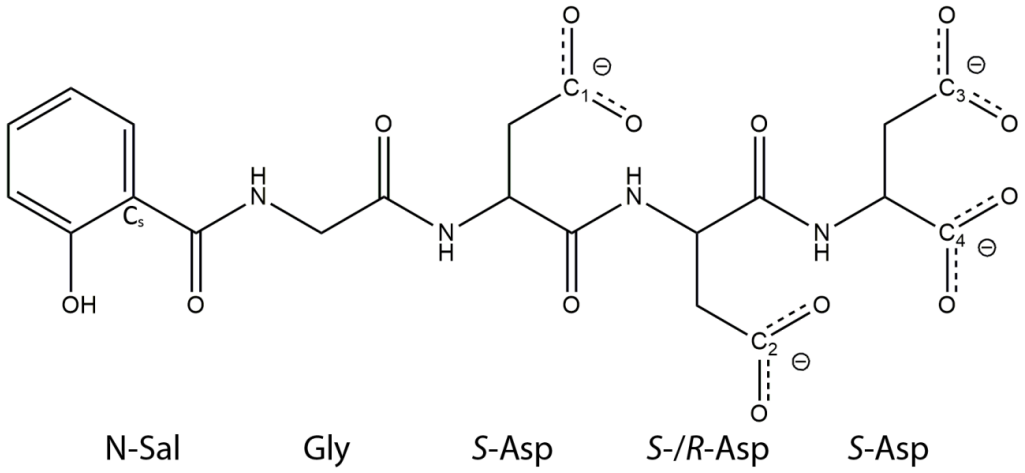


Figure 4-1. Structure of Sal peptides, difference between *S*-Sal and *R*-Sal being only in the fourth residue (denoted with *S*-/*R*-Asp, respectively). The overall charge of Sal peptides is -4*e*, due to four negatively charged carboxyl groups (C_1 to C_4, denoting carbon atoms present in four carboxylic groups).

particular protein fragment, named GP, extracted from the shell of the red abalone and especially rich in the aspartic acid content, was the compound responsible for the inhibition of the calcite growth.[30] In this respect, a recently conducted experimental study found that small peptides, namely the derivatives of salicylic amino acid containing acidic amino acid residues, in particular aspartic acid residues, show a strong propensity toward the inhibition and modification of the calcite growth,[31] which, due to the relatively small size of these additives/species, represents an ideal model system for the investigation of calcite-bioorganic material interactions using computational methods.

The theoretical tool of choice readily used to investigate the interactions of the biomolecules with mineral surfaces is molecular dynamics. In this respect, MD simulations have been used to investigate the influence of explicit water molecules[32,33] relative energies, and dynamics of protein adsorption on the biomineralization process.[33-36] In the case of calcite, molecular dynamics studies have been performed for nearly two decades.[37] The morphology of calcite and its dependence on the presence of certain ionic species has been calculated in vacuum,[38] thereby showing the dominance of the (104) calcite face, showcasing this surface as the most stable calcite surface when no extrinsic ions/adsorbents are present.[37] Moreover, with biomineralization and biomineralization-inspired drug design in mind, both the interaction and adsorption properties of organic molecules with varying functional groups to the calcite surfaces have been examined with a combination of a number of different parameterizations/force fields.[39] Recently, Freeman et al.[40] established a systematic procedure used to produce a set of parameters/potentials which can be used to model biomineralization processes on calcite. This represented a necessary step, as force fields usually used to characterize macromolecules differ significantly from the ones employed to describe the behavior of the crystal; to be able to investigate calcite – biomolecule interactions – the potentials for biomolecules, the mineral, the solvent and their cross terms need to be constructed.[40,41] A number of theoretical biomineralization

studies have been successfully conducted using this methodology, namely the binding of the ovocleidin-17 to both the stable (104) calcite surface and the stepped calcite surfaces[42] was investigated using classical MD, elucidating the role of this protein in the eggshell formation. Moreover, the conformational behavior and binding patterns of 16- and 17-residue peptides bound to (104) calcite surface were investigated using the same potential-building methodology.[43] The interaction of aspartic acid rich peptides with (104) calcite surface were also investigated using classical MD, helping to explain the stronger inhibition of calcite growth observed by polyaspartic compared to the polyacrylic acid.[37] Very recently, an advanced molecular dynamics simulation study investigated the adsorption behavior of oligomers of polystyrene sulfonate on the stable (104) and the polar (001) calcite surfaces, successfully explaining experimentally observed stabilization of the (001) calcite surface mediated by polystyrene sulfonate.[7] Due to the use of an advanced sampling technique this study was able to tackle both the qualitative and the quantitative aspects of the biomineralization process, giving rise to the free energy profiles of binding of polystyrene sulfonate to calcite, otherwise unattainable using „normal" classical MD.

In spite of the combined experimental and theoretical efforts that occurred in recent years, our understanding of the interactions between biomolecules and minerals, governing the process of biomineralization and playing a critical role in the drug design, is still far from complete. One of the most important, yet uninvestigated, aspects of this process is the influence of the chirality of peptides/proteins on biomineralization, as the change of the chirality of one or more amino acids in the protein/peptide sequence could give us a mean to obtain a detailed tuning of the biomineral – active biomolecule interactions, thereby, on one hand, achieving a higher level of the control of biomineralization process and, on the other hand, enabling us to modify and change pharmacological activity of the tentative drug delivery system. In this respect, it was very recently found that, somewhat unexpectedly, a difference in chirality of only one amino acid can change the binding behavior of peptides rather significantly.[31] More precisely, by performing crystal growth kinetic experiments and using determined Langmuir adsorption constants, it was concluded that *R*-Sal (N-Sal-Gly-*S*-Asp-*R*-Asp-*S*-Asp, N-Sal represents salicylic acid residue, Figure 4-1) peptide binds significantly more strongly to calcite compared to its epimeric counterpart, namely *S*-Sal (N-Sal-Gly-*S*-Asp-*S*-Asp-*S*-Asp, Figure 4-1). Motivated by these findings, we decided to theoretically probe the calcite – peptide interactions in order to investigate the influence of chirality on the biomineralization process and biomineralization-inspired drug design, thereby also exploring whether molecular dynamics techniques provide a fine enough tool to observe differences in biomineralization between two highly related/similar biomolecules. Therefore, we chose to model the interaction between calcite and the two epimeric peptides used in [31], namely *R*- and *S*-Sal, both of which are rich in aspartic acid content. More precisely, we parameterized biomineral – peptide systems using already mentioned methodology of Freeman *et al.*,[40] combining AMBER force field description of peptide and solvent

interactions with the parameterization of calcite mineral by Pavese *et al*.[44,45] Finally, to properly investigate the biomineral – biomolecule interactions one needs to carefully choose the mineral surface with which the biomolecule interacts. We thus model two calcite surfaces, first one representing the stable (104) calcite surface, while the polar (001) surface was used to model the growing surface of calcite. We chose the latter surface for this purpose due to the fact that it, together with (012) calcite surface, represents the most relevant growth plane during calcite crystallization,[46-48] and was found to play a crucial role in the biomineralization studies of calcite involving similar additives.[7] To obtain a detailed description of the interactions present in this system we decided to use advanced molecular dynamics technique, namely umbrella sampling, which allowed us to investigate in detail the free energy profiles along the direction normal to the investigated calcite surfaces. We find that the advanced molecular dynamics simulations indeed provide us with a valuable tool for investigation of these complex systems, finding that *R*-Sal epimer binds more strongly to the growing (001) calcite surface compared to its *S*- counterpart, in agreement with the experimental work performed in [31].

4.2. Methodology

4.2.1. Modeling the Peptides

To model the two epimeric peptides we used the Duan et al. *ff03*, thereby staying in agreement with the treatment of the other peptide species studied in this work. While the parameters for aspartic acid and glycine residues are readily available in the *ff03* force field, parameters for the salicylic residue (N-Sal) are missing from the force field. These parameters are thus obtained in a manner consistent with the *ff03* force field. Specifically, the charges were obtained from a restrained (RESP) two-

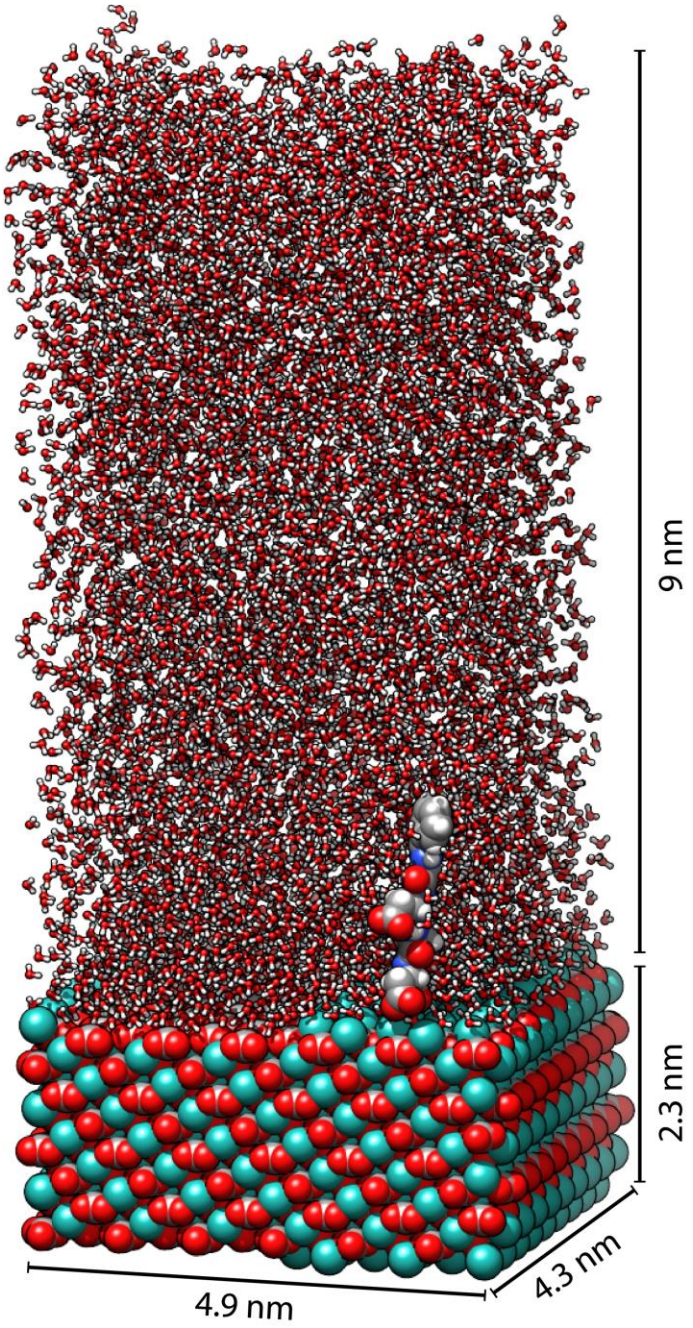


Figure 4-2. Unit cell box of the (001) calcite surface with the *S*-Sal peptide and approximately 8000 water molecules.

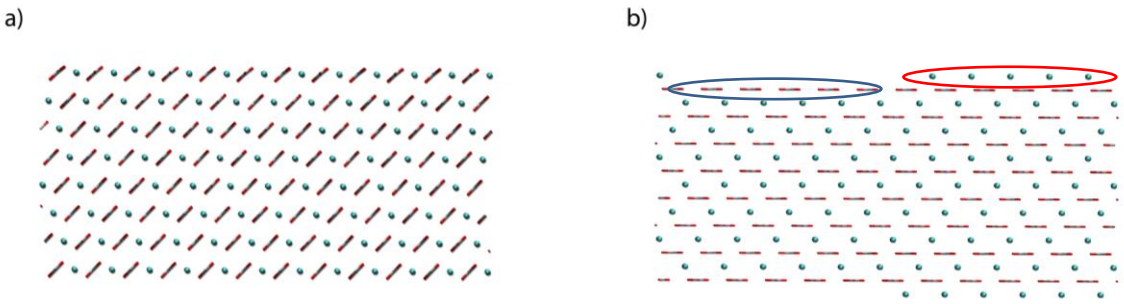


Figure 4-3. Side view of the a) stable (104) and b) polar (001) calcite surface used in this work. The (001) surface contains a valley rich in the carboxyl groups (blue ellipse) and the calcium ion-rich step (red ellipse).

conformer fit to the electrostatic potential. The potential was obtained from the B3LYP/cc-pVTZ//HF/6-31G(d,p) quantum mechanical method with the IEFPCM (ε = 4.335) model representing a (low-polarity) polarizable organic continuum. Prior to calculating the electrostatic potential, both Sal conformers were optimized without any constraints using the HF/6-31G(d,p) level of theory. All aforementioned calculations were performed using Gaussian 09 package.[49]

4.2.2. Modeling the Calcite Surfaces

In this study we investigated adsorption of two epimeric peptides, namely *R*- and *S*-Sal to both (104) and (001) types of calcite surfaces, producing four different simulation systems. The obtained systems contain a calcite slab (laying parallel to the *xy* plane) where the respective surface is in contact with a water layer, whose thickness is set to approximately 9 nm, corresponding to 9000 water molecules. The thickness of the water layer has to be such that one can obtain a thick enough layer of water that shows the converged bulk properties. This bulk layer of water starts at approximately 30 Å from the (001) surface and already at around 10 Å from the stable (104) calcite surface. Above the described water layer a 22 nm thick vacuum layer can be found, used to prevent the interactions of the systems with their periodic image in the *z*-direction, as the periodic boundary conditions were imposed in all three spatial directions (the final setup for the (001) system is shown in Figure 4-2). Moreover, the calcite slabs for the (104) and the (001) surface systems were composed of 960 and 864 $CaCO_3$ formula units respectively, resulting in a surface size of approximately 4.9×4.8 nm^2 and 5.0×4.3 nm^2, and a thickness of approximately 2.5 and 2.4 nm, respectively. The alternating arrangement of Ca^{2+} and CO_3^{2-} ions in the (104) surface makes it overall nonpolar and the most stable calcite surface under normal conditions. This is not the case with the (001) calcite surface, where separate layers of cations and anions produce strong macroscopic dipole moment in the *z*-direction, which can introduce severe artifacts into electrostatic calculations due to the periodic boundary conditions and finite system size. This effect can be accounted for by moving half of Ca^{2+} ions from the upper layer to the bottom layer, thereby removing the strong dipole moment in the direction normal to the (001) surface. After this

treatment one-half of the upper layer of the (001) calcite surface is rich in Ca^{2+} ions, while the negatively charged layer of CO_3^{2-} ions is exposed in the other half (Figure 4-3b).

4.2.3. Modeling the Interaction Between Peptides and Surfaces

Crystal non-bonding parameters were taken from Freeman *et al.*,[40] which are based on the Pavese *et al.* force field,[44,45] while the combination parameters between calcite constituents, more precisely oxygen atoms from the carbonate groups, and the two peptides were constructed using combination rules also proposed in the reference.[40] Buckingham potentials for non-bonding interactions between calcium ions in calcite, and oxygen and nitrogen atoms from the peptides were determined by using the charges found in amino acids and applying them to the appropriate crystal (the $CaCO_3$ crystal for interactions with oxygen, or the Ca_3N_2 crystal for interactions with nitrogen), while rescaling the remaining charges to keep the system neutral. Afterwards we performed a fit with the new charges to obtain the nonbonding potentials necessary to keep the lattice parameters constant. This was performed with the General Utility Lattice Program (GULP).[50] This method for obtaining the missing inorganic-organic parameters was proposed by Schröder *et al*.[51]

4.2.4. Simulation Details

All MD simulations were performed using GROMACS 4.5.5[52] package, with periodic boundary conditions applied in all directions. The Particle Mesh Ewald method (PME) was used with a real space cutoff of 1.2 nm, with the same distance being also set as a cutoff distance for the Lennard-Jones interactions. The temperature in all performed simulations was maintained at 300 K via Langevin dynamics (NVT ensemble) with a time coupling constant of 0.5 ps. The time step in all performed simulations was set to 2 fs. The overall non-zero charge of the simulation box due to the -4 *e* charge of the peptides was neutralized using uniform neutralizing plasma.[52] In the case of the stable (104) calcite surface simulations, the first three layers of calcite were left free to move, while the bottom five layers were kept frozen to obtain the bulk-like properties of the calcite crystal. On the other hand, in the case of the unstable/polar (001) surface, both Ca^{2+} and CO_3^{2-} ions belonging to the upper layers were positionally restrained via a harmonic potential with a force constant of 100 $kJ \cdot mol \cdot Å^{-2}$, while the bottom four layers, similarly to the case of the (104) calcite surface, were maintained frozen. In both cases the snapshots from the simulations were saved every 1 ps.

To calculate the potential of the mean force (PMF) of the peptides on both calcite surfaces we performed a set of umbrella sampling simulations. We chose the distance in the *z*-direction (normal to the surface) between the center of mass of the entire peptide and the surface (the center of the bottom frozen layer) as the reaction coordinate of interest. More specifically, in the case of the (104) surface, the part of the PMF ranging between 5 and 12 Å was generated from 15 umbrella windows, with the spacing between the umbrella windows set to 0.5 Å and a spring constant of 40 $kJ \cdot mol^{-1} \cdot Å^{-2}$. For each

distance from the surface 5 distinct starting conformations were used, and were propagated for 12 ns (the first 2 ns were removed to take into account the equilibration time), which, overall, gave rise to 50 ns production run per umbrella window (750 ns for a production run, in total, per epimer). The remaining part of the PMF (the distance of the center of mass of the peptides is set to lie at approximately 3 to 5 Å from the surface) was generated from 11 umbrella windows, with the spacing between the windows set to 0.2 Å, where the spring constant was set to 100 $kJ \cdot mol^{-1} \cdot Å^{-2}$. The starting conformations for these simulations were generated by pulling the representative structures of *R*- and *S*-Sal (found in the bulk) toward the surface. Each window in this region was propagated for 6 ns (equilibration time accounted for by removing first 2 ns from the simulations), giving rise to an overall time of 44 ns for the production run covering this region of the PMF (histograms produced by the described procedure are shown in Figure 4-5c).

In the case of the (001) calcite surface, for the distances to the surface between 7 and 23 Å, the interval between umbrella windows was set to 0.03 Å, and a spring constant of 250 $kJ \cdot mol^{-1} \cdot Å^{-2}$ was used. For every umbrella window in this region a 12 ns NVT simulation was carried out. For the larger distances (z > 23 Å, representing bulk), a significantly weaker spring constant was used, namely 10 $kJ \cdot mol^{-1} \cdot Å^{-2}$, enabling us to cover distances from 28 to 50 Å utilizing only 10 umbrella windows (the interval between umbrella windows in this region was set to 2.5 Å). The bulk simulations were prepared in a similar fashion to what was used to generate the 4.5 to 12 Å region of the (104) surface case (5 starting conformations per similar distance), with overall 60 ns of simulation time per window. The part of the PMF corresponding to the bound states of peptides (distance to the surface between approximately 4 and 7 Å), was generated in the following way: a) 50 different starting configurations from the umbrella widow at approximately 15 Å distance from the surface were chosen (per peptide), b) unbiased simulations were propagated for 5 ns (all 50 peptide conformations spent at least 3 ns being adsorbed on the surface during the course of the unrestrained simulations), c) restrained simulations of all 50 systems that were obtained from the unrestrained simulations were propagated for 6 ns each, giving rise to a total of 200 ns of simulation time per peptide for the part of the reaction coordinate in which peptides bind to the (001) calcite surface (as previously, equilibration time was accounted for by removing the first 2 ns from the simulations).

Finally, the construction of the free energy profiles for all simulated scenarios was carried out using the weighted histogram analysis method (WHAM).[53,54]

4.3. Results

4.3.1. Attempt with Unbiased MD Simulations

We begin our exploration of the interactions of *R*- and *S*-Sal with the calcite surfaces by employing classical MD simulations, which we used to provide us with the initial insight into the adsorption process occurring between the epimers and the surface. Initially, we attempted to extract useful thermodynamic and structural data of the system without utilizing more advanced and more time consuming biased techniques. The simulation box was set up as explained in the Section 4.2.2, enabling us to propagate MD simulations of both epimers on both calcite surfaces. We thereby

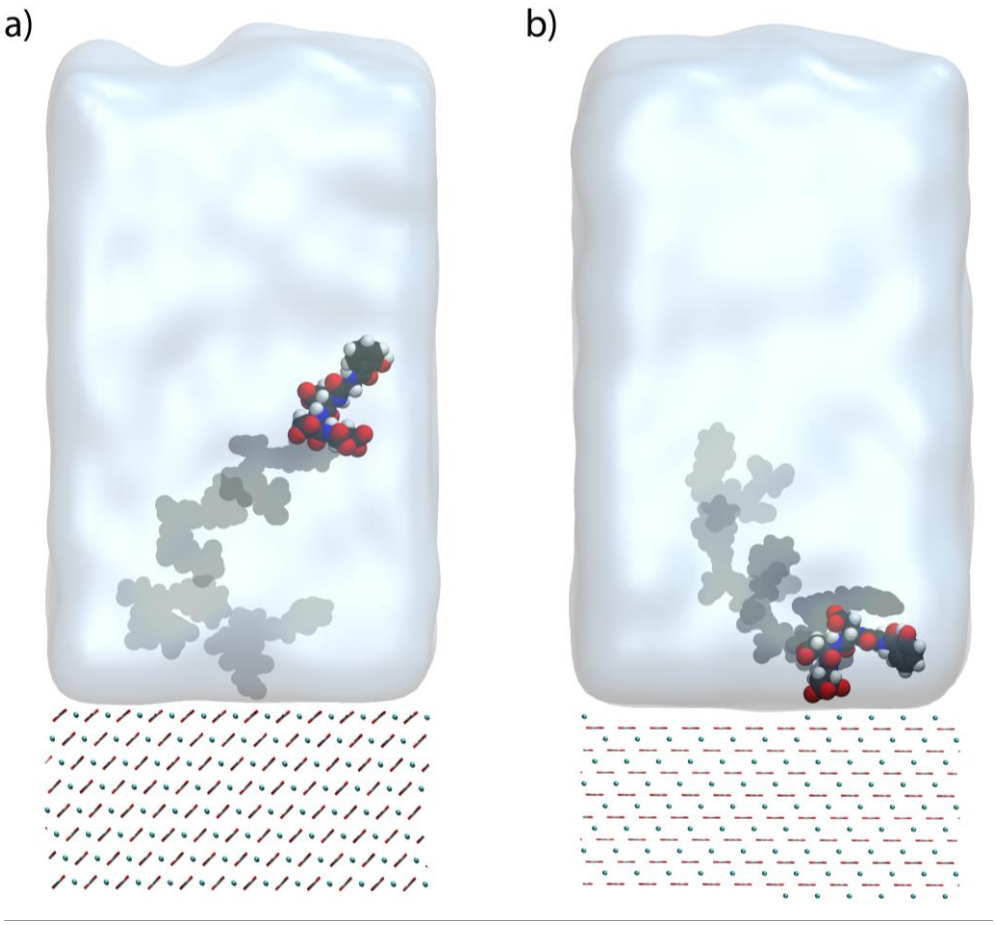


Figure 4-4. a) Snapshots of the representative MD simulation of *R*-Sal (an analogous picture holds for *S*-Sal) on the stable (104) calcite surface. The simulations started close to the calcite crystal, approximately at the calcite-water interface. As the simulation progresses, the peptides move away from the (104) surface, whereby the path traversed by the molecule is represented by the shadowy projections of the peptides (the intensity of the projections follows the progression of the peptide in time). b) The opposite trend is observed when the peptides are exposed to the polar (001) surface, namely, the peptides start in the bulk of water, and are quickly drawn toward the surface.

performed 10 simulations per epimer–surface combination, with peptides being completely unrestrained, in the duration of 15 ns (see Section 4.2.4 for simulation details). The starting positions of the peptides on the stable (104) calcite surface were chosen so that the peptides lie relatively close to the surface, with their centers of mass set to approximately 5 to 10 Å from the surface. The initial conformations were chosen from the trajectories obtained by simulating peptides in water. On the other hand, the epimers were placed at 30 to 40 Å distance from the polar (001) surface. We encountered two drastically different behaviors for the two cases. Namely, neither of the peptides showed preference toward the stable calcite surface, spending in general a rather minute amount of time close to the (104) surface, with a strong tendency toward the bulk of the water (Figure 4-4a). On the other hand, both peptides adsorbed very strongly to the polar (001) surface, pointing to the fact that the binding free energy profiles for both epimers are very steep. We were therefore unable to properly investigate the adsorption process occurring on the stable and polar calcite surfaces using classical MD. In the former case, the peptides tend to spend the majority of the time away from the surface, thereby significantly undersampling the region of interest, while in the latter one, the adsorption process is so dominant that it is not possible to properly sample the behavior of the peptides in the vicinity of the (001) surface – nor to investigate their binding modes to the polar surface, as the peptides become kinetically trapped in the deep minima of the free energy landscape, from which they cannot escape, thereby hindering and biasing the obtained statistics. Thus, even though the classical/unbiased MD simulations gave us a qualitative insight into the behavior of the investigated systems, to adequately research the adsorption process and to elucidate the differences between not just the behaviors of epimers on the two different interfaces, but also between the epimers, it is necessary to apply more advanced sampling technique.

4.3.2. PMFs/Free Energy Profiles

We employed the umbrella sampling method in an attempt to obtain the free energy profiles of binding along the direction normal to the calcite surfaces (along the *z*-axis), for both investigated epimers. We conclusively find that neither peptide binds to the energetically stable (104) surface, but rather that both peptides exhibit a repulsive trend towards the (104) surface, as visible from their respective free energy profiles (Figure 4-5a), with the repulsion between the two growing rapidly when the peptides come at the distance of approximately 5 Å from the surface. The opposite situation is observed in the case of the (001) calcite surface (Figure 4-5b), where free energy profiles of both peptides show a significant and deep minimum appearing at ≈ 5 Å from the (001) surface of calcite, which was already qualitatively indicated by the unbiased MD simulations (Figure 4-4b). More precisely, while both peptides bind extremely strongly to the calcium rich step of the (001) surface ($\Delta G \approx$ -1000 kJ·mol^{-}1 in both cases) (see Figure 4-5b for the depiction of the *S*-Sal and *R*-Sal structures bound to the surface), it is quite obvious that the *R*-Sal binds more strongly to the (001) surface compared to its epimeric counterpart, with the difference between the two being $\Delta\Delta G$(*R*-Sal - *S*-Sal) ≈ -84 kJ·mol^{-1}. In both

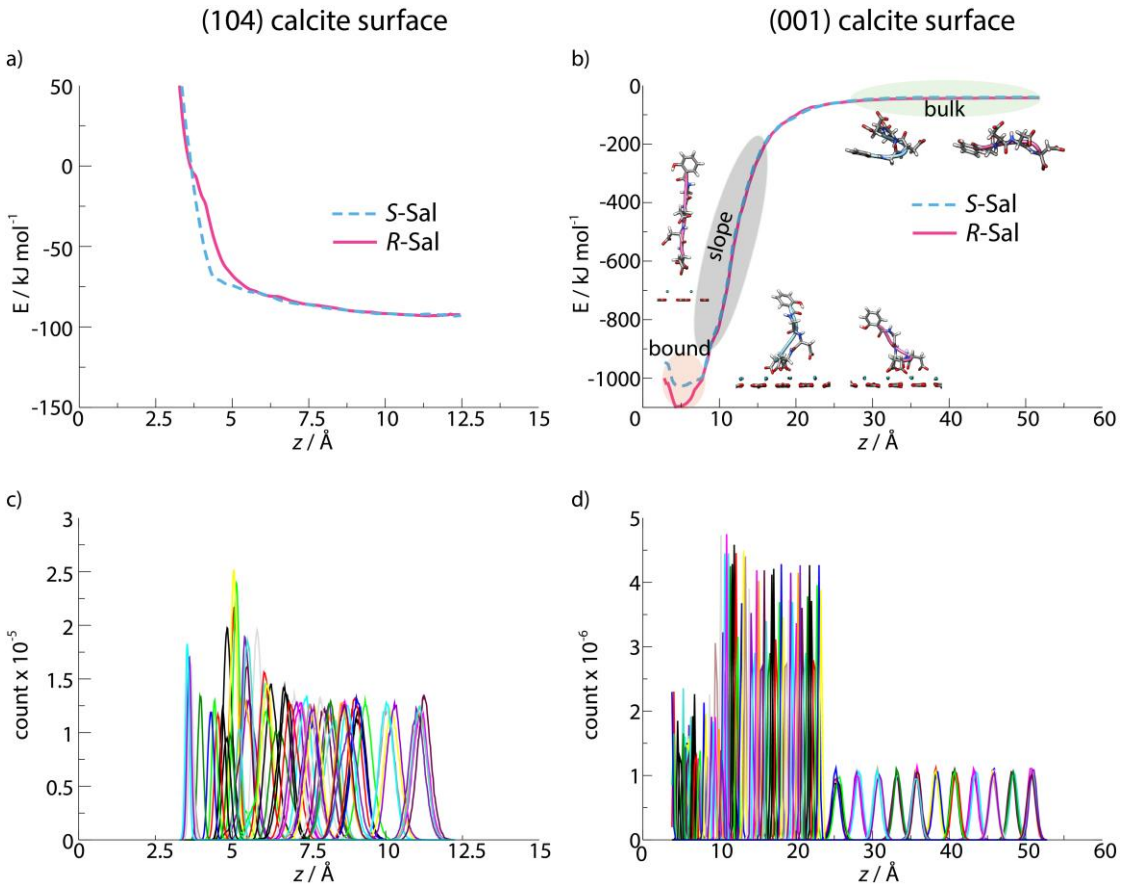


Figure 4-5. a) Free energy profiles of *S*-Sal and *R*-Sal along the normal direction to the (104) calcite and b) (001) calcite surface. Structures of *S*-Sal (blue ribbon) and *R*-Sal (pink ribbon) representing different parts of the free energy profile (bulk, bound, and the slope regions) are additionally shown in b). Histograms used to produce the free energy profiles using the WHAM method for the c) (104) and d) (001) calcite surface.

cases peptides predominantly bind through the interaction of the negatively charged carboxylic groups they possess with each group carrying an effective charge of *-e* (Figure 4-1) with the positively charged (calcium-rich) step of the (001) calcite surface. The interaction between both Sal peptides and (001) surface is so strong that the peptides become completely elongated as they approach the surface (Figure 4-5b, slope part of the PMF), namely the potential between the positively charged surface and negatively charged carboxylic acids in *S*- and *R*-Sal is enough to denature/unfold the peptides from their folded/bulk conformations, on distances as large as 18 Å from the (001) calcite surface. Interestingly, we find that neither peptide exhibits an additional binding mode with the first/second water layer, making them different from the cases of similar negatively charged compounds.[7] We will now delve more deeply in the investigation of the binding process on the (001) surface, firstly by investigating the internal conformational phase spaces of the peptides in bulk and in the bound regions (Figure 4-6).

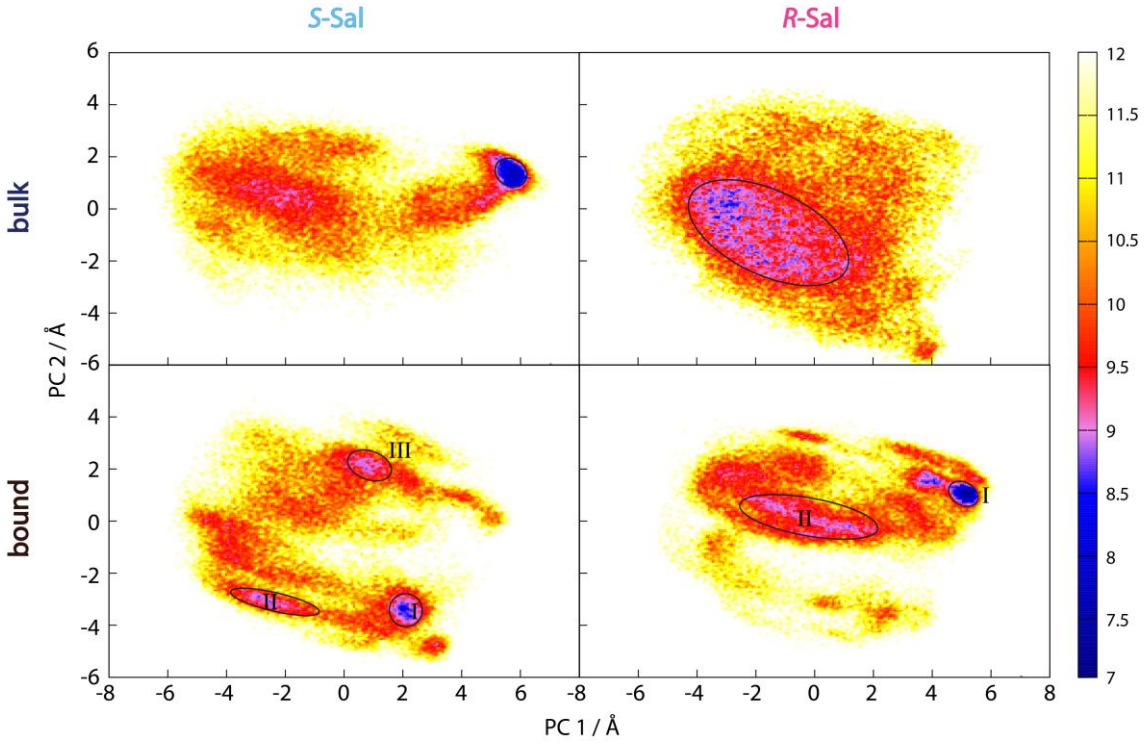


Figure 4-6. Free energy surfaces of *S*-Sal and *R*-Sal (left and right column) in the bulk and bound regions (top and bottom panels, respectively). The most important/predominantly occupied parts of the conformational phase space are highlighted using the ellipses.

4.3.3. Free Energy Surfaces of Peptides in the Bulk and Bound Regions

In order to explore the conformational phase spaces in a comparative fashion, 50000 structures found in the bulk (extracted from the umbrella window set at 45 Å from the surface) and equivalently, the ones gathered from the restrained simulations on the (001) surface (obtained from 50 simulations representing the bound region, generated as explained in the Simulation Details, Section 4.2.4) were combined into a single set of structures, giving overall rise to 200000 structures. This set was used to determine a common set of principal components, based on the positions of α-carbon atoms and an additional carbon atom from the salicylic residue (atom C_s from the salicylic moiety, Figure 4-1). Each structure was then projected onto the first two principal components, which, similarly to the case of Met- and Ada-enkephalins from the third chapter/first scenario of this thesis, could be associated with the end-to-end distance in the peptide and the twist of the backbone, respectively. Upon obtaining the conformational phase spaces we calculated the associated free energy surfaces from the local density of points in the representation of principal components, analogously to the procedure used in the case of enkephalins (Figure 3-17).

We are now in a position to generally comment on the conformational preferences of both epimers in the regions of interest. Firstly, let us focus on the bulk region. While it is possible to observe that both *S*- and *R*-Sal possess a rather wide conformational phase space, with a relatively

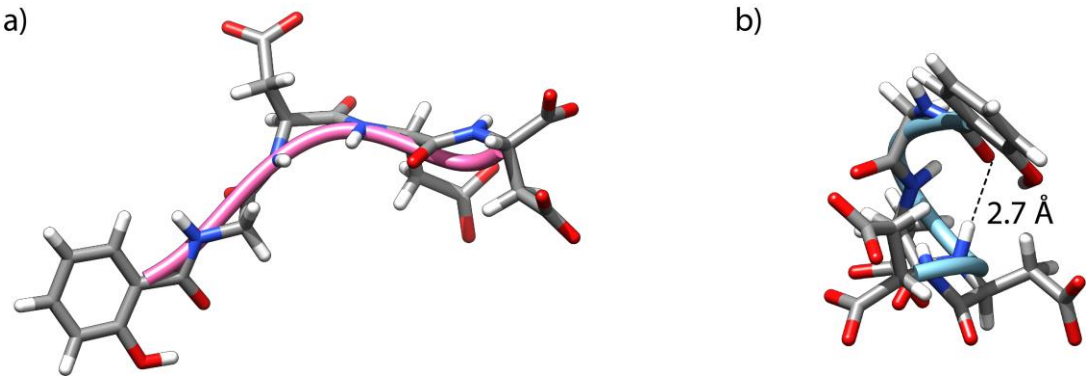


Figure 4-7. Structures representing the minima in the free energy surface of a) *S*-Sal and b) *R*-Sal, as found in the bulk region.

wide range of structures that they can achieve, one can in both cases observe the existence of a more pronounced minima in their respective free energy surfaces. Specifically, more than 55% of *R*-Ada structures are found in a wide shallow minimum denoted with the ellipse (Figure 4-6, *R*-Sal, bulk), which is predominantly occupied with the "open" backbone structures (Figure 4-7a), where the main stabilizing factors are the intermolecular interactions between the carboxyl groups and surrounding water molecules, which are screening the negatively charged carboxylic residues. On the other hand, secondary structures in about 30% of all *S*-Sal conformations belong to an island of conformations denoted with a small ellipse (Figure 4-6, *S*-Sal, bulk, found approximately at PC1 value of 6 Å), giving rise to a deep minimum in the free energy surface of *S*-Sal. This region, which is not visited by *R*-Sal conformations, is predominantly occupied with the so-called quasi α-helical conformations, representative of whose is shown in the Figure 4-7b. Thus, while *S*-Sal exhibits a strong tendency towards α-helical conformations, being stabilized via intramolecular interactions, more precisely, through the $i + 4 \rightarrow i$ type hydrogen bond between Gly2 and Asp5 residues (Figure 4-7b), the majority of *S*-Sal conformations lives in a wider conformational phase space occupied predominantly with the "open" type structures which are stabilized via the intermolecular rather than intramolecular interactions.

We now in a similar fashion investigate the conformational preferences of peptides bound to the (001) surface. We thereby observe that, in the case of both epimers, their behavior tends to show rather drastically different patterns when compared to their predominant conformations in the bulk. More precisely, free energy surface of *S*-Sal can be described by the three predominant and separate minima (Figure 4-6, *S*-Sal, bound, minima denoted with the Roman numerals I-III), neither of which is as strongly pronounced as the quasi α-helical region found in the bulk, with the percentage of structures found in the minima I to III being approximately 15%, 11% and 10% of all structures, respectively (representative structures of these regions are shown in the Figure 4-8a,b,c). Interestingly, although we find that the representative structure of the minimum I (found at PC1 value of about 2, Figure 4-6, *S*-Sal, bound) exhibits the U turn secondary structure (see Figure 4-8a), being

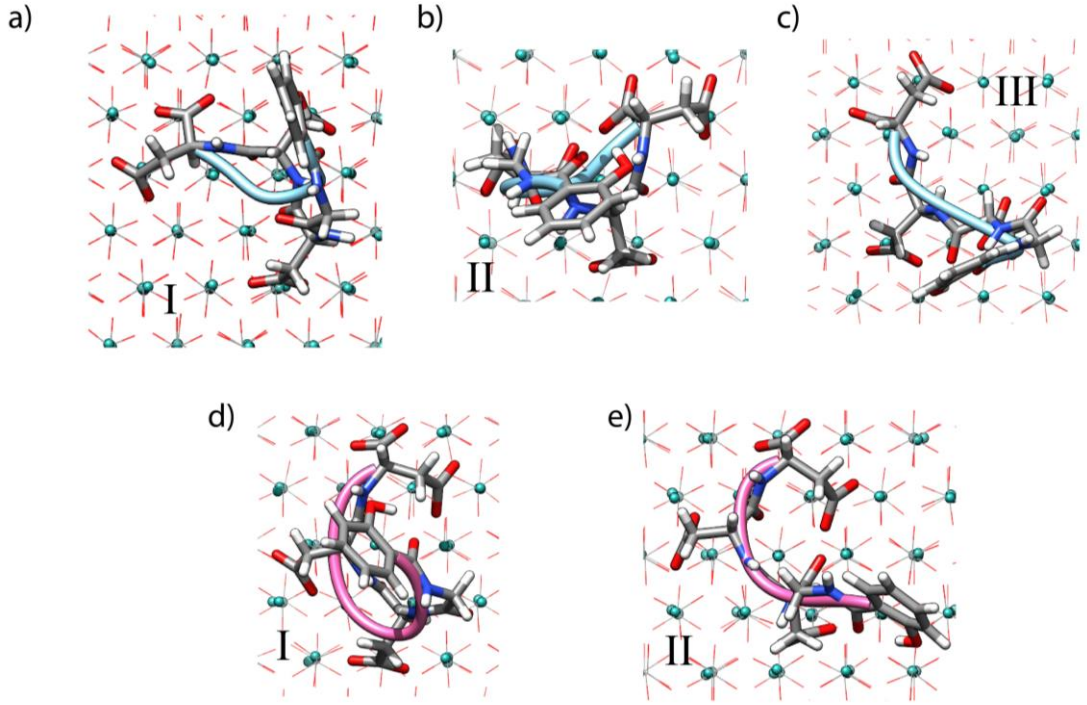


Figure 4-8. Top view on the structures representing the minima in the free energy surface for the bound region of a), b) and c) *S*-Sal and d) and e) *R*-Sal. The Roman numerals denoting each structure correspond to the regions in Figure 4-6.

bound with all four carboxylic groups to the (001) surface, the island of quasi α-helices previously found in the bulk completely vanishes when *S*-Sal binds to the surface, with both minima II and III containing structures similar to the ones found in the already described wide minimum of the bulk region of *R*-Sal (compare Figure 4-7a with the Figure 4-8b and 4-8c). Analogously, conformational behavior of *R*-Sal is changed drastically after binding to the surface, now possessing the free energy surface similar to *S*-Sal surface found in the bulk region (compare *S*-Sal, bulk and *R*-Sal, bound in the Figure 4-6). Namely, we were able to isolate two separate minima in the free energy surface of a bound *R*-Sal, first one closely resembling the minimum in the bulk region of *S*-Sal (minimum I, PC1 value of about 5.5), containing around 20% of all bound *R*-Sal structures (representative structure shown in the Figure 4-8d). The top view shown in the aforementioned figure allows us to observe that the structures indeed tend toward the quasi α-helical types of structures, however, due to the fact that *R*-Sal contains *R*-Asp residue, the particular section of the phase space corresponding to the true α-helices remains inaccessible. Moreover, although the representative structure found in the minimum II (14% of all structures, Figure 4-8e) features a more open backbone structure, it also resembles more to the section of the free energy surface found in the bulk region of *S*-Sal (Figure 4-6, *S*-Sal, bulk) than to the minimum found in its own bulk region. Overall, we surprisingly find that the conformational preferences of the two peptides when bound to the surface "mimic" to a rather large extent the behavior of their epimeric counterparts when in the bulk. While this finding seems unexpected and

somewhat unlikely, the deeper investigation of the bound states will showcase the reason behind this, on the first glance rather strange, behavior.

4.3.4. How do Peptides Bind to the (001) Calcite Surface?

Up to this point we have investigated the internal degrees of freedom of the peptides in the bulk and when bound to the polar (001) calcite surface. However, we are yet to discern the exact nature of the interactions between the peptides and the growing (001) calcite surface. We thus analyzed *R*- and *S*-Sal conformations representing the bound region of the free energy profiles (Figure 4-5 and 4-6) in detail, thereby finding that both epimers bind to the (001) calcite surface predominantly via carboxyl groups, which strongly interact with the positively charged step of this surface through electrostatic interactions. To quantitatively characterize the bound states we inspected the distances between the carbon atoms of four carboxyl groups and the surface (the position of the surface plane was found by averaging the positions of calcium ions found in the calcium-rich step, see Figure 4-2, red ellipse). In this respect we present Table 4-1, showing the average distances of the carbon atoms for all four carboxyl groups from the (001) surface. We observe that, while the behavior of the second (C_2 atom) and the fourth (C_4 atom) carboxyl group is rather similar for both epimers, C_3- and especially C_1-surface distance differs drastically in the case of *R*- and *S*-Sal, with *R*-Sal possessing almost 0.5 Å shorter average distance between first carboxyl group (Asp1) and the surface compared to its epimer (Table 4-1). We inspected this significant difference in more depth by plotting C_n-surface distances, thereby finding that carboxyl groups exhibit overall three distinct binding modes, following the trend shown for Asp1 carboxylic group (C_1-surface distance) of *S*-Sal shown in Figure 4-9. We thereby found that carboxyl groups in the first binding mode are situated exclusively at the distances smaller than 3 Å from the surface, in the case of both epimers and irrespective of the carboxyl group investigated. We were thus in a position to set a cutoff distance for the first bound state of the carboxyl groups with (001) calcite surface at 3 Å, simplifying the bound state analysis and enabling us to discern the predominant difference between the two epimers. More precisely, we were now able to find the fraction of the structures having *n* carboxyl groups simultaneously bound to (001) calcite surface, with *n* ranging from 1 to 4. We thus observed that both epimers bind preferentially with all 4

Table 4-1. Distance of carbon atoms of four carboxyl groups (C_1-C_4 atoms) present in *R*- and *S*-Sal from the (001) surface (see Figure 4-1 for the nomenclature of carbon atoms).

Average C_n-(001) surface distance / Å	*R*-Sal	*S*-Sal
C_1	2.12	2.61
C_2	2.06	2.11
C_3	1.88	2.09
C_4	1.79	1.77

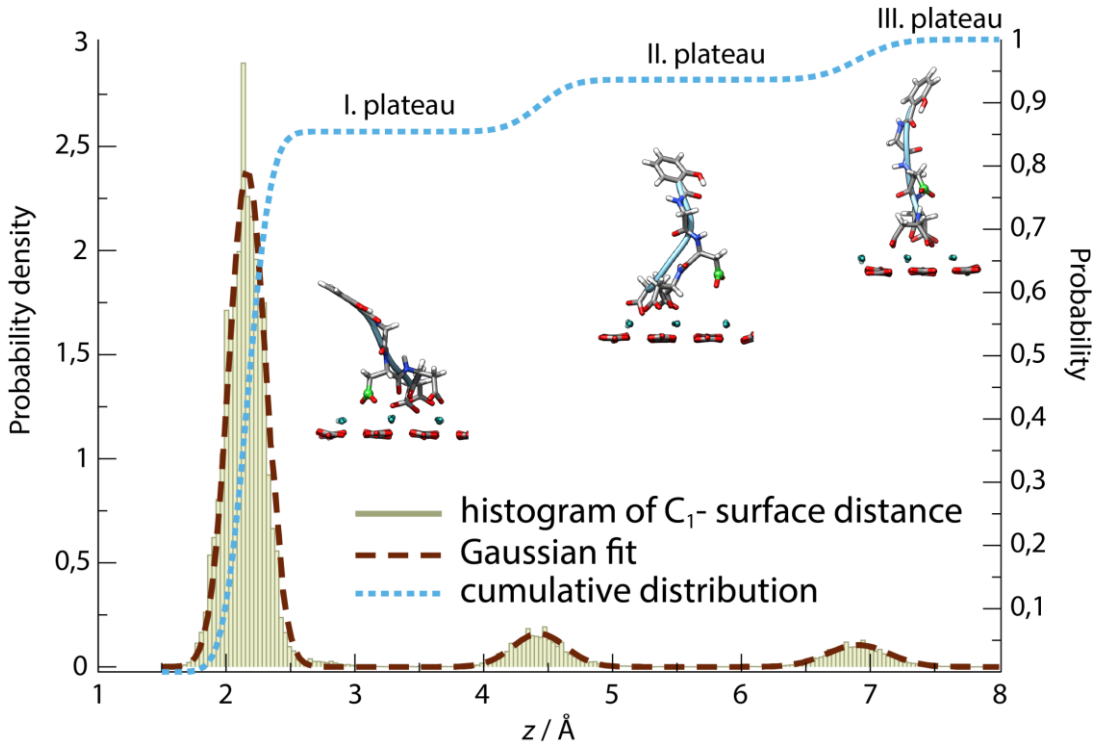


Figure 4-9. Gaussian fit (brown dashed line) of the histogram of C_1-surface distance (green rectangles) for the conformations of *S*-Sal peptide found in the bound region (Figure 4-6). The cumulative distribution of the distances is also shown (blue dotted line). Representative structures corresponding to three observed regions are shown below the three corresponding plateaus, where the inspected atom (C_1, Figure 4-1) is shown in green.

carboxyl groups, with neither epimer being bound with less than 3 carboxyl groups to the growing surface of calcite (Table 4-2). We can now explain the interesting behavior observed in the previous section, where we noticed that both epimers experience a rather large change in their conformational phase space when moving from the bulk into the bound region. In this respect, the fact that both epimers preferentially bind with all four available carboxyl group to the calcite surface puts a heavy restraint on the conformational phase space, as only certain geometries of the epimers enable this type of adsorption. Geometrically, we can test whether a certain structure is eligible to bind with all four carboxyl group to the calcite surface by inspecting the distance of an arbitrary chosen C_n atom from the plane determined by the positions of the three remaining C_n atoms. If the distance of the chosen atom from the constructed plane is larger than 3 Å, the particular structure does not satisfy the geometric requirement, thus being unable to form four bonds with the calcite surface. In this respect, we found that in both cases only a marginal percentage of peptide structures in the bulk (< 2%) fulfils this requirement, implying that the conformational phase space of both epimers needs to change drastically in order to enable this preferential binding of the epimers with the (001) calcite surface, thereby explaining the results shown in Section 4.3.3.

Table 4-2. Percentage of bound *R*-Sal and *S*-Sal structures that bind through *n* carboxyl groups, with *n* ranging from 1 to 4.

Number of carbonate groups bound to the (001) surface	*R*-Sal	*S*-Sal
1	-	-
2	-	-
3	4 %	16 %
4	96 %	84 %

Finally, using the performed analysis of the bound states we can shed light on the observed difference in the free energies of binding of the two epimers on the (001) calcite surface (Figure 4-5), where we found that *R*-Sal binds more strongly ($\Delta\Delta G$(*R*-Sal - *S*-Sal) ≈ -84 kJ·mol^{-1}). From Table 4-2 we can see that, although both epimers tend toward binding modes in which all four carboxyl groups bind to the calcite surface, *R*-Sal structures show higher preference towards this binding mode, with 96% of all its structures being adsorbed through all four carboxyl group to the (001) surface, with 12% less conformations *S*-Sal conformations belonging to the same binding pattern. In particular, the difference stems mostly from the behavior the carboxyl group of Asp1 residue, with 16% of all *S*-Sal structures not being bound to calcite surface via this group (see Figure 4-9), which can also be observed from the average distance of C_1 atoms from the surface (see Table 4-1, *S*-Sal), whereas only 4% of bound *R*-Sal structures exhibits this behavior. We can thus tentatively conclude that the observed difference in the free energies of binding of *R*-Sal and *S*-Sal stems from the larger enthalpic contribution to the binding process, which in turn arises due to the stronger average electrostatic interaction formed between *R*-Sal and the (001) calcite surface compared to its epimer.

4.4. Conclusion

We investigated the adsorption behavior of two epimeric peptides, namely *R*- and *S*-Sal, on calcite surfaces using advanced molecular dynamics simulations, namely umbrella sampling, enabling us to obtain the free energies of binding of the two peptides. We found that both *R*- and *S*-epimer adsorb very strongly to the calcite (001) surface, both directly binding to the calcite surface via either three of four available carboxyl groups with the Ca^{2+} rich part of the mineral surface. The adsorption free energy of both peptides on the calcite (001) surface is around 1000 kJ·mol^{-1}, with the binding of *R*-Sal being more favorable for about 85 kJ·mol^{-1}. We explain the observed difference in the free energies of binding by finding that the *R*-epimer has a stronger electrostatic interaction with the (001) calcite surface, exhibiting predominantly (96% of all bound structures) the adsorption mode in which all four carboxyl groups interact with Ca^{2+} rich surface, while approximately 16% of *S*-Sal structures are adsorbed through only three carboxyl groups. Interestingly, while studies of similar negatively charged compounds, such as PSS, showed the existence of two binding modes on the calcite (001) surface, one of them being the solvent-mediated one, in which the negatively charged moiety forms very strong hydrogen bonds with water molecules in the first and second solvation layers above calcite surface, thereby giving rise to an additional minimum in the free energy profile,[7] no such behavior was observed in the case of *R*- and *S*-Sal peptides. In contrast to the (001) case, neither peptide binds to the stable (104) calcite surface, whereby both peptides show a monotonously rising repulsive potential with respect to the (104) surface.

While we have not been able to directly tackle important chemical and physical processes readily occurring on the calcite surface, namely nucleation and crystallization, our study serves to show that advanced MD simulations can indeed provide valuable insight into both the thermodynamic and the structural information governing the adsorption process of biomolecules on inorganic interfaces, thereby unveiling the chemo-physical driving forces behind biomineralization. In this respect, our finding that *R*-epimer binds more strongly to the growing (001) surface agrees very well with the very recently conducted experiments, where it was found that *R*-epimer more strongly inhibits the growth of the calcite mineral compared to its epimeric counterpart.[31] This study serves to show that a difference in a single chiral center can change conformational preferences of the investigated compounds, thereby giving rise to small differences in their binding/adsorption patterns, in turn being able to change the quantitative aspects of biomineralization. Overall, our study is the first theoretical work directly suggesting that chirality can indeed be used as a fine-tuning tool via which one can control the vitally important phenomenon of biomineralization.

4.5. References

(1) Ukrainczyk, M.; Gredičak, M.; Jerić, I.; Kralj, D. *J. Colloid Interface Sci.* **2012**, *365*, 296.

(2) Fujiwara, M.; Shiokawa, K.; Morigaki, K.; Zhu, Y; Nakahara, Y. *Chem. Eng. J.* **2008**, *137*, 14.

(3) Wang, C.; He, C.; Tong, Z.; Liu, X.; Ren, B.; Zeng, F. *Int. J. Pharm.* **2006**, *308*, 160.

(4) Ueno, Y.; Futagawa, H.; Takagai, Y.; Ueno, A.; Mizushima, Y. *J. Controlled Release* **2005**, *103*, 93.

(5) Ogomi, D.; Serizawa, T.; Akashi, M. *J. Controlled Release* **2005**, *103*, 315.

(6) Gopal, K.; Lu, Z.; de Villiers, M. M.; Lvov, Y. *J. Phys. Chem. B* **2006**, *110*, 2471.

(7) Shen, J.-W.; Li, C.; van der Vegt, N. F. A.; Peter, C. *J. Phys. Chem. C* **2013**, *117*, 6904.

(8) Nassralla-Aboukais, N.; Boughriet, A.; Fischer, J. C.; Wartel, M.; Langelin, H. R.; Aboukais, A. *J. Chem. Soc.* **1996**, *92*, 3211.

(9) Katz, J. L.; Reick, M. R.; Herzog, R. E.; Parsiegla, K. I. *Langmuir* **1993**, *9*, 1423.

(10) Compton, R. G.; Brown, C. A. J. *Colloid Interface Sci.* **1994**, *165*, 445.

(11) Brečević, L.; Nothig-Laslo, V.; Kralj, D.; Popović, S. *J. Chem. Soc.* **1996**, *92*, 1017.

(12) Deleuze, M.; Brantley, S. L. *Geochim. Cosmochim. Acta* **1997**, *61*, 1475.

(13) Pasarin, I. S.; Yang, M.; Bovet, N.; Glyvradal, M.; Nielsen, M. M.; Bohr, J.; Feidenhans'l, R.; Stipp, S. L. S. *Langmuir* **2012**, *28*, 2545.

(14) Orme, C. A.; Noy, A.; Wierzbicki, A.; McBride, M. T.; Grantham, M.; Teng, H. H.; Dove, P. M.; DeYoreo, J. J. *Nature* **2001**, *411*, 775.

(15) Chen, C. L.; Qi, J. H.; Zuckermann, R. N.; DeYoreo, J. J. *J. Am. Chem. Soc.* **2011**, *133*, 5214.

(16) Metzler, R. A.; Tribello, G. A.; Parrinello, M.; Gilbert, P. U. P. A. *J. Am. Chem. Soc.* **2010**, *132*, 11585.

(17) Wang, X. Q.; Wu, C. M.; Tao, K.; Zhao, K.; Wang, J. Q.; Xu, H.; Xia, D. H.; Shan, H. H.; Lu, J. R. *J. Phys. Chem. B* **2010**, *114*, 5301.

(18) Marin-Garcia, L.; Frontana-Uribe, B. A.; Reyes-Grajeda, J. P.; Stojanoff, V.; Serrano-Posada, H. J.; Moreno, A. *Cryst. Growth Des.* **2008**, *8*, 1340.

(19) Masica, D. L.; Schrier, S. B.; Specht, E. A.; Gray, J. J. *J. Am. Chem. Soc.* **2010**, *132*, 12252.

(20) Sigel, A.; Sigel, H.; Sigel, R. K. O. *Biomineralization. From Nature to Application*; Wiley: New York, 2008; Vol. 4.

(21) Mann, S. *Nature* **1988**, *332*, 119.

(22) Jiang, H.; Liu, X.-Y.; Lim, C. T.; Hsu, C. Y. *Appl. Phys. Lett.* **2005**, *86*, 163901.

(23) Schrier, S. B.; Sayeg, M. K.; Gray, J. J. *Langmuir*, **2010**, *27*, 11520.

(24) Elhadj, S.; De Yoreo, J. J.; Hoyer, J. R.; Dove, P. M. *Proc. Natl. Acad. Sci. U.* **2006**, *103*, 19237.

(25) Elhadj, S.; Salter, E. A.; Wierzbicki, A.; De Yoreo, J. J.; Han, N.; Dove, P. M. *Cryst. Growth Des.* **2006**, *6*, 197.

(26) Mann, S.; Didymus, J. M.; Sanderson, N. P.; Heywood, B. R.; Samper, E. J. A. *J. Chem. Soc.* **1990**, *86*, 1873.

(27) Wei, H.; Shen, Q.; Zhao, Y.; Wang, D.; Xu, D. *J. Cryst. Growth* **2004**, *260*, 545.

(28) Chien, Y. C.; Hincke, M. T.; Vali, H.; McKee, M. D. *J. Struct. Biol.* **2008**, *163*, 84.

(29) Fu, G.; Valiyaveettil, S.; Wopenka, B.; Morse, D. E. *Biomacromolecules* **2005**, *6*, 1289.

(30) Lin, A.; Meyers, M. A. *Mater. Sci. Eng. A* **2005**, *390*, 27.

(31) Ukrainczyk, M.; Gredičak, M.; Jerić, I.; Kralj, D. Unpublished results.

(32) Sun, Y., Dominy, B.N. and Latour, R. A. *J. Comput. Chem.* **2007**, *28*, 1883.

(33) Wierzbicki, A.; Dalal, P.; Cheatham III, T. E.; Knickelbein, J. E.; Haymet, A. D. J.; Madura, J. D. *Biophys. J.* **2007**, *93*, 1442.

(34) Raut, V. P.; Agashe, M. A.; Stuart, S. J.; and Latour, R. A. *Langmuir* **2005**, *21*, 1629.

(35) Gerbaud, V.; Pignol, D.; Loret, E.; Bertrand, J. A.; Berland, Y.; Fontecilla-Camps, J. C. *J. Biol. Chem.* **2000**, *275*, 1057.

(36) Huq, N. L.; Cross, K. J.; Reynolds, E. C. *J. Mol. Modeling* **2000**, *6*, 35.

(37) Aschauer, U.; Spagnoli, D.; Bowen, P.; Parker, S. C. *J. Colloid Interface Sci.* **2010**, *346*, 226.

(38) Titiloye, J. O.; Parker, S. C.; Mann, S. *J. Cryst. Growth* **1993**, *131*, 533.

(39) Duffy, D. M.; Harding, J. H. *Langmuir* **2004**, *18*, 7630.

(40) Freeman, C. L.; Harding, J. H.; Cooke, D. J.; Elliott, J. A.; Lardge, J. S.; Duffy, D. M. *J. Phys. Chem. C* **2007**, *32*, 11943.

(41) Freeman, C. L.; Harding, J. H.; Quigley D.; Rodger, P. M. *J. Phys. Chem. C* **2011**, *115*, 8175.

(42) Freeman, C. L.; Harding, J. H.; Quigley D.; Rodger, P. M. *Phys. Chem. Chem. Phys.* **2012**, *28*, 7287.

(43) Yang, M.; Rodger, P. M.; Harding, J. H.; Stipp, S. L. S. *Mol.Simul.* **2009**, *35*, 547.

(44) Pavese, A.; Catti, M.; Price, G. D.; Jackson, R. A. *Phys. Chem. Miner.* **1992**, *19*, 80.

(45) Pavese, A.; Catti, M.; Parker, S. C.; Wall, A. *Phys. Chem. Miner.* **1996**, *23*, 89.

(46) Weiner, S.; Addadi, L. J. *Mater. Chem.* **1997**, *7*, 68.

(47) Addadi, L. I. A.; Politi, Y.; Nudelman, F.; Weiner, S. *Design Strategies and Mechanisms of Mineral Formation: Operating at the Edge of Instability*; Springer: Netherlands, 2008.

(48) Berman, A.; Ahn, D. J.; Lio, A.; Salmeron, M.; Reichert, A.; Charych, D. *Science* **1995**, *269*, 515.

(49) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J.

B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg,J.J.;Dapprich,S.;Daniels,A.D.;Farkas,O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, revision A.02*; Gaussian, Inc.: Wallingford, CT, 2009.

(50) Gale, J. D. *J.C.S. Faraday Trans.* **1997**, *93*, 629.

(51) Schröder, K.-P.; Sauer, J.; Leslie, M.; Catlow, C. R. A.; Thomas, J. M. *Chem. Phys. Lett.* **1992**, *188*, 320.

(52) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435.

(53) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13* , 1011.

(54) Hub, J. S.; de Groot, B. L.; van der Spoel, D. *J. Chem. Theory Comput.* **2010**, *6*, 3713.

5. Conclusion and Outlook

In this thesis we used *state-of-the-art* theoretical methods to investigate two cornerstones of modern chemistry, namely molecular flexibility and chirality, and the interplay between the two. In this respect, we tackled two important scenarios, whereby we investigated the behavior of flexible and chiral moieties both in solution and at the interfaces of inorganic surfaces.

Firstly, we investigated the phenomenon of CD spectroscopy, which represents a potentially powerful tool for structural characterization of optically active chiral molecules. As a CD spectrum arises as an ensemble average, in order to properly understand and assign experimental spectral features of flexible molecular species, a theoretical framework becomes necessary. We thus developed a methodology which combines *advanced classical molecular dynamics simulations*, with which one generates the conformational phase space of a molecule, *statistical tools*, used to characterize the phase space and extract conformations representing the molecular ensemble, and *quantum calculations*, via which one obtains the CD spectra of small and highly flexible chiral molecules. Using our methodology we were able to determine the absolute configuration of two organic compounds, which are presented in Chapter 1 (Case 1). Moreover, and in spite of inherent complexity, we were able to obtain good agreement between theory and experiment in the case of highly flexible pentapeptides (Chapter 3, Case 2), which represent an extremely difficult test for the methodology, in terms of both classical simulations and quantum calculations.

Our methodology will benefit from the development of novel, more exact TD-DFT functionals, which should describe the excited states with a higher quality. In this respect, the excited-state calculations should also gain from the application of multireference methods (such as DFT/MRCI and MRMP2) and/or more sophisticated *ab initio* methods, which are finally becoming feasible for larger molecules. Moreover, developments of even more precise atomistic force fields should also enable us to obtain even finer details with regard to the conformational phase space of a molecule, thus generating more precise and improved molecular ensembles. The increase in computer power will enable inclusion of solvent effects quantum mechanically, which could additionally improve the performance of our methodology. Taken together, these improvements can only give rise to a better agreement between the theory and the experiment. Perhaps most importantly, as our methodology represents a completely general procedure of ensemble generation and characterization for highly flexible molecules, becoming specialized only at the step involving CD spectra calculations, one can readily use it to model a general experimental observable arising from the ensemble average, such as UV, VCD or NMR spectra. As such an approach can be beneficially used by the larger scientific community; we can conclude that our methodology represents a step in the right direction, enabling us to link conformational phase space and experimentally observable spectrum of flexible chiral molecules for the first time.

In the second part of this thesis we investigated the phenomenon of biomineralization and biomineralization-inspired drug design, on the molecular-level, using advanced classical molecular dynamics simulations, encompassing extensive *umbrella sampling*. We focused on the calcite – peptide composite systems, representing the most promising bioinorganic systems for pharmaceutical purposes, possessing a unique combination of properties, such as high activity and specificity, biocompatibility and nontoxicity, which make them ideal candidates for controlled drug release in humans. Nevertheless, the chirality and flexibility, underlying the interactions between peptides and calcite surfaces, which ultimately lead to the bioactivity of the obtained composite, are rather scarcely investigated by theoretical means. We thus chose to model the interaction of two epimeric peptides which have been shown experimentally to strongly interact with calcite, thereby significantly inhibiting and modulating its growth. The composites of the chosen peptides with calcite serve as a prototypical example of novel biodegradable drug delivery systems, with calcite biomineral playing the role of a carrier, while the peptides represent both the drug (N-terminal salicylic acid) and the linker through which the drug adheres to the inorganic carrier. We found that, using theoretical methods, it is not only possible to discriminate between the adsorptive behavior of the peptides on two distinct surfaces, namely stable (104) and growing (001) calcite surface, but that it is also possible to arrive at an explanation of the experimentally observed differences in the activity/growth inhibition exhibited by the two peptides. We thereby found that rather small differences, such as the change in the chirality of only one constituent amino acid, changes the conformational behavior/flexibility of the peptide to an extent significant enough to induce different binding patterns and interactions on mineral surfaces, leading to an overall different qualitative and quantitative adsorption of active biomolecules/peptides. This in turn enabled us to illuminate the roles of flexibility and chirality on biomineralization and biomineralization-inspired drug design.

Even though we successfully applied *state-of-the-art* theoretical methods to the problem at hand, our study represents only a first stepping stone toward a complete theoretical understanding of the phenomenon of biomineralization and its applications in the form of targeted drug delivery systems. This is most obviously reflected in the fact that our study was unable to directly address the crystal growth and the process of nucleation. In this respect, one could model growing islands of different sizes (and belonging to different surface terminations) on the stable surface of calcite, which would enable the investigation of biomineral – active biomolecule interactions on model systems more similar to the mineral/calcite morphologies that occur in nature. Furthermore, to properly understand the chemistry occurring on the surface of a biomineral, which potentially governs a number of aspects of biomineralization and/or drug delivery capabilities of these biocomposites, one could consider modeling these systems using the so-called reactive force fields, which, although computationally very expensive, account for explicit bond formation and/or bond breaking and thus enable a significantly deeper insight into the growing and nucleation phenomena occurring on the mineral surface. While it is not yet computationally fully feasible to tackle systems of this size and complexity, a

natural extension of the aforementioned approach comes in terms of either QM/MM hybrid methods or pure quantum molecular dynamics simulations, which would even more accurately describe the chemistry and the desired interactions, being already successfully employed in the cases of smaller and less complex systems. While the proposed molecular dynamics methods enable us to obtain both structural and thermodynamic information about the investigated biomineral – biomolecule complexes, it becomes necessary, in the end, to compare them with the experimental measurements. In this respect, we find that the combination of the two scenarios investigated in this thesis represents an ideal starting point for further research. Namely, as spectroscopic approaches, including CD spectroscopy, represent one of the most sensitive methods for probing the structural changes and properties of chiral biomolecules, one could conceive its application to interfacial problems. Thus, by calculating spectra as a function of distance from the surface one would, on one hand, build a theoretical monitoring „device“ for the process of biomineralization and biomineralization-inspired drug design while, on the other hand, enabling a direct comparison of the theory with the experiment.

The results presented in this thesis illustrate the power of molecular modelling, which, despite its limitations, is able to successfully tackle problems of a wide range and depth, arising from the fields of physics, chemistry and biology. We find that the combination of theory and experiment is not just beneficial, but rather necessary. On one hand, theory enables the interpretation of the experimental measurements, which often cannot be obtained in an unambiguous manner. On the other hand, experiment both validates and is used to further develop existing models and theories. With the rise of computational power, enabling more accurate and faster calculations, the interplay between the two can only become more spectacular, in turn enhancing our understanding of the universe and of the pale blue dot on which we live. There is indeed (circularly polarized) light at the end of the tunnel.

6. Appendix

A.1. CD spectra of Met-enkephalin and Ada-enkephalin with an N:Z ratio of 4:1

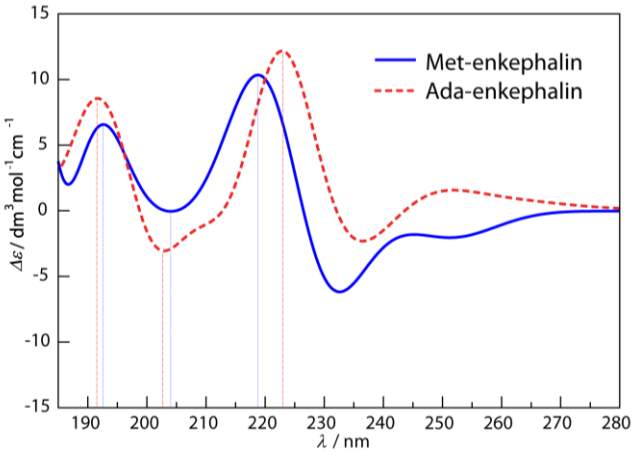
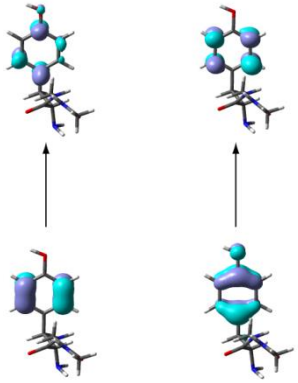
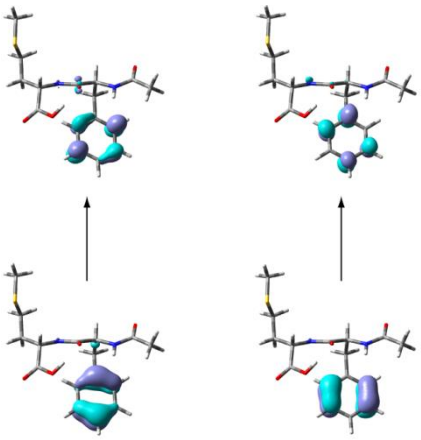
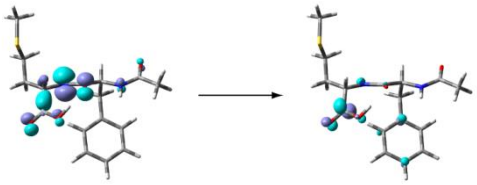
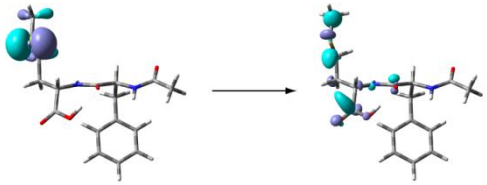
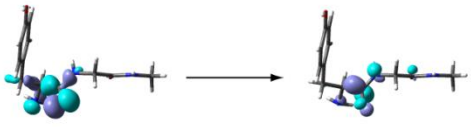
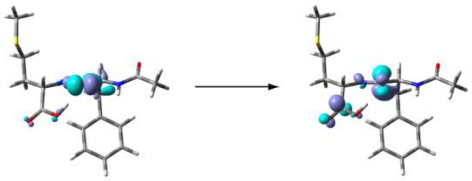
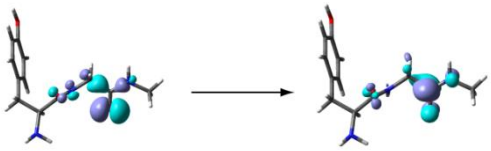
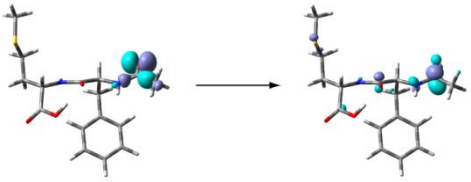
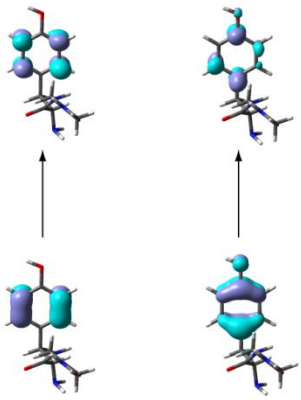
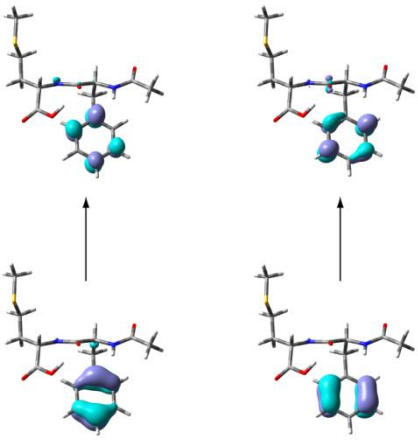
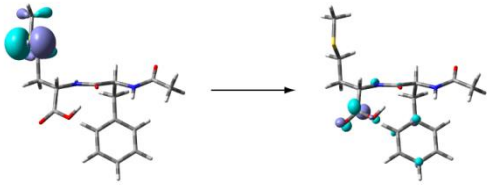


Figure A-1. Met-enkephalin and Ada-enkephalin CD spectra, obtained analogously to those shown in Figure 3-27a, the only difference is that in this case the ratio of neutral to zwitterionic forms is set to 4:1.

A.2. Orbital depiction of the main transition in model systems.

Table A-1. Orbital presentations of the investigated excitations.

Excited state	Orbital presentation
Tyrosine ^1L_b	
Phenylalanine ^1L_b	
$n_O(COOH) + \pi(F\text{-}M) \rightarrow \pi^*(COOH)$	
$n_S(M) \rightarrow n_S(M)^*$	
$n_O(Y) \rightarrow \pi^*(Y\text{-}G)$	

$n_O(F) \rightarrow \pi^*(F\text{-}M)$	
$n_O(G) \rightarrow \pi^*(G\text{-}Nme)$	
$n_O(Ace) \rightarrow \pi^*(Ace\text{-}F)$	
Tyrosine ^1L_a	
Phenylalanine ^1L_a	
$n_S(M) \rightarrow \pi^*(COOH)$	

7. Additional Material

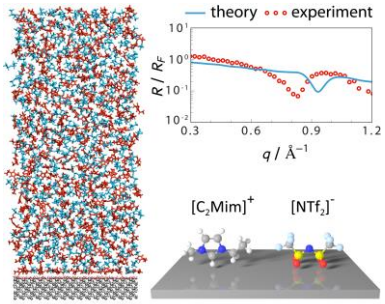
Here we present additional work performed during this doctorate, namely the combined theoretical and experimental study of ionic liquids at solid interfaces. We thereby include an article published in the *Journal of Physical Chemistry Letters*, where we used molecular dynamics simulations to investigate the behavior of an imidazolium-based ionic liquid at a neutral sapphire interface by a means of a direct comparison of the theoretical results with the experimentally measured X-ray reflectivity.

Complementary Molecular Dynamics and X-Ray Reflectivity Study of an Imidazolium-Based Ionic Liquid at a Neutral Sapphire Interface

Zlatko Brkljača,[†,Δ] Michael Klimczak,[‡,Δ] Zoran Miličević,[†] Matthias Weisser,[‡] Nicola Taccardi,[ǂ] Peter Wasserscheid,[ǂ] David M. Smith,[§,ǁ,*] Andreas Magerl,[‡] Ana-Sunčana Smith[†,§]

[†]Institute for Theoretical Physics, FAU Erlangen-Nürnberg, Staudtstrasse 7, Erlangen, 91058, Germany
[‡]Crystallography and Structural Physics, FAU Erlangen-Nürnberg, Staudtstrasse 3, Erlangen, Germany
[ǂ]Chemical Reaction Engineering, FAU Erlangen-Nürnberg, Egerlandstrasse 3, Erlangen, Germany
[§]Ruđer Bošković Institute, Bijenička 54, 10000, Zagreb, Croatia
[ǁ]Center for Computational Chemistry, FAU Erlangen-Nürnberg, Nägelsbachstrasse 25, Erlangen, Germany

[Δ] both authors contributed equally to the manuscript



ABSTRACT: Understanding the molecular-level behavior of ionic liquids (ILs) at IL-solid interfaces is of fundamental importance with respect to their application in, e.g., electrochemical systems and electronic devices. Using a model system, consisting of an imidazolium-based IL ($[C_2Mim][NTf_2]$) in contact with sapphire substrate, we have approached this problem using complementary experimental and theoretical methodologies. We employed high-resolution X-ray reflectivity measurements, capable of probing buried IL-solid interfaces, and atomistic molecular dynamics (MD) simulations, which can be used to interpret experimental data in atomistic detail. Our strategy enabled us to compare experimental and theoretically calculated reflectivities in a direct manner, thereby critically assessing the applicability of force-field variants we implemented. On the other hand, using the best-matching MD description, we are able to describe the nature of the model IL-solid interface in appreciable detail. More specifically, we find that characteristic interactions between the surface hydroxyl groups and donor and acceptor sites on the IL constituents have a dominant role in inducing a multidimensional layering profile of the cations and anions.

KEYWORDS: IL-solid interface, double layer, atomistic molecular dynamics, simulation, 2D ordering

Ionic liquids (ILs) are molten salts which, by common definition, have a melting point below 100 ºC.[1,2] While known for a century,[3] interest in these compounds has exploded over the last decade, when it was found that ILs are suitable for a large range of applications,[4] including gas storage,[5] electrodeposition,[6] biocatalysis,[7] and as lubricants.[8] ILs have been found to be especially promising in the field of electrochemistry,[9] as they exhibit a wide and selective electrochemical window and can be tailored to specific requirements by pairing different ions.[10] In this respect, one can tune the physical properties of ILs, such as thermal stability, viscosity and conductivity to obtain a task-specific IL with significant potential for application in batteries/fuel cells, electrochemical double-layer capacitors and conductive composite materials. As the behavior of ILs at the boundary between the solid and liquid phases plays a crucial role in the development of novel, small, more powerful electronic devices, one of the major goals in recent years has been to obtain a better understanding of the properties of this interface.[11-14]

Ionic liquid–solid interfaces have been investigated with a number of different surface-sensitive techniques, including X-ray[15,16] and neutron reflectivity,[17] surface force measurements[18,19] and XPS.[20] While molecular-level layering, which defines the electrochemical properties of these interfaces, was observed using all of the aforementioned techniques, only X-ray reflectivity can both access the buried solid substrate – bulk liquid interface and define the structural chemistry at the interface with submolecular accuracy.[15] However, X-ray results are usually interpreted with the aid of parameterized structure models, which give rise to a "best fit model" paradigm. This indirect approach yields good qualitative and quantitative results when the behavior of the IL of interest is reasonably well-understood and the structures arising at the interface are simple. However, with more complex systems and limited initial information on the expected structures, results can become ambiguous and several physically sound models, or several parameterizations of the same model, can be found to adequately describe the experimental data.[21] Thus, while general features such as layer thickness are usually unambiguously described across the models, and can therefore be extracted with confidence, the detailed molecular structure of the interface is more difficult to obtain.

Classical molecular dynamics (MD) simulations are a powerful tool for the investigation of complex systems in atomistic detail. This technique has been readily used to investigate the behavior of "pure" ILs,[22] either as a complement to experiments or predictively, with demonstrated success for properties such as densities, diffusion coefficients, enthalpies of vaporization and viscosities, which are accurate to within a few percent of experimental values over an extended temperature range.[23-26] Similarly, classical simulations have been used to understand prospective applications of ILs in combination with molecular liquids,[27] the interaction of ILs with carbon nanotubes and fullerenes,[28-31] and the liquid/vapor interface of ILs.[32-35]

The more idealized interfaces between planar solids and ILs have also been investigated using MD simulations.[11] Because of the importance of electrical double layer capacitors, much of this work has been focused on charged

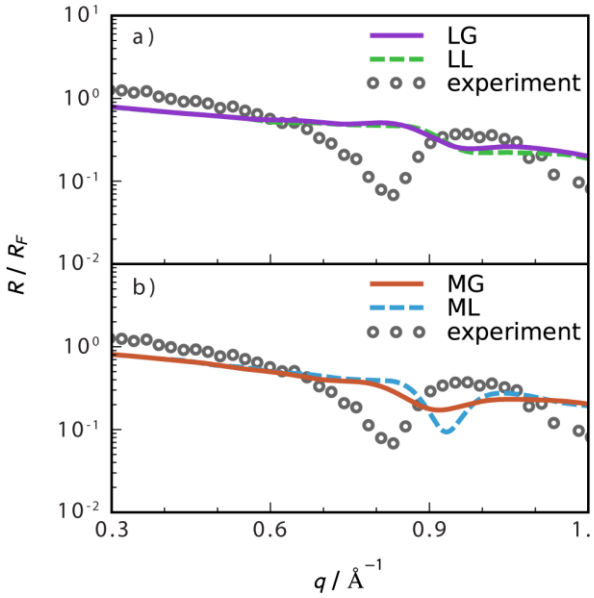


Figure 1. Comparison of the experimental and theoretically calculated reflectivities obtained taking into account the last 30 ns from the 100 ns long simulations (largest window of convergence), for a) force fields of set L and b) force fields of set M, where q and R/R_F denote the wave vector transfer and the reflectivity normalized by Fresnel reflectivity, respectively.

surfaces[36] and the associated possibilities of higher-dimensional electrical double layers.[12,13,37,38] Other successful applications[39,40] include the qualitative reproduction of the layering found in surface force experiments,[18,19] the provision of insight into the behavior of ILs on topologically distinct graphite surfaces,[41,42] and the revelation of additional information regarding ion orientation in the innermost layer.[36] With an increasing need for better and more reliable MD simulations, aiming to provide not just a qualitative, but also a quantitative description of the IL-solid interface, the development and validation of the requisite force-field parameters becomes a crucial task.

Because X-ray reflectivity is the most sensitive experimental probe of IL-solid interfaces, its application in force-field validation would seem prudent. Indeed, the comparison of the simulated electron density profile with the fitted model from experiment represents a step in the right direction.[43] Such an approach, however, only indirectly and partially utilizes the experimentally available information. Calculating the reflectivity from the simulations, on the other hand, allows for a more direct, unbiased and complete validation of the MD results, both qualitatively and quantitatively. While this approach has been reported for the well-understood calcite-water interface,[21] no such studies have been undertaken for ILs.

In this study, using an archetypical model system, consisting of an imidazolium-based IL ([C_2Mim][NTf_2]) in contact with a fully hydroxylated sapphire substrate, we present and compare the results of high-resolution X-ray reflectivity measurements and extensive MD simulations. In addition to the direct comparison between the calculated and experimental reflectivities, we have also compared the

results from the simulations to a fitted model of the measured data.

The experimental measurements, which are described in detail in the Experimental Section (SI), were performed at Advanced Photon Source beamline 33-ID using a custom-built solid-liquid sample cell designed to facilitate access to the sapphire-IL interface. Using fresh samples of $[C_2Mim][NTf_2]$ and freshly prepared sapphire, the reflectivity $R(q)$ of the system (Figure 1) was recorded as a function of momentum transfer q ($q = 4\pi/\lambda\sin(\theta)$, where λ and θ represent the X-ray wavelength and the incident angle, respectively). The measurement was repeated several times using fresh substrates and liquid on each occasion. The $R(q)$ values presented in Figure 1 are thus averaged over multiple data sets. It has been proposed that the interfacial layering of ionic liquids at the sapphire interface seen in X-ray reflectivity is mainly driven by negative surface charge caused by X-ray illumination of the substrate.[15] We minimized this possibility by a lateral translation of the sample such that every data point of the measured reflectivity curve is recorded at a different position of the substrate, thus minimizing X-ray exposition. Furthermore, by a time resolved measurement of each point (total exposure of 2 s split into 100 ms steps) we were able to confirm that the reflectivity was not impacted by the X-ray illumination times.

The MD simulations, which are also described in the Computational Details section (SI), were performed for 100 ns with four different parameterizations. The simulated systems consisted of a slab of sapphire (82.4 Å x 62.9 Å x 21.2 Å optimized in GULP[44]), with a fully hydroxylated (0001) xy surface, described by the CLAYFF[45] force field. A volume of $[C_2Mim][NTf_2]$, (82.4 Å x 62.9 Å x 140.0 Å, consisting of 1686 ion pairs) was placed adjacent to the surface, resulting in the system shown in Figure S2. The ionic liquid was described using fixed-charge force fields, with either the Ludwig *et al.*[24] (denoted as L) or the Maginn[46]/Canongia Lopes[23] *et al.* (denoted as M) parameters. The IL and sapphire force fields were coupled using the two most commonly applied mixing rules, either geometric (denoted as G) and Lorentz-Berthelot mixing rules (denoted as L), thus producing four distinct parameterizations, denoted as LG, LL, MG and ML. One could also employ polarizable force fields to describe the ILs[47] and, although this approach has its merits, it is significantly more computationally expensive and complex. We have therefore chosen to carefully explore the performance of the fixed-charge approach before considering the inclusion of polarization.

The simulations, which were all performed in GROMACS 4.5.5[48] with a time step of 2 fs, van der Waals and short range Coulomb cut-offs of 12 Å, three dimensional periodic boundary conditions and the Particle Mesh Ewald procedure, were carried out as follows: (a) The pure ionic liquid was first minimized and then equilibrated for 0.6 ns at 600 K, which was followed by simulated annealing from 600 to 300 K for 2 ns. (b) The obtained IL box was combined with the sapphire slab such that the smallest distance between the liquid and the surface was approximately 3 Å. Four semi-isotropic NPT simulations (P = 1 atm, $\beta = 4.8\times10^{-5}$ MPa^{-1}) corresponding to the four parameterizations were each performed for 1.4 ns, where the box size was only allowed to vary in the z direction. The average values of the z component over the final 1 ns (convergence seen after 0.3 ns) were used in the four subsequent NVT simulations. (c) The four model systems were then subjected to NVT simulated annealing (600 to 300 K for 12 ns). (d) The obtained model systems (see, e.g. Figure S2) were further simulated under NVT conditions, at T = 300 K, for 100 ns.

The experimental reflectivity curve shows a significant and deep dip at approximately q = 0.83 Å^{-1} (Figure 1). The calculated reflectivity curves for the set L force fields (Figure 1a) do not exhibi t particularly good agreement with respect to this feature, with both producing much smaller minima, centered around q = 0.96 Å^{-1} and q = 0.98 Å^{-1} for LG and LL, respectively. The set M force fields (Figure 1b) fare significantly better, with both MG and ML showing a more pronounced minimum in the q region of interest (found at approximately q = 0.91 and q = 0.93 Å^{-1} for MG and ML, respectively). While both force fields from the set M clearly outperform their counterparts from the set L, the ML combination appears to provide the best agreement with the experiment, producing a strongly pronounced dip whose shape-wise quality strongly resembles the experimental curve. Despite the strong resemblance, the absolute position of the minimum in the ML reflectivity curve is somewhat shifted towards larger q values. The reasons for this discrepancy, as well as the origin of the poorer performance of the LG, LL, and MG parameterizations, can be understood with the help of a careful analysis of the measured and simulated data, as is undertaken below.

Before embarking on that analysis, however, it is instructive to address the sampling efficiency of MD simulations as it pertains to reflectivity data. To this end we have evaluated $R(q)$ (see SI, Computational Details) for four distinct 10 ns time windows of the ML simulation, shown in Figure 2 (analogous data for LG, LL and MG are shown in Figure S3a). We observe that the calculated reflectivity converges only after approximately 70 ns, with the result from the first 50 ns (Figure 2, 40-50 ns window) drastically differing from the converged cases. Interestingly, we found that the innermost layers form already during the first 5-10 ns of simulation. This can be observed from the behaviour of interface-normal number density profile of cations within approximately 7.5 Å from the surface (Figure S4, profiles for LG, LL and MG are shown in Figure S3b). Smaller variations are observed out to distances of 15 Å while equilibration in the transition zone between the interface and the bulk regions (15 Å - 60 Å) is particularly slow, which can be indirectly concluded from the behavior of the reflectivity curves (Figure 2).

We now turn to a more detailed analysis of the structuring of the ionic liquid in response to the presence of the sapphire surface. In this connection, it is instructive to discuss a model-fit approach to interpreting the experimental data. For example, using a variation of the distorted crystal model,[15,49] where the cation and anion contributions to the electron density are each represented by a series of Gaussian profiles with equal layer spacing and increasing broadening, one can reconstruct the reflectivity, which arises as the Fourier transform of the derivative of the total electron density (sum of cation, anion and substrate electron densities, see Computational Details). Thus, by fitting/matching the modeled and experimental reflectivity, one arrives at a decomposition representing cation and anion electron

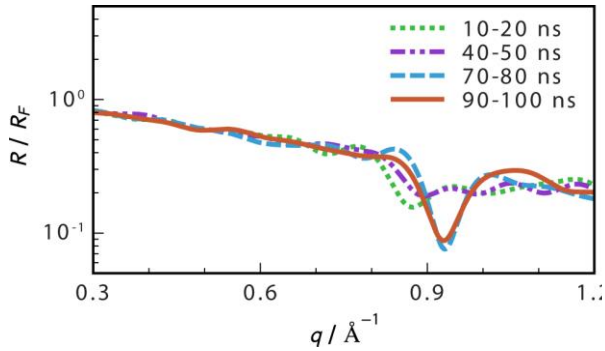


Figure 2. Time evolution of reflectivity, as obtained from MD simulation using the ML parameterization.

density profiles of the investigated system. In the present case we employed interface-normal number density distribution of cations and anions, which we obtained from their corresponding electron densities. This allows us to distinguish whether the IL likely contains no layers, alternating cation/anion layers, or checkerboard-style layers (equal number of cations and anions per layer), representing the most commonly found structures at IL-solid interfaces.[15,16,18] Figure 3 shows that the fitted model is consistent with a double-layer arrangement of cations and anions, with five clearly discernible ionic bilayers (defined as those layers whose peak values deviate by at least 5% from the bulk density) extending up to 30-40 Å into the bulk s olution, with cations ($[C_2Mim]^+$) forming the first layer at the sapphire surface.

In order to compare with the model fit approach, as well as to analyze the molecular details of the computational results, we constructed interface-normal number density profiles from the MD simulations (Figure 4, LG and MG profiles are shown in Figure S5). We find that all investigated force fields provide the same qualitative picture of the investigated IL-sapphire interface, showing both a well pronounced bilayering profile of $[C_2Mim][NTf_2]$ at the interfacial region (first 40-50 Å from the interface) and a predominance of cations at the sapphire surface. Moreover, the first cation layer is found to lie at approximately 3 Å from the hydroxylated sapphire surface, in all four simulations, which virtually coincides with the fitted-model

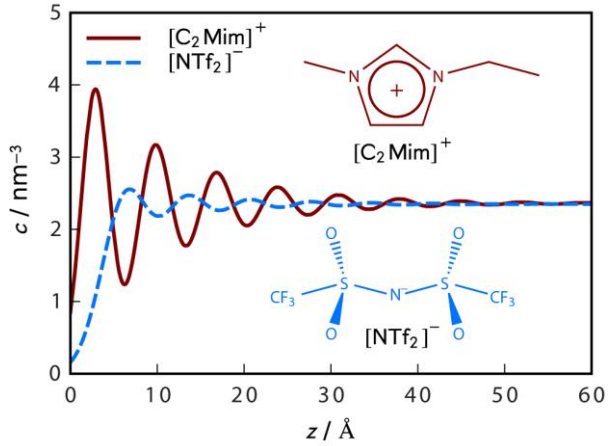


Figure 3. Interface-normal number density profiles of $[C_2Mim]^+$ (red line) and $[NTf_2]^-$ (blue dashed line), as obtained from the model dependent fit to the experiment. Structures of $[C_2Mim]^+$ and $[NTf_2]^-$ are shown in red and blue, respectively.

picture (2.8 Å). Interestingly, only the ML parameterization shows well pronounced bilayers extending up to 45 Å into the bulk (5-6 discernible bilayers), while the layering profiles in the remaining force fields die out 10-15 Å before that point, showing only 3-4 clearly observable bilayers (Figures 4 and S5).

The largest distinction between the fitted model and simulations is observed in the behavior of anion layers. The model-dependent fit predicts a wider and less pronounced layering profile, while somewhat richer and more structured aggregation of anions is observed in the simulations. Interestingly, the interfacial anion layer also marks a distinction between the set L (LG and LL) and the set M (MG and ML) parameterizations, with the 2-10 Å interfacial region being broken into three narrowly separated anion peaks for set M, with the first of the aforementioned maxima coinciding with the innermost cation layer. On the other hand, set L gives rise to a much wider and more homogeneous feature in the same region (Figure 4). Nevertheless, in agreement with set M, set L also shows a coexistence of anions and cations in the innermost layers.

We are now in a position to comment on the observed shift in the reflectivities of the simulated systems toward larger q values compared to the experiment (Figure 1). Number density profiles (Figure 4), showing both more pronounced anion layering and more rapid dampening in the density of the cation layers compared to the model-fit (Figure 3), led us to examine the smoothed electron density profiles of the simulated systems (Figure S6). Interestingly, though we found a rather strong resemblance between the model fit and the simulations with respect to the cation and anion electron density profiles, which is especially evident for ML (≈ 7 Å gaps between the neighboring cation/anion layers in both cases), the differences are significantly more pronounced in the total electron density. This arises from the fact that, while the model-fit exhibits perfect periodicity of the cation/anion layers (Figure 3), the simulated systems show a more disordered picture (Figure 4). This imperfect alignment results in a narrower oscillation of the total electron density (Figure S6). In turn, this shifts the dominant minimum in the theoretically obtained reflectivities towards larger q values.

We have also investigated the $[C_2Mim][NTf_2]$-sapphire interface by analyzing the orientational preferences of anions and cations close to the surface from the MD simulations. The $[NTf_2]^-$ anions in the region $z < 11$ Å from the interface are arranged such that the S-S vector is predominantly aligned parallel to the surface. In the immediate vicinity of the surface, the *cis* conformation predominates, often with all four oxygens accepting hydrogen bonds from the sapphire (Figure 5b). This coordination persists irrespective of the differences in the anion number density profiles observed for the two investigated sets of force fields (Figure 4 and S5). Similarly, we find that, for all four parameterizations, the alkyl tails of the cations lie predominantly parallel to the sapphire surface (Figure 5b, Figure S7), resembling the behavior observed on charged mica.[36] For the simulations performed with the set L parameterization, there is no marked preference for the further orientation of the imidazolium rings (Figure S7). On the other hand, the set M parameterization reveals a strong preference for the imidazolium rings to be oriented such that the majority of the C2 protons (most acidic in $[C_2Mim]^+$)

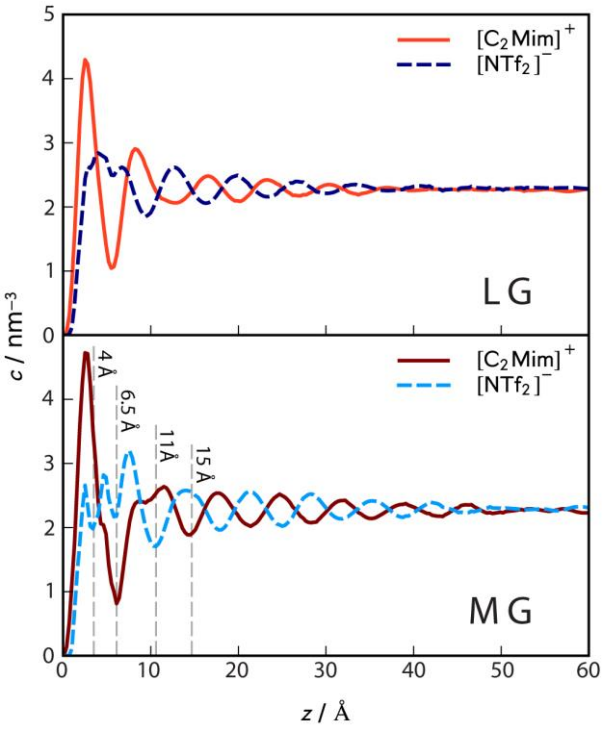


Figure 4. Interface-normal number density profiles of $[C_2Mim]^+$ ($c^+(z)$) and $[NTf_2]^-$ ($c^-(z)$) obtained from last 30 ns of a 100 ns long simulation, for two parameterizations of the IL-sapphire system. To account for the actual size of the molecules, the density distributions were derived on the basis of the positions of all atoms rather than the centers of mass of the ions.

point towards the surface and engage in hydrogen bonds with the sapphire interface. This preference is slightly stronger in the ML simulations than it is in the MG data and is potentially connected to the ML simulations exhibiting a somewhat larger proportion of cations (by 5%) and a relatively strongly depleted proportion of anions (by 15%) in the first cation enriched layer ($z < 4$ Å).

This stronger demixing of cations and anions in the region closest to the interface,[11] observed in the case of ML, propagates through the system to produce a layering density profile that extends significantly more towards the bulk solution than the other parameterizations (see Figure 4 and S5). In turn, this extended layering produces a reflectivity curve that much more closely resembles the experiment and suggests that the ML parameterization is by far the most realistic of the fixed-charge alternatives presented herein (Figure 1).

The strong coordination of anions and cations at the surface is responsible for a distinct charge density distribution $\rho^{\pm}(z)$ (Figure 5a). The positively charged acidic protons of $[C_2Mim]^+$ accumulate closest to the surface in order to access the buried hydroxyl oxygens. This is counterbalanced by the sulfonyl oxygens of a lower number of $[NTf_2]^-$ ions, accepting hydrogen bonds slightly further away from the surface. As a result, the integrated total charge density is neutral already at $z = 2.2$ Å (within the first molecular layer). The structure in $\rho^+(z)$ and $\rho^-(z)$ between 2.2 and 6.5 Å is predominantly a result of compensatory partial atomic charges (initially C2 in $[C_2Mim]^+$, and S in $[NTf_2]^-$), connected to the surface-bound atoms. While it is possible to find additional charge neutral points, by 6.5 Å, the surplus cations produce an excess charge of 0.2 e/nm^2 at the edge of the first molecular layer. Actually, z = 6.5 Å coincides with the minimum in both cation and anion densities $c^+(z)$ and $c^-(z)$ (Figure 4), suggesting the boundary of the first double layer. Interestingly, the adjacent peak in $c^-(z)$ (6.5 - 11 Å) forms a charge-neutral shell, as does the region between 11 and 15 Å. Despite the long-range layering in $c^+(z)$ and $c^-(z)$, the charge densities ($\rho^+(z)$ and $\rho^-(z)$) show no fluctuations beyond 15 Å. One should note that this decoupling of density and charge fluctuations would not be apparent had the analysis treated the ions as point-like objects.

In the context of viewing the electric double layers in ionic liquids as 3D entities,[12] our previous analysis can be considered as a refined version of a 1D picture of the double layer at zero charge, characterized with $c^{\pm}(z)$ and $\rho^{\pm}(z)$. While this is entirely appropriate for comparison with reflectivity data, it is nevertheless interesting to consider the surface-induced ordering in the xy planes, parallel to the neutral sapphire surface. To this end, we analyze the lateral details of different z-slabs, from the ML simulations (Figure 5c-f, as well Figure S8a-e).

The first slab ($z < 4$ Å) accounts for the first anion peak in $c^-(z)$, as well as for the first cation enriched region as seen in $c^+(z)$ (Figure 4). In the ML parametrization, at the sapphire surface we find, on average, $\langle c^-(z < 4\text{ Å})\rangle = 2.1$ $[NTf_2]^-$/nm^3 compared to $\langle c^+(z < 4\text{ Å})\rangle = 4.6$ $[C_2Mim]^+$/nm^3. Importantly, nearly all cations and only 50% of the anions have their centers of mass in that region, on average. The fluctuations around these average values are nearly symmetric, with maximal deviations of 13% and 105% for $[C_2Mim]^+$ and $[NTf_2]^-$, respectively. This suggests that the mobility of cations in the direction perpendicular to the sapphire surface is much smaller than that of the anions, the latter exchanging relatively frequently with slabs at larger z values.

In order to understand lateral organization of ions we calculate the in-plane total correlation functions h_{xy}^{ss} ($h_{xy}^{ss}(r) = g_{xy}^{ss}(r) - 1$, where $g_{xy}^{ss}(r)$ is the lateral radial distribution function, with each s superscript denoting an ion species). We find that the in-plane $[C_2Mim]^+$ and $[NTf_2]^-$ density correlation functions for the first slab (h_{xy}^{++} and h_{xy}^{--}, Figure 5c) show significant lateral ordering, where the first peak of the former precedes that of the latter. This means that the average distance between cations is somewhat smaller then between anions, although both are close to 1 Å. Most likely, this effect is induced by the crowding of cations, despite the mutually repulsive Coloumb interactions. Surprisingly, stronger long-range correlations (with a periodicity of 1 Å) are evident in the h_{xy}^{--} curve, while the h_{xy}^{++} and h_{xy}^{+-} correlations are more damped at larger distances. Nevertheless, the h_{xy}^{+-} curve possesses a sharp peak at 0.5 Å, suggesting that, despite their mobility, the anions correlate strongly with the cations (Figure 5c), intercalating the cation distribution.

To compensate for the net overscreening of the surface by the cations, the anions are organized to produce a sublayer, indicated by a peak in $c^-(z)$ between 4 and 6.5 Å. Despite the larger anion concentration ($\langle c^-(4 < z < 6.5\text{ Å})\rangle = 2.3$ $[NTf_2]^-$/nm^3), the lateral organization, inferred from the

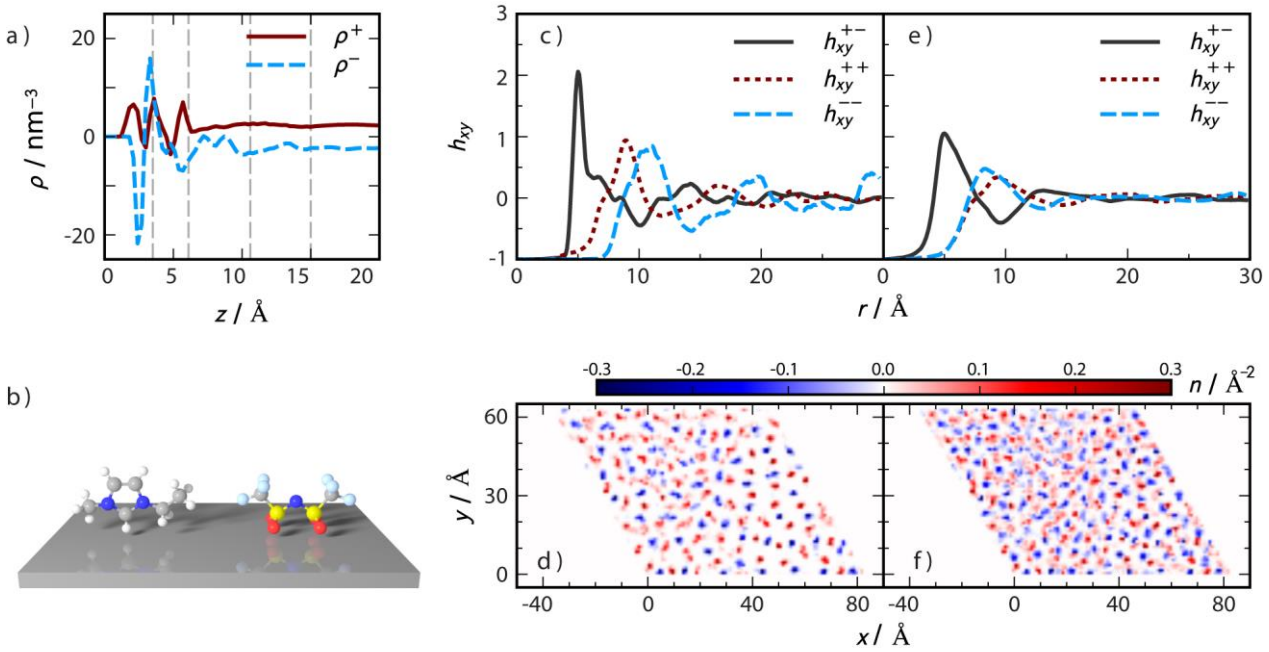


Figure 5. a) Charge distribution for cations ($\rho^+(z)$) and anions ($\rho^-(z)$) as a function of the distance from the sapphire surface. Vertical lines correspond to those shown in Figure 4. b) Depiction of the predominant orientations of anions (for both L and M sets) and cations (for the set M) in their first layers, respectively. c) The anion-anion ($\boldsymbol{h_{xy}^{--}}(\boldsymbol{r})$), cation-anion ($\boldsymbol{h_{xy}^{+-}}(\boldsymbol{r})$) and the cation-cation ($\boldsymbol{h_{xy}^{++}}(\boldsymbol{r})$) total in-plane correlation functions for $z < 4$ Å. d) $\boldsymbol{h_{xy}^{--}}(\boldsymbol{r})$, $\boldsymbol{h_{xy}^{+-}}(\boldsymbol{r})$ and $\boldsymbol{h_{xy}^{++}}(\boldsymbol{r})$ for $6.5 < z < 11$ Å. e) 2D histogram of the in-plane positions of the ions in the first double layer ($z < 6.5$ Å). f) Analogous 2D histogram for $z < 11$ Å. Correlation functions and the histograms are generated on the basis of the centers of mass of the constituent ions over the last 30 ns of the ML simulation.

density distribution functions in the second slab, is still very similar to the one found at the surface (Figure S8a), with the difference that correlations diminish at shorter separations. This is consistent with the second slab being fully within the first cation peak, albeit in the depleted region ($\langle c^+(4 < z < 6.5\ \text{Å})\rangle = 1.6$ $[C_2Mim]^+/nm^3$). Nevertheless, the cation distribution at the surface seems to be the determining structural feature of this slab which, together with the first one, forms the first double layer.

The above description of the first double layer is perhaps best represented in the two-dimensional histogram of the center-of-mass positions of the ions, plotted over the last 30 ns of simulation time (Figure 5d). Here, the average concentration of ions clearly shows the intercalating nature of the anticipated structure, as well as a large degree of lateral order that minimizes repulsive ion interactions. Given the neutral nature of the sapphire, this level of 2D ordering, which consists of a nearly immobile cation grid (with a relatively small amount of excess positive charge very close to the sapphire), intercalated with diffusive and much less concentrated anions (directing more negative charge towards the surface per molecule), cannot be understood from a simplistic balance of ionic interactions. Indeed, the entire structure appears to be stabilized by hydrogen bonding to the substrate. Given the low in- and out-of-plane cation mobility (Figure 5d), and the capacity of anions to leave the surface, the H-bonds seem to be significantly stronger for $[C_2Mim]^+$.

We further analyze the lateral organization in the neutral shell ($6.5 < z < 11$ Å) adjacent to the first double layer. It consists of a well-pronounced anion peak in $c^-(z)$ and a small cation peak overlapping with structures at larger separations from the surface. Despite the charge neutrality, the overall density of anions ($\langle c^-(6.5 < z < 11\ \text{Å})\rangle = 2.8$ $[NTf_2]^-/nm^3$), is slightly higher than that of cations ($\langle c^+(6.5 < z < 11\ \text{Å})\rangle = 2.4$ $[C_2Mim]^+/nm^3$), which partially compensates for the cation excess in the first double layer. Nevertheless, the lateral organization, inferred from the density distribution functions appears more diffusive or liquid-like (Figure 5e). However, this region is structurally dominated by the anions being, on average, more closely packed than they are on either the surface (Figure 5c) or in the bulk (Figure S8d), again, probably due to crowding. This argument is supported by the fact that in the layers further away from the substrate, where the densities of both species are same, the correlation functions h_{xy}^{++} and h_{xy}^{--} have peaks at the same positions. To this end, it is worth noting that the preferred distance between anions and cations remains to be 0.5 Å (see the peak of h_{xy}^{+-}), which is independent of the z position of the molecules.

To conclude our analysis of the 3D structure of the IL layers at the interface with the surface, we determine the 2D density map of the centers of mass of all ions in the region that includes the second double layer (Figure 5f). While the intercalating structure of anions and cations was evident in the first double layer (Figure 5d), the histogram for $0 < z < 11$ Å shows that the anions and cations both continue to develop a complex network, that projected onto a 2D plane resembles a non-overlapping grid. This is in contrast with the bulk (Figure S8e), where the low intensity of colors, and the disperse patterns show that the probability to find an anion and cation on the same lateral position is nearly the same.

Our present study reveals the fundamental importance of the direct comparison of the MD simulations with the reflectivity experiments, which enables one to use the detailed description obtained from simulations with confidence. In particular, we found that, while the indirect comparison does not provide a successful means for distinguishing between the various parameterizations, the direct comparison showcases significant distinctions in their performances, enabling us to clearly identify the best performing fixed-charge force field (ML). The validated simulations enabled us to investigate the nature of $[C_2Mim][NTf_2]$-sapphire interface in detail. We find that the hydrogen bonding between the surface hydroxyl groups and the constituent ions induces a laterally ordered structure at the interface, characterized by an excess of cations. The overall result is the spontaneous charging of an otherwise neutral surface, which induces a multidimensional bilayering profile of cations and anions that extends deeply toward the bulk solution.

ASSOCIATED CONTENT

AUTHOR INFORMATION

Corresponding Author

*Email: dsmith@irb.hr

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENT

This work was financially supported by the EAM Cluster of Excellence at Friedrich-Alexander University Erlangen-Nürnberg. This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. We acknowledge the help of local contact Christian M. Schlepütz and the staff of APS beamline 33-ID in running the synchrotron experiments.

Supporting Information. Computational Details, Experimental Section, Supporting Figures S1-S9, Force Field Parameters. This material is available free of charge via the Internet http://pubs.acs.org.

REFERENCES

(1) Welton, T. Room-Temperature Ionic Liquids. Solvents for Synthesis and Catalysis. *Chem. Rev.* **1999**, *99*, 2071–2084.
(2) Wasserscheid, P.; Welton, T. (eds.). *Ionic Liquids in Synthesis*; Wiley-VCH: Weinheim, Germany; 2008.
(3) Walden, P. Molecular Weights and Electrical Conductivity of Several Fused Salts. *Bull. Russian Acad. Sci.* **1914**, *1800*, 405–422.
(4) Plechkova, N. V.; Seddon, K. R. Applications of Ionic Liquids in the Chemical Industry. *Chem. Soc. Rev.* **2008**, *37*, 123–150.
(5) Tempel D. J.; Henderson P. B.; Brzozowski J. R.; Pearlstein R. M.; Cheng H. High Gas Storage Capacities for Ionic Liquids Through Chemical Complexation. *J. Am. Chem. Soc.* **2008**, *130*, 400–401.
(6) Endres, F. Ionic Liquids: Solvents for the Electrodepostion of Metals and Semiconductors. *ChemPhysChem* **2002**, *3*, 144–154.
(7) van Rantwijk, F.; Sheldon, R. A. Biocatalysis in Ionic Liquids. *Chem. Rev.* **2007**, *107*, 2757–2785.
(8) Ye, C.; Liu, W.; Chen, Y.; Yu, L. Room-Temperature Ionic Liquids: a Novel Versatile Lubricant. *Chem. Commun.* **2001**, *21*, 2244–2245.
(9) Armand, M.; Endres, F.; MacFarlane, D. R.; Ohno, H.; Scrosati, B. Ionic-Liquid Materials for the Electrochemical Challenges of the Future. *Nat. Mater.* **2009**, *8*, 621–629.
(10) Rogers, R. D.; Seddon, K. R. Ionic Liquids--Solvents of the Future? *Science* **2003**, *302*, 792–793.
(11) Fedorov, M. V.; Kornyshev, A. A. Ionic Liquids at Electrified Interfaces. *Chem. Rev.* **2014**, *114*, 2978–3036.
(12) Kornyshev, A. A.; Qiao, R. Three-Dimensional Double Layers. *J. Phys. Chem. C* **2014**, *118*, 18285−18290.
(13) Merlet, C.; Rotenberg, B.; Madden, P. A.; Salanne, M. Computer Simulations of Ionic Liquids at Electrochemical Interfaces. *Phys. Chem. Chem. Phys.* **2013**, *15*, 15781–15792.
(14) Simon P.; Gogotsi, Y. Materials for Electrochemical Capacitors. *Nat. Mater.* **2008**, *7*, 845–854.
(15) Mezger, M.; Schröder H.; Reichert, H.; Schramm, S.; Okasinski, J. S.; Schöder, S.; Honkimäki, V.; Deutsch, M.; Ocko, B. M.; Ralston. J.; *et al.* Molecular Layering of Fluorinated Ionic Liquids at a Charged Sapphire (0001) Surface. *Science* **2008**, *322*, 424–428.
(16) Mezger, M.; Schramm, S.; Schröder, H.; Reichert, H.; Deutsch, M.; De Souza, E. J.; Okasinski, J. S.; Ocko B. M.; Honkimäki, V.; Dosch H. Layering of $[BMIM]^+$- Based Ionic Liquids at a Charged Sapphire Interface. *J. Chem. Phys.* **2009**, *131*, 094701.
(17) Lauw, Y.; Horne, M. D.; Rodopoulos, T.; Lockett, V.; Akgun, B.; Hamilton, W. A.; Nelson, A. R. Structure of [C4mpyr][NTf2] Room-Temperature Ionic Liquid at Charged Gold Interfaces. *Langmuir* **2012**, *28*, 7374–7381.
(18) Hayes, R.; El Abedin, S. Z.; Atkin, R. Pronounced Structure in Confined Aprotic Room-Temperature Ionic Liquids. *J. Phys. Chem. B* **2009**, *113*, 7049–7052.
(19) Perkin, S.; Crowhurst, L.; Niedermeyer, H.; Welton, T.; Smith, A. M.; Gosvami, N. N. Self-Assembly in the Electrical Double Layer of Ionic Liquids. *Chem. Commun.* **2011**, *47*, 6572–6574.
(20) Cremer, T.; Stark, M.; Deyko, A.; Steinrück, H.-P.; Maier, F. Liquid/Solid Interface of Ultrathin Ionic Liquid Films: $[C_1C_1Im][Tf_2N]$ and $[C_8C_1Im][Tf_2N]$ on Au(111). *Langmuir* **2011**, *27*, 3662–3671.
(21) Fenter, P.; Kerisit, S.; Raiteri, P.; Gale, J. D. Is the Calcite-Water Interface Understood? Direct Comparisons of Molecular Dynamics Simulations with Specular X-Ray Reflectivity Data. *J. Phys. Chem. C* **2013**, *117*, 5028–5042.
(22) Maginn, E. J. Molecular Simulation of Ionic Liquids: Current Status and Future Opportunities. *J. Phys. Condens. Matter* **2009**, *21*, 373101.
(23) Canongia Lopes, J. N.; Deschamps, J.; and Pádua, A. A. H. Modeling Ionic Liquids Using a Systematic All-Atom Force Field. *J. Phys. Chem. B* **2004**, *108*, 2038–2047.
(24) Chaban, V. V. Polarizability Versus Mobility: Atomistic Force Field for Ionic Liquids. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16055−16062.
(25) Köddermann, T.; Paschek, D.; Ludwig, R. Molecular Dynamic Simulations of Ionic Liquids: A Reliable Description of Structure, Thermodynamics and Dynamics. *ChemPhysChem* **2007**, *8*, 2464–2470.
(26) Köddermann, T.; Reith, D.; Ludwig, R. Comparison of Force Fields on the Basis of Various Model Approaches-How To Design the Best Model for the $[C_nMIM][NTf_2]$ Family of Ionic Liquids. *ChemPhysChem* **2013**, *14*, 3368–3374.
(27) Chaban, V. V.; Prezhdo, O. V. Ionic and Molecular Liquids: Working Together for Robust Engineering. *J. Phys. Chem. Lett.* **2013**, *4*, 1423−1431.
(28) Chaban, V. V.; Prezhdo, O. V. Nanoscale Carbon Greatly Enhances Mobility of a Highly Viscous Ionic Liquid. *ACS Nano* **2014**, *8*, 8190–8197.
(29) Ohba, T.; Chaban, V. V. A Highly Viscous Imidazolium Ionic Liquid Inside Carbon Nanotubes. *J. Phys. Chem. B* **2014**, *118*, 6234−6240.
(30) Fileti, E. E.; Chaban, V. V. Structure and Supersaturation of Highly Concentrated Solutions of Buckyball in 1-Butyl-3-Methylimidazolium Tetrafluoroborate. *J. Phys. Chem. B* **2014**, *118*, 7376−7382.
(31) Fileti, E. E.; Chaban, V. V. Imidazolium Ionic Liquid Helps to Disperse Fullerenes in Water. *J. Phys. Chem. Lett.* **2014**, *5*, 1795−1800.
(32) Chaban, V. V.; Prezhdo, O. V. How Toxic Are Ionic Liquid/Acetonitrile Mixtures? *J. Phys. Chem. Lett.* **2011**, *2*, 2499–2503.
(33) Chaban, V. V.; Prezhdo, O. V. Ionic Vapor: What Does It Consist Of? *J. Phys. Chem. Lett.* **2012**, *3*, 1657−1662.
(34) Chaban, V. V.; Prezhdo, O. V. Water Phase Diagram Is Significantly Altered by Imidazolium Ionic Liquid. *J. Phys. Chem. Lett.* **2014**, *5*, 1623−1627.
(35) Neto, B. A. D.; Meurer, E. C.; Galaverna, R.; Bythell, B. J.; Dupont, J.; Cooks, R. G.; Eberlin, M. N. Vapors from Ionic Liquids:

Reconciling Simulations with Mass Spectrometric Data. *J. Phys. Chem. Lett.* **2012**, *3*, 3435−3441.
(36) Payal, R. S.; Balasubramanian, S. Orientational Ordering of Ionic Liquids Near a Charged Mica Surface. *ChemPhysChem* **2012**, *13*, 1764–1771.
(37) Merlet, C.; Limmer, D. T.; Salanne, M.; van Roij, R.; Madden, P. A.; Chandler, D.; Rotenberg, B. The Electric Double Layer Has a Life of Its Own. *J. Phys. Chem. C* **2014**, *118*, 18291−18298.
(38) Ivaništšev, V.; Fedorov, M. V.; Lynden-Bell, R. M. Screening of Ion−Graphene Electrode Interactions by Ionic Liquids: The Effects of Liquid Structure. *J. Phys. Chem. C* **2014**, *118*, 5841–5847.
(39) Köhler, R.; Restolho, J.; Krastev, R.; Shimizu, K.; Canongia Lopes, J. N.; Saramago, B. Liquid- or Solid-Like Behavior of [omim][BF_4] at a Solid Interface? *J. Phys. Chem. Lett.* **2011**, *2*, 1551–1555.
(40) Shimizu, K.; Pensado, A.; Malfreyt, P.; Pádua, A. H.; Canongia Lopes, J. N. 2D or not 2D: Structural and Charge Ordering at the Solid-Liquid Interface of the 1-(2-hydroxyethyl)-3-Methylimidazolium Tetrafluoroborate Ionic Liquid. *Faraday Discuss.* **2012**, *154*, 155–169.
(41) Wang, S.; Li, S.; Cao, Z.; Yan, T. Molecular Dynamics Simulations of Ionic Liquids at Graphite Surface. *J. Phys. Chem. C* **2010**, *114*, 990–995.
(42) Rajput, N. N.; Monk, J.; Singh, R.; Hung, F. R. On the Influence of Pore Size and Pore Loading on Structural and Dynamical Heterogeneities of an Ionic Liquid Confined in a Slit Nanopore. *J. Phys. Chem. C* **2012**, *116*, 5169–5181.
(43) Zhou, H.; Rouha, M.; Feng, G.; Lee, S. S.; Docherty, H.; Fenter, P.; Cummings, P. T.; Fulvio, P. F.; Dai, S.; McDonough, J.; *et al.* Nanoscale Perturbations of Room Temperature Ionic Liquid Structure at Charged and Uncharged Interfaces. *ACS Nano* **2012**, *6*, 9818–9827.
(44) Gale, J. D.; Rohl, A. L. The General Utility Lattice Program. *Mol. Simul.* **2003**, *29*, 291–341.
(45) Cygan, R. T.; Liang, J.-J.; Kalinichev, A. G. Molecular Models of Hydroxide, Oxyhydroxide, and Clay Phases and the Development of a General Force Field. *J. Phys. Chem. B* **2004**, *108*, 1255–1266.
(46) Kelkar, M.; Maginn, E. Effect of Temperature and Water Content on the Shear Viscosity of the Ionic Liquid 1-Ethyl-3-Methylimidazolium bis (Trifluoromethanesulfonyl) Imide as Studied by Atomistic Simulations. *J. Phys. Chem. B* **2007**, *111*, 4867–4876.
(47) Borodin, O. Polarizable Force Field Development and Molecular Dynamics Simulations of Ionic Liquids. *J. Phys. Chem. B* **2009**, *113*, 11463–11478.
(48) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comp.* **2008**, *4*, 435–447.
(49) Magnussen, O.; Ocko, B.; Regan, M. X-ray Reflectivity Measurements of Surface Layering in Liquid Mercury. *Phys. Rev. Lett.* **1995**, *74*, 4444–4447.

List of Figures

List of Tables and Schemes

Abbreviations

Ada_N	Neutral Form of Ada-Enkephalin
Ada_Z	Zwitterionic Form of Ada-Enkephalin
ALDA	Adiabatic Local Density Approximation
AMBER	Assisted Model Building with Energy Refinement simulation package
APBS	Adaptive Poisson-Boltzmann Solver programming package
ASEC	Average Solvent Electrostatic Configuration
ASEP	Averaged Solvent Electrostatic Potential
B3	Becke's Three-Parameter Exchange Functional
CAM-B3LYP	Coulomb-Attenuated B3LYP functional
CC	Coupled-Cluster Theory
CD	Circular Dichroism
CGTO	Contracted Gaussian Type Orbitals
CHARMM	Chemistry at HARvard Molecular Mechanics simulation package
CI	Configuration Interaction Approach
$CMet_N$	C-terminal Model Peptide Prepared from Neutral Form of Met-Enkephalin
$CMet_Z$	C-terminal Model Peptide Prepared from Zwitterionic Form of Met-Enkephalin
CT	Charge Transfer
DFT	Density Functional Theory
DMSO	Dimethyl Sulfoxide
ESP	Electrostatic Potential
GAFF	General AMBER force field
GGA	Generalized Gradient Approximation
GHF	Global Hybrid Functional
GROMACS	Groeningen Machine for Computer Simulations simulation package
GTO	Gaussian Type Orbitals
GULP	General Utility Lattice Program
HF	Hartree-Fock

HMF	Hybrid Meta-GGA Functional
IR	Infrared
LCAO	Linear Combination of Atomic Orbitals
LCF	Long-Range Corrected Functional
LD	Langevin Dynamics
LDA	Local Density Approximation
LF	Leap-Frog Algorithm
LID	Linear Dichroism
LR-TD-DFT	Linear Response Time-Dependent Density Functional Theory
LSDA	Local Spin Density Approximation
LYP	Lee-Yang-Parr's Correlation Functional
MC	Monte Carlo
MD	Molecular Dynamics
Met_N	Neutral Form of Met-Enkephalin
Met_Z	Zwitterionic Form of Met-Enkephalin
MP	Møller-Plesset perturbation theory
$NMet_N$	N-terminal Model Peptide Prepared from Neutral Form of Met-Enkephalin
NMR	Nuclear Magnetic Resonance
OGF	Opioid Growth Factor
OGFr	Opioid Growth Factor-receptor
OPLS	Optimized Potentials for Liquid Simulations
PC	Principal Component
PCA	Principal Component Analysis
PGTO	Primitive Gaussian Type Orbitals
PME	Particle Mash Ewald
PMF	Potential of Mean Force
QM	Quantum Mechanics
R.E.DD.B	RESP ESP charge DDataBase
REMD	Replica Exchange Molecular Dynamics

RESP	Restrained Electrostatic Potential
RICC2	Resolution-of-Identity Second-Order Approximate Coupled-Cluster
SCF	Self-Consistent Field
STO	Slater Type Orbitals
TD	Time-Dependent
TD-DFT	Time-Dependent Density Functional Theory
TFE	Trifluoroethanol
UV	Ultraviolet
WHAM	Weighted Histogram Analysis Method

Bibliography

Literature is presented in alphabetical order.

[1] Abramavicius, D.; Jiang, J.; Bulheller, B. M.; Hirst, J. D., Mukamel, S. Simulation Study of Chiral Two-Dimensional Ultraviolet Spectroscopy of the Protein Backbone. *J. Am. Chem. Soc.* **2010**, *132*, 7769.

[2] Addadi, L. I. A.; Politi, Y.; Nudelman, F.; Weiner, S. *Design Strategies and Mechanisms of Mineral Formation: Operating at the Edge of Instability*; Springer: Netherlands, 2008.

[3] Amadei, A.; Linssen, A. B.; Berendsen, H. J. Essential Dynamics of Proteins. *Proteins* **1993**, *17*, 412.

[4] Andrec, M.; Felts, A. K.; Gallicchio, E.; Levy, R. M. Protein Folding Pathways from Replica Exchange Simulations and a Kinetic Network Model. *Proc. Natl. Acad. Sci. U.* **2005**, *19*, 6801.

[5] Andrew, C. D.; Bhattacharjee, S.; Kokkoni, N.; Hirst, J. D., Jones, G. R.; Doig, A. J. Stabilizing Interactions between Aromatic and Basic Side Chains in Alpha-helical Peptides and Proteins. Tyrosine Effects on Helix Circular Dichroism. *J. Am. Chem. Soc.* **2002**, *124*, 12706.

[6] Aquino, A. J. A.; Nachtigallova, D.; Hobza P.; Truhlar, D. G.; Hättig, C.; Lischka, H. The Charge-Transfer States in a Stacked Nucleobase Dimer Complex: A Benchmark Study. *J. Comput. Chem.* **2011**, *32*, 1217.

[7] Ariens E. J.; Soudijn W.; Timmermans PBMWM, editors. *Stereochemistry and Biological activity of Drugs.* Oxford: Blackwell Scientific; 1983. pp. 11–33.

[8] Aschauer, U.; Spagnoli, D.; Bowen, P.; Parker, S. C. Growth Modification of Seeded Calcite Using Carboxylic Acids: Atomistic Simulations. *J. Colloid Interface Sci.* **2010**, *346*, 226.

[9] Atkins, P.; Friedman, R. *Molecular Quantum Mechanics*, 4th ed., Oxford University Press, Oxford, 2005.

[10] Autschbach, J. Introduction to the Computation of Chiroptical Properties with First–Principles Theoretical Methods: Background and Illustrative Examples. *Chirality* **2009**, *21*, 116.

[11] Autschbach, J.; Jorge, F. E.; Ziegler, T. Density Functional Calculations on Electronic Circular Dichroism Spectra of Chiral Cobalt(III) Complexes. *Inorg. Chem.* **2003**, *42*, 2867.

[12] Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. *Proc. Natl. Acad. Sci. U. S. A* **2001**, *98*, 10037.

[13] Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 1372.

[14] Berman, A.; Ahn, D. J.; Lio, A.; Salmeron, M.; Reichert, A.; Charych, D. Total Alignment of Calcite at Acidic Polydiacetylene Films: Cooperativity at the Organic-Inorganic Interface. *Science* **1995**, *269*, 515.

[15] Berova, N.; Nakanishi, K.; Woody, R. W., eds. In *Circular Dichroism : Principles and Applications*, 2nd edition, Wiley-VCH, New York, 2000; p. 75.

[16] Besley, N. A.; Hirst, J. D. Theoretical Studies toward Quantitative Protein Circular Dichroism Calculations. *J. Am. Chem. Soc.* **1999**, *121*, 9636.

[17] Bhattacharjee, S.; Tóth, G.; Lovas, S.; Hirst, J. D. Influence of Tyrosine on the Electronic Circular Dichroism of Helical Peptides. *J. Phys. Chem. B* **2003**, *107*, 8682.

[18] Brahms S.; Brahms, J. Determination of Protein Secondary Structure in Solution by Vacuum Ultraviolet Circular Dichroism. *J. Mol. Biol.* **1980**, *138*, 149.

[19] Brečević, L.; Nothig-Laslo, V.; Kralj, D.; Popović, S. Effect of Divalent Cations on the Formation and Structure of Calcium Carbonate Polymorphs. *J. Chem. Soc.* **1996**, *92*, 1017.

[20] Brkljača, Z.; Čondić-Jurkić, K.; Smith, A-S.; Smith, D. M. Calculation of the CD Spectrum of a Peptide from Its Conformational Phase Space: The Case of Met-enkephalin and Its Unnatural Analogue. *J. Chem. Theory Comput.* **2012**, *8*, 1694.

[21] Brkljača, Z.; Mališ, M.; Smith, D. M.; Smith, A.-S. Calculating CD Spectra of Flexible Peptides: An Assessment of TD-DFT functionals. *J. Chem. Theory Comput.*, **2014**, *10* , 3270.

[22] Bunn, C. W. Molecular Structure and Rubber-Like Elasticity. II. The Stereochemistry of Chain Polymers. *Proc. R. Soc. Lond. A* **1942**, *180*, 67.

[23] Bunn, C. W. Molecular Structure and Rubber-Like Elasticity. III. Molecular Movements in Rubber-Like Polymers. *Proc. R. Soc. Lond. A* **1942**, *180*, 82.

[24] Carsky, P. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Clark, N. L., Gasteiger, J., H. F. S., III, Schreiner, P. R., Eds.; Wiley: Chichester, U.K., 1998; p 485.

[25] Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668.

[26] Case, D. A.; Darden, T. A.; Cheatham, T.E. III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsich, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman , P. A. The FF12SB Force Field, AmberTools 13 Reference Manual (2013) 27–29.

[27] Case, D. A.; Darden, T. A.; Cheatham, T.E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman ,P. A. *AMBER 10*, University of California, San Francisco 2008.

[28] Casida, M. E. in the *Recent Advances in Density Functional Methods*, editor D.P. Chong, World Scientific, Singapure, 1995, p 155.

[29] Casida, M. E.; Huix-Rotllant, M. Progress in Time-Dependent Density-Functional Theory. *Annu. Rev. Phys. Chem.* **2012**, *63*, 287.

[30] Chai, J.-D.; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom-Atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615.

[31] Chandler, D. *Introduction to Statistical Mechanics*, Oxford University Press, New York, 1987.

[32] Charaf-Eddin, A.; Planchat, A; Mennucci, B.; Adamo, C.; Jacquemin, D. Choosing a Functional for Computing Absorption and Fluorescence Band Shapes with TD-DFT. *J. Chem. Theory Comput.* **2013**, *9*, 2749.

[33] Chen, C. L.; Qi, J. H.; Zuckermann, R. N.; DeYoreo, J. J. Engineered Biomimetic Polymers as Tunable Agents for Controlling CaCO3 Mineralization. *J. Am. Chem. Soc.* **2011**, *133*, 5214.

[34] Cheng, F.; McLaughlin, P. J.; Verderame, M. F.; Zagon, I. S. The Opioid Growth Factor (OGF)-OGF Receptor Axis Uses the p16 Pathway to Inhibit Head and Neck Cancer. *Cancer Res.* **2007**, *67*, 10511.

[35] Cheng, F.; McLaughlin, P. J.; Verderame, M. F.; Zagon, I. S. The OGF-OGFr Axis Utilizes the p21 Pathway to Restrict Progression of Human Pancreatic Cancer. *Mol. Cancer* **2008**, *7*, 5.

[36] Chien, Y. C.; Hincke, M. T.; Vali, H.; McKee, M. D. Ultrastructural Matrix-Mineral Relationships in Avian Eggshell, and Effects of Osteopontin on Calcite Growth in Vitro. *J. Struct. Biol.* **2008**, *163*, 84.

[37] Christiansen, O.; Koch, H.; Jørgensen, P. The Second-Order Approximate Coupled Cluster Singles and Doubles Model CC2. *Chem. Phys. Lett.* **1995**, *243*, 409.

[38] Christiansen, O.; Koch, H.; Jørgensen, P. Response Functions in the CC3 Iterative Triple Excitation Model. *J. Chem. Phys.* **1995**, *103*, 7429.

[39] Compton, R. G.; Brown, C. A. J. The Inhibition Of Calcite Dissolution Precipitation - Mg^{2+} Cations. *Colloid Interface Sci.* **1994**, *165*, 445.

[40] Condon, E. U.; Altar, W.; Eyring, H. One Electron Rotatory Power. *J. Chem. Phys.* **1937**, *5*, 753.

[41] Coutinho, K.; Georg, H. C.; Fonseca, T. L.; Ludwig, V.; Canuto, S. An Efficient Statistically Converged Average Configuration for Solvent Effects. *Chem. Phys. Lett.* **2007**, *437*, 148.

[42] Cramer, C. J. *Essentials of Computational Chemistry – Theories and Models* 2004, Second Edition, J.Wiley & Sons Inc., New York.

[43] Crawford, T. D.; Stephens, P. J. Comparison of Time-Dependent Density-Functional Theory and Coupled Cluster Theory for the Calculation of the Optical Rotations of Chiral Molecules. *J. Phys. Chem. A* **2008**, *112*, 1339.

[44] Deleuze, M.; Brantley, S. L. Inhibition of Calcite Crystal Growth by Mg^{2+} at 100°C and 100 Bars: Influence of Growth Regime. *Geochim. Cosmochim. Acta* **1997**, *61*, 1475.

[45] Drayer D. E. Clin. Pharmacodynamic and Pharmacokinetic Differences between Drug Enantiomers in Humans: An Overview. *Pharmacol. Ther.* **1986,** *40*, 125.

[46] Dreuw, A.; Head-Gordon, M. Single-Reference ab Initio Methods for the Calculation of Excited States of Large Molecules. *Chem. Rev.* **2005**, *105*, 4009.

[47] Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J. W.; Wang, J.; Kollman, P. A. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *23*, 1999.

[48] Duffy, D. M.; Harding, J. H. Simulation of Organic Monolayers as Templates for the Nucleation of Calcite Crystals. *Langmuir* **2004**, *18*, 7630.

[49] Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007.

[50] Dupradeau, F-Y.; Cézard, C.; Lelong, R.; Stanislawiak, E.; Pecher, J.; Delepine, J. C.; Cieplak, P. R.E.DD.B.: A Database for RESP and ESP Atomic Charges, and Force Field Libraries. *Nucleic Acids Res.* **2008**, *36*, 360.

[51] Elhadj, S.; De Yoreo, J. J.; Hoyer, J. R.; Dove, P. M. Role of Molecular Charge and Hydrophilicity in Regulating the Kinetics of Crystal Growth. *Proc. Natl. Acad. Sci. U.* **2006**, *103*, 19237.

[52] Elhadj, S.; Salter, E. A.; Wierzbicki, A.; De Yoreo, J. J.; Han, N.; Dove, P. M. Peptide Controls on Calcite Mineralization: Polyaspartate Chain Length Affects Growth Kinetics and Acts as a Stereochemical Switch on Morphology. *Cryst. Growth Des.* **2006**, *6*, 197.

[53] Engel, E., Dreizler, R. M. *Density Functional Theory*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[54] Ernzerhof, M.; Scuseria, G. E. Assessment of the Perdew-Burke-Ernzerhof Exchange-Correlation Functional. *J. Chem. Phys.* **1999**, *110*, 5029.

[55] Fidler, J.; Rodger, P. M.; Rodger, A. Circular Dichroism as a Probe of Chiral Solvent Structure around Chiral Molecules. *J. Chem. Soc., Perkin Trans. 2* **1993**, *2*, 235.

[56] Fock, V. Z. Näherungsmethode zur Lösung des Quantenmechanischen Mehrkörperproblems. *Phys.* **1930**, *61*, 126.

[57] Forrey, C.; Douglas, J. F.; Gilson, M. K. The Fundamental Role of Flexibility on the Strength of Molecular Binding. *Soft Matter* **2012**, *8*, 6385.

[58] Freeman, C. L.; Harding, J. H.; Cooke, D. J.; Elliott, J. A.; Lardge, J. S.; Duffy, D. M. New Forcefields for Modeling Biomineralization Processes. *J. Phys. Chem. C* **2007**, *32*, 11943.

[59] Freeman, C. L.; Harding, J. H.; Quigley D.; Rodger, P. M. Simulations of Ovocleidin-17 Binding to Calcite Surfaces and Its Implications for Eggshell Formation. *J. Phys. Chem. C* **2011**, *115*, 8175.

[60] Freeman, C. L.; Harding, J. H.; Quigley D.; Rodger, P. M. Protein Binding on Stepped Calcite Surfaces: Simulations of Ovocleidin-17 on Calcite {31.16} and {31.8}. *Phys. Chem. Chem. Phys.* **2012**, *28*, 7287.

[61] Frey, P. A. Radical Mechanisms of Enzymatic Catalysis. *Annu. Rev. Biochem.* **2001**, *70*, 121.

[62] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg,J.J.;Dapprich,S.;Daniels,A.D.;Farkas,O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, revision A.02*; Gaussian, Inc.: Wallingford, CT, 2009.

[63] Fu, G.; Valiyaveettil, S.; Wopenka, B.; Morse, D. E. $CaCO_3$ Biomineralization: Acidic 8-kDa Proteins Isolated from Aragonitic Abalone Shell Nacre Can Specifically Modify Calcite Crystal Morphology. *Biomacromolecules* **2005**, *6*, 1289.

[64] Fujiwara, M.; Shiokawa, K.; Morigaki, K.; Zhu, Y; Nakahara, Y. Calcium Carbonate Microcapsules Encapsulating Biomacromolecules. *Chem. Eng. J.* **2008**, *137*, 14.

[65] Gale, J. D. GULP - A Computer Program for the Symmetry Adapted Simulation of Solids. *J. C. S. Faraday Trans.* **1997**, *93*, 629.

[66] Garcia, A. E.; Herce, H.; Paschek, D. Simulations of Temperature and Pressure Unfolding of Peptides and Proteins with Replica Exchange Molecular Dynamics. *Annu. Rep. Comput. Chem.* **2006**, *2*, 83.

[67] Gauss, J. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Clark, N. L., Gasteiger, J., H. F. S., III, Schreiner, P. R., Eds.; Wiley: Chichester, U.K., 1998; p 615.

[68] Gerbaud, V.; Pignol, D.; Loret, E.; Bertrand, J. A.; Berland, Y.; Fontecilla-Camps, J. C. Mechanism of Calcite Crystal Growth Inhibition by the N-Terminal Undecapeptide of Lithostathine. *J. Biol. Chem.* **2000**, *275*, 1057.

[69] Gerothanassis, I. P.; Birlirakis, N.; Karayannis, T.; Tsikaris, V.; Sakarellos-Daitsiotis, M.; Sakarellos, S.; Vitouxb, B.; Marraud, M. ^{17}O NMR and FT-IR Study of the Ionization State of Peptides in Aprotic Solvents. Application to Leu-enkephalin. *FEBS Lett.* **1992**, *298*, 188.

[70] Gilbert, A. T. B.; Hirst, J. D. Charge-Transfer Transitions in Protein Circular Dichroism Spectra. *J. Mol. Struct.: THEOCHEM* **2004**, *675*, 53.

[71] Glättli, A.; Daura, X.; Seebach, D.; van Gunsteren, W. F. Can One Derive the Conformational Preference of a Beta-Peptide from Its CD spectrum? *J. Am. Chem. Soc.* **2002**, *124*, 12972.

[72] Goerigk, L.; Grimme, S. Calculation of Electronic Circular Dichroism Spectra with Time-Dependent Double-Hybrid Density Functional Theory. *J. Phys. Chem. A* **2009**, *113*, 767.

[73] Gopal, K.; Lu, Z.; de Villiers, M. M.; Lvov, Y. Composite Phospholipid-Calcium Carbonate Microparticles: Influence of Anionic Phospholipids on the Crystallization of Calcium Carbonate. *J. Phys. Chem. B* **2006**, *110*, 2471.

[74] Gopal, R.; Park, J. S.; Seo, C. H.; Park, Y. Applications of Circular Dichroism for Structural Analysis of Gelatin and Antimicrobial Peptides. *Int. J. Mol. Sci.* **2012**, *13*, 3229.

[75] Gredičak, M.; Supek, F.; Kralj, M.; Majer, Z.; Hollosi, M.; Šmuc, T.; Mlinarić-Majerski, K.; Horvat, Š. Computational Structure-Activity Study Directs Synthesis of Novel Antitumor Enkephalin Analogs. *Amino Acids* **2009**, *38*, 1185.

[76] Greenfield, N. J. Methods to Estimate the Conformation of Proteins and Polypeptides from Circular Dichroism Data. *Anal. Biochem.* **1996**, *235*, 1.

[77] Greenfield, N. J. Using Circular Dichroism Spectra to Estimate Protein Secondary Structure. *Nature Protoc.* **2007**, *1*, 2876.

[78] Guido, C. A.; Knecht, S.; Kongsted, J; Mennucci, B. Benchmarking Time-Dependent Density Functional Theory for Excited State Geometries of Organic Molecules in Gas-Phase and in Solution. *J. Chem. Theory Comput.* **2013**, *9*, 2209.

[79] Hald, K.; Hättig, C.; Yeager, D. L.; Jørgensen, P. Linear Response CC2 Triplet Excitation Energies. *Chem. Phys. Lett.* **2000**, *328*, 291.

[80] Hall, G. C. The Molecular Orbital Theory of Chemical Valency. VIII. A method of Caculating Ionization Potentials. *Proc. Roy. Soc. (London)* **1951**, *A205*, 541.

[81] Hammes, G. G.; Benkovic, S. J.; Hammes-Schiffer, S. Flexibility, Diversity, and Cooperativity: Pillars of Enzyme Catalysis. *Biochemistry* **2011**, *50*, 10422.

[82] Hartree, D. R. The Wave Mechanics of an Atom with Non-Coulombic Central Field: Parts I, II, III. *Proc. Cam. Phil. Soc.* **1928**, *24*, 246.

[83] Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comp.* **2008**, *4*, 435.

[84] Hirst J. D.; Bhattacharjee S.; Onufriev A. V. Theoretical Studies of Time-Resolved Protein Folding. *Faraday Discuss.* **2002**, *122*, 253.

[85] Hirst, J. D.; Colella, K.; Gilbert, A. T. B. Electronic Circular Dichroism Spectra of Proteins from First Principles Calculations. *J. Phys. Chem. B* **2003**, *107*, 11813.

[86] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65*, 712.

[87] Horvat, Š; Mlinarić-Majerski, K.; Glavaš-Obrovac, Lj.; Jakas, A.; Veljković, J.; Marczi, S.; Kragol, G.; Roščić, M.; Matković, M.; Milostić-Srb, A. Tumor-cell Targeted Methinine-Enkephalin Analogues Containing Unnatural Amino Acids: Design, Synthesis and in Vitro Antitumor Activity. *J. Med. Chem.* **2006**, *49*, 3136.

[88] Hotelling, H. Analysis of a Complex of Statistical Variables into Principal Components. *J. Educ. Psychol*. **1933**, *4*, 417.

[89] Hou, T.; Wang, J.; Zhang, W.; Xu, X. ADME Evaluation in Drug Discovery. 6. Can Oral Bioavailability in Humans be Effectively Predicted by Simple Molecular Property-Based Rules? *J. Chem. Inf. Model*. **2007**, *47*, 460.

[90] https://en.wikipedia.org/wiki/Chirality.

[91] Hub, J. S.; de Groot, B. L.; van der Spoel, D. g_wham – A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.* **2010**, *6*, 3713.

[92] Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33.

[93] Huq, N. L.; Cross, K. J.; Reynolds, E. C. Molecular Modelling of a Multiphosphorylated Sequence Motif Bound to Hydroxyapatite Surfaces. *J. Mol. Modeling* **2000**, *6*, 35.

[94] Jacob, C. R. (2011) *Computational Strategies for Spectroscopy: from Small Molecules to Nano Systems* (1st ed.), edited by V. Barone, Wiley, New York.

[95] Jacquemin, D.; Wathelet, V.; Perpète, E. A.; Adamo, C. Extensive TD-DFT Benchmark: Singlet-Excited States of Organic Molecules. *J. Chem. Theory Comput.* **2009**, *5*, 2420.

[96] Jain, A. K.; Murty, M. N.; Flynn, P. J. Data Xlustering: A Review. *ACM Comput. Surv.* **1999**, *31*, 264.

[97] Jamorski, C.; Foresman, J. B.; Thilgen, C.; Lüthi, H.-P. Assessment of TDDFT for the Calculation of Critical Features in the Absorption Spectra of a Series of Aromatic Donor-acceptor Systems. *J. Chem. Phys.* **2002**, *116*, 8761.

[98] Janin, J.; Sternberg, M. J. E. Protein Flexibility, not Disorder, is Intrinsic to Molecular Recognition *F1000 Biol. Reports* **2013**, *5*, 2.

[99] Jaya, N.; Garcia, V.; Vierling, E. Substrate Binding Site Flexibility of the Small Heat Shock Protein Molecular Chaperones. *Proc. Natl. Acad. Sci. U. S. A* **2009**, *106*, 15604.

[100] Jensen, F. *Introduction To Computational Chemistry*, J. Wiley & Sons Inc., New York, 2007.

[101] Jensen, L.; Swart, M.; van Duijnen, P. T.; Autschbach, J. The Circular Dichroism Spectrum of $[Co(en)3]^{3+}$ in water: A Discrete Solvent Reaction Field Study. *Int. J. Quant. Chem.* **2006**, *106*, 2479.

[102] Jiang, H.; Liu, X.-Y.; Lim, C. T.; Hsu, C. Y. Ordering of Self-Assembled Nanobiominerals in Correlation to Mechanical Properties of Hard Tissues. *Appl. Phys. Lett.* **2005**, *86*, 163901.

[103] Jiang, J.; Abramavicius, D.; Bulheller, B. M.; Hirst, J. D.; Mukamel, S. Ultraviolet Spectroscopy of Protein Backbone Transitions in Aqueous Solution: QM/MM Simulations. *J. Phys. Chem. B* **2010**, *114*, 8270.

[104] John, G.; Jadhav, S. R.; Menon, V. M.; John, V. T. Flexible Optics: Recent Developments in Molecular Gels. *Angew. Chem. Int. Ed*. **2012**, *51*, 1760.

[105] Jolliffe, I. T. *Principal Component Analysis.* New York: Springer; 2002.

[106] Kamali, R.; Kharazmi, A. Molecular Dynamics Simulation of Surface Roughness Effects on Nanoscale Flows. *Int. J. Ther. Sci.* **2011**, *50*, 226.

[107] Kaminský, J.; Kubelka, J.; Bouř, P. Theoretical Modeling of Peptide α-Helical Circular Dichroism in Aqueous Solution. *J. Phys. Chem. A* **2011**, *115*, 1734.

[108] Kästner, J. Umbrella Sampling. *Comput. Mol. Sci.* **2011**, *1*, 932.

[109] Katz, J. L.; Reick, M. R.; Herzog, R. E.; Parsiegla, K. I. Calcite Growth Inhibition by Iron. *Langmuir* **1993**, *9*, 1423.

[110] Kelly, S. M.; Price, N. C. The Application of Circular Dichroism to Studies of Protein Folding and Unfolding. *Biochim. Biophys. Acta* **1997**, 1*338*, 161.

[111] Kendrew, J. C. In *The Encyclopedia of Molecular Biology*; Blackwell Science: Oxford, United Kingdom, 1994; pp. 772-781.

[112] Knipp, G. T.; Vander Velde, D. G.; Siahaan, T. J.; Borchardt, R. T. The Effect of Beta-Turn Structure on the Passive Diffusion of Peptides across Caco-2 Cell Monolayers. *Pharm. Res.* **1997**, *14*, 1332.

[113] Kohn W.; Sham L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, 1133.

[114] Kohn, H.; Becke, A. D.; Parr, R. G. Density Functional Theory of Electronic Structure. *J. Phys. Chem.* **1996**, *100*, 12974.

[115] Kohn, W.; Hohenberg, P. Inhomogeneous Electron Gas. *Phys. Rev. B* **1964**, *136*, 864.

[116] Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comp. Chem.* **1992**, *13* , 1011.

[117] Kuritz, N.; Stein, T.; Baer, R.; Kronik, L. Charge-Transfer-Like $\pi \rightarrow \pi^*$ Excitations in Time-Dependent Density Functional Theory: A Conundrum and Its Solution. *J. Chem. Theory Comput* **2011**, *7*, 2408.

[118] Kuroda, Y.; Ueda, H.; Nozawa, H.; Ogoshi, H. Adamantyl Amino Acid as γ-Turn Inducer for Peptide. *Tetrahedron Lett.* **1997**, *38*, 7901.

[119] Kypr, J.;Kejnovská, I.; Renčiuk, D.; Vorličková, M. Circular Dichroism and Conformational Polymorphism of DNA. *Nucleic Acids Res.* **2009**, *37*, 1713.

[120] Landoni M. F.; Soraci A. Pharmacology of Chiral Compounds: 2-Arylpropionic Acid Derivatives. *Curr. Drug Metab.* **2001**, *2*, 37.

[121] Laurent, A. D.; Jacquemin, D. TD-DFT Benchmarks: A Review. *Int. J. Quant. Chem.* **2013**, *113*, 2019.

[122] Leach, A. R. *Molecular Modelling*, Prentice Hall, 2001.

[123] Leschziner, A. E.; Nogales, E. Visualizing Flexibility at Molecular Resolution: Analysis of Heterogeneity in Single-Particle Electron Microscopy Reconstructions. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 43.

[124] Li, R.; Zheng, J.; Truhlar, D. G. Density Functional Approximations for Charge Transfer Excitations with Intermediate Spatial Overlap. *Phys. Chem. Chem. Phys.* **2010**, 12, 12697.

[125] Limão-Vieira, P.; Giuliani, A.; Delwiche, J.; Parafita, R.; Mota, R.; Duflot, D.; Flament, J.-.P.; Drage, E.; Cahillane, P.; Mason, N. J.; Hoffmann, S. V.; Hubin-Franskin, M.-J. Acetic Acid Electronic State Spectroscopy by High-Resolution VUV Photo-Absorption, Electron Impact, He(I) Photoelectron Spectroscopy and Ab Initio Calculations. *Chem. Phys.* **2006**, *324*, 339.

[126] Lin, A.; Meyers, M. A. Growth and Structure in Abalone Shell. *Mater. Sci. Eng. A* **2005**, *390*, 27.

[127] Ling, L.; Fan, M.; Wang, B.; Zhang, R. Application of Computational Chemistry in Understanding the Mechanisms of Mercury Removal Technologies: A Review. *Energy Environ. Sci.* **2015**, Advance Article.

[128] Maass, A.; Tekin, E. D.; Schüller, A.; Palazoglu, A.; Reith, D.; Faller, R. Folding and Unfolding Characteristics of Short Beta Strand Peptides under Different Environmental Conditions and Starting Configurations. *Biochim. Biophys. Acta*. **2010**, *1804*, 2003.

[129] Mann, S. Molecular Recognition in Biomineralization. *Nature* **1988**, *332*, 119.

[130] Mann, S.; Didymus, J. M.; Sanderson, N. P.; Heywood, B. R.; Samper, E. J. A. Morphological Influence of Functionalized and Non-Functionalized-Alpha, Omega-Dicarboxylates on Calcite Crystallization. *J. Chem. Soc.* **1990**, *86*, 1873.

[131] Marcotte, I.; Separovic, F.; Auger, M.; Gagne, S. M. A Multidimensional 1H NMR Investigation of the Conformation of Methionine-Enkephalin in Fast-Tumbling Bicelles. *Biophys. J.* **2004**, *86*, 1587.

[132] Marina, M.; Safont-Sempere, M.; Osswald, P.; Stolte, M.; Grüne, M.; Renz, M.; Kaupp, M.; Radacki, K.; Braunschweig, H.; Würthner, F. Impact of Molecular Flexibility on Binding Strength and Self-Sorting of Chiral π-Surfaces. *J. Am. Chem. Soc.* **2011**, *133*, 9580.

[133] Marin-Garcia, L.; Frontana-Uribe, B. A.; Reyes-Grajeda, J. P.; Stojanoff, V.; Serrano-Posada, H. J.; Moreno, A. Chemical Recognition of Carbonate Anions by Proteins Involved in Biomineralization Processes and Their Influence on Calcite Crystal Growth. *Cryst. Growth Des.* **2008**, *8*, 1340.

[134] Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *J. Comp. Chem.* **2009**, *30*, 2157.

[135] Masica, D. L.; Schrier, S. B.; Specht, E. A.; Gray, J. J. De Novo Desigh of Peptide-Calcite Biomineralization Systems. *J. Am. Chem. Soc.* **2010**, *132*, 12252.

[136] Mavrikakis, M. Computational Methods: A Search Engine for Catalysts. *Nature Mater.* **2006**, *5*, 847.

[137] Mearns, D.; Kohn, W. Frequency-Dependent *v*-Representability in Density-Functional Theory. *Phys. Rev. A* **1987**, *35*, 4796.

[138] Metzler, R. A.; Tribello, G. A.; Parrinello, M.; Gilbert, P. U. P. A. Asprich Peptides are Occluded in Calcite and Permanently Disorder Biomineral Crystals. *J. Am. Chem. Soc.* **2010**, *132*, 11585.

[139] Michalka, J. R.; McIntyre, P. W.; Daniel Gezelter, J. Molecular Dynamics Simulations of the Surface Reconstructions of Pt(557) and Au(557) under Exposure to CO. *J. Phys. Chem. C* **2013**, *117*, 14579.

[140] Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618.

[141] Nassralla-Aboukais, N.; Boughriet, A.; Fischer, J. C.; Wartel, M.; Langelin, H. R.; Aboukais, A. Electron Paramagnetic Resonance (EPR) Study of Cu^{2+} and Mn^{2+} Ions Interacting as Probes with Calcium Carbonate during the Transformation of Vaterite into Cubic Calcite. *J. Chem. Soc.* **1996**, *92*, 3211.

[142] Ogomi, D.; Serizawa, T.; Akashi, M. Controlled Release Based on the Dissolution of a Calcium Carbonate Layer Deposited on Hydrogels. *J. Controlled Release* **2005**, *103*, 315.

[143] Ohkura, K.; Shinohara, Y.; Hori, H. Efficiency of Antimicrobial Defense: Molecular Flexibility of Natural Defensin and Artificial Bis-Quaternary Ammonium Compound. *Anticancer Res.* **2011**, *31*, 2561.

[144] Orme, C. A.; Noy, A.; Wierzbicki, A.; McBride, M. T.; Grantham, M.; Teng, H. H.; Dove, P. M.; DeYoreo, J. J. Formation of Chiral Morphologies through Selective Binding of Amino Acids to Calcite Surface Steps. *Nature* **2001**, *411*, 775.

[145] Pace, C. N. Determination and Analysis of Urea and Guanidine Hydrochloride Denaturation Curves. *Methods Enzymol.* **1986**, *131*, 266.

[146] Papaleo, E.; Tiberti, M.; Invernizzi, G.; Pasi, M.; Ranzani, V. Molecular Determinants of Enzyme Cold Adaptation: Comparative Structural and Computational Studies of Cold- and Warm-Adapted Enzymes. *Curr. Protein Pept. Sci.* **2011**, *12*, 657.

[147] Parac, M.; Grimme, S. Comparison of Multireference Møller-Plesset Theory and Time-Dependent Methods for the Calculation of Vertical Excitation Energies of Molecules. *J. Phys. Chem. A* **2002,** *106*, 6844.

[148] Park, S.; Wang, G.; Cho, B.; Kim, Y.; Song, S.; Ji, Y.; Yoon, M.-H.; Lee, T. Flexible Molecular-Scale Electronic Devices. *Nature Nanotech.* **2012**, *7*, 438.

[149] Parr, R. G.; Yang, W., *Density Functional Theory of Atoms and Molecules*; Oxford Univ. Press, New York, 1989.

[150] Pasarin, I. S.; Yang, M.; Bovet, N.; Glyvradal, M.; Nielsen, M. M.; Bohr, J.; Feidenhans'l, R.; Stipp, S. L. S. Molecular Ordering of Ethanol at the Calcite Surface. *Langmuir* **2012**, *28*, 2545.

[151] Patocka J.; Dvorak A. Biomedical Aspects of Chiral Molecules. *J. Applied Med.* **2004**, *2*, 95.

[152] Pavese, A.; Catti, M.; Parker, S. C.; Wall, A. Modelling of the Thermal Dependence of Structural and Elastic Properties of Calcite, $CaCO_3$. *Phys. Chem. Miner.* **1996**, *23*, 89.

[153] Pavese, A.; Catti, M.; Price, G. D.; Jackson, R. A. Interatomic Potentials for $CaCO_3$ Polymorphs (Calcite and Aragonite), Fitted to Elastic and Vibrational Data. *Phys. Chem. Miner.* **1992**, *19*, 80.

[154] Peach, M. J. G.; Benfield, P.; Helgaker, T.; Tozer, D. J. Excitation Energies in Density Functional Theory: An Evaluation and a Diagnostic Test. *J. Chem. Phys.* **2008**, *128*, 044118.

[155] Pearson, K. *Philos. Mag.* On Lines and Planes Closest Fit to Systems of Points in Space. **1901**, *2*, 559.

[156] Ponder, J. W.; Case D. A. Force Fields for Protein Simulations. *Adv. Protein Chem.* **2003**, *66*, 27.

[157] Qi, S.; Krogsgaard, M.; Davis, M. M.; Chakraborty, A. K. Molecular Flexibility can Influence the Stimulatory Ability of Receptor-Ligand Interactions at Cell-Cell Junctions. *Proc. Natl. Acad. Sci. U. S. A* **2006**, *103*, 4416.

[158] Raikumar S. V. Thalidomide: Tragic Past and Promising Future. *Mayo Clin. Proc.* **2004**, *79*, 899.

[159] Raut, V. P.; Agashe, M. A.; Stuart, S. J.; and Latour, R. A. Molecular Dynamics Simulations of Peptide–Surface Interactions. *Langmuir* **2005**, *21*, 1629.

[160] Rogers, D. M.; Hirst, J. D. Calculations of Protein Circular Dichroism from First Principles. *Chirality* **2004**, *16*, 234.

[161] Roothan, C. C. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **1951**, *23*, 69.

[162] Rudolph, M; Autschbach, J. Performance of Conventional and Range-Separated Hybrid Density Functionals in Linear Response Calculations of Electronic Circular Dichroism Spectra of Transition Metal Complexes. *J. Phys. Chem. A* **2011**, *115*, 14677.

[163] Runge E.; Gross, E. K. U. Density-Functional Theory for Time-Dependent Systems. *Phys. Rev. Lett.* **1984**, *59*, 997.

[164] Sanbonmatsu, K. Y.; Garcia, A. E. Structure of Met-Enkephalin in Explicit Aqueous Solution using Replica Exchange Molecular Dynamics. *Proteins* **2002,** *46*, 225.

[165] Sanchez, M. L.; Aguilar, M. A.; Olivares del Valle, F. J. Study of Solvent Effects by Means of Averaged Solvent Electrostatic Potentials Obtained from Molecular Dynamics Data. *J. Comput. Chem.* **1997**, *18*, 313.

[166] Schellman, J. A. Circular Dichroism and Optical Rotation. *Chem. Rev.* **1975**, *75*, 323.

[167] Schreiber, M.; Silva-Junior, M. R.; Sauer, S. P. A.; Thiel, W. Benchmarks for Electronically Excited States: CASPT2, CC2, CCSD, and CC3. *J. Chem. Phys.* **2008**, *128*, 134110.

[168] Schrier, S. B.; Sayeg, M. K.; Gray, J. J. Prediction of Calcite Morphology from Computational and Experimental Studies of Mutations of a De Novo-Designed Peptide. *Langmuir*, **2010**, *27*, 11520.

[169] Schröder, K.-P.; Sauer, J.; Leslie, M.; Catlow, C. R. A.; Thomas, J. M. Bridging Hydroxyl Groups in Zeolitic Catalysts: A Computer Simulation of Their Structure, Vibrational Properties and Acinidity in Protonated Faujasites H-Y Zeolites). *Chem. Phys. Lett.* **1992**, *188*, 320.

[170] Schuler, B.; Hofmann, H. Single-Molecule Spectroscopy of Protein Folding Dynamics – Expanding Scope and Timescales. *Curr. Opin. Struct. Biol.* **2013**, *23*, 36.

[171] Sebastián, N.; López, D. O.; Diez-Berart, S.; de La Fuente, M. R.; Salud, J.; Pérez-Jubindo, M. A.; Ros, M. B. Effect of Molecular Flexibility on the Nematic-to-Isotropic Phase Transition for Highly Biaxial Molecular Non-Symmetric Liquid Crystal Dimers. *Materials* **2011**, *4*, 1632.

[172] Shao, J.; Tanner, S. W.; Thompson, N.; Cheatham, T. E. Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *J. Chem. Theory Comput.* **2007**, *3*, 2312.

[173] Shen, J.-W.; Li, C.; van der Vegt, N. F. A.; Peter, C. Understanding the Control of Mineralization by Polyelectrolyte Additives: Simulation of Preferential Binding to Calcite Surfaces. *J. Phys. Chem. C* **2013**, *117*, 6904.

[174] Shou, Q.; Smith, J. E.; Mon, H.; Brkljača, Z., Smith, A.-S.; Smith, D. M.; Griesser, H. J.; Wohlmuth, H. Rhodomyrtals A–D, Four Unusual Phloroglucinol-Sesquiterpene Adducts from Rhodomyrtus Psidioides. *RSC Adv.* **2014**, *4*, 13514.

[175] Sigel, A.; Sigel, H.; Sigel, R. K. O. *Biomineralization. From Nature to Application*; Wiley: New York, 2008; Vol. 4.

[176] Silva-Junior, M. R.; Schreiber, M.; Sauer, S. P. A.; Thiel, W. Benchmarks for Electronically Excited States: Time-Dependent Density Functional Theory and Density Functional Theory Based Multireference Configuration Interaction. *J. Chem. Phys.* **2008**, *129*, 104103.

[177] Simha, R. The Influence of Molecular Flexibility on the Intrinsic Viscosity, Sedimentation, and Diffusion of High Polymers. *J. Chem. Phys.* **1945**, *13*, 188.

[178] Stryer, L.; Berg, J. M.; Tymoczko, J. L. (2002), *Biochemistry* (5th ed.), New York: W. H. Freeman

[179] Su, L.; Cukier, R. I. Hamiltonian and Distance Replica Exchange Method Studies of Met-Enkephalin. *J. Phys. Chem. B* **2007**, *111*, 12310.

[180] Sugita Y.; Okamoto. Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141.

[181] Sun, Y., Dominy, B.N. and Latour, R. A. Comparison of Solvation-Effect Methods for the Simulation of Peptide Interactions with a Hydrophobic Surface *J. Comp. Chem.* **2007**, *28*, 1883.

[182] Tan, P. N.; Steinbach, M.; Kumar, V. *Introduction to Data Mining*, Addison-Wesley, chapter *Cluster Analysis: Basic Concepts and Algorithms*. pp. 487–568, 2006.

[183] Tanaka, T.; Kodoma, T. S.; Morita, H. E.; Ohno, T. Electronic and Vibrational Circular Dichroism of Aromatic Amino Acids by Density Functional Theory. *Chirality* **2006**, *18*, 652.

[184] Teixeira, C.; Serradji, N.; Amroune, S.; Storck, K.; Rogez-Kreuz, C.; Clayette, P.; Barbault, F.; Maurel, F. Is the Conformational Flexibility of Piperazine Derivatives Important to Inhibit HIV-1 Replication? *J. Mol. Graphics Modell.* **2013**, *44*, 91.

[185] Terenius, L. Alcohol Addiction (Alcoholism) and the Opioid System. *Alcohol* **1996**, *13*, 31.

[186] Titiloye, J. O.; Parker, S. C.; Mann, S. Atomistic Simulation of Calcite Surfaces and the Influence of Growth Additives on Their Morphology. *J. Cryst. Growth* **1993**, *131*, 533.

[187] Tozer, D. J.; Amos, R. D.; Handy, N. C.; Roos, B. O.; Serrano-Andrés, L. Does Density Functional Theory Contribute to the Understanding of Excited States of Unsaturated Organic Compounds? *Mol. Phys.* **1999**, *97*, 859.

[188] TURBOMOLE V6.3 2011, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH.

[189] Ueno, Y.; Futagawa, H.; Takagai, Y.; Ueno, A.; Mizushima, Y. Drug-Incorporating Calcium Carbonate Nanoparticles for a New Delivery System. *J. Controlled Release* **2005**, *103*, 93.

[190] Ukrainczyk, M.; Gredičak, M.; Jerić, I.; Kralj, D. Interactions of Salicylic Acid Derivatives with Calcite Crystals. *J. Colloid Interface Sci.* **2012**, *365*, 296.

[191] Ukrainczyk, M.; Gredičak, M.; Jerić, I.; Kralj, D. Unpublished results.

[192] Ullrich, C. A. *Time-dependent Density-functional Theory - Concepts and Applications*, Oxford University Press, 2012.

[193] van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; Kastenholz, M. A.;Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F. A.; Yu, H. B. Biomolecular Modelling: Goals, Problems, Perspectives. *Angew. Chem. Int. Ed.* **2006**, *45*, 4064.

[194] van Leeuwen, R. Mapping from Densities to Potentials in Time-Dependent Density-Functional Theory. *Phys. Rev. Lett.* **1999**, *82*, 3863.

[195] Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular Properties that Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem.* **2002**, *45*, 2615.

[196] Wakai, A.; Fukasawa, H.; Yang C.; Mori, T.; Inoue Y. Theoretical and Experimental Investigations of Circular Dichroism and Absolute Configuration Determination of Chiral Anthracene Photodimers. *J. Am. Chem. Soc.* **2012**, *134*, 4990.

[197] Wang, C.; He, C.; Tong, Z.; Liu, X.; Ren, B.; Zeng, F. Combination of Adsorption by Porous $CaCO_3$ Microparticles and Encapsulation by Polyelectrolyte Multilayer Films for Sustained Drug Delivery. *Int. J. Pharm.* **2006**, *308*, 160.

[198] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157.

[199] Wang, X. Q.; Wu, C. M.; Tao, K.; Zhao, K.; Wang, J. Q.; Xu, H.; Xia, D. H.; Shan, H. H.; Lu, J. R. Influence of Ovalbumin on $CaCO_3$ Precipitation during in Vitro Biomineralization. *J. Phys. Chem. B* **2010**, *114*, 5301.

[200] Wang, Y.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comp. Chem.* **2000**, *21*, 1049.

[201] Warnke, I.; Furche, F. Circular Dichroism: Electronic. *WIREs Comput. Mol. Sci.* **2012**, *2*, 150.

[202] Wei, H.; Shen, Q.; Zhao, Y.; Wang, D.; Xu, D. Crystallization Habit of Calcium Carbonate in Presence of Sodium Dodecyl Sulfate and/or Polypyrrolidone. *J. Cryst. Growth* **2004**, *260*, 545.

[203] Weiner, S.; Addadi, L. J. Design Strategies in Mineralized Biological Materials. *Mater. Chem.* **1997**, *7*, 68.

[204] Wierzbicki, A.; Dalal, P.; Cheatham III, T. E.; Knickelbein, J. E.; Haymet, A. D. J.; Madura, J. D. Antifreeze Proteins at the Ice/Water Interface: Three Calculated Discriminating Properties for Orientation of Type I Proteins. *Biophys. J.* **2007**, *93*, 1442.

[205] Wolf, A.; Kirschner, K. N. Principal Component and Clustering Analysis on Molecular Dynamics Data of the Ribosomal L11·23S Subdomain. *J. Mol. Model.*, **2013**, *19*, 539.

[206] Woody, R. W. The Exciton Model and the Circular Dichroism of Polypeptides. *Monatsh. Chem.* **2005**, *136*, 347.
[207] Yanai, T.; Tew, D. P.; Handy, N. C. A New Hybrid Exchange–Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51.
[208] Yang, M.; Rodger, P. M.; Harding, J. H.; Stipp, S. L. S. Molecular Dynamics Simulations of Peptides on Calcite Surface. *Mol. Simul.* **2009**, *35*, 547.
[209] Zagon, I. S.; Verderame, M. F.; McLaughlin, P. J. The Biology of the Opioid Growth Factor Receptor (OGFr). *Brain Res. Rev.* **2002**, *38*, 351.
[210] Zhang, L.; Peslherbe, G. H.; Muchall, H. M. Ultraviolet Absorption Spectra of Substituted Phenols: A Computational Study. *Photochem. Photobiol.* **2006**, *82*, 324.
[211] Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals. *Theor. Chem. Acc.* **2008**, *120*, 215.
[212] Zhou, R. Replica Exchange Molecular Dynamics Method for Protein Folding Simulation. *Methods Mol. Biol.* **2007**, *350*, 205.

Curriculum Vitae

Personal Information:

Family name:	Brkljača
Name:	Zlatko
Date of birth:	11. March 1986
Place of birth:	Zadar, Croatia
Nationality:	Croatian

Education:

Since May 2010	PhD studies in theoretical chemistry, Institute for Theoretical Physics and Cluster of Excellence: Engineering of Advanced Materials (EAM), University Erlangen-Nürnberg Member of Graduate School "Advanced Materials and Processes" Topic: Application of Computational Methods to the Structural and Functional Properties of Flexible Chiral Molecules Superviser: Prof. Dr. Ana-Sunčana Smith Co-supervisor: Prof. Dr. David M. Smith
Sept. 2009	Graduated as „Diplomirani inženjer kemije" (Dipl. -Ing. in Chemistry) Grade: 4.97 (max. 5.0)
Oct. 2004 - Sept. 2009	Student at the Department of Chemistry, Faculty of Mathematics and Natural Sciences, University of Zagreb, Croatia.
Mar. 2008 - Sept. 2009	Completion of the diploma project at Institute Ruđer Bošković, Zagreb, Croatia Thesis: "Structure of Methionine-enkephalin and Its Unnatural Analogue Using Molecular Dynamics" Supervisor: Prof. Dr. David M. Smith Grade: Excellent
May 2004	School leaving examination, grade: 5.0 (best grade)
Sept. 2000 - May 2004	Secondary school, Gimnazija Pula, Pula, Croatia
Sept. 1992 - Juli 2000	Primary school, OŠ Šijana, Pula, Croatia

Work experience and training:

October 2015 -	Research assistant on the ESF project MIPoMat, RBI, Zagreb, Group of Dr. David M. Smith
March 2015 – Sept. 2015	Research assistant on the UKF project "Time resolved photoelectron spectroscopy as a probe for ultrafast dynamics", RBI, Zagreb, Group of Dr. Nađa Došlić

List of Publications

1. Brkljača, Z.; Čondić-Jurkić, K.; Smith, A.-S.; Smith, D. M. Calculation of a CD Spectrum of a Peptide from its Conformational Phase Space: The Case of Met-enkephalin and Its Unnatural Analogue. *J. Chem. Theory Comput.*, **2012**, *8*, 1694.

2. Brkljača, Z.; Mališ, M.; Smith, D. M.; Smith, A.-S. Calculating CD Spectra of Flexible Peptides: An Assessment of TD-DFT functionals. *J. Chem. Theory Comput.*, **2014**, *10* , 3270.

3. Shou, Q.; Smith, J. E.; Mon, H.; Brkljača, Z., Smith, A.-S.; Smith, D. M.; Griesser, H. J.; Wohlmuth, H. Rhodomyrtals A–D, four unusual phloroglucinol-sesquiterpene adducts from Rhodomyrtus psidioides. *RSC Adv.* **2014**, *4*, 13514.

4. Brkljača, Z.; Klimczak M.; Miličević, Z.; Weisser M.; Taccardi, N.; Wasserscheid, P.; Smith, D. M.; Magerl, A.; Smith, A.-S. Complementary Molecular Dynamics and X-ray Reflectivity Study of an Imidazolium-Based Ionic Liquid at a Neutral Sapphire Interface. *J. Phys. Chem. Lett.* **2015**, *6*, 549.

Contributed Talks and Posters at Conferences

Talks:

15 - 20 Mar. 2015 Spring Meeting of the Deutsche Physikalische Gesellschaft, Berlin, Germany.

29 Aug. - 1 Sept. 2013 Computational Life Sciences on Open Shell Species, Primošten, Croatia.

10 - 15 Mar. 2013 Spring Meeting of the Deutsche Physikalische Gesellschaft, Regensburg, Germany.

4 - 6 Apr. 2011 25th Molecular Modelling Workshop, Erlangen, Germany.

15 - 18 Mar. 2011 EAM Winter School 2011, Kirchberg, Austria 15 – 20 Mar. 2015 Spring Meeting of the Deutsche Physikalische Gesellschaft, Berlin, Germany.

Posters:

22 July 2010 EAM Young Researchers' Day, Erlangen, Germany.

11 July 2012 2nd EAM Young Researchers' Day, Erlangen, Germany.

17 - 22 July 2011 WATOC, Santiago de Compostela, Spain.

12 - 14 Mar. 2012 26th Molecular Modelling Workshop, Erlangen, Germany.

25 - 30 Mar. 2012 Spring Meeting of the Deutsche Physikalishe Gesellschaft, Berlin, Germany.

25 - 27 Feb. 2013 27th Molecular Modelling Workshop, Erlangen, Germany.

22 - 26 Sept. 2013 49th Symposium on Theoretical Chemistry, Erlangen, Germany.

30 Mar.-4 Apr. 2014 Spring Meeting of the Deutsche Physikalishe Gesellschaft, Dresden, Germany.

26 - 27 Feb. 2015 Symposium Physics & Medicine, Erlangen, Germany.