

Effiziente Fluidphasentheorie für Protonierungsprozesse in komplexen Systemen

Dissertation zur Erlangung des Doktorgrads vorgelegt von

Dipl.-Ing. Jochen Heil

Die Dissertation wurde im Zeitraum von 09/2008 bis 09/2009 am Eduard-Zintl-Institut für Anorganische und Physikalische Chemie der Technischen Universität Darmstadt und von 10/2009 bis 12/2016 an der Fakultät für Chemie und Chemische Biologie der Technischen Universität Dortmund angefertigt.

Erstgutachter: Prof. Dr. Stefan M. Kast

Zweitgutachter: Prof. Dr. Alfons Geiger

Dortmund 2016

„The fundamental laws necessary for the mathematical treatment of a large part of physics and the whole of chemistry are thus completely known, and the difficulty lies only in the fact that application of these laws leads to equations that are too complex to be solved. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation.“

P. A. M. Dirac

„Shut up and calculate!“

N. D. Mermin zugeschrieben

Inhaltsverzeichnis

Danksagung	4
Zusammenfassung/Abstract.....	5
Fremdanteile in dieser Arbeit	7
Digitale Unterstützungsmaterialien	8
1 Einleitung	10
2 Grundlagen	12
2.1 Integralgleichungstheorie molekularer Fluide.....	12
2.1.1 Klassische Dichtefunktionaltheorie und das OZ-/Closure-Gleichungssystem.....	12
2.1.2 Erweiterung auf Solvat-Solvens-Systeme	16
2.1.3 Erweiterungen für anisotrope Fluide	17
2.1.4 Das 3D <i>Reference Interaction Site Model</i> (3D-RISM).....	18
2.2 Das EC-RISM-Solvatationsmodell für quantenchemische Berechnungen.....	24
3 Ergebnisse.....	27
3.1 Beschleunigung der 3D-RISM-Methodologie.....	27
3.1.1 Adaption des <i>Particle-Mesh-Ewald</i> Ansatzes für 3D-RISM.....	27
3.1.2 Schnelle Berechnung des chemischen Exzesspotentials im reziproken Raum.....	32
3.1.3 OpenMP-Parallelisierung des 3D-RISM Löasers.....	36
3.1.4 Genauigkeits- und Geschwindigkeitstests	38
3.2 QM/MM-Berechnungen mit EC-RISM-Solvens.....	45
3.3 Schnelle Approximation der Solvens-Elektrostatik für EC-RISM.....	47
3.4 Tautomerenverhältnisse, pK_a und Speziesverteilung in Lösung.....	53
3.4.1 Reaktions-Gibbs-Energien und das Potential der mittleren Kraft	54
3.4.2 Tautomeren- und Dissoziationsgleichgewichte	58
3.4.3 pH-abhängige Speziesverteilung in Lösung	59
3.5 Kraftfeld-Geometrieoptimierung mit AMBER/3D-RISM	62

3.6	<i>ab initio</i> p <i>K</i> _a -Vorhersage kleiner organischer Moleküle	69
3.7	Tautomerenverhältnisse von Guaninen	80
3.8	pH-abhängige Speziesverteilung von Histamin	84
3.9	Charakterisierung des Protonierungszustands des RTA/PTA-Komplexes	95
3.9.1	Ligand.....	99
3.9.2	Komplex	123
4	Zusammenfassung und Ausblick.....	134
5	Anhang	138
5.1	Details der p <i>K</i> _a -Modelle	138
5.1.1	Gibbs-Energien aller Spezies	138
5.2	Gibbs-Energien aller Histaminkonformere	140
5.3	Aufbau des QM/MM-Subsystems des RTA-PTA-Komplexes	152
6	Literaturverzeichnis.....	155

Danksagung

- Herrn Prof. Dr. Stefan M. Kast danke ich für die Möglichkeit, diese Arbeit unter seiner Leitung anzufertigen, seine geduldige kontinuierliche Unterstützung und viele gute Gespräche.
- Meinen Eltern danke ich für die jahrelange finanzielle und emotionale Unterstützung. Die vorliegende Arbeit widme ich meinem verstorbenen Vater.
- Meinen Kollegen, Diplomanden, Master- und Bacheloranden und Praktikanten danke ich für die gute Zusammenarbeit. Das betrifft in (halbwegs) chronologischer Reihenfolge: Michael Kreim, Sebastian Ehrhardt, Karen Kuhn, Leonard Henkes, Roland Frach, Daniel Tomazic, Franziska Hoffgaard, Florian Mrugalla, Patrick Kibies, Simon Egbers, Nicolas Tielker, Martin Urban, Raj Schneider und Lukas Eberlein. Es war mir ein Privileg, mit euch arbeiten zu dürfen! Ich hoffe nur, ich habe niemanden vergessen... Herausheben möchte ich meinen Zimmerkollegen Patrick Kibies, der mir speziell in der letzten sehr stressigen Zeit mit Rat und Tat zur Seite stand, sowie Leonhard Henkes für die mentale Unterstützung zu ungewöhnlicher Uhrzeit.
- Herrn Prof. Dr. Alfons Geiger danke ich für den Tipp, bei der Nächster-Nachbar-Suche im „Punktladungskompressor“ über das Thema Voronoi-Zerlegung nachzudenken, sein allgemeines Interesse an meiner Arbeit und speziell für seine Bereitschaft, das Koreferat zu übernehmen, obwohl er bereits seit längerem in Rente ist.
- Herrn Dr. Stefan Güssregen danke ich für die Bereitstellung mittels Chemoinformatiktools erzeugter Ausgangsstrukturen und seine Bereitschaft zur Diskussion damit verbundener Problematiken.
- Herrn Dr. Jan Treibig danke ich für die Unterstützung bei der Parallelisierung des DIIS-Verfahrens.
- Meinem Arbeitgeber GNS Systems GmbH danke ich für die Flexibilität, die mir die Fertigstellung dieser Arbeit ermöglicht hat.
- Ganz spezieller Dank gebührt allen, die Teile dieser Arbeit oder auch der Fachartikel korrekturgelesen oder wichtige Diskussionsbeiträge geliefert haben.

Zusammenfassung/Abstract

Im Rahmen der vorliegenden Arbeit wurden effiziente Methoden entwickelt, um den Protonierungszustand kleiner Moleküle in der Bindetasche von Proteinen mittels physikbasierter computergestützter Berechnungen in Form der *Embedded-Cluster Reference Interaction Site Model* (EC-RISM-) Methode zu charakterisieren, die quantenchemische Berechnungen mit einem Modell zur Quantifizierung von Lösungsmittelleffekten („Solvatationsmodell“) auf Basis der *3D Reference Interaction Site Model* (3D-RISM-) Integralgleichungstheorie verknüpft. Hierzu waren zwei verschiedene Arten von Herausforderungen zu überwinden: Um den Zeitaufwand für die Berechnungen akzeptabel zu halten mussten die verwendeten Computerprogramme beschleunigt werden und für die Beschreibung komplexer Protonierungsszenarien musste ein für die EC-RISM-Methode adäquates theoretisches Fundament gefunden werden. Um diesen Herausforderungen zu begegnen wurden der Quellcode des 3D-RISM-Lösers parallelisiert und optimiert, ein schnelles Summationsverfahren aus dem Bereich der Partikelsimulation für 3D-RISM adaptiert, aufwendig zu berechnende langreichweitige Realraumterme für die Berechnung thermodynamischer Observablen von großen uniformen 3D-Gittern auf für die Computerauswertung vorteilhafte radiale logarithmische Gitter verlagert und ein adaptives Verfahren zur Reduktion der Anzahl der für EC-RISM benötigten 1-Elektronenintegrale, ein EC-RISM-Quantenmechanik/Molekülmechanik- (QM/MM-) Verfahren sowie ein statistisch-mechanischer Formalismus zur Behandlung gekoppelter Ionisierungs- und Tautomerengleichgewichte auf der Basis von Gibbs-Energien entwickelt. Im EC-RISM-QM/MM-Verfahren können Systeme mit tausenden Atomen wie z.B. Protein-Ligand-Komplexe traktiert werden, da die Teile dieser Systeme, die vom Ort reaktiver Veränderungen weiter entfernt sind, mit Hilfe einfacher klassischer Kraftfelder beschrieben werden. Mit dem statistisch-mechanischen Formalismus kann der jeweilige Anteil unterschiedlich protonierter chemischer Spezies in Abhängigkeit des pH-Werts berechnet werden. Die Methodik wurde anhand mehrerer unterschiedlicher Testsysteme mit zunehmender Komplexität validiert. Als Anwendungsbeispiel wurde der Protonierungszustand von Pteroinsäure im Komplex mit dem stark toxischen Protein Rizin charakterisiert.

The present thesis deals with the creation of efficient methods for the characterization of protonation states of small molecules inside the binding pockets of proteins using physics-based computer-assisted calculations with the *Embedded-Cluster Reference Interaction Site Model* (EC-RISM-) method which couples quantum-chemical calculations with a model for the quantification

of solvent effects („solvation model“) based on 3D Reference Interaction Site Model (3D RISM) integral equation theory. For achieving this objective, two different types of obstacles needed to be overcome: To keep computing times reasonable, the computer codes needed acceleration and a statistical-mechanical formalism adequate for EC-RISM needed to be developed. To this end, the 3D RISM software was parallelized and optimized, a fast summation technique from the area of particle simulations was adapted for 3D RISM, computationally expensive long-range real-space terms needed to compute thermodynamic observables were shifted from uniform 3D grids to computationally expedient radial logarithmic grids and an adaptive method to reduce the number of 1-electron-integrals needed for EC-RISM, an EC-RISM-Quantum Mechanics/Molecular Mechanics- (QM/MM-) method as well as a statistical-mechanical framework for the treatment of coupled ionization and tautomerization equilibria was developed. The EC-RISM-QM/MM-method can be used to treat systems containing thousands of atoms like e.g. protein-ligand-complexes since parts of the system far removed from the place of reactive events can be described using simple force-fields. Using the statistical-mechanical framework, we proceed to calculate the fractions of different chemical species exhibiting diverse protonation pattern as a function of pH. The methodology was validated on several test systems with increasing complexity. As an application, the protonation state of pteric acid in complex with the highly toxic protein ricin was characterized.

Fremdanteile in dieser Arbeit

- Die Berechnung von Ionisierungsstufen und Tautomerenverhältnissen des Histamins wurde unter Anleitung des Autors und unter Verwendung des vom Autor entwickelten Formalismus von Herrn Lukas Eberlein durchgeführt.
- Ein Teil der pK_a -Modelle wurde bereits im Rahmen der Diplomarbeit des Autors vorgestellt.
- Die ursprüngliche Methodologie um mittels EC-RISM einfache Tautomerenverhältnisse zu berechnen wurde von Prof. Stefan M. Kast im Rahmen des SAMPL2-Wettbewerbs etabliert.
- Die Entwicklung des Quellcodes zur Berechnung des für die AMBER/3D-RISM Geometrieoptimierung benötigten Gradienten des chemischen Exzesspotentials bezüglich der Kernkoordinaten erfolgte in Zusammenarbeit mit Herrn Dr. Roland Frach.
- Die für die Durchmusterung der Konformationsräume des Histamins und des Pteroinsäure-Liganden verwendeten Skripte wurde von Patrick Kibies entwickelt. Die Ausgangsstrukturen für die verschiedenen Ionisierungszustände und Tautomere wurden von Herrn Dr. Stefan Güssregen zur Verfügung gestellt.
- Die optimierten Strukturen der Purin-Analoga stammen von Herrn Dr. Nico J. R. van Eikema Hommes. Die biochemische Argumentationslinie geht auf Herrn Prof. Nigel Richards, Herrn Dr. S. A. Benner, Herrn Prof. Tim Clark und ihre jeweiligen Mitarbeiter zurück. Der Anteil des Autors beschränkt sich auf die Durchführung der EC-RISM-Rechnungen und die Teilnahme an Diskussionen über das technische Vorgehen.

Teile der vorliegenden Arbeit sind oder werden in folgenden Publikationen veröffentlicht:

- J. Heil, S. M. Kast, „3D RISM theory with fast reciprocal-space electrostatics“, *J. Chem. Phys.* **2015**, *142*, 114107.
- N. J. R. van Eikema Hommes, J. Heil, N. B. Karalkar, S. A. Benner, S. M. Kast, T. Clark, N. G. J. Richards, „Tautomeric Equilibria of Purine Analogs used in Expanded Genetic Alphabets“, eingereicht.
- N. Tielker, D. Tomazic, J. Heil, T. Kloss, S. Ehrhart, S. Güssregen, K. F. Schmidt, S. M. Kast, „Solvation free energies, aqueous pK_a , and cyclohexane-water log D from embedded-cluster integral equation theory“, *J. Comput.-Aided Mol. Des.* **2016**. Doi:10.1007/s10822-016-9939-7.

Weitere Publikationen unter Beteiligung des Autors mit indirektem Bezug zu dieser Arbeit:

- T. Kloss, J. Heil, S. M. Kast, „Quantum chemistry in solution by combining 3D integral equation theory with a cluster embedding approach“, *J. Phys. Chem. B* **2008**, *112*, 4337-4343.
- S. M. Kast, J. Heil, S. Güssregen, K. F. Schmidt, „Prediction of tautomer ratios by embedded-cluster integral equation theory“, *J. Comput.-Aided Mol. Des.* **2010**, *24*, 343-353.
- F. Hoffgaard, J. Heil, S. M. Kast, „Three-Dimensional RISM Integral Equation Theory for Polarizable Solute Models“, *J. Chem. Theory Comput.* **2013**, *9*, 4718-4726.
- J. Heil, D. Tomazic, S. Egbers, S. M. Kast, „Acidity in DMSO from the embedded cluster integral equation quantum solvation model“, *J. Mol. Model* **2014**, *20*, 2161.
- J. Engel, A. Richters, M. Getlik, S. Tomassi, M. Keul, M. Termathe, J. Lategahn, C. Becker, S. Mayer-Wrangowski, C. Grütter, N. Uhlenbrock, J. Krüll, N. Schaumann, S. Eppmann, P. Kibies, F. Hoffgaard, J. Heil, S. Menninger, S. Ortiz-Cuaran, J. M. Heuckmann, V. Tinnefeld, R. P. Zahedi, M. L. Sos, C. Schultz-Fademrecht, R. K. Thomas, S. M. Kast, D. Rauh, „Targeting Drug Resistance in EGFR with Covalent Inhibitors: A Structure-Based Design Approach“, *J. Med. Chem.* **2015**, *58*, 6844-6863.
- S. M. Kast, J. Heil, F. Hoffgaard in „*Computational Trends in Solvation and Transport in Liquids*“ (Hrsg.: G. Sutmann, J. Grotendorst, G. Gompper, D. Marx), Schriften des Forschungszentrums Jülich, IAS Series Vol. 28, Jülich, **2015**.
- R. Frach, J. Heil, S. M. Kast, „Structure and Thermodynamics of Nondipolar Molecular Liquids and Solutions from Integral Equation Theory“, *Mol. Phys.* **2016**, *114*, 2461-2476.
- S. Güssregen, H. Matter, G. Hessler, E. Lionta, J. Heil, S. M. Kast, „Thermodynamic Characterization of Hydration Sites from Integral Equation-Derived Free Energy Densities: Application to Protein Binding Sites and Ligand Series“, eingereicht.

Digitale Unterstützungsmaterialien

In dieser Arbeit wurde eine Vielzahl von Molekülstrukturen aus Geometrieoptimierungen erhalten. Aufgrund dieser Vielzahl erscheint es nicht sinnvoll, die Lage der Atome direkt in dieser Arbeit in Schriftform anzugeben. Daher sind die entsprechenden Daten im weit verbreiteten „XYZ“-Format beigefügt und können mit den üblichen Anzeigeprogrammen wie z.B. Avogadro^[1] betrachtet werden. Konformationen des Histamins sind dem elektronischen Anhang der Masterarbeit von Lukas Eberlein^[2]

zu entnehmen. Strukturen des RTA-PTA-Komplexes sind als AMBER Koordinatendateien hinterlegt, zusätzlich sind AMBER „Parameter-Topologie“-Dateien mit Kraftfeldparametern sowie die Koordinaten des QM-Subsystems im XYZ-Format beigefügt.

1 Einleitung

Das Ziel der vorliegenden Arbeit ist die Entwicklung von Berechnungsmethoden, um den Protonierungszustand eines Liganden in der Bindetasche eines Proteins auf der Basis der Embedded-Cluster Reference Interaction Site Model- (EC-RISM-) Methode^[3] zu ermitteln. Der Protonierungszustand beeinflusst in entscheidendem Maße die Bildung von Wasserstoffbrücken und damit auch die Stärke der Wechselwirkung zwischen Ligand und Protein. Die Fragestellung ist daher von erheblichem Interesse für alle Bereiche, bei denen ein möglichst gut bindender Ligand gezielt auf der Basis rationaler Überlegungen konstruiert werden soll, also insbesondere für molekulare Erkennungsprozesse z.B. im Bereich der medizinischen Chemie oder der synthetischen Biologie. Die EC-RISM-Methode auf dem Stand der Technik vor dem Beginn dieser Arbeit eignet sich noch nicht zur Untersuchung von Protein-Ligand-Komplexen. Zur Betrachtung solcher Systeme war eine Reihe von weiteren Überlegungen und Erweiterungen entweder nötig oder zumindest hilfreich. Auf diese soll im Folgenden eingegangen werden.

Die EC-RISM-Methode koppelt die auf der klassischen Dichtefunktionaltheorie^[4-8] aufbauende *3D Reference Interaction Site Model-* (3D-RISM-) Integralgleichungstheorie^[9-11] mit quantenchemischen Methoden, z.B. um Gibbs-Energien zu berechnen. Dies ermöglicht im Gegensatz zu den traditionell eingesetzten dielektrischen Kontinuumsmodellen auf der Basis der Poisson- bzw. für Elektrolytlösungen der Poisson-Boltzmann-Gleichung^[12-14] gleichzeitig die Berechnung der lokalen Dichte des Solvens rund um das Solvat und erlaubt die Einbeziehung von Granularitätseffekten wie z.B. der Ausbildung von Packungen oder Wasserstoffbrückennetzwerken im Solvens. Direkte EC-RISM-Berechnungen an großen Systemen mit mehreren tausend Atomen wie die meisten Protein-Ligand-Komplexe sie besitzen sind derzeit prohibitiv aufwendig. Hauptgrund dafür sind die quantenchemischen Berechnungen, die ein ungünstiges Skalierungsverhalten bezüglich der Systemgröße zeigen. Einen Ausweg bieten Quantenmechanik/Molekülmechanik- (QM/MM-) Verfahren, die nur einen kleinen Teil des Solvats quantenmechanisch und den Rest mit klassischen Kraftfeldmethoden behandeln. Ein solches Verfahren unter Einbeziehung von EC-RISM wurde hier entwickelt. Aufwendig sind aber auch die 3D-RISM-Berechnungen, die durch Anwendung der *Particle-Mesh-Ewald-* (PME-) Methode zur schnellen Potentialberechnung, Verlagerung von Berechnungsschritten in den reziproken Raum und Parallelisierungsmaßnahmen deutlich beschleunigt werden konnten. Schließlich bildet noch die Kopplung von 3D-RISM an die Quantenchemie, die mittels eines auf einem Gitter diskretisierten Ladungsfeldes durchgeführt wird, einen „Flaschenhals“ bezüglich der Programmlaufzeit. Für dieses

Ladungsfeld wurde eine Methode entwickelt, um das Gitter adaptiv zu gestalten, was das EC-RISM-Verfahren ebenfalls erheblich beschleunigt.

Die Berechnung des Protonierungszustands in der Proteinbindetasche stellt ein pH-abhängiges Konformeren-/Tautomeren-/Ionisierungsproblem dar. Im Rahmen dieser Arbeit wurde hierfür ein statistisch-mechanischer Ansatz entwickelt und getestet, der bei gegebenen Bedingungen die Stoffmengenanteile aller Konformere, Tautomere und Ionisierungszustände liefert. Damit können Speziesverteilungsdiagramme erstellt werden, die ein intuitives Verständnis des Protonierungsverhaltens des Solvats ermöglichen.

Die Methodik wurde dann schrittweise anhand zunehmend komplexer Testprobleme erprobt. Konkret wurden Arbeiten aus der Diplomarbeit des Autors fortgeführt und pK_a -Modelle für unterschiedliche Klassen kleiner Moleküle erstellt, die pH-abhängige Speziesverteilung von Histamin ermittelt und in einem Komplex des toxischen Proteins Rizin die pH-abhängige Speziesverteilung eines Pteroinsäure-Liganden berechnet. Außerdem konnten die hier sowie im Rahmen des SAMPL2-Wettbewerbs gewonnenen Erfahrungen^[15] genutzt werden, um die Gruppen von Nigel G. J. Richards, Stephen A. Benner und Timothy Clark bei dem Ziel der Etablierung von Berechnungsmethoden für Tautomerprobleme im Bereich der Synthetischen Biologie zu unterstützen. Hier wurden auf der Basis zur Verfügung gestellter Strukturen von Tautomer von 5 Derivaten der Nukleobase Guanin EC-RISM-Berechnungen durchgeführt, um die Tautomererhältnisse zu berechnen und durch Vergleich mit experimentellen Beobachtungen die Methodik zu validieren.

Diese Arbeit ist wie folgt organisiert: In Kapitel 2 werden die mathematischen und physikalischen Grundlagen der hier verwendeten Methoden 3D-RISM und EC-RISM vorgestellt. In Kapitel 3 folgen die im Rahmen des vorliegenden Werks erarbeiteten technischen Verbesserungen und numerischen Verfahren zur Beschleunigung der Berechnungen, methodischen Innovationen sowie die Testprobleme und die Bearbeitung der finalen Fragestellung. In Kapitel 4 wird ein Resümee gezogen und Anknüpfungspunkte für weitere Forschungen werden diskutiert.

2 Grundlagen

2.1 Integralgleichungstheorie molekularer Fluide

Die im Rahmen dieser Arbeit verwendeten statistisch-mechanischen Modelle klassischer Fluide besitzen eine erhebliche Komplexität und eine vollständige Herleitung würde den Rahmen der vorliegenden Arbeit sprengen. Daher kann an dieser Stelle nur eine kurze Herleitungsskizze bereitgestellt werden und für weitergehende Recherchen muss auf Fachartikel^[10,11,16–25] und Monografien^[26,27] verwiesen werden.

2.1.1 Klassische Dichtefunktionaltheorie und das OZ-/Closure-Gleichungssystem

Ausgehend von der quantenmechanischen Dichtefunktional-Beschreibung des Elektronengases nach Hohenberg und Kohn^[28] konnte Mermin durch eine Erweiterung auf endliche Temperaturen ein Variationsprinzip für das großkanonische Potential Ω als Funktional der Elektronendichte herleiten.^[29] Überträgt man dieses Prinzip auf klassische Fluide zeigt sich, dass die lokale Teilchendichte $\rho(\mathbf{r})$ im Gleichgewicht das großkanonische Potential minimiert.^[30] Da Minima durch Nullstellen der ersten Ableitung charakterisiert sind, gilt dementsprechend, dass die Funktionalableitung des großkanonischen Potentials nach der lokalen Teilchendichte verschwinden muss, d.h.

$$\left. \frac{\delta\Omega}{\delta\rho(\mathbf{r})} \right|_{\rho=\rho_{\text{eq}}} = 0. \quad 2.1.1$$

Das großkanonische Potential ist

$$\Omega = F - \mu N = F[\rho(\mathbf{r})] - \mu \int d\mathbf{r} \rho(\mathbf{r}), \quad 2.1.2$$

mit der Helmholtz-Energie F , dem chemischen Potential μ und der Teilchenzahl N , die bei Integration der Teilchendichte über den gesamten Raum erhalten wird. Im allgemeinen Fall befindet sich das Fluid in einem Potential $U(\mathbf{r})$. Die Helmholtz-Energie besteht dann aus einem intrinsischen Anteil und eben diesem externen Potential. Damit wird die vorhergehende Gleichung zu

$$\Omega = F_{\text{int}}[\rho(\mathbf{r})] + \int d\mathbf{r} \rho(\mathbf{r})U(\mathbf{r}) - \mu \int d\mathbf{r} \rho(\mathbf{r}). \quad 2.1.3$$

Für homogene Fluide, d.h. solche Fluide, die sich nicht innerhalb externer Felder befinden, kann $U(\mathbf{r})$ das Potentialfelds eines einzelnen fixierten Fluidteilchens sein. Dies ist Teil einer Argumentationslinie, die in der Literatur unter dem Namen „Percus Trick“ bekannt geworden ist.^[21] Die intrinsische Freie Helmholtz-Energie wird zweckmäßigerweise weiter nach

$$F_{\text{int}}[\rho(\mathbf{r})] = F_{\text{int}}^{\text{id}}[\rho(\mathbf{r})] + F_{\text{int}}^{\text{ex}}[\rho(\mathbf{r})] \quad 2.1.4$$

in einen Ideal- und einen Exzessanteil aufgespalten. Damit wird aus 2.1.1-2.1.4 mit dem intrinsischen chemischen Potential $\psi(\mathbf{r}) = \mu - U(\mathbf{r})$ die Gleichgewichtsbedingung

$$\left. \frac{\delta F_{\text{int}}^{\text{id}}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} \right|_{\rho=\rho_{\text{eq}}} + \left. \frac{\delta F_{\text{int}}^{\text{ex}}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} \right|_{\rho=\rho_{\text{eq}}} - \psi(\mathbf{r}) = 0. \quad 2.1.5$$

Der Idealteil des Helmholtz-Energie-Funktional kann einfach aus der Zustandssumme des idealen Gases nach

$$F_{\text{int}}^{\text{id}} = -\beta^{-1} \ln(Q_{\text{id}}) = -\beta^{-1} \ln\left(\frac{1}{N!} \frac{V^N}{\Lambda^{3N}}\right) \quad 2.1.6$$

berechnet werden. Mit Hilfe der Stirling-Näherung, $\rho = N/V$ und der Einsicht, dass Gl. 2.1.6 bei Feldern mit variierender Dichte jeweils individuell für infinitesimal kleine Volumenelemente gilt, ergibt sich dann nach etwas Algebra

$$F_{\text{int}}^{\text{id}}[\rho(\mathbf{r})] = \beta^{-1} \int d\mathbf{r} [\rho(\mathbf{r}) \ln(\rho(\mathbf{r}) \Lambda^3) - 1], \quad 2.1.7$$

und bei funktionalem Ableiten nach der Dichte $\rho(\mathbf{r})$

$$\frac{\delta F_{\text{int}}^{\text{id}}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} = \beta^{-1} [\ln(\rho(\mathbf{r}) \Lambda^3)]. \quad 2.1.8$$

Der Exzessanteil erweist sich als deutlich komplexer. Entwickelt man $F_{\text{int}}^{\text{ex}}[\rho(\mathbf{r})]$ in eine funktionale Taylorreihe bis zur zweiten Ordnung in der Dichte des homogenen Referenzsystems, so erhält man den Ausdruck

$$\begin{aligned} F_{\text{int}}^{\text{ex}}[\rho(\mathbf{r})] &\approx F_{\text{int}}^{\text{ex}}[\rho_0(\mathbf{r})] + \int \left. \frac{\delta F_{\text{int}}^{\text{ex}}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} \right|_{\rho=\rho_0} (\rho(\mathbf{r}) - \rho_0(\mathbf{r})) d\mathbf{r} \\ &+ \frac{1}{2!} \iint \left. \frac{\delta^2 F_{\text{int}}^{\text{ex}}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r}) \delta \rho(\mathbf{r}')} \right|_{\rho=\rho_0} (\rho(\mathbf{r}) - \rho_0(\mathbf{r})) (\rho(\mathbf{r}') - \rho_0(\mathbf{r}')) d\mathbf{r} d\mathbf{r}' \\ &+ O(\rho(\mathbf{r})^3). \end{aligned} \quad 2.1.9$$

Mit $\Delta\rho(\mathbf{r}) = \rho(\mathbf{r}) - \rho_0(\mathbf{r})$ und den Definitionen der 1- und 2-Teilchen direkten Korrelationsfunktion $c^{(1)}(\mathbf{r})$ und $c^{(2)}(\mathbf{r}, \mathbf{r}')$,

$$\frac{\delta F_{\text{int}}^{\text{ex}}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} = -\beta^{-1} c^{(1)}(\mathbf{r}) = \mu^{\text{ex}} \quad 2.1.10$$

und

$$\frac{\delta^2 F_{\text{int}}^{\text{ex}}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r}) \delta \rho(\mathbf{r}')} = -\beta^{-1} c^{(2)}(\mathbf{r}, \mathbf{r}') \quad 2.1.11$$

folgt bei Abbruch nach dem 2. Glied

$$\begin{aligned} F_{\text{int}}^{\text{ex}}[\rho(\mathbf{r})] &\approx F_{\text{int}}^{\text{ex}}[\rho_0(\mathbf{r})] - \beta^{-1} \int c_0^{(1)}(\mathbf{r}) \Delta \rho(\mathbf{r}) d\mathbf{r} \\ &\quad - \frac{\beta^{-1}}{2} \iint c_0^{(2)}(\mathbf{r}, \mathbf{r}') \Delta \rho(\mathbf{r}) \Delta \rho(\mathbf{r}') d\mathbf{r} d\mathbf{r}' \\ &= F_{\text{int}}^{\text{ex}}[\rho_0(\mathbf{r})] - \mu_0^{\text{ex}} \int \Delta \rho(\mathbf{r}) d\mathbf{r} \\ &\quad - \frac{\beta^{-1}}{2} \iint c_0^{(2)}(\mathbf{r}, \mathbf{r}') \Delta \rho(\mathbf{r}) \Delta \rho(\mathbf{r}') d\mathbf{r} d\mathbf{r}'. \end{aligned} \quad 2.1.12$$

Durch Kombination von Gl. 2.1.8, 2.1.12 und der Gleichgewichtsbedingung 2.1.5 und Umstellen nach $\rho(\mathbf{r})$ kann gezeigt werden, dass die Dichte

$$\rho(\mathbf{r}) = \rho_0(\mathbf{r}) \exp(-\beta U(\mathbf{r}) + \int d\mathbf{r}' [\Delta \rho(\mathbf{r}') c_0^{(2)}(\mathbf{r}, \mathbf{r}')]) \quad 2.1.13$$

die Gleichgewichtsdichte, d.h. diejenige, die mit einem Minimum der Helmholtz-Energie einhergeht, approximiert. Wir nehmen im Sinne des „Percus Tricks“ an, dass ein beliebig herausgegriffenes Partikel eines uniformen atomaren Fluides der makroskopischen Dichte ρ , in dessen Inertialsystem wir uns begeben, die Quelle des externen Potentials ist, und dass dieses Potential ein Paarpotential zwischen dem herausgegriffenen Partikel und einem Zweiten ist. Unter diesen Bedingungen werden die Funktionen in Gl. 2.1.13 radialsymmetrisch bezüglich des Ortes des herausgegriffenen Partikels. Weiterhin wird ausgenutzt, dass im Rahmen des „Percus Tricks“ nicht mehr zwischen der Dichte des eigentlichen Systems und der des Referenzsystems unterschieden werden kann. Somit folgt aus Gl. 2.1.13 nach einigen weiteren Schritten die *Hypernetted-Chain-* (HNC-) Integralgleichung

$$g(\mathbf{r}) = g(r) = \exp(-\beta u(r) + \rho \int dr' [h(r') c(r - r')]). \quad 2.1.14$$

Wird die Reihenentwicklung in Gl. 2.1.9 nicht nach dem Term 2. Ordnung abgebrochen, so folgt in diesem Fall die allgemeine Form

$$g(r) = \exp(-\beta u(r) + \rho \int dr' [h(r')c(r-r')] + B(r)). \quad 2.1.15$$

Die Brückenfunktion $B(r)$ ist nicht in geschlossener Form bekannt und die direkte Auswertung der Reihenentwicklung wird aufgrund der Mehrfachintegrale für Glieder höherer Ordnung schnell prohibitiv aufwendig. $g(r)$ ist die Paarverteilungsfunktion, die im Falle radialsymmetrischer Partikel oder Zentren auch als radiale Verteilungsfunktion bezeichnet wird, und $h(r)$ ist die totale Korrelationsfunktion. Beide unterscheiden sich nur in der Normierung: $g(r)$ läuft im großkanonischen Ensemble asymptotisch gegen 1, $h(r)$ gegen 0. Allgemein gilt für $g(r)$, $h(r)$ und die makroskopische Dichte des Fluides ρ' der Zusammenhang

$$g(r) = h(r) + 1 = \rho(r) / \rho'. \quad 2.1.16$$

Die HNC-Integralgleichung begründet für sich alleine genommen noch keine lösbare Theorie. Sowohl die direkte als auch die totale Korrelationsfunktion sind unbekannt, und bei zwei Unbekannten werden auch zwei Gleichungen benötigt, die diese in Bezug zueinander setzen. Die Ornstein-Zernike-(OZ-) Gleichung

$$h(r) = c(r) + \rho \int dr'' c(|r'-r''|)h(|r-r''|) = c(r) + \rho(c * h)(r). \quad 2.1.17$$

wurde von Ornstein und Zernike im Zusammenhang mit Beugungsexperimenten an Flüssigkeiten experimentell gefunden, kann aber auch durch den DFT-Formalismus erhalten werden. Dazu wird ausgenutzt, dass die 2. funktionale Ableitung des großkanonischen Potentials nach dem intrinsischen chemischen Potential $\psi(\mathbf{r}) = \mu - U(\mathbf{r})$ nach

$$\frac{\delta^2 \Omega[\rho(\mathbf{r})]}{\delta \psi(\mathbf{r}) \delta \psi(\mathbf{r}')} = -\beta (h(\mathbf{r}, \mathbf{r}') \rho(\mathbf{r}) \rho(\mathbf{r}') + \rho(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r}')) = -\beta H(\mathbf{r}, \mathbf{r}'), \quad 2.1.18$$

mit der Dichte-Dichte-Korrelationsfunktion $H(\mathbf{r}, \mathbf{r}') = h(\mathbf{r}, \mathbf{r}') \rho(\mathbf{r}) \rho(\mathbf{r}') + \rho(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r}')$ identifiziert werden kann. Mit der funktionalanalytischen Darstellung der Deltafunktion und der Definition der „Funktionalinversen“ lässt sich dann nach einer Reihe weiterer komplexer algebraischer und funktionalanalytischer Transformationen die OZ-Gleichung herleiten. Durch Einsetzen in 2.1.14 und 2.1.15 folgen die Abschlussbeziehungen in den in der Literatur gängigen Formen

$$g(r) = \exp(-\beta u(r) + h(r) - c(r)) = \exp(-t^R(r)), \quad 2.1.19$$

bzw.

$$g(r) = \exp(-\beta u(r) + h(r) - c(r) + B(r)) = \exp(-t^R(r) + B(r)), \quad 2.1.20$$

jeweils für HNC und den allgemeinen Fall. $t(r) = h(r) - c(r)$ ist die indirekte Korrelationsfunktion, $t^R(r) = t(r) - \beta u(r)$ die renormierte indirekte Korrelationsfunktion. Diese und einige andere Abschlussbeziehungen werden in der englischsprachigen Fachliteratur und im Folgenden als „*Closure*“ bezeichnet. Eine weitere *Closure*, die in dieser Arbeit verwendet wurde, basiert auf einer partiellen Taylorreihenentwicklung der HNC-*Closure* für $t^R > 0$ nach

$$g(r) = \begin{cases} \sum_{i=0}^n (t^R(r))^i / i! & \Leftrightarrow t^R(r) > 0 \\ \exp(-t^R(r)) & \Leftrightarrow t^R(r) \leq 0 \end{cases}, \quad 2.1.21$$

und wird als *Partial Series Expansion n-ter Ordnung* (PSE- n) bezeichnet.^[31] Für $n \rightarrow \infty$ wird die PSE-*Closure* zur HNC-*Closure*. PSE-1 ist auch als Kovalenko-Hirata- (KH-) *Closure* bekannt.^[32] Je niedriger n , desto besser ist typischerweise das Konvergenzverhalten bei numerischen Lösungen.

2.1.2 Erweiterung auf Solvat-Solvens-Systeme

Die OZ-Gleichung lässt sich auf Lösungen mit einem bei unendlicher Verdünnung vorliegenden Solvat (Index „u“) in einem Solvens (Index „v“) erweitern. In diesem Fall wird aus der einzelnen OZ-Gleichung das Gleichungssystem

$$\begin{pmatrix} h_{vv} & h_{uv} \\ h_{uv} & h_{uu} \end{pmatrix} - \begin{pmatrix} c_{vv} & c_{uv} \\ c_{uv} & c_{uu} \end{pmatrix} = \begin{pmatrix} c_{vv} & c_{uv} \\ c_{uv} & c_{uu} \end{pmatrix} * \begin{bmatrix} \rho_v & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{pmatrix} h_{vv} & h_{uv} \\ h_{uv} & h_{uu} \end{pmatrix}. \quad 2.1.22$$

Dieses zerfällt in drei Gleichungen für die totalen Korrelationsfunktionen für den Solvens-Solvens- („vv-“), den Solvat-Solvens- („uv-“) und den Solvat-Solvat- („uu-“) Fall:

$$h_{vv} = c_{vv} + (c_{vv} * \rho_v h_{vv}), \quad 2.1.23$$

$$h_{uv} = c_{uv} + (c_{uv} * \rho_v h_{vv}), \quad 2.1.24$$

$$h_{uu} = c_{uu} + (c_{uv} * \rho_v h_{uv}). \quad 2.1.25$$

In der vorliegenden Arbeit ist primär der uv-Fall von Interesse. In 2.1.24 taucht die Solvens-Solvens direkte Korrelationsfunktion h_{vv} auf, die durch Lösen von Gl. 2.1.23 für das jeweilige Lösungsmittel einmalig vorberechnet werden kann. Analog müsste für uu-Berechnungen nach Gl. 2.1.25 vorher eine uv-Berechnung nach Gl. 2.1.24 durchgeführt werden. Die Gleichungen gelten sowohl für Lösungen radialsymmetrischer als auch anisotroper Teilchen, also z.B. molekulare Fluide.

Im letztgenannten Fall sind dann die OZ-artigen Gleichungen durch ihre im Folgenden eingeführten MOZ-, 3D-RISM-, 1D-RISM- oder die dielektrisch konsistenten DRISM-Äquivalente zu ersetzen.

2.1.3 Erweiterungen für anisotrope Fluide

Die OZ-Gleichung lässt sich auch auf anisotrope Systeme verallgemeinern. Dann besitzen alle Korrelationsfunktionen und das Potential zusätzlich zur Abhängigkeit vom Abstand noch eine Orientierungsabhängigkeit. Die resultierende Theorie ist in der Literatur als molekulare Ornstein-Zernike- (MOZ-) Theorie bekannt.^[20] In der Praxis gestaltet sich die Lösung des MOZ/*Closure*-Gleichungssystems schwierig und benötigt substantielle Rechenzeitressourcen.^[33] Eine praxisgerechte Alternative stellen orientierungsgemittelte Theorien dar, die in der Literatur unter dem Begriff *Reference Interaction Site Model* (RISM) geführt werden.^[16,19] Bei der 1D-RISM-Näherung wird das Fluid als Gemisch einzelner Wechselwirkungszentren betrachtet, wobei als Wechselwirkungszentren typischerweise die Atomkernpositionen gewählt werden. Daher wird diese Methodik auch als *Site-Site*-Ornstein-Zernike (SSOZ) bezeichnet. Für diese Zentren wird die direkte Korrelationsfunktion als Summe zentrumsspezifischer direkter Korrelationsfunktionen nach

$$c \approx \sum_{\alpha} \sum_{\gamma} c_{\alpha\gamma}(|\mathbf{R}_{\alpha} - \mathbf{R}_{\gamma}|) \quad 2.1.26$$

angesetzt, woraus eine Mittelung über Orientierungsfreiheitsgrade resultiert.^[16]

Die 1D-RISM-Analoga zu 2.1.23-2.1.25 sind dann die Matrixgleichungen

$$\boldsymbol{\rho}_v \mathbf{h}_{vv} = \boldsymbol{\omega}_v * \mathbf{c}_{vv} * \boldsymbol{\chi}_{vv}, \quad 2.1.27$$

$$\boldsymbol{\rho}_v \mathbf{h}_{uv} = \boldsymbol{\omega}_u * \mathbf{c}_{uv} * \boldsymbol{\chi}_{vv}, \quad 2.1.28$$

$$\mathbf{h}_{uu} = \boldsymbol{\omega}_{u(1)} * \mathbf{c}_{uu} * \boldsymbol{\omega}_{u(2)} + \boldsymbol{\omega}_{u(1)} * \mathbf{c}_{u(1)v} * \boldsymbol{\rho}_v \mathbf{h}_{u(2)v}. \quad 2.1.29$$

Hierbei ist

$$\boldsymbol{\chi}_{vv} = \boldsymbol{\rho}_v \boldsymbol{\omega}_v + \boldsymbol{\rho}_v^2 \mathbf{h}_{vv} \quad 2.1.30$$

die Solvesssuszeptibilität. Die Indizes der Matrizen beziehen sich jeweils auf die Wechselwirkungszentren, also z.B. $\mathbf{h} = (h_{\alpha\gamma})$, $\boldsymbol{\rho}_v$ sind Diagonalmatrizen mit den Teilchendichten der einzelnen Zentren. Mit Hilfe der intramolekularen Korrelationsfunktionen $\boldsymbol{\omega} = (\omega_{\alpha\gamma})$ mit $\omega_{\alpha\gamma} = \delta(|r_{\alpha\gamma} - l_{\alpha\gamma}|)/(4\pi l_{\alpha\gamma}^2)$ mit den Abständen zwischen zwei Wechselwirkungszentren im gleichen Molekül $l_{\alpha\gamma}$ wird die dreidimensionale Struktur der beteiligten Moleküle berücksichtigt.

Die Orientierungsmittelung hat Konsequenzen für die *closure*-Beziehungen, die prinzipiell angepasst werden müssten („*diagrammatically proper/improper*“).^[34] Die dadurch hervorgerufene fehlerbehaftete Orientierung und damit auch Orientierungspolarisation des Lösungsmittels ist der Grund dafür, dass Dielektrizitätskonstanten aus 1D-RISM-Rechnungen stark von experimentellen Befunden abweichen.^[35] Empirische Korrekturen führen zu den beiden Methoden ARISM^[36] und DRISM.^[22,23] Bei ARISM werden elektrostatische Wechselwirkungen mit einem empirischen Skalierungsfaktor versehen, um die gewünschte Dielektrizitätskonstante zu erhalten, was jedoch bei endlichen Elektrolytlösungen zu Problemen führt. Bei DRISM wird die gewünschte Dielektrizitätskonstante durch eine Modifikation der Integralgleichung in Form einer Brückenkorrektur erzwungen. Die in dieser Arbeit verwendete Solvenssuszeptibilitätsfunktion wurde mit Hilfe der DRISM-Korrektur erstellt.

2.1.4 Das 3D Reference Interaction Site Model (3D-RISM)

Die 3D-RISM uv-Gleichungen^[9,10]

$$\rho_\gamma h_\gamma(\mathbf{r}) = \sum_{\gamma'} \int d\mathbf{r}' c_{\gamma'}(\mathbf{r}-\mathbf{r}')^* \chi_{\gamma\gamma'}(|\mathbf{r}'|), \quad 2.1.31$$

im Folgenden kurz als 3D-RISM-Gleichungen bezeichnet, entstehen dadurch, dass ausgehend von der MOZ-Theorie die direkte Korrelationsfunktion als Summe direkter Korrelationsfunktionen der Lösungsmittelzentren nach

$$c \approx \sum_{\gamma} c_{\gamma}(\mathbf{r}_{\gamma}) \quad 2.1.32$$

angesetzt wird. Dabei bleibt im Gegensatz zu 1D-RISM die Anisotropie des Solvats erhalten, die des Lösungsmittels jedoch nicht. Das führt dazu, dass die direkten, indirekten und totalen Korrelationsfunktionen und die Potentialfunktionen im Gegensatz zu 1D-RISM nicht mehr Matrix-, sondern nur noch Vektorform haben, wobei der Index ein Wechselwirkungszentrum des Solvens bezeichnet. Das bedeutet auch, dass es sich jetzt nicht mehr um radialsymmetrische Funktionen handelt. Stattdessen sind diese Funktionen hier skalare Felder im dreidimensionalen Raum. Für die *Closures* gilt Entsprechendes. Damit werden aus Gl. 2.1.19 für HNC

$$g_\gamma(\mathbf{r}) = \exp(-\beta u_\gamma(\mathbf{r}) + h_\gamma(\mathbf{r}) - c_\gamma(\mathbf{r})) = \exp(-t_\gamma^R(\mathbf{r})) \quad 2.1.33$$

und 2.1.21 für PSE-*n*

$$g_\gamma(\mathbf{r}) = \begin{cases} \sum_{i=0}^n (t_\gamma^R(\mathbf{r}))^i / i & \Leftrightarrow t_\gamma^R(\mathbf{r}) > 0 \\ \exp(-t_\gamma^R(\mathbf{r})) & \Leftrightarrow t_\gamma^R(\mathbf{r}) \leq 0 \end{cases} \quad 2.1.34$$

Die in Gleichung 2.1.31 enthaltene Solvenssuszeptibilitätsfunktion $\chi_{\gamma\gamma'}$ wird typischerweise mittels der 1D-RISM Gleichung 2.1.27 oder, im Fall von Elektrolytlösungen, ihrem DRISM-Analogen berechnet und ist radialsymmetrisch. Alternativ kann Sie auch aus einer Moleküldynamik- oder Monte-Carlo-Simulation extrahiert werden.^[37]

Das in den *closures* auftauchende Potential $u_\gamma(\mathbf{r})$ wird als Paarpotential zwischen den Wechselwirkungszentren des Solvats an den Orten \mathbf{R}_α und jeweils einem Solvenszentrum γ am Ort \mathbf{r} aufgefasst. Typischer- aber nicht notwendigerweise werden dabei apolare Wechselwirkungen durch ein Lennard-Jones- (LJ-) 12-6-Potential und polare Wechselwirkungen durch ein Coulomb-Potential beschrieben. Damit gilt

$$u_\gamma(\mathbf{r}) = 4 \sum_{\alpha=1}^{N_\alpha} \varepsilon_{\alpha\gamma} \left[\left(\frac{\sigma_{\alpha\gamma}}{|\mathbf{r} - \mathbf{R}_\alpha|} \right)^{12} - \left(\frac{\sigma_{\alpha\gamma}}{|\mathbf{r} - \mathbf{R}_\alpha|} \right)^6 \right] + q_\gamma \sum_{\alpha=1}^{N_\alpha} \frac{q_\alpha}{|\mathbf{r} - \mathbf{R}_\alpha|}. \quad 2.1.35$$

In dieser Arbeit werden die LJ-Parameter $\sigma_{\alpha\gamma}$ durch Lorentz-Berthelot-Mischungsregeln, d.h. durch Bildung des arithmetischen Mittels $\sigma_{\alpha\gamma} = (1/2)(\sigma_\alpha + \sigma_\gamma)$ von typspezifischen Parametern σ_α und σ_γ erhalten. Das geometrische Mittel $\sigma_{\alpha\gamma} = \sqrt{\sigma_\alpha \sigma_\gamma}$, das z.B. bei Berechnungen mit dem OPLS-Kraftfeld^[38] benutzt wird, kann jedoch ebenfalls verwendet werden. Die LJ-Parameter $\varepsilon_{\alpha\gamma}$ werden immer durch geometrische Mittelung erhalten.

2.1.4.1 Chemische Exzesspotentiale

Ein Vorteile der OZ-basierten Integralgleichungstheorien mit HNC-artigen *Closures* ist die Verfügbarkeit analytischer Ausdrücke für das chemische Exzesspotential μ^{ex} auf Basis der Korrelationsfunktionen. Das chemische Exzesspotential ergibt sich formal aus einem Kopplungsparameterintegral mit dem Kopplungsparameter λ im Sinne der thermodynamischen Integration im 3D-RISM-Fall als^[17,18,39,40]

$$\mu^{\text{ex}} = \sum_\gamma \rho_\gamma \int_0^1 d\lambda \int d\mathbf{r} [u(\mathbf{r}, \lambda) g(\mathbf{r}, \lambda)]. \quad 2.1.36$$

Bei direkter Auswertung der vorhergehenden Gleichung müsste die Integralgleichung für verschiedene Werte des Kopplungsparameters gelöst, und dann eine numerische Integration durchgeführt werden. Singer und Chandler konnten jedoch zeigen, dass für die HNC-*Closure* die Kopplungsparameter-Integration analytisch ausgeführt werden kann.^[18] Man erhält dann den Ausdruck

$$\mu_{\text{HNC}}^{\text{ex}} = \beta^{-1} \sum_{\gamma} \rho_{\gamma} \int d\mathbf{r} [h_{\gamma}^2(\mathbf{r}) - c_{\gamma}(\mathbf{r}) - \frac{1}{2} h_{\gamma}(\mathbf{r}) c_{\gamma}(\mathbf{r})]. \quad 2.1.37$$

Für *Closures* mit Brückenfunktion kommt noch ein zusätzlicher Term hinzu. Damit erhält man z.B. für Brückenfunktionen vom Typ $B_{\gamma} = B_{\gamma}(t_{\gamma}^R)$

$$\mu^{\text{ex}} = \mu_{\text{HNC}}^{\text{ex}} + \mu_B^{\text{ex}} = \mu_{\text{HNC}}^{\text{ex}} + \beta^{-1} \sum_{\gamma} \rho_{\gamma} \int d\mathbf{r} \left[h(\mathbf{r}) - \int_0^{t^R} dt^R (h(t^R, \mathbf{r}) + 1) \right]. \quad 2.1.38$$

Fasst man die Differenz zwischen HNC- und PSE-*n-Closure* als Brückenkorrektur auf, so kann man auf diesem Weg den Ausdruck

$$\mu_{\text{PSE-}n}^{\text{ex}} = \mu_{\text{HNC}}^{\text{ex}} - \beta^{-1} \sum_{\gamma} \rho_{\gamma} \int d\mathbf{r} [\Theta(h(\mathbf{r})) (t^R(\mathbf{r}))^{n+1} / (n+1)!] \quad 2.1.39$$

für das chemische Exzesspotential der PSE-*Closures* herleiten. ^[31]

Dieser einfache Zugang zur Energetik von Solvatationsprozessen steht im Gegensatz zu Moleküldynamiksimulationen, wo aufwendige Einzelrechnungen für verschiedene Werte eines Kopplungsparameters durchgeführt werden müssen. ^[39,41]

Für Geometrieoptimierungen in Lösung werden die Gradienten des chemischen Exzesspotentials bezüglich der Koordinaten von Solvatzentren benötigt. Der Gradient lässt sich unabhängig von der verwendeten *Closure* als

$$\nabla_{\mathbf{R}_{\alpha}} \mu^{\text{ex}}(\{\mathbf{R}\}) = \sum_{\gamma} \rho_{\gamma} \int d\mathbf{r} [g_{\gamma}(\mathbf{r}; \{\mathbf{R}\}) \nabla_{\mathbf{R}_{\alpha}} u_{\gamma}(\mathbf{r}; \{\mathbf{R}\})] \quad 2.1.40$$

formulieren. ^[42] Hierbei bezeichnet $\{\mathbf{R}\}$ den Satz von Koordinaten von Solvatzentren \mathbf{R}_{α} mit

$$\mathbf{R}_{\alpha} = (R_{\alpha,x}, R_{\alpha,y}, R_{\alpha,z})^T. \quad 2.1.41$$

2.1.4.2 Numerische Lösungsverfahren

Das einfachste Verfahren zur numerischen Lösung des 3D-RISM/*Closure*-Gleichungssystems ist die Fixpunktiteration nach Picard. ^[43] Für unpolare Systeme werden dabei einfach *Closure*-Beziehung und Integralgleichung wechselseitig ineinander eingesetzt, und ausgehend von einem geeigneten Startwert (typischerweise $c_{\gamma,0}(\mathbf{r}) = 0$) mit der durch

$$t_{\gamma,i}(\mathbf{r}) = \sum_{\gamma'} \mathfrak{F}^{-1} [\mathfrak{F}[c_{\gamma',i}(\mathbf{r})] \hat{\chi}'_{\gamma\gamma'}(|\mathbf{k}|)] - c_{\gamma,i}(\mathbf{r}) \quad 2.1.42$$

und

$$c_{\gamma,i+1}(\mathbf{r}) = \exp[u_\gamma(\mathbf{r}) + t_{\gamma,i}(\mathbf{r}) + B_\gamma(\mathbf{r})] - 1 - t_{\gamma,i}(\mathbf{r}) \quad 2.1.43$$

gegebenen Iterationsvorschrift bis zur Selbstkonsistenz iteriert. Hierbei werden die Faltungsintegrale in Gl. 2.1.31 mittels der Fourier-Transformation ausgewertet. Dabei wird ausgenutzt, dass eine Faltung zweier Funktionen im Realraum einem Produkt im Reziprokraum, d.h. dem Bildraum der Fourier-Transformation, entspricht. Für zwei Funktionen a und b gilt dann

$$a * b = \mathfrak{F}^{-1} [\mathfrak{F}[a]\mathfrak{F}[b]]. \quad 2.1.44$$

Hierbei ist zu beachten, dass die Verwendung der Fourier-Transformation automatisch eine Periodisierung der Funktionen impliziert. In der Praxis wird die Fourier-Transformation durch eine diskrete Fourier-Transformation approximiert, die mittels des *Fast-Fourier-Transform*- (FFT-) Verfahrens^[44] mit einer algorithmischen Komplexität von $O(n_x n_y n_z \log(n_x n_y n_z))$ durchgeführt wird. Für die Fourier-Transformation wird eine Reihe unterschiedlicher Definitionen verwendet. Im Rahmen dieser Arbeit werden die Fourier-Transformation und ihre Inverse als

$$\hat{f}(\mathbf{k}) = \mathfrak{F}[f(\mathbf{r})] = \int d\mathbf{r} f(\mathbf{r}) \exp(-i\mathbf{k} \cdot \mathbf{r}) \quad 2.1.45$$

und

$$f(\mathbf{r}) = \mathfrak{F}^{-1}[\hat{f}(\mathbf{k})] = \frac{1}{(2\pi)^3} \int d\mathbf{k} \hat{f}(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{r}) \quad 2.1.46$$

definiert.

Für polare Systeme muss das Verfahren modifiziert werden, da das Coulomb-Potential nur sehr langsam abfällt und bei der diskreten Fourier-Transformation aufgrund einer endlichen Simulationszelle Abschneideartefakte auftreten. Dies kann vermieden werden, indem direkte und indirekte „renormierte“ Korrelationsfunktionen

$$c_\gamma^S(\mathbf{r}) = c_\gamma(\mathbf{r}) + \beta u_\gamma^L(\mathbf{r}) \quad 2.1.47$$

und

$$t_\gamma^S(\mathbf{r}) = t_\gamma(\mathbf{r}) - \beta u_\gamma^L(\mathbf{r}) \quad 2.1.48$$

eingeführt werden, die eine kurzreichweitige Asymptotik aufweisen.^[45] Damit ändert sich die Iterationsvorschrift in

$$t_{\gamma,i}^S(\mathbf{r}) = \sum_{\gamma'} \mathfrak{F}^{-1} [\{\mathfrak{F}[c_{\gamma',i}^S(\mathbf{r})] - \beta \hat{u}_{\gamma'}^L(\mathbf{k})\} \hat{\chi}_{\gamma\gamma'}(|\mathbf{k}|)] - c_{\gamma,i}^S(\mathbf{r}) \quad 2.1.49$$

und

$$c_{\gamma,i+1}^S(\mathbf{r}) = \exp[u_\gamma^S(\mathbf{r}) + t_{\gamma,i}^S(\mathbf{r}) + B_\gamma(\mathbf{r})] - 1 - t_{\gamma,i}^S(\mathbf{r}). \quad 2.1.50$$

Zu diesem Zweck muss das Potential ebenfalls in einen langreichweitigen Teil mit Coulomb-Asymptotik und einen schnell abfallenden, kurzreichweitigen Teil aufgespalten werden. Typischerweise nutzt man dazu, analog zu Ewald-Summationstechniken für die Moleküldynamik^[46], die Gauss'sche Fehlerfunktion

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x dy \exp(-y^2) \quad 2.1.51$$

und die komplementäre Fehlerfunktion

$$\operatorname{erfc}(x) = 1 - \operatorname{erf}(x). \quad 2.1.52$$

Damit ergeben sich aus Gl. 2.1.35

$$u_\gamma^S(\mathbf{r}) = 4 \sum_{\alpha=1}^{N_\alpha} \varepsilon_{\alpha\gamma} \left[\left(\frac{\sigma_{\alpha\gamma}}{|\mathbf{r} - \mathbf{R}_\alpha|} \right)^{12} - \left(\frac{\sigma_{\alpha\gamma}}{|\mathbf{r} - \mathbf{R}_\alpha|} \right)^6 \right] + q_\gamma \sum_{\alpha=1}^{N_\alpha} \frac{q_\alpha \operatorname{erfc}(\kappa |\mathbf{r} - \mathbf{R}_\alpha|)}{|\mathbf{r} - \mathbf{R}_\alpha|} \quad 2.1.53$$

und

$$u_\gamma^L(\mathbf{r}) = q_\gamma \sum_{\alpha=1}^{N_\alpha} \frac{q_\alpha \operatorname{erf}(\kappa |\mathbf{r} - \mathbf{R}_\alpha|)}{|\mathbf{r} - \mathbf{R}_\alpha|}. \quad 2.1.54$$

Die Auswertung der Abstände erfolgt periodisch im Rahmen der „*Minimum-Image*“-Konvention. Der apolare Anteil des Potentials wird hier vollständig dem kurzreichweitigen Potential zugeschlagen. In der 3D-RISM-Gleichung wird direkt die Fourier-Transformierte von Gl. 2.1.54,

$$\hat{u}_\gamma^L(\mathbf{k}) = q_\gamma \frac{4\pi}{k^2} \exp\left(-\frac{k^2}{4\kappa^2}\right) \sum_{\alpha=1}^{N_\alpha} q_\alpha \exp(-i\mathbf{k} \cdot \mathbf{R}_\alpha), \quad 2.1.55$$

verwendet. Für den Beweis, dass die rechten Seiten von Gl. 2.1.54 und Gl. 2.1.55 ein Fourier-Transformationspaar bilden, wird auf die Unterstützungsmaterialien von Ref. ^[47] verwiesen.

Für Solvate, die nicht nur polar sind, sondern auch eine von Null verschiedene Gesamtladung tragen, reicht die bisher beschriebene Modifikation nicht aus, damit alle diskreten Fourier-Transformationen ohne große Abschneideartefakte durchgeführt werden können. Von der langreichweitigen Potentialfunktion $u_\gamma^L(\mathbf{r})$ muss ein Term abgezogen werden, der das asymptotische Verhalten des Potentials der Gesamtladung besitzt. Dieser Term wird dann separat behandelt. Zu diesem Zweck wurden bisher 3 unterschiedliche Verfahren publiziert. Die Methode von Kovalenko und

Hirata^[48], die im Folgenden als „Hintergrund-Renormierung“ bezeichnet wird, verwendet dazu das Potential eines homogenen entgegengesetzt geladenen uniformen Plasmas. Damit wird Gl. 2.1.49 zu

$$t_{\gamma,i}^S(\mathbf{r}) = \sum_{\gamma'} \mathfrak{F}^{-1} [\{ \mathfrak{F}[c_{\gamma',i}^S(\mathbf{r})] - \beta \hat{u}_{\gamma'}^L(\mathbf{k}) \} \hat{\chi}_{\gamma\gamma'}'(|\mathbf{k}|)] - c_{\gamma,i}^S(\mathbf{r}) - \frac{4\pi\beta}{V} \left(\sum_{\alpha=1}^{N_a} q_\alpha \right) \lim_{k \rightarrow 0} \sum_{\gamma'} q_{\gamma'} \frac{\hat{\chi}_{\gamma\gamma'}(k)}{k^2}. \quad 2.1.56$$

Alternativ kann die „Monopol-Renormierung“ von Kast und Kloss^[49] verwendet werden, bei der eine einzelne Punktladung mit der entgegengesetzten Gesamtladung des Systems an einem beliebigen Ort im Simulationsvolumen platziert wird. In diesem Fall wird aus Gl. 2.1.49

$$t_{\gamma,i}^S(\mathbf{r}) = \sum_{\gamma'} \mathfrak{F}^{-1} [\{ \mathfrak{F}[c_{\gamma',i}^S(\mathbf{r})] - \beta \hat{u}_{\gamma'}^L(\mathbf{k}) \} \hat{\chi}_{\gamma\gamma'}'(|\mathbf{k}|)] - c_{\gamma,i}^S(\mathbf{r}) - \beta \sum_{\gamma'} \mathfrak{F}_{\text{ID}}^{-1} [\hat{u}_{\gamma'}^{L(0)}(k) \hat{\chi}_{\gamma\gamma'}(k)], \quad 2.1.57$$

wobei $\mathfrak{F}_{\text{ID}}^{-1}$ eine inverse Fourier-Bessel-Transformation bezeichnet, die in dieser Arbeit numerisch mit einem Verfahren von Talman^[50] auf einem eindimensionalen logarithmischen Gitter durchgeführt wird. Hier ist zu beachten, dass im Faltungsprodukt anstelle von $\hat{u}_{\gamma'}^L(\mathbf{k})$ das um den Monopolterm bereinigte Potential

$$\hat{u}_{\gamma'}^L(\mathbf{k}) = \hat{u}_{\gamma'}^L(\mathbf{k}) - \hat{u}_{\gamma'}^{L(0)}(k), \quad 2.1.58$$

verwendet werden muss. Im Folgenden wird nicht zwischen $\hat{u}_{\gamma'}^L(\mathbf{k})$ und $\hat{u}_{\gamma'}^L(\mathbf{k})$ unterschieden, die Anwendung von Gl. 2.1.58 ist impliziert. Der Vorteil der Monopol-Renormierung liegt darin begründet, dass hiermit auch für Elektrolytlösungen korrekt renormierte Paarverteilungsfunktionen berechnet werden können. Dies ist mit der Hintergrund-Korrektur nicht direkt möglich. Kürzlich ist jedoch auch eine diesbezügliche Erweiterung erschienen.^[51] Die dritte Methode ist eine Erweiterung der Monopol-Renormierung von Pettitt und Mitarbeitern^[52], die jedes Solvat-Wechselwirkungszentrum einzeln renormiert, dabei aber einen erheblichen Rechenaufwand verursacht und deshalb für schnelle Berechnungen ungeeignet erscheint. Formal ermöglicht diese Methode das schnellstmögliche Abfallen der kurzreichweitigen Potentialfunktionen. Eigene vorläufige Tests ergaben aber keine praxisrelevanten Vorteile im Vergleich zu den anderen Methoden.

2.2 Das EC-RISM-Solvatationsmodell für quantenchemische Berechnungen

Das chemische Exzesspotential, das mittels Gl. 2.1.36-2.1.39 evaluiert werden kann, ist die Gibbs-Energie, die beim Lösen eines Moleküls eines Stoffes entsteht, wenn elektronische und Kernfreiheitsgrade des Moleküls beim Lösungsvorgang unverändert bleiben. Mit anderen Worten: Das chemische Exzesspotential beinhaltet lediglich die Änderung der Gibbs-Energie durch die Störung des Fluides, aber nicht die durch die Polarisierung des Solvats durch die Solvensumgebung. Das im Rahmen einer Diplomarbeit^[53] und weiterer vorhergehender Arbeiten von Stefan M. Kast, Thomas Kloss und dem Autor entwickelte EC-RISM Verfahren^[3] ist eine Möglichkeit, die 3D-RISM Methode so zu erweitern, dass die Änderung der elektronischen Freiheitsgrade (und im Prinzip auch der Kernfreiheitsgrade) ebenfalls berücksichtigt werden kann. Dazu wird die 3D-RISM-Methode in einem iterativen Verfahren mit Quantenchemie-Software gekoppelt. Ähnliche Verfahren werden auch von anderen Gruppen eingesetzt.^[54-61] Im Detail funktioniert das EC-RISM-Verfahren wie folgt:

1. Die Wellenfunktion des Solvats wird im Vakuum evaluiert.
2. Auf der Basis der Wellenfunktion werden wahlweise das elektrostatische Potential des Solvats auf einem Gitter oder an das elektrostatische Potential angepasste atomzentrierte Punktladungen berechnet.
3. Potential oder Punktladungen bilden zusammen mit Lennard-Jones-Parametern aus einer externen Quelle ein intermolekulares Kraftfeld. Mit diesem wird das 3D-RISM/*Closure*-Gleichungssystem gelöst. Es werden die 3D Paarverteilungsfunktion $g_\gamma(\mathbf{r})$ auf dem Gitter und das chemische Exzesspotential μ^{ex} nach Gl. 2.1.36-2.1.39 erhalten.
4. Aus den Paarverteilungsfunktionen und Partialladungen des Solvens wird eine gitterbasierte Ladungsdichteverteilung

$$\rho_q(\mathbf{r}) = \sum_\gamma q_\gamma \rho_\gamma g_\gamma(\mathbf{r}) \quad 2.2.1$$

berechnet, die die Elektrostatik des Lösungsmittels repräsentiert. In dieser Umgebung wird dann die Wellenfunktion des Solvats evaluiert und ein neues elektrostatisches Potential $\phi(\mathbf{r})$ bzw. neue Punktladungen q_α generiert. Außerdem wird die quantenmechanische Energie des Solvats E_{Sol} erhalten.

5. Wenn die Gibbs-Energie

$$G = E_{\text{Sol}} + \mu^{\text{ex}} \quad 2.2.2$$

sich nicht mehr ändert, so wird die Iteration beendet. Ansonsten wird das Verfahren bei Schritt 3 fortgesetzt.

Unter der Annahme, dass das chemische Exzesspotential bereits alle intra-Solvens-Wechselwirkungen beinhaltet, so ergibt sich die quantenmechanische Energie des aus den Kernen (n), und Elektronen (e) aufgebauten Solvats formal aus allen Anteilen die sich nur auf diese beziehen, sowie auf deren Wechselwirkung mit dem Solvens, d.h. als Erwartungswert

$$E_{\text{Sol}} = \langle \Psi_{\text{tot}} | \hat{H}_1 | \Psi_{\text{tot}} \rangle, \quad 2.2.3$$

wobei

$$\hat{H}_1 = \hat{H}_{\text{ne}} + \hat{H}_{\text{ee}} + \hat{H}_{\text{nn}} \quad 2.2.4$$

der übliche molekulare Hamiltonoperator des Solvats im Vakuum ist, und Ψ_{tot} eine effektive Wellenfunktion des Solvats in Lösung, die sich als Eigenfunktion der Schrödingergleichung

$$\hat{H}_{\text{tot}} \Psi_{\text{tot}} = E_{\text{tot}} \Psi_{\text{tot}} \quad 2.2.5$$

ergibt. Der Hamiltonoperator in Lösung \hat{H}_{tot} lässt sich als

$$\hat{H}_{\text{tot}} = \hat{H}_1 + \hat{H}_2 = \hat{H}_1 + \hat{H}_{\text{nq}} + \hat{H}_{\text{eq}} + \hat{H}_{\text{qq}} \quad 2.2.6$$

schreiben, wobei \hat{H}_{qq} , \hat{H}_{nq} und \hat{H}_{eq} die Selbstwechselwirkung der Solvens-Ladungsverteilung sowie deren Wechselwirkung mit Kernen und Elektronen des Solvats beschreiben. Damit lässt sich E_{Sol} als

$$E_{\text{Sol}} = E_{\text{tot}} - E_2 = \langle \Psi_{\text{tot}} | \hat{H}_{\text{tot}} | \Psi_{\text{tot}} \rangle - \langle \Psi_{\text{tot}} | \hat{H}_2 | \Psi_{\text{tot}} \rangle, \quad 2.2.7$$

evaluieren. Hier tritt das Problem auf, dass die Berechnung von \hat{H}_{qq} quadratisch mit der Anzahl der Gitterpunkte der Solvens-Ladungsverteilung skaliert, was bei den im EC-RISM-Verfahren auftretenden Gittergrößen insbesondere mit Gaussian 03^[62] prohibitive Rechenzeitkosten verursacht. Neuere Revisionen von Gaussian 09^[63] implementieren die linear skalierende schnelle Multipolmethode,^[64] die Rechenzeitkosten sind jedoch noch immer hoch. Eine vergleichsweise einfache Lösung dieses Problems ist, die Berechnung von \hat{H}_{qq} durch Modifikationen im Quellcode der Programme zu unterdrücken, wodurch der entsprechende Ausdruck in beiden Termen in Gl. 2.2.7 entfällt. Die Berechnung von E_2 reduziert sich dann auf

$$E_2 = \langle \Psi_{\text{tot}} | \hat{H}_{\text{nq}} + \hat{H}_{\text{eq}} | \Psi_{\text{tot}} \rangle = \int d\mathbf{r} \rho_q(\mathbf{r}) \varphi(\mathbf{r}). \quad 2.2.8$$

Das Integral in Gl. 2.2.8 wird durch Summation über die auf dem 3D Gitter diskretisierten Funktionen approximiert. Prinzipiell müsste das elektrostatische Potential des Solvats in jeder EC-RISM-Iteration aufwendig evaluiert werden. Bisherige Erfahrungen^[15] zeigen jedoch, dass das Konvergenzverhalten sich nicht wesentlich ändert, wenn der E_2 -Term während der Iteration ausgelassen und erst nach dem Erreichen der Konvergenz berechnet wird. In dieser Arbeit wird, sofern nicht explizit anderweitig angegeben, nach dieser Methodik verfahren. Weiterhin wird während der EC-RISM-Iteration lediglich das Hartree-Fock-Verfahren verwendet, da vor allem bei Elektronenkorrelationsverfahren wie MP2^[65] die Berechnung des elektrostatischen Potentials zur Generierung der atomzentrierten Punktladungen, aber auch bereits die Berechnung der Wellenfunktion sehr aufwendig werden können.

3 Ergebnisse

3.1 Beschleunigung der 3D-RISM-Methodologie

Zur Beschleunigung des 3D-RISM-Lösers wurden drei verschiedene Teilprojekte durchgeführt. Auf der algorithmischen Seite wurde zunächst versucht, die Berechnung der Reziprokraum-Solvat-Solvens-Potentialfelder $\hat{u}_\gamma^L(\mathbf{k})$ mittels eines *Particle-Mesh*-Ansatzes, wie er in ähnlicher Form für Moleküldynamiksimulationen verwendet wird, zu beschleunigen. Weiterhin wurde die Idee verfolgt, darauf aufbauend die Berechnung des chemischen Exzesspotentials 2.1.36-2.1.39 so umzuformulieren, dass entweder die so mit geringem Aufwand berechneten Reziprokraum-Potentialfelder oder schnell berechenbare radialsymmetrische Potentialfunktionen hier verwendet werden können. Auf technischer Seite wurde eine OpenMP-Parallelisierung^[66] durchgeführt, die auf modernen *Multicore*-Systemen gegenüber einer Berechnung auf der Basis von seriellem Code oder mit nur einem einzelnen Prozessorkern eine deutliche Beschleunigung herbeiführt.

3.1.1 Adaption des Particle-Mesh-Ewald Ansatzes für 3D-RISM

Die Verwendung des langreichweitigen Reziprokraum-Potentials impliziert, dass für jeden Reziprokraum-Gittervektor \mathbf{k} für jedes Solvenszentrum γ Gl. 2.1.55 separat zu evaluieren ist, was erheblichen Rechenaufwand verursacht. Wir fragen daher nach Wegen, diesen Rechenaufwand zu verringern. Dazu ist es zunächst sinnvoll, Gl. 2.1.55 als

$$\hat{u}_\gamma^L(\mathbf{k}) = q_\gamma \hat{\phi}(k = |\mathbf{k}|) \hat{s}(\mathbf{k}) = q_\gamma \hat{a}(k) \hat{b}(k) \hat{s}(\mathbf{k}), \quad 3.1.1$$

zu schreiben. Hierbei ist

$$\hat{\phi}(k) = \hat{a}(k) \hat{b}(k), \quad 3.1.2$$

eine in k radialsymmetrische Funktion, die in der englischsprachigen Fachliteratur als „*influence function*“ bezeichnet wird. Diese setzt sich zusammen aus der Green'schen Funktion des Coulomb-Potentials

$$\hat{a}(k) = \frac{4\pi}{k^2} \quad 3.1.3$$

und einer „Verbreiterungsfunktion“

$$\hat{b}(k) = \exp\left(-\frac{k^2}{4\kappa^2}\right). \quad 3.1.4$$

Der elektrostatische Strukturfaktor

$$\hat{s}(\mathbf{k}) = \sum_{\alpha=1}^{N_\alpha} q_\alpha \exp(-i\mathbf{k} \cdot \mathbf{R}_\alpha) \quad 3.1.5$$

berücksichtigt die Verschiebung der Solvatzentren bezüglich des Ursprungs und beinhaltet eine Summe über alle Solvatzentren. Außerdem ist er nicht radialsymmetrisch und muss daher in vollem Umfang auf einem 3D Gitter evaluiert werden, wobei eine Komplexität von $O(N_\alpha n_x n_y n_z)$ entsteht. Bei genauere Betrachtung fällt auf, dass Gl. 3.1.5 der Definition der diskreten Fouriertransformation entsprechen würde, wenn die Solvatzentren auf den Knoten eines regulären Gitters lägen, wie es z.B. im Festkörperbereich für kubisch- oder orthorhombisch-primitive Kristalle der Fall ist. Hier könnte der Strukturfaktor mittels der 3D FFT mit einer Komplexität von $O(n_x n_y n_z \log(n_x n_y n_z))$ berechnet werden. Um die Komplexitätsreduktion durch die FFT hier auszunutzen muss dementsprechend eine gitterbasierte Ladungsdichte erzeugt werden, deren Potential das der ursprünglichen irregulären Ladungsverteilung möglichst gut approximiert. Diese Problemstellung tritt in ähnlicher Form bei N -Körper-Simulationen in unterschiedlichen Bereichen auf. Beispiele sind astrophysikalische Galaxiedynamik-Simulationen^[67] oder die Moleküldynamik.^[68] Entsprechende Techniken sind unter Namen und Akronymen wie *Particle-Mesh Ewald* (PME)^[69], *smooth Particle-Mesh Ewald* (SPME),^[70] *Particle-Particle Particle-Mesh* (P3M)^[70], *Particle-Particle Nonuniform FFT* (P2NFFT)^[71] oder *Ewald Summation based on Non-Uniform FFT* (ENUF)^[72] bekannt geworden. In neuerer Zeit hat sich gezeigt, dass viele dieser häufig völlig unabhängig entwickelten Verfahren eine sehr ähnliche mathematische Struktur aufweisen und sich teilweise gar nicht oder nur in Details unterscheiden.^[73,74] Die im Weiteren beschriebene Technik basiert auf dem ursprünglichen PME-Verfahren von Darden, York und Pedersen.^[69] Dieses ist vergleichsweise einfach zu implementieren, lässt die *influence function* unverändert, was z.B. bei SPME und P3M nicht der Fall ist und liefert, wie im Folgenden gezeigt wird, sehr genaue Ergebnisse.

Wir schreiben zunächst Gl. 3.1.5 als

$$\hat{s}(\mathbf{k}) = \sum_{\alpha=1}^{N_\alpha} q_\alpha \exp(-ik_x R_{\alpha,x}) \exp(-ik_y R_{\alpha,y}) \exp(-ik_z R_{\alpha,z}) \quad 3.1.6$$

und approximieren die komplexen Exponentialfunktionen nach

$$\exp(-ikx) \approx \sum_{m=0}^p L_m^p(x) \exp(-ikx_m) \quad 3.1.7$$

als Linearkombination von Lagrange-Basispolynomen der Ordnung p ,

$$L_m^p(x) = \prod_{j=0, j \neq m}^p \frac{x - x_j}{x_m - x_j}, \quad 3.1.8$$

wobei die Indizes $m = [0, p]$ Gitterpunkte um ein Solvatzentrum α referenzieren. Prinzipiell gilt

$$x_m = \begin{cases} x' + d_x \left(m - \frac{p}{2} \right) & \Leftrightarrow p \text{ gerade} \\ x' + d_x \left(m - \frac{p-1}{2} \right) & \Leftrightarrow p \text{ ungerade} \end{cases}, \quad 3.1.9$$

wobei x' für gerade p der zu $R_{\alpha,x}$ nächstliegende Gitterpunkt, für ungerade p der von $R_{\alpha,x}$ aus gesehen nächste Gitterpunkt auf der Seite mit den kleineren Indizes ist. Damit ergibt sich für den Strukturfaktor

$$\begin{aligned} \hat{s}(\mathbf{k}, \{\mathbf{R}_\alpha\}) &\approx \sum_{\alpha=1}^{N_\alpha} q_\alpha \sum_{l,m,n} L_l^p(R_{\alpha,x}) L_m^p(R_{\alpha,y}) L_n^p(R_{\alpha,z}) \exp(-i\mathbf{k} \cdot \mathbf{r}_{l,m,n}) \\ &= \sum_{l,m,n} \psi_{l,m,n}(\{\mathbf{R}_\alpha\}) \exp(-i\mathbf{k} \cdot \mathbf{r}_{l,m,n}) \end{aligned}, \quad 3.1.10$$

wobei

$$\psi_{l,m,n}(\{\mathbf{R}_\alpha\}) = \sum_{\alpha=1}^{N_\alpha} q_\alpha L_l^p(R_{\alpha,x}) L_m^p(R_{\alpha,y}) L_n^p(R_{\alpha,z}) \quad 3.1.11$$

die gitterbasierte Ladungsdichte der Solvatzentren ist. Der konkrete Algorithmus zur Berechnung von $\psi_{l,m,n}$ wird in Abbildung 1 noch einmal verdeutlicht. Da die Lagrange-Basispolynome einen „kompakten Träger“ besitzen, also nur in einem kleinen Bereich von Null verschiedene Werte annehmen, lässt sich $\psi_{l,m,n}$ in $O(N_\alpha)$ berechnen. Die gesamte Berechnung des Strukturfaktors besitzt somit eine Komplexität von $O(N_\alpha + n_x n_y n_z \log(n_x n_y n_z))$. Durch Einsetzen von Gl. 3.1.10 folgt für das langreichweitige Reziprokraumpotential mit den um $\mathbf{k} = (0,0,0)^T$ zentrierten Reziprokraum-Gittervektoren

$$\mathbf{k}_{o,p,q} = 2\pi \left(\frac{(o - n_x/2)}{n_x d_x}, \frac{(p - n_y/2)}{n_y d_y}, \frac{(q - n_z/2)}{n_z d_z} \right)^T \quad 3.1.12$$

mit den ganzen Zahlen $o, p, q = [0, n_x - 1]$

$$\begin{aligned} \hat{u}_\gamma^L(\mathbf{k}_{o,p,q}) &\approx q_\gamma \hat{\phi}(|\mathbf{k}_{o,p,q}|) \sum_{l,m,n} \psi_{l,m,n}(\{\mathbf{R}_\alpha\}) \exp(-i\mathbf{k}_{o,p,q} \cdot \mathbf{r}_{l,m,n}) \\ &= q_\gamma \hat{\phi}(|\mathbf{k}_{o,p,q}|) \sum_{l=0}^{n_x-1} \sum_{m=0}^{n_y-1} \sum_{n=0}^{n_z-1} \psi_{l,m,n}(\{\mathbf{R}_\alpha\}) \exp \left[-2\pi i \left(\frac{ol}{n_x} + \frac{pm}{n_y} + \frac{qn}{n_z} \right) \right] \\ &= q_\gamma \hat{\phi}(|\mathbf{k}_{o,p,q}|) \text{FFT}[\psi(\mathbf{r}_{l,m,n})]. \end{aligned} \quad 3.1.13$$

Die Expansion der komplexen Exponentialfunktion in Lagrange-Basispolynome stellt praktisch eine invertierte Polynominterpolation nach Lagrange dar. Grundsätzlich tritt bei einer Polynominterpolation mit äquidistanten Stützstellen das Runge'sche Phänomen auf.^[75] Dieses äußert sich in einer mit zunehmender Ordnung des Interpolationspolynoms bzw. der Anzahl der Stützstellen abnehmenden Interpolationsgüte. Entsprechende Artefakte konnten jedoch bei den bisher durchgeführten Berechnungen bis zur Ordnung 10 praktisch nicht beobachtet werden.

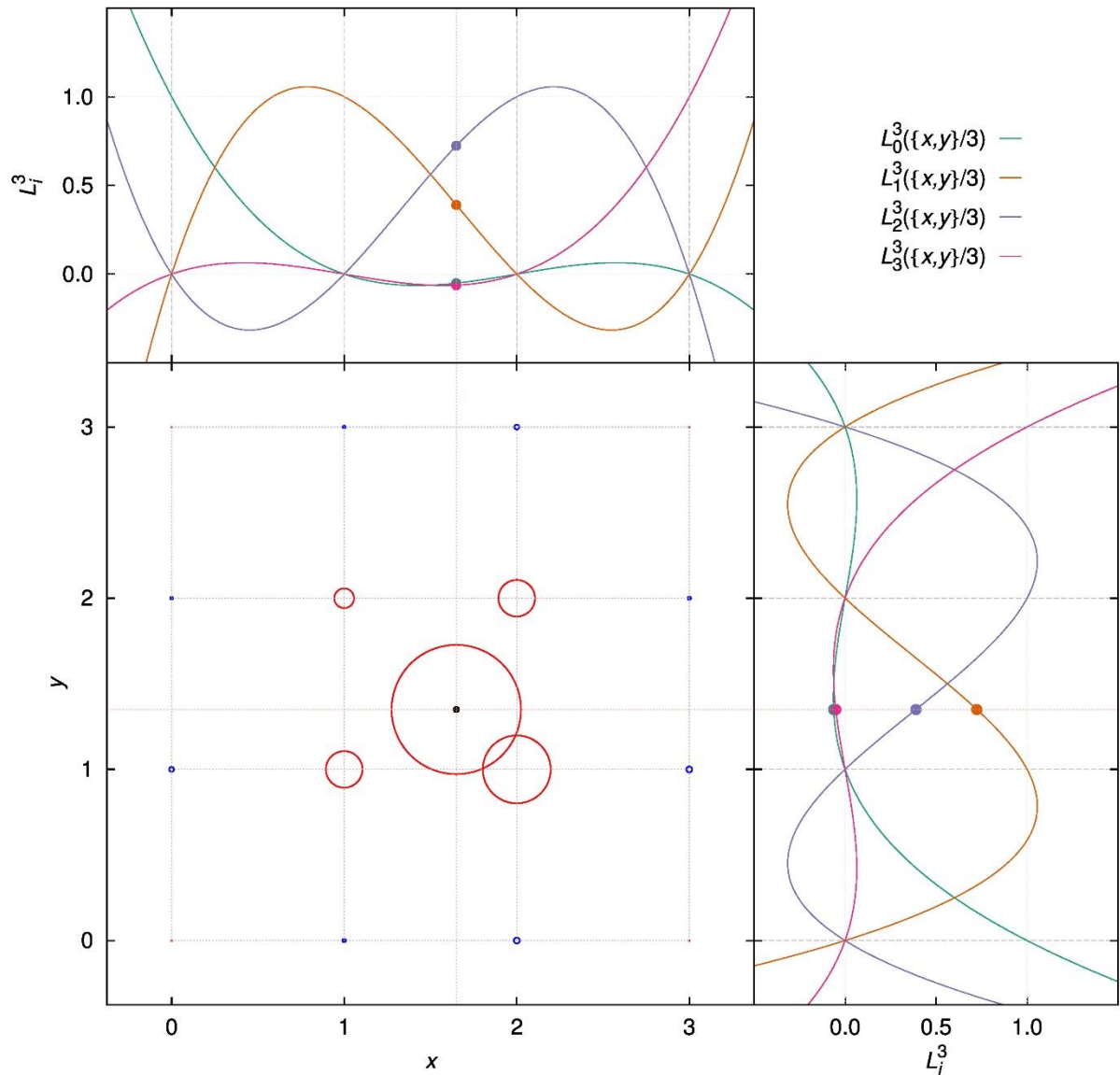


Abbildung 1: Schematische Darstellung der PME Ladungsverteilungsprozedur in zwei Dimensionen für Lagrange-Basispolynome 3. Ordnung. Der große quadratische Bereich zeigt ein Partikel am Ort $(x = 0.55, y = 0.45)$ mit der Ladung $q = 1$ und die nächsten 4 Gitterpunkte in x - und y -Richtung rund um den Partikelort, auf die die Ladung des Partikels verteilt wird. Die Größe der Kreise um die einzelnen Gitterpunkte und den Partikelort ist proportional zur Ladung am Gitterpunkt bzw. Partikelort. Die Farbe Rot zeigt eine positive Ladung an, die Farbe Blau negative Ladung. Jeweils für x und y wird ein auf das Intervall $[0,3]$ skaliertes Basispolynom 3. Ordnung am x - bzw. y -Ort des Partikels ausgewertet, wie im oberen und im rechten Teil des Schemas gezeigt. Die Ladung an einem Gitterpunkt (a, b) ist dann durch ein Produkt der Ladung des Partikels mit dem a -ten Lagrangepolynom am x -Ort und dem b -ten Lagrangepolynom am y -Ort gegeben. Die PME Ladungsverteilungsprozedur in 3 Dimensionen erfolgt analog. Alle verwendeten Orte sind entsprechend der Gitterauflösung zu skalieren.

3.1.2 Schnelle Berechnung des chemischen Exzesspotentials im reziproken Raum

Wir schreiben den HNC-Anteil des chemischen Exzesspotentials Gl. 2.1.37 als

$$\mu_{\text{HNC}}^{\text{ex}} = \mu^{\text{ex},(1)} + \mu^{\text{ex},(2)} \quad 3.1.14$$

mit

$$\mu^{\text{ex},(1)} = -\beta^{-1} \sum_{\gamma} \rho_{\gamma} \int d\mathbf{r} \left[\frac{1}{2} h_{\gamma}(\mathbf{r}) c_{\gamma}(\mathbf{r}) \right] \quad 3.1.15$$

und

$$\mu^{\text{ex},(2)} = \beta^{-1} \sum_{\gamma} \rho_{\gamma} \int d\mathbf{r} [h_{\gamma}^2(\mathbf{r}) - c_{\gamma}(\mathbf{r})] = \beta^{-1} \sum_{\gamma} \rho_{\gamma} \int d\mathbf{r} [h_{\gamma}^2(\mathbf{r}) - c_{\gamma}^{\text{S}}(\mathbf{r})], \quad 3.1.16$$

wobei die zweite Identität in Gl. 3.1.16 die Folge der Ladungsneutralität des Solvens ist. Es zeigt sich, dass $\mu^{\text{ex},(2)}$ schnell mit der Zellengröße abfällt, $\mu^{\text{ex},(1)}$ jedoch nicht. Der Grund dafür ist die nicht renormierte direkte Korrelationsfunktion $c_{\gamma}(\mathbf{r})$. Da nach der Iteration nur $c_{\gamma}^{\text{S}}(\mathbf{r})$ verfügbar ist, und eine numerische Fouriertransformation von $\hat{u}_{\gamma}^{\text{L}}(\mathbf{k})$ an der Coulomb-Asymptotik scheitert, wird $u_{\gamma}^{\text{L}}(\mathbf{r})$ im Realraum nach Gl. 2.1.54 auf dem 3D Gitter berechnet, sinnvollerweise gleichzeitig mit $u_{\gamma}^{\text{S}}(\mathbf{r})$ nach Gl. 2.1.53 um die doppelte Berechnung eines gemeinsamen Zwischenergebnisses, des Abstands zwischen Gitterpunkt und Solvatzentrum, zu vermeiden. Weiterhin kann Gl. 2.1.52 genutzt werden, um die Berechnung entweder der Fehlerfunktion oder der komplementären Fehlerfunktion einzusparen. Trotzdem ist die Berechnung von $u_{\gamma}^{\text{L}}(\mathbf{r})$ aufwendig und kann auch nicht durch Abschneiden des Coulomb-Potentials beschleunigt werden, ohne das erhebliche Abweichungen im chemischen Exzesspotential beobachtet werden. Es zeigt sich jedoch, dass sich $\mu^{\text{ex},(1)}$ in drei Terme zerlegen lässt, von denen zwei schnell abfallen und der dritte eine Summe effektiver solvensvermittelter Wechselwirkungen zwischen Solvatzentren darstellt. Diese Summe kann aufgrund der relativ niedrigen Anzahl der Solvatzentren schnell evaluiert werden. Dazu setzen wir die 3D-RISM Integralgleichung 2.1.31 für $h_{\gamma}(\mathbf{r})$ ein und verwenden das Parseval-Theorem

$$\int d\mathbf{r} [y(\mathbf{r}) z(\mathbf{r})] = \frac{1}{(2\pi)^3} \int d\mathbf{k} [\hat{y}(\mathbf{k}) \hat{z}^*(\mathbf{k})]. \quad 3.1.17$$

Hierbei kennzeichnet der Stern die komplexe Konjugation. Damit wird aus 3.1.15

$$\begin{aligned}
\mu^{\text{ex,(1)}} &= -\frac{1}{2}\beta^{-1}\sum_{\gamma}\frac{\rho_{\gamma}}{(2\pi)^3}\int d\mathbf{k}\left[\hat{c}_{\gamma}^*(\mathbf{k})\hat{h}_{\gamma}(\mathbf{k})\right] \\
&= -\frac{1}{2}\beta^{-1}\sum_{\gamma}\frac{1}{(2\pi)^3}\int d\mathbf{k}\left[\hat{c}_{\gamma}^*(\mathbf{k})\sum_{\gamma'}\hat{c}_{\gamma'}(\mathbf{k})\hat{\chi}_{\gamma\gamma'}(k)\right].
\end{aligned} \tag{3.1.18}$$

Aufspalten von \hat{c}_{γ} nach 2.1.47 und ausmultiplizieren liefert

$$\begin{aligned}
\mu^{\text{ex,(1)}} &= \mu^{\text{ex,(1a)}} + \mu^{\text{ex,(1b)}} + \mu^{\text{ex,(1c)}} \\
&= -\frac{1}{2}\beta^{-1}\sum_{\gamma}\frac{1}{(2\pi)^3}\int d\mathbf{k}\left[\hat{c}_{\gamma}^{\text{S}*}(\mathbf{k})\sum_{\gamma'}\hat{c}_{\gamma'}^{\text{S}}(\mathbf{k})\hat{\chi}_{\gamma\gamma'}(k)\right] \\
&\quad + \sum_{\gamma}\frac{1}{(2\pi)^3}\int d\mathbf{k}\left[\hat{c}_{\gamma}^{\text{S}*}(\mathbf{k})\sum_{\gamma'}\hat{u}_{\gamma'}^{\text{L}}(\mathbf{k})\hat{\chi}_{\gamma\gamma'}(k)\right] \\
&\quad - \frac{1}{2}\beta\sum_{\gamma}\frac{1}{(2\pi)^3}\int d\mathbf{k}\left[\hat{u}_{\gamma}^{\text{L}*}(\mathbf{k})\sum_{\gamma'}\hat{u}_{\gamma'}^{\text{L}}(\mathbf{k})\hat{\chi}_{\gamma\gamma'}(k)\right].
\end{aligned} \tag{3.1.19}$$

Die Asymptotik von $\mu^{\text{ex,(1a)}}$ und $\mu^{\text{ex,(1b)}}$ wird durch die kurzreichweitige direkte Korrelationsfunktion dominiert, was die direkte Summation im reziproken Raum und damit die Verwendung der in jedem Fall für die Iteration benötigten Funktion $\hat{u}_{\gamma}^{\text{L}}(\mathbf{k})$ an Stelle von $u_{\gamma}^{\text{L}}(\mathbf{r})$ erlaubt. Damit verbleibt der problematische dritte Term $\mu^{\text{ex,(1c)}}$. Wir setzen Gl. 2.1.55 für $\hat{u}_{\gamma}^{\text{L}}(\mathbf{k})$ ein und erhalten mit dem Fourier-Verschiebungstheorem

$$\int d\mathbf{r}[f(\mathbf{r}-\mathbf{r}_0)\exp(-i\mathbf{k}\cdot\mathbf{r})] = \exp(-i\mathbf{k}\cdot\mathbf{r}_0)\mathfrak{F}[f(\mathbf{r})] \tag{3.1.20}$$

den Ausdruck

$$\begin{aligned}
\mu^{\text{ex,(1c)}} &= -\frac{1}{2}\beta\sum_{\gamma,\gamma'}\frac{1}{(2\pi)^3}\int d\mathbf{k}\left[\sum_{\alpha=1}^{N_{\alpha}}\sum_{\alpha'=1}^{N_{\alpha}}q_{\alpha}q_{\alpha'}q_{\gamma}q_{\gamma'}\exp\{i\mathbf{k}\cdot(\mathbf{R}_{\alpha}-\mathbf{R}_{\alpha'})\}\hat{\phi}^2(k)\hat{\chi}_{\gamma\gamma'}(k)\right] \\
&= -\frac{1}{2}\beta\sum_{\alpha=1}^{N_{\alpha}}\sum_{\alpha'=1}^{N_{\alpha}}q_{\alpha}q_{\alpha'}\frac{1}{(2\pi)^3}\int d\mathbf{k}\left[\exp\{i\mathbf{k}\cdot(\mathbf{R}_{\alpha}-\mathbf{R}_{\alpha'})\}\hat{\phi}^2(k)\sum_{\gamma,\gamma'}q_{\gamma}q_{\gamma'}\hat{\chi}_{\gamma\gamma'}(k)\right] \\
&= -\frac{1}{2}\beta\sum_{\alpha=1}^{N_{\alpha}}\sum_{\alpha'=1}^{N_{\alpha}}q_{\alpha}q_{\alpha'}\theta(|\mathbf{R}_{\alpha}-\mathbf{R}_{\alpha'}|).
\end{aligned} \tag{3.1.21}$$

Die Funktion $\theta(r)$ kann mittels einer Fourier-Bessel-Transformation nach

$$\theta(r) = \mathfrak{F}[\hat{\phi}^2(k)\sum_{\gamma,\gamma'}q_{\gamma}q_{\gamma'}\hat{\chi}_{\gamma\gamma'}(k)] = \frac{1}{2\pi^2}\int_0^{\infty} dk[v(k,r)] \tag{3.1.22}$$

mit

$$v(k, r) = k^2 j_0(k, r) \hat{\phi}^2(k) \sum_{\gamma, \gamma'} q_\gamma q_{\gamma'} \hat{\chi}_{\gamma\gamma'}(k) \quad 3.1.23$$

berechnet werden, wobei $j_0(k, r) = \sin(kr)/(kr)$ die Bessel-Funktion nullter Ordnung des ersten Typs ist. Damit ist die Berechnung von $\mu^{\text{ex},(1c)}$ auf Operationen mit radialsymmetrischen Funktionen beschränkt, die im Gegensatz zu Operationen auf 3D Gittern problemlos in den schnellen Zwischenspeichern („Caches“) moderner Mikroprozessoren durchgeführt werden können. Es bleibt zu bemerken, dass die Integrale in Gl. 3.1.21 reell sind, wie man sieht, wenn man die Euler'sche Formel auf die komplexen Exponentialfunktionen anwendet und die einzelnen Terme der resultierenden trigonometrischen Funktionen separat integriert.

Für Integrale über Korrelationsfunktionen im Kontext von OZ- und RISM-Theorien haben sich logarithmische Gitter bewährt.^[76] Diese beginnen aber nicht bei $k = 0$, daher muss Gl. 3.1.22 in zwei Teile zerlegt werden, von 0 bis zum ersten Gitterpunkt k_0 und von k_0 bis unendlich:

$$\theta(r) = \theta_1(r) + \theta_2(r) = \frac{1}{2\pi^2} \int_0^{k_0} dk [v(k, r)] + \frac{1}{2\pi^2} \int_{k_0}^{\infty} dk [v(k, r)]. \quad 3.1.24$$

Im Folgenden wird $\theta_1(r)$ durch eine einfache Rechteckregel nach

$$\theta_1(r) \approx \frac{k_0 v(k_0, r)}{2\pi^2} \quad 3.1.25$$

approximiert. $\theta_2(r)$ scheint prinzipiell in einem Schritt mit Hilfe des logarithmischen FFT-Algorithmus von Talman^[50] evaluierbar, der im Kontext des in dieser Arbeit genutzten RISM-Lösers bereits Verwendung findet. Wie aus Abbildung 2 ersichtlich scheitert dies jedoch in der Praxis daran, dass $\theta(r)$ für kleine r starke artifizielle Oszillationen zeigt, was insbesondere für die Diagonalelemente $\alpha = \alpha'$ ein Problem darstellt. Dies kann vermieden werden, indem mit $v(k, r)$ eine *Spline*-Interpolation mit einem kubischen natürlichen B-*Spline* durchgeführt und das Integral dann für jeden r -Gitterpunkt analytisch evaluiert wird. Mit den so erhaltenen diskreten Funktionswerten im Realraum wird dann wieder eine *Spline*-Interpolation durchgeführt und für alle nicht-Diagonalelemente $\alpha \neq \alpha'$ eine sehr gute Approximation für $\theta_2(r)$ erhalten. Für die Diagonalelemente $\alpha = \alpha'$ werden die ersten r -Gitterpunkte von $\theta_2(r)$ berechnet und eine Polynomextrapolation auf $r = 0$ durchgeführt. In der vorliegenden Arbeit wurde hierzu ein Polynom 4. Ordnung mit dementsprechend 5 Stützstellen verwendet.

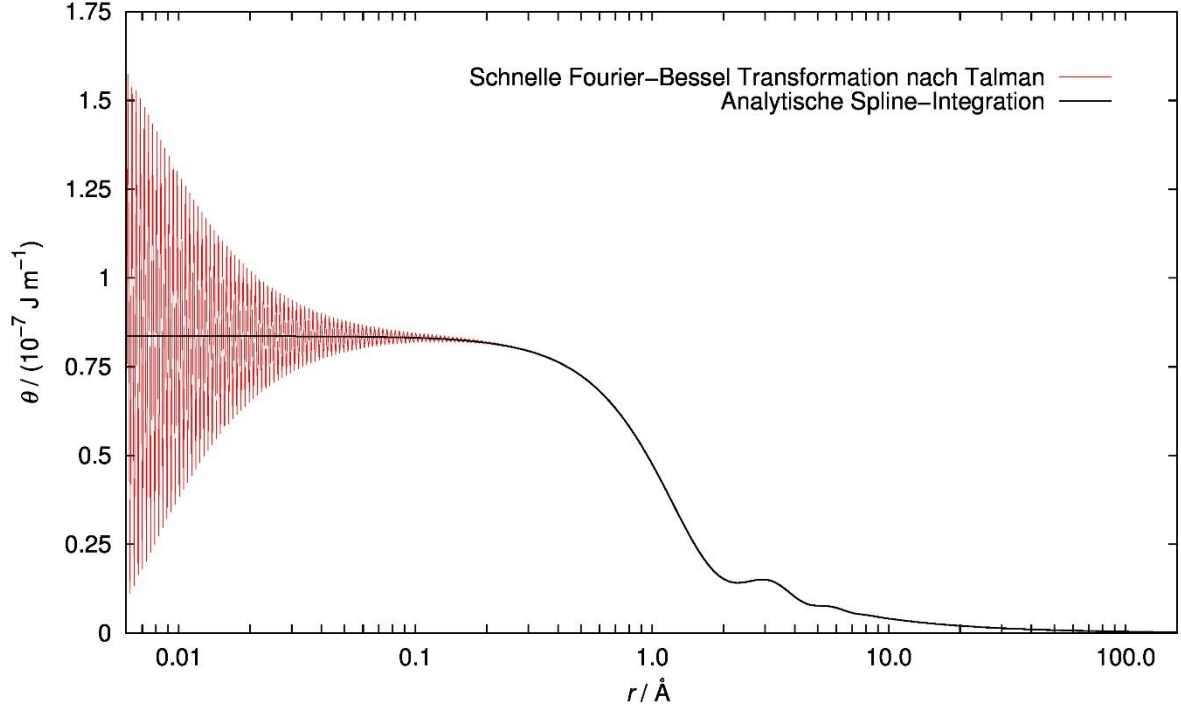


Abbildung 2: $\theta(r)$ nach Gl. 3.1.22 für modifiziertes SPC/E-Wasser als Funktion des Abstands r zwischen zwei Zentren, berechnet mittels Talmans schneller Fourier-Bessel-Transformation und mittels einer analytischen *Spline*-Integration.

Der hier entwickelte Algorithmus findet auch auf die Solvat-Solvens Innere Exzessenergie U^{ex} Anwendung. Diese lässt sich nach

$$U^{\text{ex}} = \sum_{\gamma} \rho_{\gamma} \int d\mathbf{r} [g_{\gamma}(\mathbf{r}) \{u_{\gamma}^{\text{S}}(\mathbf{r}) + u_{\gamma}^{\text{L}}(\mathbf{r})\}] = U^{\text{ex,S}} + U^{\text{ex,L}} \quad 3.1.26$$

ebenfalls in einen langreichweitigen und einen problemlos direkt evaluierbaren kurzreichweitigen Teil zerlegen. Der langreichweitige Teil lässt sich dann als Summe zweier bereits bekannter Terme schreiben:

$$\begin{aligned} U^{\text{ex,L}} &= \sum_{\gamma} \rho_{\gamma} \int d\mathbf{r} [(h_{\gamma}(\mathbf{r}) + 1)u_{\gamma}^{\text{L}}(\mathbf{r})] = \sum_{\gamma} \frac{\rho_{\gamma}}{(2\pi)^3} \int d\mathbf{k} [\hat{h}_{\gamma}^*(\mathbf{k}) + \hat{1}] \hat{u}_{\gamma}^{\text{L}}(\mathbf{k}) \\ &= \sum_{\gamma} \frac{1}{(2\pi)^3} \int d\mathbf{k} \left[\hat{u}_{\gamma}^{\text{L}*}(\mathbf{k}) \sum_{\gamma'} \{\hat{c}_{\gamma'}^{\text{S}}(\mathbf{k}) - \beta \hat{u}_{\gamma'}^{\text{L}}(\mathbf{k})\} \hat{\chi}_{\gamma\gamma'}(k) \right] \\ &= \frac{1}{(2\pi)^3} \left\{ \sum_{\gamma} \int d\mathbf{k} \left[\hat{c}_{\gamma}^{\text{S}*}(\mathbf{k}) \sum_{\gamma'} \hat{u}_{\gamma'}^{\text{L}}(\mathbf{k}) \hat{\chi}_{\gamma\gamma'}(k) \right] - \beta \sum_{\gamma} \int d\mathbf{k} \left[\hat{u}_{\gamma}^{\text{L}*}(\mathbf{k}) \sum_{\gamma'} \hat{u}_{\gamma'}^{\text{L}}(\mathbf{k}) \hat{\chi}_{\gamma\gamma'}(k) \right] \right\} \\ &= \mu^{\text{ex,(1b)}} + 2\mu^{\text{ex,(1c)}}. \end{aligned} \quad 3.1.27$$

3.1.3 OpenMP-Parallelisierung des 3D-RISM Löasers

In den letzten Jahren werden, vom Konsumenten-PC aus dem Elektronik Einzelhandel bis zu den Rechnern in den Bundeshöchstleistungsrechenzentren in Jülich, München und Stuttgart, zunehmend Prozessoren verwendet die mehrere Prozessorkerne auf einem Chip vereinigen. In noch stärkerem Maße wird dieses Prinzip durch die „Xeon Phi“-Beschleunigerkarten von Intel repräsentiert, deren Prozessoren (derzeit) 60-72 Kerne besitzen. Die Rechenkapazität mehrerer Kerne steht jedoch dem Benutzer nicht ohne Weiteres zur Verfügung, sondern der Programmierer muss i.d.R. das Zusammenspiel der Prozessorkerne explizit koordinieren. Dazu können verschiedene Hilfsmittel wie z.B. Cilk Plus^[77] und OpenMP^[66] verwendet werden, wobei sich im wissenschaftlichen Bereich insbesondere OpenMP durchgesetzt hat. OpenMP spezifiziert eine Reihe von Compilerdirektiven und Bibliotheksfunktionen, die eine vergleichsweise einfache Parallelisierung hauptsächlich auf der Ebene von Schleifen ermöglichen. Ein einfaches Beispiel ist hier in Listing 1 gezeigt.

```
program omp_example
integer :: i,n

!$OMP PARALLEL DO
do i=1,n
    call do_something_with(i)
end do
!$OMP END PARALLEL DO

end program omp_example
```

Trotz der relativen Simplität von OpenMP unterliegt der Programmierer einer Reihe von Einschränkungen, die die parallele Skalierbarkeit deutlich vermindern können. Dazu gehört insbesondere die „Speicheraffinität“ der Prozessorkerne. Darunter versteht man, dass unterschiedliche Prozessorkerne unterschiedlich schnell auf verschiedene Bereiche des Hauptspeichers zugreifen können. Ziel des Programmierers muss es demnach sein, dafür zu sorgen, dass jeder Prozessorkern möglichst nur auf Daten zugreift, die in seinem Affinitätsbereich liegen. Dies wird typischerweise dadurch ermöglicht, dass bereits das Initialisieren eines Speicherbereichs von demjenigen Kern durchgeführt wird, der später auch auf diesem Speicherbereich arbeiten soll, da der Speicher dann in

den Affinitätsbereich des jeweiligen Kerns verlegt wird („*first touch policy*“). Nichtbeachtung dieser Thematik ist ein häufiger Grund für Skalierungsprobleme.^[78]

Zu Beginn der vorliegenden Arbeit war der in der Arbeitsgruppe Kast hauptsächlich von Thomas Kloss entwickelte 3D-RISM-Löser in Teilen bereits mit OpenMP-Anweisungen versehen, parallele Skalierbarkeit war jedoch nicht gegeben und gegenüber der seriellen Ausführung konnte praktisch kein Geschwindigkeitsgewinn erzielt werden. Als Hauptproblem stellte sich die Nichtbeachtung der Speicheraffinität heraus. Nach der Behebung dieses Problems konnte für das Testsystem „*Trp Cage*“-Miniprotein^[79] mit 337 Atomen auf einem Intel Xeon E5-4640 „Sandy Bridge“-System mit 32 Prozessorkernen auf 4 Sockeln das in Abbildung 3 gezeigte Verhalten beobachtet werden.

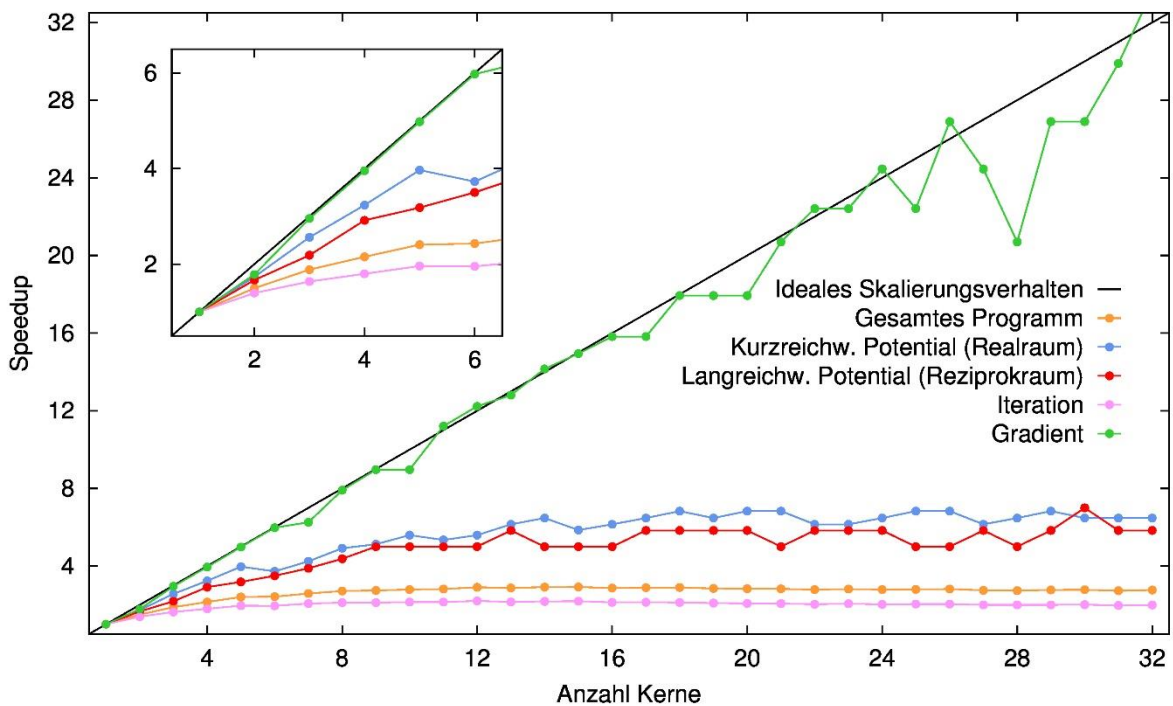


Abbildung 3: Skalierungsverhalten des 3D-RISM-Lösers am Beispiel eines Miniproteins mit 337 Atomen. Der *Speedup* ist definiert als der Quotient der auf einem einzelnen Prozessorkern benötigten Programmlaufzeit und der bei paralleler Ausführung benötigten Programmlaufzeit. Verwendet wurden ein 200^3 -Gitter, 7 MDIIS-Vektoren, die Monopol-Renormierung und das Reziprokraumverfahren. Die Programmdatei wurde mit dem Intel Compiler 15.0.3.187 und der Intel MKL-Bibliothek 11.2.3.187 erstellt.

Die für das 3D-RISM-Verfahren benötigte Zeit kann also in etwa halbiert werden. Eine Betrachtung der einzelnen Programmteile offenbart jedoch ein differenziertes Bild. Die Gradientenberechnung skaliert im Wesentlichen linear, während für die 3D-RISM-Iteration der

maximale *Speedup* bereits bei wenigen Kernen erreicht ist. Die Routinen für die Potentialberechnung liegen dazwischen. Dazu ist zu bemerken, dass der Quellcode für die Berechnung des Gradienten im Rahmen dieser Arbeit vom Autor in Zusammenarbeit mit Roland Frach neu geschrieben wurde und in hohem Maße von den Datenstrukturen des restlichen Programms unabhängig ist. Diese Datenstrukturen beruhen auf Überlegungen von Thomas Kloss bezüglich der Maximierung der Wiederverwendbarkeit von einzelnen Bausteinen für einen 1D-RISM-Löser. Da, wie in Kapitel 2.1 angesprochen, die 1D-RISM-Gleichungen Matrixform, die 3D-RISM-Gleichungen aber Vektorform haben, besitzen die Felder zur Darstellung der Korrelations- und Potentialfunktionen einen zusätzlichen hier aber nicht benötigten Index, was möglicherweise die konsekutive Ablage der Daten im Speicher stört und die Optimierungsfunktionalität des *Compilers* behindert. In der Tat konnte in ersten vorläufigen Experimenten mit einem neuen und entsprechend optimierten, hauptsächlich von Franziska Hoffgaard unter Beteiligung des Autors erstellten Programm, dass statt OpenMP auf das „Message Passing Interface“ (MPI)^[80] setzt und ein explizites Gebietszerlegungsschema verwendet, praktisch perfektes Skalierungsverhalten auf einem einzelnen Rechner mit 16 Kernen beobachtet werden.^[81]

3.1.4 Genauigkeits- und Geschwindigkeitstests

An dieser Stelle sollen anhand einiger weniger Beispiele durch PME und Reziprokraumverfahren erreichbare Geschwindigkeitsvorteile sowie die Approximationsgüte betrachtet werden. Zunächst untersuchen wir, ob die schnelle Berechnung des chemischen Exzesspotentials im reziproken Raum für ein sphärisches Ion in Wasser im Grenzfall eines Gitters mit sehr kleiner Maschenweite das gleiche Ergebnis liefert wie die Lösung der 1D-RISM-Gleichung auf einem radialen logarithmischen Gitter. Das Ergebnis ist in Tabelle 1 dargestellt. Wie man sieht, nähert sich das Ergebnis für kleinere Gitter und für das Reziprokraumverfahren immer weiter dem 1D-RISM-Referenzergebnis an, während für das Realraumverfahren ein „Überschießen“ beobachtet wird, das vermutlich auf residuale Gitterartefakte zurückzuführen ist. Diese werden beim Reziprokraumverfahren dadurch vermindert, dass die Berechnung von $\mu^{\text{ex,(lc)}}$ auf den radialen Gittern nahezu ohne Artefakte durchgeführt werden kann. Weiterhin wurde die Abhängigkeit des chemischen Exzesspotentials von der Zellengröße und dem Renormierungsverfahren getestet. Das Reziprokraumverfahren ist im Vergleich zum Realraumverfahren absolut kompetitiv und zeigt speziell gegenüber dem Realraumverfahren mit Monopolrenormierung eine dramatisch geringere Abhängigkeit von der Zellengröße, unabhängig davon ob es selbst mit der Hintergrund- oder der Monopolrenormierung kombiniert wird (s. Abbildung 4).

Tabelle 1: Abweichung des chemischen Exzesspotentials und der Solvat-Solvens inneren Exzessenergie eines atomaren Ions in Wasser vom 1D-RISM-Referenzwert, berechnet im Realraum (a) und im reziproken Raum (b).

Gitterweite/Å	$\Delta\mu_a^{\text{ex}}/(\text{kcal mol}^{-1})$	$\Delta\mu_b^{\text{ex}}/(\text{kcal mol}^{-1})$	$\Delta U_a^{\text{ex}}/(\text{kcal mol}^{-1})$	$\Delta U_b^{\text{ex}}/(\text{kcal mol}^{-1})$
0.05	-0.03	0.01	-0.07	0.02
0.1	-0.01	0.03	-0.04	0.05
0.2	0.26	0.31	0.46	0.55
0.4	0.23	0.27	0.40	0.49
0.8	-4.92	-2.72	-9.58	-5.17

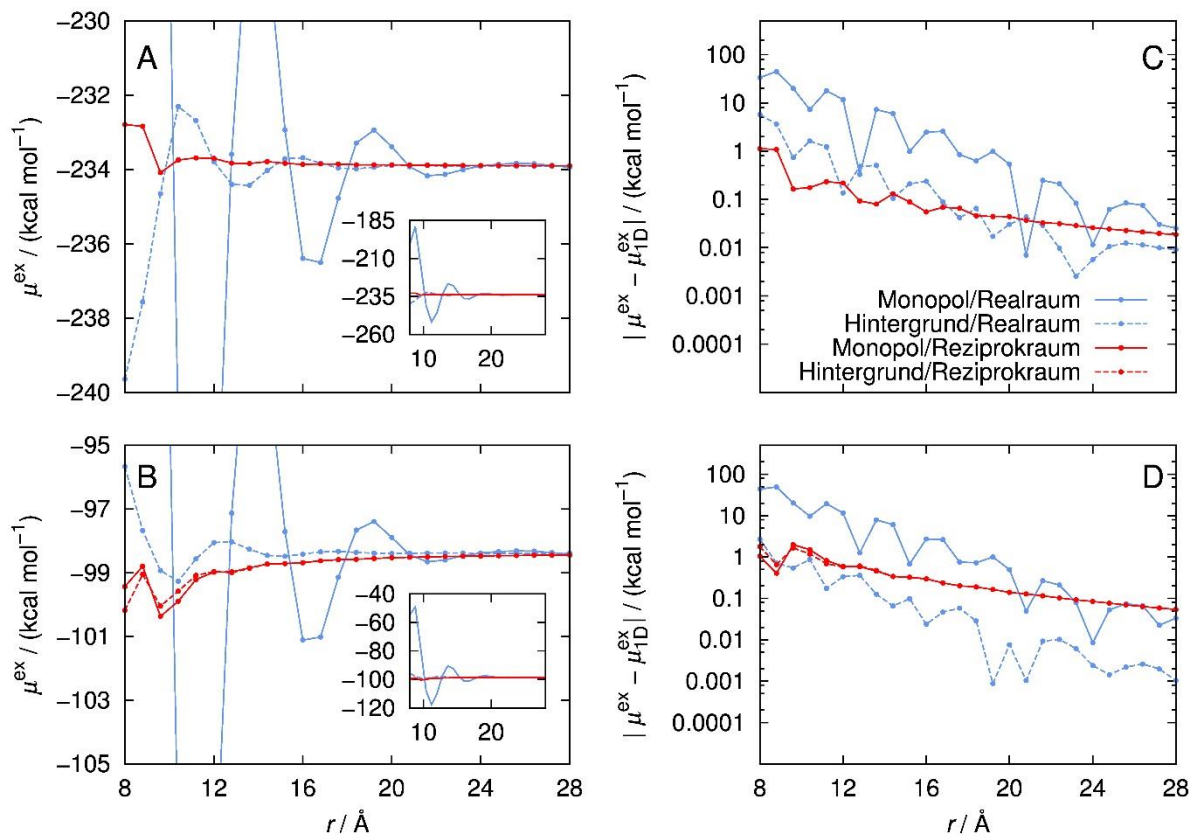


Abbildung 4: μ^{ex} (A, B) und Abweichung von 1D-RISM Referenzergebnissen (C, D) als Funktion der Zellengröße (Gitterauflösung: 0.05 Å) für ein sphärisches Ion in Wasser (A/C: Anion, B/D: Kation).

Das Ion ist ein sinnvoller erster Test, jedoch tragen für einzelne sphärosymmetrische Teilchen nur die Diagonalelemente in Gl. 3.1.21 bei. Daher werden als nächstes das Acetatanion und das Methylammoniumkation betrachtet. Auch muss getestet werden, ob das Reziprokrumverfahren mit Elektrolytlösungen funktioniert. Dementsprechend wurde hier neben Wasser auch eine 0.1 M KCl-Lösung verwendet. Die Ergebnisse in Abbildung 5 sind ähnlich zu den Ergebnissen für das sphärische Ion und auch für die Elektrolytlösungen funktioniert das Reziprokrumverfahren gut.

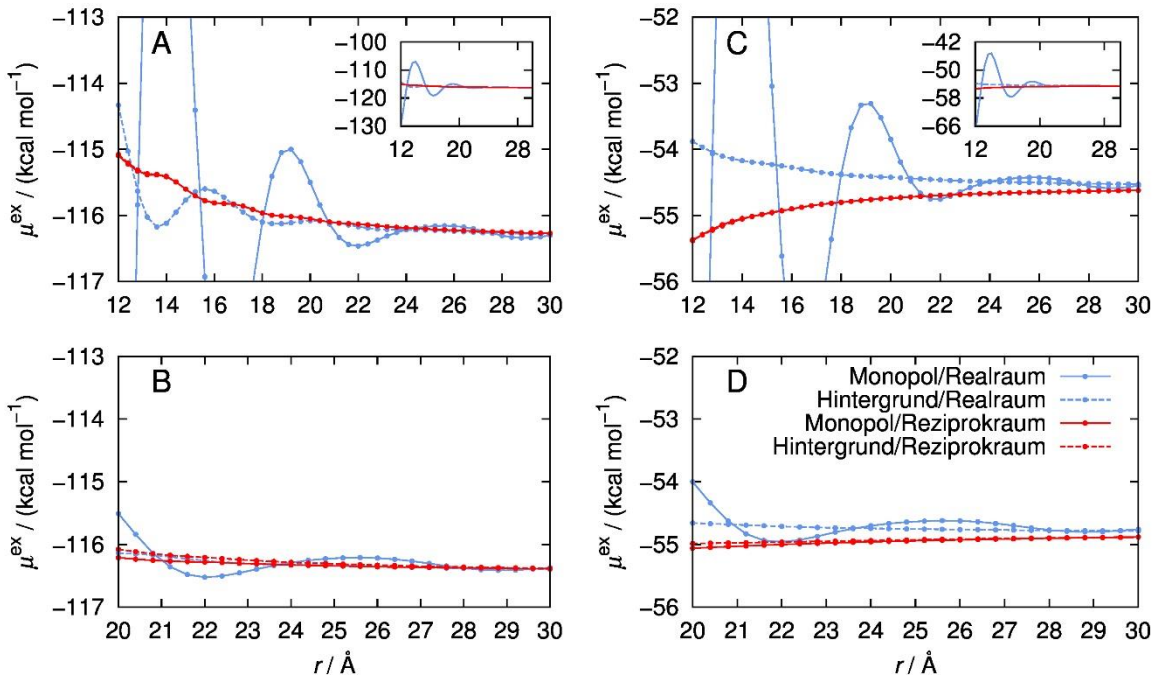


Abbildung 5: μ^{ex} des Acetatanions (A, B) und des Methylammoniumkations (C, D) in Wasser (A, C) und einer 0.1 M KCl-Lösung (B, D). Die Maschenweite des Gitters beträgt 0.2 Å.

Aufgrund der ermutigenden Ergebnisse für die kleinen Testsysteme wurde überprüft, inwiefern das Reziprokrumverfahren zusammen mit dem PME-Ansatz aus Kapitel 3.1.1 für Konformere des Proteins Thioredoxin mit 1655 Atomen in Wasser ähnliche relative chemische Exzesspotentiale liefert wie das etablierte Realraumverfahren. Dazu wurden mit Hilfe einer Moleküldynamik-Simulation durch periodisches Abkühlen 10 Konformere erzeugt und ihre relativen chemischen Exzesspotentiale und Solvat-Solvens Exzessenergien auf verschiedene Art und Weise evaluiert. Mit den so gewonnenen Daten wurden eine statistische Auswertung durchgeführt und mittlere Absolutabweichungen

$$\text{MAA}(\{\mu^{\text{ex}}\}) = \frac{2}{N_{sn}(N_{sn} + 1)} \sum_{j>i}^{N_{sn}} |\mu_i^{\text{ex}} - \mu_j^{\text{ex}} - (\mu_{i,\text{ref}}^{\text{ex}} - \mu_{j,\text{ref}}^{\text{ex}})| \quad 3.1.28$$

und mittlere Relativabweichungen (MRA, Abweichung als Anteil des Betrags von $\Delta\mu^{\text{ex}}$)

$$\text{MRA}(\{\mu^{\text{ex}}\}) = \frac{2}{N_{sn}(N_{sn} - 1)} \sum_{j>i}^{N_{sn}} \left| \frac{\mu_i^{\text{ex}} - \mu_j^{\text{ex}} - (\mu_{i,\text{ref}}^{\text{ex}} - \mu_{j,\text{ref}}^{\text{ex}})}{\mu_{i,\text{ref}}^{\text{ex}} - \mu_{j,\text{ref}}^{\text{ex}}} \right| \quad 3.1.29$$

berechnet.

Die in Tabelle 2 aufgeführten Ergebnisse zeigen, dass spätestens mit Lagrange-Basispolynomen der Ordnungen 8 und 10 und Gitterabständen unter 0.5 Å Energien erhalten werden können, die nur wenig von der Referenz (Realraumverfahren ohne PME) abweichen, obwohl die gleichzeitige Anwendung von PME und Reziprokraumverfahren zu einem Fehlerverstärkungseffekt führt. Die Abweichung der Exzess Inneren Energie ist immer knapp doppelt so groß wie die des chemischen Exzesspotentials, was sich leicht mit dem Faktor 2 im zweiten Term in Gl. 3.1.27 erklären lässt.

Tabelle 2: Mittlere Absolutabweichungen MAA (mittlere Relativabweichungen MRA in Klammern dahinter) für chemisches Exzesspotential und Solvat-Solvens innere Exzessenergie für Thioredoxin-Konformere. a: PME/Realraumverfahren, b: PME/Reziprokraumverfahren.

Polynom- ordnung	Gitter- weite/Å	MAA($\Delta\mu_a^{\text{ex}}$)/ (kcal mol⁻¹)	MAA($\Delta\mu_b^{\text{ex}}$)/ (kcal mol⁻¹)	MAA(U_a^{ex})/ (kcal mol⁻¹)	MAA(U_b^{ex})/ (kcal mol⁻¹)
10	0.3	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.01)
10	0.4	0.00 (0.00)	0.02 (0.00)	0.00 (0.00)	0.05 (0.02)
10	0.5	0.00 (0.01)	0.33 (0.01)	0.01 (0.01)	0.66 (1.27)
10	0.6	0.02 (0.03)	1.56 (0.05)	0.05 (0.02)	3.12 (2.43)
8	0.3	0.00 (0.00)	0.01 (0.00)	0.00 (0.00)	0.03 (0.07)
8	0.4	0.00 (0.00)	0.11 (0.00)	0.00 (0.00)	0.22 (0.16)
8	0.5	0.01 (0.02)	1.02 (0.02)	0.02 (0.02)	2.05 (3.38)
8	0.6	0.04 (0.06)	3.29 (0.08)	0.08 (0.05)	6.58 (5.25)
6	0.3	0.00 (0.01)	0.11 (0.00)	0.00 (0.00)	0.23 (0.33)
6	0.4	0.01 (0.01)	0.57 (0.01)	0.01 (0.02)	1.15 (1.25)
6	0.5	0.02 (0.03)	3.56 (0.04)	0.04 (0.03)	7.12 (10.13)
6	0.6	0.07 (0.10)	7.55 (0.13)	0.13 (0.09)	15.10 (12.13)
4	0.3	0.01 (0.08)	1.03 (0.02)	0.02 (0.02)	2.06 (1.22)
4	0.4	0.04 (0.07)	3.61 (0.08)	0.08 (0.13)	7.21 (9.83)
4	0.5	0.06 (0.05)	14.34 (0.12)	0.12 (0.08)	28.67 (37.59)
4	0.6	0.16 (0.14)	19.28 (0.35)	0.35 (0.18)	38.55 (30.19)
2	0.3	0.14 (0.83)	12.52 (0.29)	0.29 (0.22)	25.02 (34.48)
2	0.4	0.27 (0.45)	28.69 (0.59)	0.59 (1.01)	57.36 (89.63)
2	0.5	0.50 (0.54)	71.03 (1.06)	1.06 (0.74)	142.01 (184.36)
2	0.6	0.97 (1.50)	61.97 (1.94)	1.94 (1.09)	123.90 (87.49)

Es bleiben noch der Einfluss der PME-Näherung auf die Korrelationsfunktionen und der erzielbare Geschwindigkeitszuwachs zu untersuchen. Dazu wurden die Sauer- und Wasserstoffpaarverteilungsfunktionen von Wasser um den im Vergleich zum relativ kleinen Thioredoxin mit 5604 Atomen wesentlich größeren PARP-Komplex mit dem PDB Code 4PAX^[82]

berechnet. Wie in Abbildung 6 zu sehen ist der Unterschied bereits mit Lagrange-Basispolynomen zweiter Ordnung kaum noch zu erkennen. Noch besser lässt sich dies verdeutlichen, indem man die räumlichen Paarverteilungsfunktionen durch Integration über den Winkelraum in radiale Verteilungsfunktionen um einzelne Solvatzentren überführt und diese betrachtet. Wie aus Abbildung 7 ersichtlich, sind die Abweichungen auch hier minimal.

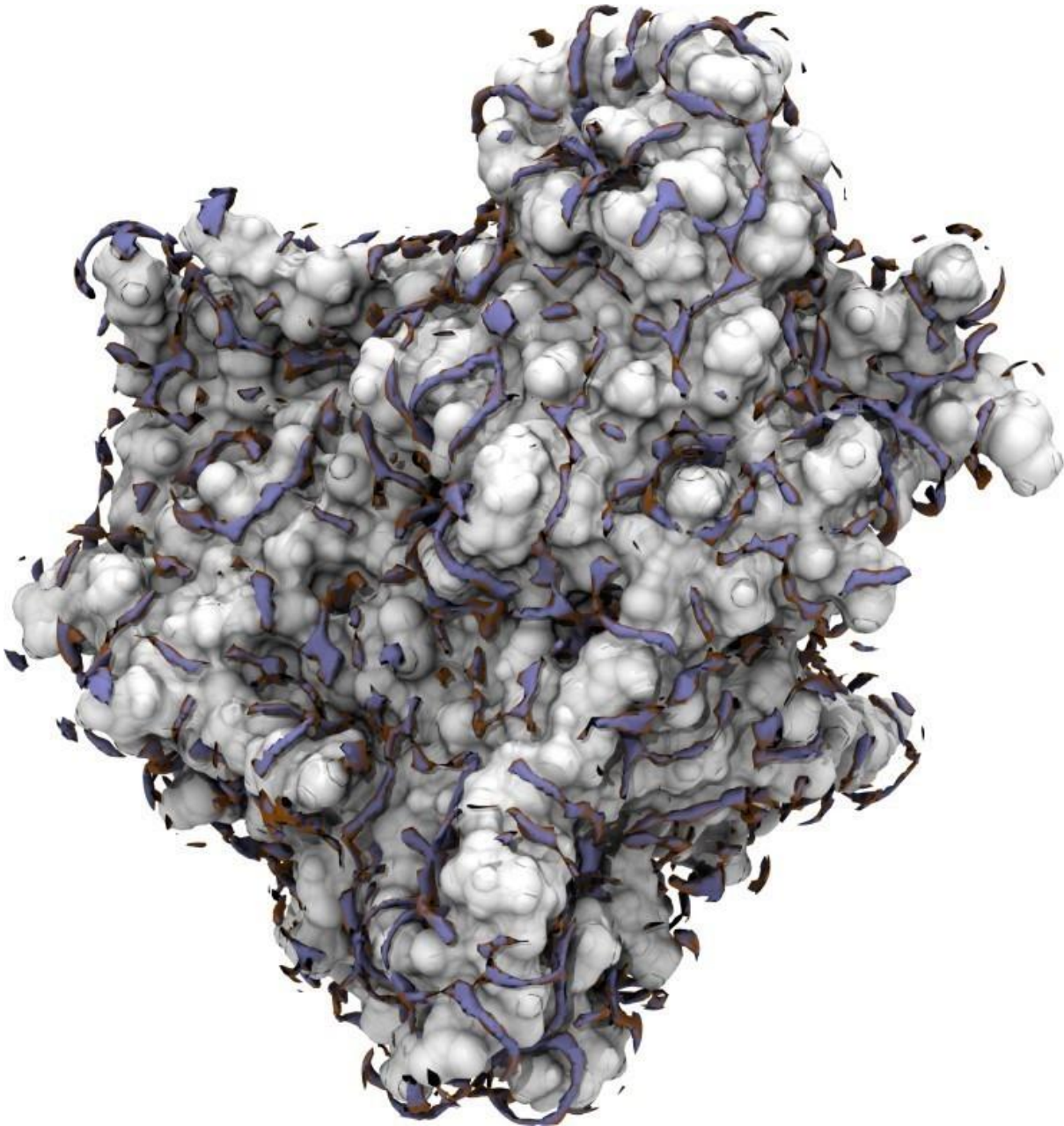


Abbildung 6: 3D Wasserverteilung (Isoflächen der 3-fachen *Bulk*-Sauerstoffdichte) um den PARP Komplex mit dem PDB Code 4PAX. Orange: PME mit Lagrange Basispolynomen zweiter Ordnung und einem 12.0 Å Abscheideradius für das Realraumpotential. Violett: Normale Ewald-Summmation, kein Abschneiden des Realraumpotentials. Das Bild wurde mit Hilfe von VMD 1.9.1^[83] erzeugt.

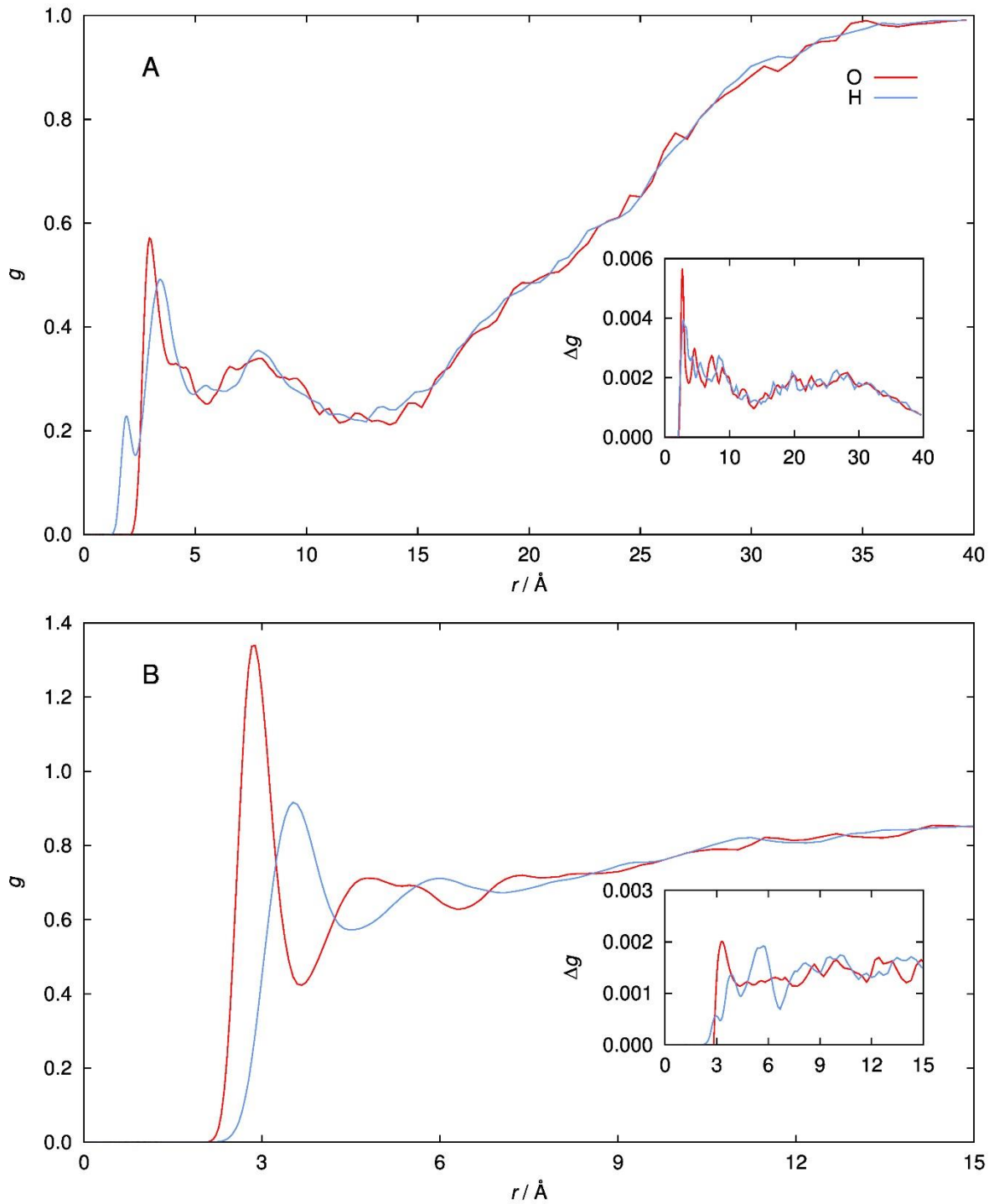


Abbildung 7: Radiale Verteilungsfunktionen von Wasser-Sauerstoff (Rot) und Wasserstoff (Blau) um (A) die Hydroxygruppe des Liganden und (B) das N-terminale Stickstoffatom des PARP Proteins. Inset-Plots zeigen die Abweichung vom Referenzergebnis ohne PME/Abschneiden des Realraumpotentials und den Ergebnissen mit PME (Lagrange-Basispolynome zweiter Ordnung) und einem 12 \AA Abschneideradius für das Realraumpotential.

Wie hoch der Geschwindigkeitsvorteil ist, der durch PME und Reziprokrumverfahren erreicht werden kann, hängt von der Größe des jeweiligen Systems ab. Generell beobachtet man, dass der Anteil der Potentialberechnung mit der Größe des Systems stark zunimmt. Der Grund ist, dass die in der 3D-RISM-Iteration auftretenden FFT-Operationen mit $O(n_x n_y n_z \log n_x n_y n_z)$ skalieren und der Rest der Iteration gitterpunktlokal verläuft, also linear mit der Systemgröße skaliert. Nimmt man an, dass die Zellengröße linear mit der Anzahl der Solvatzentren ansteigt, so skaliert die Potentialberechnung quadratisch. Im Fall von 4PAX bedeutet das, dass die Rechenzeit ohne die 3D-RISM-Iteration auf dem 6 Prozessorkernen des Testsystems (Intel Xeon X5690) durch den Einsatz von PME und Reziprokrumverfahren von 18.5 auf 6.7 Minuten gesenkt werden konnte, wie in Abbildung 8 gezeigt wird. Wird zusätzlich noch ein Abschneideradius von 12 Å für das Realraumpotential verwendet, kann die Rechenzeit noch stärker auf 1.6 Minuten gesenkt werden. Die Rechenzeit für die Iteration beträgt ca. 21 Minuten. Damit entfällt für das 4PAX-System fast die Hälfte der Rechenzeit.

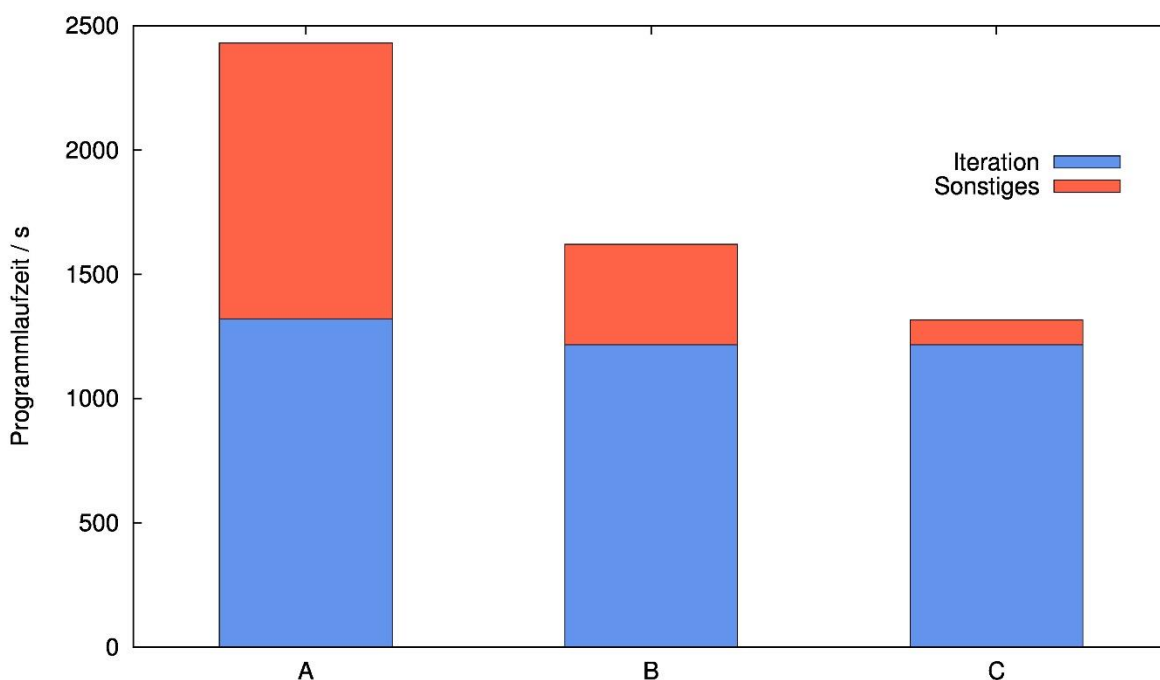


Abbildung 8: Verteilung der Rechenzeit zwischen 3D-RISM-Iteration und sonstigem (i.W. Potential- und μ^{ex} -Berechnung) für den PARP Komplex mit dem PDB Code 4PAX. A: Realraumverfahren/Ewald Summation. B: Reziprokrumverfahren/PME (Lagrange-Basispolynome 10. Ordnung). C: Reziprokrumverfahren/PME (Lagrange-Basispolynome 10. Ordnung)/Realraum-Abschneideradius 12 Å.

3.2 QM/MM-Berechnungen mit EC-RISM-Solvens

Für die Berechnung der Eigenschaften sehr großer Systeme wie z.B. von Protein-Ligand-Komplexen mit mehreren tausend Atomen ist die EC-RISM-Methode wie in Kapitel 2.2 beschrieben nicht unmittelbar geeignet. Hauptgrund dafür ist, dass der Aufwand für typische *ab-initio*- und Dichtefunktional-Quantenchemieverfahren nur für maximal wenige hundert Atome tragbar ist. Aus diesem Dilemma bieten sich prinzipiell zwei unterschiedliche Auswege. Zum einen können stärker approximative quantenchemische Verfahren wie semiempirische *Neglect of Diatomic Differential Overlap*- (NDDO-) Methoden wie AM1^[84] oder *Density-Functional based Tight Binding* (DFTB)^[85,86] eingesetzt werden, zum anderen besteht die Möglichkeit, nur einen Teil des Systems quantenmechanisch (QM) zu behandeln, den anderen Teil aber mit klassischer Molekülmechanik (MM) zu beschreiben. Diese Methoden sind unter dem Akronym QM/MM bekannt geworden.^[87,88] Für die im Kontext dieser Arbeit studierten Tautomer- und Ionisierungsprobleme bestehen erhebliche Genauigkeitsanforderungen. So verursacht eine Abweichung in der Gibbs-Energie von 1 kcal/mol („Chemische Genauigkeit“) bereits eine Abweichung von einer pK_a -Einheit. Insofern haben bereits kleine Kompromisse bei der quantenchemischen Beschreibung des Solvats erhebliche Folgen. QM/MM-Methoden erlauben die kompromisslose Behandlung des QM-Teilsystems, d.h. im vorliegenden Fall des Solvats, weshalb hier diese zweite Möglichkeit, die QM/MM-Beschreibung, gewählt wurde. Der Preis dafür ist, dass die Kompromisse stattdessen im Übergangsbereich zwischen der quantenmechanisch zu behandelnden Zone und dem MM-Bereich gemacht werden müssen, und dass im MM-Bereich quantenmechanische Effekte vernachlässigt werden. Auch wenn die QM/MM-Grenze keine kovalenten Bindungen durchschneidet ist im Fall von durchschnittlichen Wasserstoffbrücken, z.B. zwischen einem Protein und einem Liganden, mit Artefakten zu rechnen, weil einer der Bindungspartner durch Punktladungen und LJ-Parameter repräsentiert wird.

Die Kopplung der QM/MM-Methodologie mit impliziten Solvationsmodellen ist eine neuere Entwicklung. Bekannt geworden ist hier das in Gaussian 09 implementierte ONIOM-PCM-Modell.^[89] Eigene vorläufige Tests zeigen aber, dass die Geschwindigkeit des Verfahrens für Protein-Ligand-Komplexe völlig ungeeignet ist. Problematisch scheint hier zu sein, dass die Größe des bei der Diskretisierung der Poisson-Gleichung mit der Randelementmethode (engl. *Boundary Element Method*, BEM) im Rahmen von PCM und COSMO entstehende lineare Gleichungssystem kubisch mit der Anzahl der Solvatzentren skaliert, und dass die in diesem Gleichungssystem auftretenden Matrizen i.A. dicht besetzt sind.^[90] Vor kurzem wurden von einer Reihe von Autoren aus dem Umfeld der Gaussian-Entwickler Arbeiten veröffentlicht, die BEM mit einem Gebietszerlegungsverfahren kombinieren, aus

dem ein dünn besetztes Gleichungssystem resultiert, das mit linearem Aufwand gelöst werden kann.^[90-94] Die Softwareimplementierung wurde bislang aber nicht für die Öffentlichkeit freigegeben.

Im Gegensatz zu PCM und COSMO liegt der Aufwand für die Lösung des 3D-RISM/*Closure*-Gleichungssystems auch für große Systeme mit mehreren tausend Atomen im vertretbaren Bereich von einigen Minuten bis Stunden. Insofern bietet sich EC-RISM als Solvationsmodell für große QM/MM-Systeme sowohl in Bezug auf die Ausführungsgeschwindigkeit als auch den höheren Realitätsgrad an. Zu diesem Zweck muss der in Kapitel 2.2 eingeführte EC-RISM-Formalismus erweitert werden. Aus dem intra-Solvat-Hamiltonoperator in Gl. 2.2.4 wird

$$\hat{H}_1 = \hat{H}_{ne} + \hat{H}_{ee} + \hat{H}_{nn} + \hat{H}_{QM/MM,elec} + \hat{H}_{QM/MM,LJ} + \hat{H}_{MM}. \quad 3.2.1$$

Die Punktladungen aus dem hier verwendeten AMBER-Kraftfeld werden analog zu den gitterbasierten Punktladungen für den Solvenseffekt semi-quantenmechanisch behandelt, d.h. der Term

$$\hat{H}_{QM/MM,elec} = \sum_{i=1}^{N_e} \sum_{j=1}^{N_{MM}} \frac{e q_j}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{A=1}^{N_A} \sum_{j=1}^{N_{MM}} \frac{Z_A e q_j}{|\mathbf{r}_A - \mathbf{r}_j|}. \quad 3.2.2$$

wird zum intra-Solvat-Hamiltonoperator \hat{H}_1 addiert. Hier sind e die Elementarladung, q_j eine MM-Punktladung, Z_A die Ordnungszahl von Kern A und N_e und N_{MM} die Zahl der Elektronen des QM-Systems und die Zahl der MM-Atome. Apolare Wechselwirkungen zwischen QM-Zone und MM-Atomen müssen rein klassisch behandelt werden. Dazu wird hier einfach das übliche Lennard-Jones 6-12-Potential

$$\hat{H}_{QM/MM,LJ} = 4 \sum_{\alpha=1}^{N_\alpha} \sum_{\beta=1}^{N_{MM}} \epsilon_{A_j} \left(\left(\frac{\sigma_{\alpha\beta}}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|} \right)^{12} - \left(\frac{\sigma_{\alpha\beta}}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|} \right)^6 \right). \quad 3.2.3$$

mit AMBER bzw. GAFF^[95]-Parametern verwendet. \hat{H}_{MM} ist der übliche AMBER-Potentialausdruck^[96]

$$\begin{aligned} \hat{H}_{MM} = & \sum_i^{N_{Bindung}} b_i (R_i - R_{i,eq})^2 + \sum_i^{N_{Winkel}} a_i (\theta_i - \theta_{i,eq})^2 + \sum_i^{N_{Diederwinkel}} \frac{V_{i,n}}{2} [1 + \cos(n\phi_i - \gamma_{i,n})] \\ & + \sum_{i<j}^{N_{Atom}} \dagger 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{|\mathbf{R}_i - \mathbf{R}_j|} \right)^{12} - \left(\frac{\sigma_{ij}}{|\mathbf{R}_i - \mathbf{R}_j|} \right)^6 \right] + \sum_{i<j}^{N_{Atom}} \dagger \frac{q_i q_j}{|\mathbf{R}_i - \mathbf{R}_j|}, \end{aligned} \quad 3.2.4$$

mit den Bindungs-, Winkel- und Diederwinkelkraftkonstanten b_i , a_i und $V_{i,n}$, den Abständen und Winkeln R_i und θ_i mit ihren Gleichgewichtswerten $R_{i,eq}$ und $\theta_{i,eq}$, den Diederwinkeln ϕ_i , den Phasenverschiebungen $\gamma_{i,n}$, den Periodizitäten n , und den LJ-Parametern σ_{ij} und ϵ_{ij} . Bei den Summen

über die Terme 4 und 5 symbolisiert „†“, dass diese Terme so nur für Atompaare ij gelten, bei denen die Atome i und j entweder in unterschiedlichen Molekülen liegen oder weiter als 3 Bindungen voneinander entfernt sind. Sind zwischen den Atomen innerhalb des gleichen Moleküls nur 3 Bindungen, so werden die LJ-Wechselwirkungen im 4. Term durch 2, die elektrostatischen Wechselwirkungen im 5. Term durch den Wert 1.2 dividiert.

Ein ähnliches Verfahren wie das hier Beschriebene, jedoch auf der Basis von 3D-RISM-SCF statt EC-RISM, wurde bereits von Hirata und Mitarbeitern veröffentlicht.^[97]

3.3 Schnelle Approximation der Solvens-Elektrostatik für EC-RISM

Der Aufwand von EC-RISM-Berechnungen wird maßgeblich durch die Zahl der Punktladungen, die den namensgebenden einbettenden *Cluster* um das Solvat bilden und den elektrostatischen Einfluss des Solvens repräsentieren, bestimmt. Durch Ausnutzen der Tatsache, dass weiter vom Solvat entfernte Punktladungen bei gleicher Ladung weniger stark zum Ergebnis beitragen als solche, die sich nahe am Solvat befinden, sollte sich die Anzahl dieser Punktladungen reduzieren lassen ohne die Qualität des Ergebnisses signifikant zu beeinflussen. Dazu dienen beispielsweise sogenannte hierarchische Baumverfahren wie das Barnes-Hut- (BH-) Verfahren^[98] und die schnelle Multipolmethode (engl. *Fast Multipole Method*, FMM)^[64], die im Kontext von Partikelsimulationen breite Anwendung finden.^[68] Die grundlegende Idee des BH-Verfahrens ist, dass anstatt in $O(N^2)$ die Wechselwirkung eines jeden Partikels mit jedem anderen zu berechnen, die Partikel in Abhängigkeit des Abstands zu immer größeren Partikelgruppen zusammengefasst werden. Für jedes Partikel wird dann die Wechselwirkung mit Partikel-Gruppen unterschiedlicher Größe an Stelle der einzelnen Partikel evaluiert, woraus ein Skalierungsverhalten von $O(N \log N)$ resultiert. Die linear skalierende FMM geht darüber hinaus, indem hier Wechselwirkungen zwischen den Partikel-Gruppen evaluiert werden. Basisbestandteil des BH-Verfahrens ist eine baumartige sogenannte *Octree*-Datenstruktur. Dabei handelt es sich um eine Zerlegung des Simulationsgebiets in sukzessiv kleinere Zellen, die rekursiv aufgebaut wird. Die ursprüngliche Simulationszelle („Wurzel“) wird in 8 Teile aufgeteilt. Für jedes dieser Teile wird überprüft, ob sich darin mehr als eine vorgegebene Mindestanzahl Partikel befindet. Ist das der Fall, wird in den Koordinaten dieses Teils eine Tochterzelle konstruiert, die mit der „Elternzelle“ über eine Zeigervariable verbunden ist. Falls die Anzahl der Partikel in der Tochterzelle größer als die vorgegebene Mindestanzahl ist, wird die Tochterzelle dann ihrerseits analog der ursprünglichen Simulationsbox in Tochterzellen aufgeteilt. Dieses Vorgehen wird fortgesetzt, bis keine neuen Tochterzellen mit mehr als der Mindestanzahl an Partikeln mehr konstruiert werden. Die Zellen

mit der Mindestzahl an Partikeln oder weniger werden in diesem Kontext als „Blätter“ bezeichnet. Zu jeder Box werden die Gesamtladung Q aller Partikel in der Zelle sowie ein Größenmaß d und ein Mittelpunkt \mathbf{R} (im einfachsten Fall einer würfelförmigen Box ist d die Kantenlänge und \mathbf{R} ihr geometrischer Mittelpunkt) gespeichert. Für einen Auswertungsort \mathbf{r} wird dann das elektrostatische Potential $\Phi(\mathbf{r})$ berechnet, indem der *Octree* ausgehend von der Wurzel in Richtung der Blätter durchlaufen wird. Ist die Zelle ein Blatt, wird das elektrostatische Potential aller in dieser Zelle enthaltenen Partikel explizit evaluiert und zu $\Phi(\mathbf{r})$ addiert. Ansonsten wird für jede besuchte Zelle das Akzeptanzkriterium

$$\theta \leq \frac{d}{|\mathbf{r} - \mathbf{R}|} \quad 3.3.1$$

mit einem vorgegebenen Schwellenwert θ überprüft. Erfüllt eine Zelle das Akzeptanzkriterium nicht, so wird das Potential der Ladung Q im Abstand $|\mathbf{r} - \mathbf{R}|$ auf $\Phi(\mathbf{r})$ addiert. Dies ist die Basis der Rechenzeiterparnis, da dadurch vermieden wird, dass das Potential aller einzelnen zur Zelle gehörenden Punktladungen berechnet und zu $\Phi(\mathbf{r})$ addiert werden muss.

Die Grundidee des BH-Schemas soll im Folgenden genutzt werden, um die Anzahl der *Embedded-Cluster*-Punktladungen im EC-RISM-Verfahren zu reduzieren. Im Gegensatz zu der Situation bei Partikelsimulationen muss im Fall von EC-RISM $\Phi(\mathbf{r})$ nicht an den Positionen der Partikel, sondern den Mitten der Gitterzellen eines regulären Gitters evaluiert werden. Daher macht es hier Sinn, nicht von der Simulationsbox ausgehend einen *Octree* aufzubauen, sondern stattdessen je 8 Gitterzellen hierarchisch zusammenzufassen, wenn in Analogie zu Gl. 3.3.1 das Verhältnis von d zum Abstand des jeweiligen Zellenmittelpunkt vom nächstgelegenen Atom des Solvats einen Schwellenwert unterschreitet. Eine alternative Betrachtungsweise von Gl. 3.3.1 wäre, dass ein bestimmter Abstand $|\mathbf{r} - \mathbf{R}|$ bei gegebenem θ einen maximalen Zellendurchmesser d erlaubt, bzw. dass der erlaubte maximale Durchmesser eine lineare Funktion des Abstands ist. Ob der lineare Zusammenhang zu einem optimalen Verhältnis von Rechenzeit und Genauigkeit führt, ist allerdings unklar, insbesondere da weiter entfernte Punktladungen auch zusätzlich noch durch das Abfallen der Paarkorrelationsfunktionen betragsmäßig klein sind. Für 1D-RISM- und OZ-Rechnungen hat sich ein logarithmisches Gitter bewährt.^[76] Leider lässt sich dieses Gitter nicht ohne weiteres von der 1D- auf die 3D-Situation bei uniformen kartesischen Gittern übertragen, weil die Eigenschaften des 3-dimensionalen Raums die Existenz eines logarithmischen Gitters in 3 Dimensionen verhindern. Durch ein modifiziertes Akzeptanzkriterium wie z.B.

$$\tau_1 + \tau_2 \exp(\tau_3 |\mathbf{r} - \mathbf{R}|) - \tau_2 \leq d \quad 3.3.2$$

mit drei geeignet zu wählenden Parametern τ_1 , τ_2 und τ_3 lässt sich jedoch ein adaptives Gitter konstruieren, das eine logarithmische Verteilung des Zelldurchmessers in Abhängigkeit des Abstands approximiert. Zu beachten ist hierbei aber, dass die Maschenweite des ursprünglichen für 3D-RISM verwendeten Gitters die Zellengröße nach unten hin beschränkt. Daher ist τ_1 sinnvollerweise die Maschenweite des ursprünglichen Gitters. Ein so konstruiertes Gitter für Pteroisäure mit $\tau_1 = 0.3$, $\tau_2 = 0.005$ und $\tau_3 = 0.35$ ist in Abbildung 9 gezeigt.

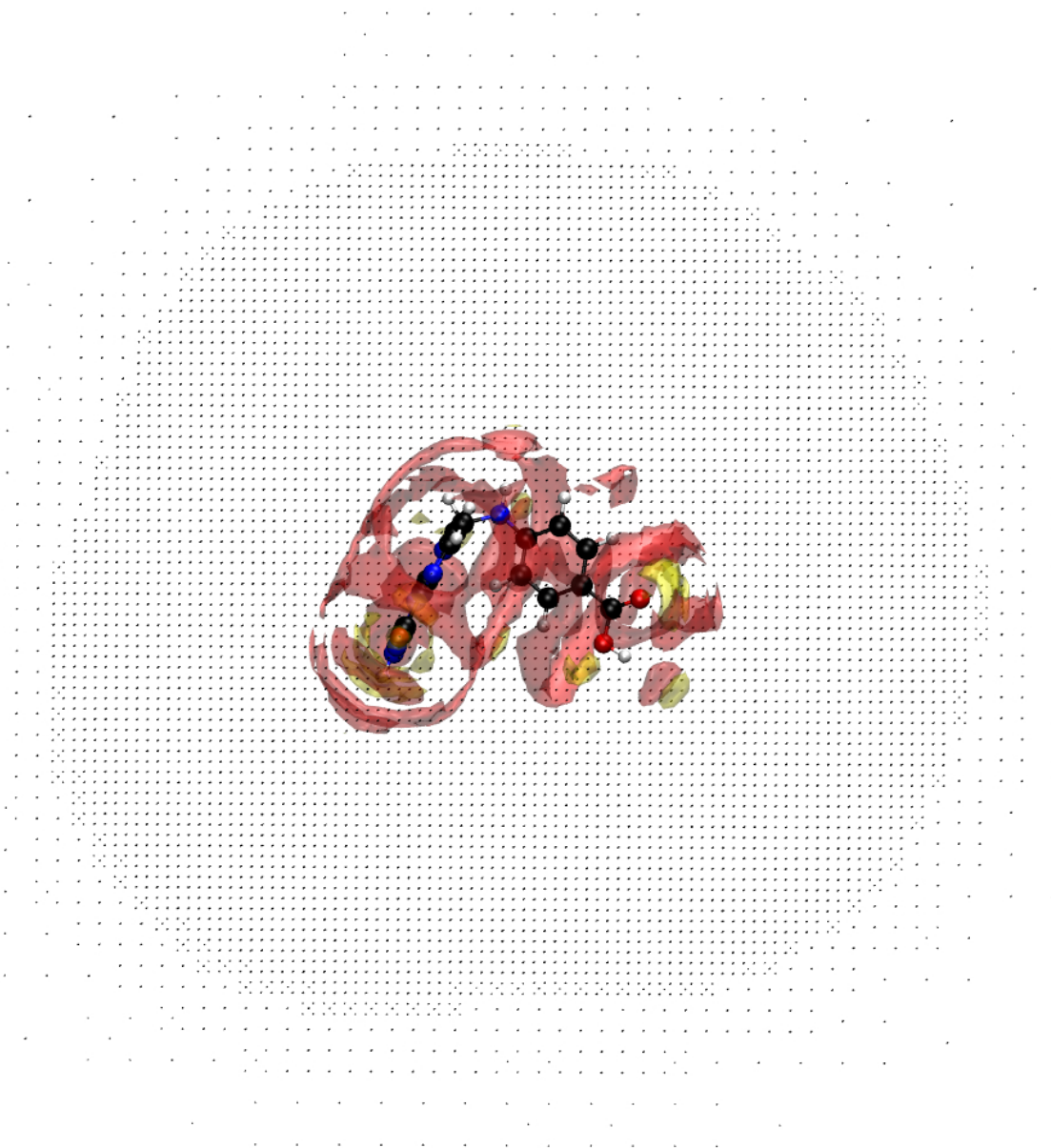


Abbildung 9: Ausschnitt aus dem adaptiven *Embedding*-Punktladungscluster für Pteroinsäure, ermittelt wie im vorliegenden Kapitel beschrieben. Zu sehen sind weiterhin Isoflächen der Paarkorrelationsfunktionen von Wasserstoff (gelb) und Sauerstoff (rot) für die 3-fache Dichte im Vergleich zu reinem Wasser, was einen Eindruck von der Größe der ersten Solvensschale vermittelt. Zur besseren Sichtbarkeit wurden ober- und unterhalb der Darstellungsebene Punktladungen ausgeblendet.

Der Algorithmus zur Konstruktion der adaptiven 3D Gitter verläuft wie folgt:

1. Das vorhandene Gitter mit den das Solvens repräsentierenden Punktladungen wird symmetrisch in alle Richtungen so vergrößert, dass in jeder Richtung die Anzahl der Gitterpunkte die nächstliegende Zweierpotenz ist. Ein Gitter mit den Gitterpunktanzahlen $z = 102 \times 108 \times 110$

würde so zu einem $z = 128^3$ -Gitter. Neu hinzugekommene Gitterzellen erhalten die Punktladung Null in ihrem Mittelpunkt zugewiesen.

2. Für jede Zweierpotenz von (in unserem Beispiel) 2-128 wird ein Gitter angelegt, in dem die Ladungen in der Mitte der Gitterzellen gespeichert werden. Zunächst wird das größte Feld mit den in Schritt 1 generierten Werten aufgefüllt. Danach wird das zweitgrößte Feld befüllt, und zwar so, dass die Zellen dieses Feldes die Ladung aus jeweils 8 einen Würfel bildenden Gitterzellen des größten Feldes erhalten. Dieses in Abbildung 10 für den 2-dimensionalen Fall gezeigte Schema wird dann sukzessive auf alle Felder angewendet, ausgehend vom größten bis zum kleinsten Feld.
3. Der *Embedded*-Punktladungscluster wird konstruiert, indem die in Abbildung 10 gezeigte Datenstruktur von oben nach unten durchlaufen wird. Für jede Zelle wird der Abstand zum dem Zellmittelpunkt nächstgelegenen Solvatatom bestimmt und die Einhaltung von Gl. 3.3.2 geprüft. Ist Gl. 3.3.2 gültig, wird der Durchlauf an dieser Stelle nicht mehr fortgesetzt und die Punktladung im Mittelpunkt der Zelle wird für den Punktladungscluster verwendet. Wenn während des Durchlaufs die unterste Ebene, also das ursprüngliche Gitter, erreicht wird, werden diese Punktladungen in jedem Fall verwendet.

Problematisch an diesem Vorgehen ist, dass für den Mittelpunkt jeder Zelle jeder Ebene das nächstgelegene Solvatatom gefunden werden muss. Wenn diese Suche mittels einer naiven 3-fach-Schleifenkonstruktion durchgeführt wird, kann die benötigte Rechenzeit für große Systeme beträchtlich sein, und die durch die Einsparung von Punktladungen gewonnene Rechenzeit aufwiegen. Glücklicherweise existieren bereits Algorithmen für eine schnelle „nächster-Nachbar-Suche“, wie das hier verwendete Verfahren auf der Basis einer Voronoi-Zerlegung. Die Voronoi-Zerlegung ist eine Methode, um für eine Gruppe von „Zentren“, im vorliegenden Fall die Positionen der Solvatatome, den Raum in Regionen aufzuteilen innerhalb derer jeder Punkt dem zu dieser Region gehörenden Zentrum näher ist als allen anderen Zentren. Ein 2-dimensionales Beispiel ist in Abbildung 11 gezeigt. Nach der Voronoi-Zerlegung muss dann nur noch diejenige Region gefunden werden, in der der Punkt, zu dem das nächste Solvatatom gefunden werden soll, liegt. Für die Voronoi-Zerlegung ist ein mit $O(N \log N)$ skalierender Algorithmus bekannt^[99], für die Suche nach dem nächsten Nachbarn kann der mit $O(\log N)$ skalierende Algorithmus von Kirkpatrick^[100] verwendet werden. In der Praxis zeigt sich, dass die beiden Aufgaben von der hier verwendeten C++-Bibliothek „Voro++“^[101,102] innerhalb weniger Sekunden für eine große Anzahl an Solvatatomen durchgeführt werden können. Wieviel Rechenzeit bereits für kleine

Systeme eingespart werden kann, wird für Pteroisäure in Abbildung 12 gezeigt. Für große Systeme sind die Einsparungen erheblich größer, z.B. kann für den Pteroisäure-Rizin-Komplex in Kapitel 3.9.2 die Anzahl der Punktladungen in etwa um den Faktor 30 reduziert werden. Ähnliche Verfahren wurden bereits im Kontext der Finite-Elemente-Methode (FEM) im Bereich der Kontinuumsmechanik diskutiert.^[103]

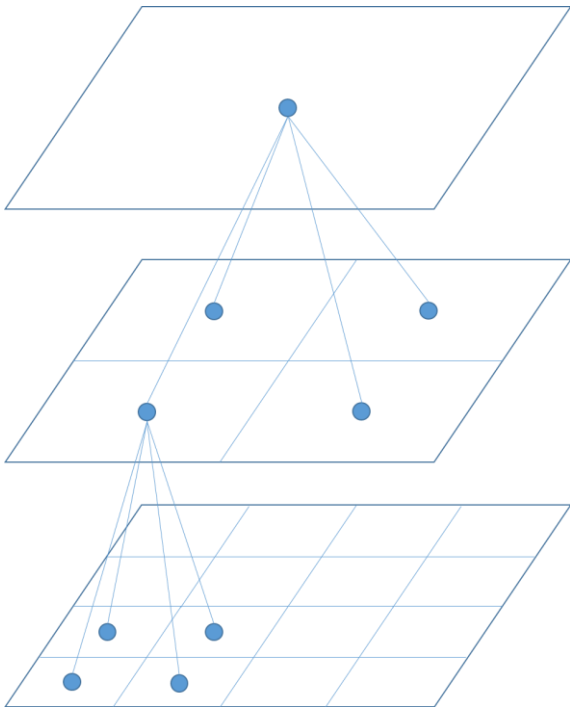


Abbildung 10: Aufbau des Hierarchiebaums für die Konstruktion des adaptiven EC-RISM-Punktladungsclusters in 2 Dimensionen.

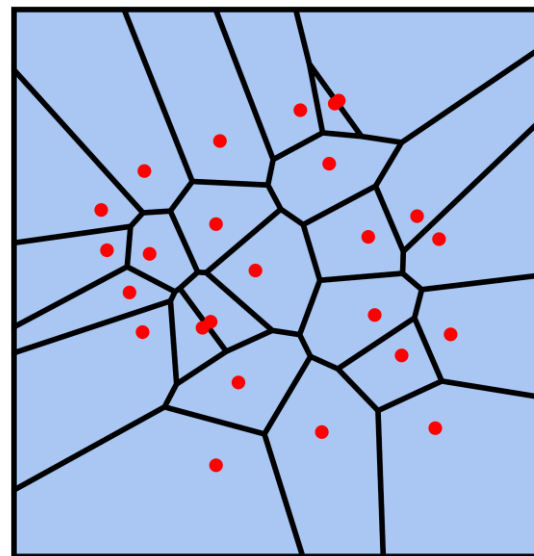


Abbildung 11: Beispiel für ein 2D-Voronoi-Diagramm. Gezeigt werden die Voronoi-Regionen mit ihren jeweiligen Zentren (rote Punkte).

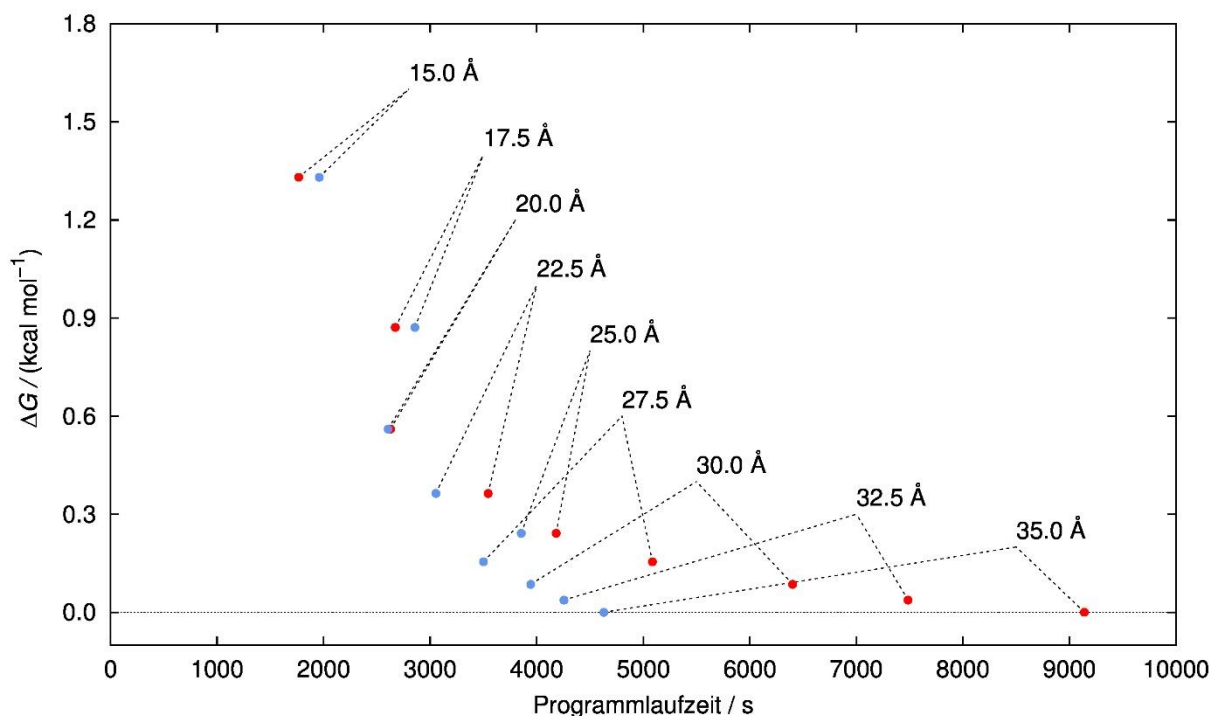


Abbildung 12: Auftragung der Programmierlaufzeit gegen den Box-Abschneidefehler für EC-RISM/PSE-2 mit B3LYP/6-31G für Pteroinsäure auf einem Gitter mit einer Maschenweite von 0.3 \AA auf 4 Prozessorkernen eines Computers mit Intel Xeon E5-2650 Prozessor mit 2 GHz Taktfrequenz bei verschiedenen Boxgrößen unter Verwendung des adaptiven Gitters für den elektrostatischen Einfluss des Solvens (blau) bzw. des Standardverfahrens (rot). Die Angaben im Inneren des Bildes geben den minimalen Abstand von äußerstem Atom des Moleküls zum Rand der Box an. Die Gibbs-Energie für das Standardverfahren bei 35 \AA wurde als Referenzwert verwendet. Wie aus der Abbildung ersichtlich, sind die Parameter $\tau_1 = 0.3$ $\tau_2 = 0.005$ und $\tau_3 = 0.35$ sehr konservativ gewählt, d.h. die mit dem adaptiven Gitter berechneten Gibbs-Energien unterscheiden sich kaum von den mit dem üblichen regulären Gitter berechneten Gibbs-Energien. Die Abweichungen liegen unterhalb von 0.0005 kcal/mol .

Nachdem in 3.1 bis 3.3 technische Voraussetzungen für schnelle EC-RISM-Rechnungen an großen Systemen geschaffen wurden, sollen nun in 3.4 und 3.5 konzeptionelle Aspekte angesprochen werden.

3.4 Tautomerenverhältnisse, pK_a und Speziesverteilung in Lösung

Im Folgenden wird eine Methodik entwickelt, um in einer Lösung mit unendlicher Verdünnung die Stoffmengenanteile verschiedener Protonierungszustände eines Solvats mit mehreren titrierbaren Zentren zu bestimmen. Ähnliche Überlegungen werden seit langem z.B. für pK_a -Berechnungen an

Proteinen verwendet^[104–108], wurden jedoch speziell in Kombination mit dielektrischen Kontinuumsmodellen formuliert und lassen sich nicht ohne Weiteres von diesen trennen.

3.4.1 Reaktions-Gibbs-Energien und das Potential der mittleren Kraft

Für das Chemische Standardpotential der Spezies i wurde von McCammon und Mitarbeiter im Kontext der biomolekularen Bindungsthermodynamik der Ausdruck

$$\mu_i^0 = -\beta^{-1} \ln \left(\frac{\eta_i O_i Z_{N,i}(V,T)}{c_i^0 \sigma_i Z_{N,0}(V,T)} \right) + p^0 \bar{V}_i \quad 3.4.1$$

hergeleitet.^[109] Hierbei sind σ_i die Symmetriezahl, O_i ein massenabhängiger Faktor, c^0 die Standardkonzentration, η_i der Orientierungsfaktor mit $\eta_i = 1$ für strukturlose Teilchen wie das Proton und $\eta_i = 8\pi^2$ für allgemeine nichtlineare Moleküle, p^0 der Standarddruck, $Z_{N,0}$ das kanonische Konfigurationsintegrale des reinen Solvens und $Z_{N,i}$ das des Solvens mit einem Solvatmolekül. $\bar{V}_i = V_{N,i} - V_{N,0}$ wird im thermodynamischen Grenzfall zum partiellen molaren Volumen. Für den zweiten Term, der Korrektur für die Arbeit gegen den Außendruck bei Insertion des Solvats in das Solvens, konnte kürzlich gezeigt werden, dass dieser aufgrund der Ensembleunabhängigkeit der Gibbs-Energie der Solvation entfällt, wenn die Insertion bei konstantem Volumen durchgeführt wird.^[110] Damit verbleibt

$$\mu_i^0 \approx -\beta^{-1} \ln \left(\frac{\eta_i O_i Z_{N,i}(V,T)}{c_i^0 \sigma_i Z_{N,0}(V,T)} \right). \quad 3.4.2$$

Das Verhältnis der Konfigurationsintegrale in Gl. 3.4.2 kann alternativ durch das sog. Potential der mittleren Kraft (engl. „*potential of mean force*“, PMF) $w_i(\Gamma_i)$ der Konformation Γ_i der Solvatspezies i ausgedrückt werden. Es ist formal definiert als diejenige Energie, die nötig ist, um ein Atom oder Molekül in einem Fluid an einen bestimmten Ort oder, wie im vorliegenden Fall, in eine bestimmte Konformation zu bringen. Es gilt

$$\frac{Z_{N,i}(V,T)}{Z_{N,0}(V,T)} = \int d\Gamma_i \exp(-\beta w_i(\Gamma_i)), \quad 3.4.3$$

und damit

$$\mu_i^0 = -\beta^{-1} \ln \left(\frac{\eta_i O_i}{c_i^0 \sigma_i} \int d\Gamma_i \exp(-\beta w_i(\Gamma_i)) \right). \quad 3.4.4$$

Die Chemischen Standardpotentiale sind mit der Reaktions-Gibbs-Energie über

$$\Delta_{\text{R}} G^0 = \sum_i \nu_i \mu_i^0, \quad 3.4.5$$

verbunden, ν_i sind stöchiometrische Faktoren. Für die Standard-Reaktions-Gibbs-Energien der beiden im vorliegenden Kontext wichtigen Prozesse Tautomerisierung- und Protonen-Dissoziationsreaktion gelten dann

$$\Delta_{\text{Tauto}} G_{\text{A} \rightarrow \text{B}}^0 = \mu_{\text{B}}^0 - \mu_{\text{A}}^0 = -\beta^{-1} \ln \left(\frac{\int d\Gamma_{\text{B}} \exp(-\beta w_{\text{B}}(\Gamma_{\text{B}}))}{\int d\Gamma_{\text{A}} \exp(-\beta w_{\text{A}}(\Gamma_{\text{A}}))} \right) \quad 3.4.6$$

und

$$\Delta_{\text{Diss}} G_{\text{AH} \rightarrow \text{A}^- + \text{H}^+}^0 = \mu_{\text{A}^-}^0 + \mu_{\text{H}^+}^0 - \mu_{\text{AH}}^0 = -\beta^{-1} \ln \left(\frac{\int d\Gamma_{\text{A}^-} \exp(-\beta w_{\text{A}^-}(\Gamma_{\text{A}^-}))}{\int d\Gamma_{\text{AH}} \exp(-\beta w_{\text{AH}}(\Gamma_{\text{AH}}))} \right) + \mu_{\text{H}^+}^0, \quad 3.4.7$$

sofern sich die Symmetrie der beteiligten Solvatspezies während der Reaktion nicht ändert, wir für alle Solvatspezies mit Ausnahme des Protons den gleichen Standardzustand verwenden und das Proton als punktförmiges Teilchen ohne innere Struktur betrachten. Offensichtlich lassen sich also die Standard-Reaktions-Gibbs-Energien allein auf Temperatur, die PMFs der unterschiedlichen Spezies und im Fall der Dissoziationsreaktion nach Gl. 3.4.4 die Standardkonzentration des Protons zurückführen. Da sich die Standardzustände der Solvatmoleküle aufheben, kann hier der Zustand unendlicher Verdünnung gewählt werden, was konsistent mit der Perspektive der Solvatationsmodelle ist.

Zu Gl. 3.4.1-3.4.7 ist anzumerken, dass, wenn wie von Borgis postuliert^[111] das Einbringen des Solvats in das Solvens bei konstantem Druck und der dazu äquivalenten Volumenänderung durchgeführt wird, ein Volumenkorrekturterm addiert werden müsste.^[110]

Als Standardzustand des Protons wird, wie üblich in Studien zur Bestimmung von $\mu_{\text{H}^+}^0$ bzw. $G_{\text{H}^+}^0$,^[112,113] eine Lösung mit einer Konzentration von 1 M angenommen. Damit sind die durchgeführten Berechnungen näherungsweise konsistent zur pH-Definition der IUPAC, wenn man davon absieht, dass diese streng genommen eine Molalität von 1 mol/kg spezifiziert.^[114] Diese ist nicht direkt in eine Konzentration umrechenbar, da die zum Umrechnen benötigte Dichte der Lösung auch von der Art der Gegenionen abhängt, die hier aber nicht berücksichtigt werden. Verwendet man die Dichte von reinem Wasser bei 298.15 K von ca. 0.997 kg/L,^[115] was bei stark verdünnten Lösungen eine gute Approximation sein dürfte, erhält man 1.003 M. Generell nehmen wir an, dass die Anzahl der Protonen und ihrer Gegenionen immer klein genug ist, dass die Solvatmoleküle von ihnen nicht gestört werden, was bei extrem großen oder kleinen pH-Werten möglicherweise nicht mehr der Fall ist.

Der Ausdruck „Potential der mittleren Kraft“ impliziert bereits, dass das PMF durch ein Kurvenintegral entlang eines beliebigen Pfades durch das Vektorfeld einer mittleren Kraft erhalten werden kann, wobei der Startpunkt durch einen sinnvoll zu wählenden Referenzzustand, und der Endpunkt durch die Konformation des Moleküls Γ_i gegeben ist. Die Frage nach der Integrationskonstante ist äquivalent zur Frage nach dem Referenzzustand. Da im Kontext dieser Arbeit quantenchemische Rechnungen in Lösung durchgeführt werden sollen, erscheint es sinnvoll, als Referenzzustand eine Situation zu wählen, in der sich alle Atomkerne und Elektronen des Solvats im Vakuum in unendlicher Entfernung zueinander befinden. In diesem Fall wäre der Lösungsprozess dann in drei Teile zerlegbar:

1. Im Vakuum wird aus den Elektronen und Kernen das zukünftige Solvat-Molekül erzeugt, wobei die Energie E_{vac} frei wird.
2. Das Solvat wird in das Solvens eingebracht und passt die Lage der Kerne und die Elektronenhülle der neuen Umgebung an, was die Reorganisationsenergie $\Delta E_{\text{reorg}} = E_{\text{Sol}} - E_{\text{vac}}$ benötigt, die nichts anderes als die Differenz der quantenmechanischen Energien in Lösung und im Vakuum ist.
3. Die Präsenz des Solvats stört die Struktur des homogenen Solvens. Die damit assoziierte Energie ist das chemische Exzesspotential μ^{ex} .

Damit könnte das PMF als

$$w_i(\Gamma_i) = E_{i,\text{Vac}}(\Gamma_{i,\text{Vac}}) + \Delta E_{\text{reorg}}(\Gamma_{i,\text{Vac}}, \Gamma_i) + \mu_i^{\text{ex}}(\Gamma_i) = E_{i,\text{Sol}}(\Gamma_i) + \mu_i^{\text{ex}}(\Gamma_i) \quad 3.4.8$$

ausgedrückt werden. Für ein einzelnes Proton entfallen die Schritte 1 und 2, da es keine innere Struktur besitzt, und es gilt $w_{\text{H}^+} = \mu_{\text{H}^+}^{\text{ex}}$.

An dieser Stelle wird darauf hingewiesen, dass einige der im Kontext der Lösungsmittelmodelle diskutierten Größen schwer gegeneinander abgrenzbar sind oder in der Literatur nicht einheitlich verwendet werden. Zum einen ist das PMF im vorliegenden Fall gleich der Gibbs-Energie in Lösung für eine gegebene Konformation. Zum anderen wird häufig nicht zwischen dem chemischen Exzesspotential und der Gibbs-Energie der Solvation unterschieden, was nur gerechtfertigt ist, wenn sich die innere Struktur des Solvats beim Lösungsvorgang nicht ändert und der Grenzfall unendlicher Verdünnung vorliegt, für den wegen

$$\mu_i = \left(\frac{\partial G}{\partial N_i} \right)_{p,T,N_{j \neq i}} \approx \left(\frac{G_{N,i} - G_{N,0}}{1} \right)_{p,T,N_{j \neq i}} = G_i \quad 3.4.9$$

Chemische Potentiale und Gibbs-Energien gleich sind.^[116] $G_{N,i}$ und $G_{N,0}$ sind hierbei die Gibbs-Energien der Lösung bzw. des reinen Solvens. Damit können die Reaktions-Gibbs-Energien in den Gleichungen 3.4.6 und 3.4.7 auch äquivalent mit den Gibbs-Energien als $\Delta_{\text{Tauto}} G_{A \rightarrow B}^0 = G_B^0 - G_A^0$ und $\Delta_{\text{Diss}} G_{\text{AH} \rightarrow \text{A}^- + \text{H}^+}^0 = G_{\text{A}^-}^0 + G_{\text{H}^+}^0 - G_{\text{AH}}^0$ formuliert werden. In diesem Zusammenhang ist wichtig, dass die in dieser Arbeit angegebenen Gibbs-Energien der Solvatspezies den in Gl. 3.4.4 enthaltenen Vorfaktor $\eta_i O_i / c_i^0 \sigma_i$ nicht berücksichtigen, da er für die hier betrachteten Prozesse wie gezeigt keine Rolle spielt. Dies dürfte, auch wenn es nicht explizit erwähnt wird, in den meisten einschlägigen Publikationen ähnlich sein. Damit ist

$$G_i^0 = -\beta^{-1} \ln \left[\int d\Gamma_i \exp(-\beta w_i(\Gamma_i)) \right]. \quad 3.4.10$$

In der Praxis stellt sich die Frage, wie das Integral in Gl. 3.4.10 zu evaluieren ist. Die einfachste Variante besteht darin, eine einzelne starre Konformation anzunehmen. In diesem Fall wird $G_i^0 = w_i$. Dabei wird auch der Anteil der Gibbs-Energie auf Grund der Schwingungsentropie vernachlässigt, der jedoch z.B. durch die Annahme eines harmonischen Potentials approximiert werden kann. Wie sich noch zeigen wird, liegen aber häufig mehrere Konformationen im bei Zimmertemperatur gut erreichbaren Energiebereich. Dann wird durch die „*Predominant States*“-Näherung von Gilson und Mitarbeitern^[108] aus Gl. 3.4.10

$$G_i^0 = -\beta^{-1} \ln \left[\sum_j \exp(-\beta w_i(\Gamma_{ij})) \right], \quad 3.4.11$$

wobei das Integral einer Summe gewichen ist, die über eine endliche Anzahl von Gruppen von Konformationen j läuft, die nach einem festzulegenden Ähnlichkeitsmaß zusammengestellt wurden.

Ein weiterer diskussionswürdiger Punkt betrifft die Wahl des Wegs vom Referenz- zum Endzustand für den Fall nettogeladener Solvatmoleküle. Grundsätzlich sind zwei Wege von Bedeutung: Der physikalische Prozess, bei dem das Solvat die Oberfläche durchtritt, sowie ein „*alchemischer*“ Prozess, bei dem die Wechselwirkung zwischen Solvat und Solvens erst im inneren des Solvens eingeschaltet wird, und der in dieser Arbeit implizit verwendet wird. Damit beide Prozesse tatsächlich äquivalent werden, muss im zweiten Fall zusätzlich ein Energiebeitrag berücksichtigt werden, der den

Durchtritt durch eine mit dem Galvanipotential φ_G polarisierte Solvat-Grenzschicht berücksichtigt. Die Gibbs-Energie wird dann zu

$$G_i^0 = -\beta^{-1} \ln \left[\sum_j \exp(-\beta w_i(\Gamma_{ij})) \right] + \varphi_G q_{\text{tot}}, \quad 3.4.12$$

wobei der erste Term in der Literatur auch als der „intrinsische“ Anteil bezeichnet wird. Für Moleküle ohne Nettoladung $q_{\text{tot}} = \sum_l q_l$ verschwindet der zweite Term. Das Galvanipotential wurde von Misin, Fedorov und Palmer für 3D-RISM und die Wassermodelle cSPC/E und cTIP3P^[117] bestimmt.^[118] Sie erhielten -13.43 kcal/(e mol) für cSPC/E und -12.55 kcal/(e mol) für cTIP3P. Setzt man Gl. 3.4.12 in Gl. 3.4.6 und Gl. 3.4.7 ein und beachtet die Ladungserhaltung während der Reaktionen, so wird für die Dissoziation wie für die Tautomerisierung deutlich, dass das Galvanipotential verschwindet und keinen Einfluss auf die Lage der Konformerengleichgewichte besitzt. Dementsprechend sind in dieser Arbeit berechneten Gibbs-Energien sofern nicht explizit anderweitig angegeben intrinsische Gibbs-Energien nach Gl. 3.4.11.

3.4.2 Tautomeren- und Dissoziationsgleichgewichte

Die Gleichgewichtskonstanten K von chemischen Reaktionen hängen mit der Standard-Reaktions-Gibbs-Energie über

$$K = \exp(-\beta \Delta_R G^0) \quad 3.4.13$$

zusammen. Damit ergibt sich unmittelbar für das Verhältnis zweier Tautomere A und B mit Aktivitäten a_A und a_B

$$K_T = \frac{a_A}{a_B} = \exp(-\beta \Delta_R G_{A \rightarrow B}^0), \quad 3.4.14$$

Für die Dissoziationskonstante einer Deprotonierungsreaktion der Säure AH gilt analog

$$K_a = \frac{a_{A^-} a_{H^+}}{a_{AH}} = \exp(-\beta \Delta_R G_{AH \rightarrow A^- + H^+}^0). \quad 3.4.15$$

Damit lässt sich wegen

$$\text{p}K_a = -\lg(K_a) = \frac{\beta \Delta_R G_{AH \rightarrow A^- + H^+}^0}{\ln(10)} = \frac{\beta(G_{A^-}^0 + G_{H^+}^0 - G_{AH}^0)}{\ln(10)} \quad 3.4.16$$

der $\text{p}K_a$ -Wert aus den Standard-Gibbs-Energien von Säure, konjugierter Base und Proton bestimmen. Problematisch ist hierbei die Gibbs-Energie des Protons, die derzeit als $G_{H^+}^0 = -265.9$ kcal/mol

angegeben wird ^[112,119], aber nicht mit hinreichender Genauigkeit bekannt ist. Ho und Coote geben einen Überblick über die derzeit verwendeten Werte.^[120] Dieses Problem kann umgangen werden, indem man ausgehend von einer Referenzsubstanz mit bekanntem pK_a -Wert nur Differenzen ΔpK_a berechnet, wobei die Gibbs-Energie des Protons entfällt. Alternativ kann die Gibbs-Energie des Protons durch eine statistische Anpassung an mehrere Referenzen bestimmt werden. Dafür bieten sich zwei Möglichkeiten. $G_{H^+}^0$ kann durch eine Minimierung der quadratisch gemittelten Abweichung entweder der berechneten pK_a -Werte oder der Dissoziations-Gibbs-Energie von experimentellen Referenzwerten $pK_{a,i,ref}$ oder $\Delta_R G_{AH \rightarrow A^- + H^+, i, ref}^0$

$$\Delta_R G_{AH \rightarrow A^- + H^+, i, ref}^0 = \beta^{-1} \ln(10) pK_{a,i,ref} \quad 3.4.17$$

jeweils als Funktion der Gibbs-Energie des Protons erhalten werden. Demzufolge gilt

$$\begin{aligned} G_{H^+}^0 &= \arg \min_{G_{H^+}^0} \left(\sqrt{\sum_i (pK_{a,i}(G_{H^+}^0) - pK_{a,i,ref})^2} \right) \\ &= \arg \min_{G_{H^+}^0} \left(\sqrt{\sum_i (G_{A_i}^0 + G_{H^+}^0 - G_{AH_i}^0 - \Delta_R G_{AH \rightarrow A^- + H^+, i, ref}^0)^2} \right). \end{aligned} \quad 3.4.18$$

Zum anderen können über eine Regressionsanalyse lineare Modelle

$$pK_a = A \frac{\beta(G_{A^-}^0 - G_{AH}^0)}{\ln(10)} + B \quad 3.4.19$$

eingeführt werden, bei denen der in Gl. 3.4.16 gezeigte lineare Zusammenhang zwischen der Differenz der Gibbs-Energien der dissoziierten und nicht dissoziierten Spezies mit einer möglicherweise stoffklassenabhängigen Steigung A modifiziert wird. Die Standard-Gibbs-Energie des Protons steht dann mit der Abszisse B über

$$B \approx \frac{\beta G_{H^+}^0}{\ln(10)} \quad 3.4.20$$

in Verbindung.

3.4.3 pH-abhängige Speziesverteilung in Lösung

Folgt man der Henry- oder Solvat-Konvention für die Aktivitätskoeffizienten, so können im Grenzfall unendlicher Verdünnung die Aktivitätskoeffizienten der Solvatspezies in Gl. 3.4.14 und 3.4.15 durch Stoffmengenanteile ersetzt werden. Zusammen mit einer Bilanzgleichung

$$1 = \sum_{m=0}^n x_m \quad 3.4.21$$

lassen sich dann Ausdrücke für die Stoffmengenanteile x_m der Ionisierungszustände in Abhängigkeit der Protonenaktivität bzw. des pH-Werts herleiten^[121], und zwar auch für Solvate mit mehreren titrierbaren Zentren. Dazu gehen wir von einer Substanz AH_{n-m} mit n titrierbaren Zentren aus, von denen $n-m$ besetzt sind, weil m Protonen fehlen. m ist dementsprechend eine Zahl zwischen 0 und n und kann zur Indizierung der Ionisationszustände genutzt werden, da n konstant ist. Die zur Dissoziation des Protons m , also der Reaktion $AH_{n-m} \rightarrow AH_{n-m-1} + H^+$, gehörende Gleichgewichtskonstante sei

$$K_{a,m} = \frac{a_{H^+} a_{m-1}}{a_m} = \frac{a_{H^+} x_{m-1}}{x_m}. \quad 3.4.22$$

Die Stoffmengenanteile für Substanzen mit einem oder zwei titrierbaren Zentren werden häufig in einschlägigen Lehrbüchern angegeben.^[121] Bei Substanzen mit einem titrierbaren Zentrum HA sind sie

$$x_0 = x_{HA} = \frac{1}{1 + (a_{H^+})^{-1} K_{a,0}} \quad 3.4.23$$

und

$$x_1 = x_{A^-} = \frac{(a_{H^+})^{-1} K_{a,0}}{1 + (a_{H^+})^{-1} K_{a,0}} \quad 3.4.24$$

Diese beiden Gleichungen lassen sich leicht durch Logarithmieren und Umstellen in die bekannte Henderson-Hasselbalch-Gleichung überführen. Bei Substanzen mit zwei titrierbaren Zentren H_2A erhält man

$$x_0 = x_{H_2A} = \frac{1}{1 + (a_{H^+})^{-1} K_{a,0} + (a_{H^+})^{-2} K_{a,0} K_{a,1}}, \quad 3.4.25$$

$$x_1 = x_{HA^-} = \frac{(a_{H^+})^{-1} K_{a,0}}{1 + (a_{H^+})^{-1} K_{a,0} + (a_{H^+})^{-2} K_{a,0} K_{a,1}}, \quad 3.4.26$$

und

$$x_2 = x_{A^{2-}} = \frac{(a_{H^+})^{-2} K_{a,0} K_{a,1}}{1 + (a_{H^+})^{-1} K_{a,0} + (a_{H^+})^{-2} K_{a,0} K_{a,1}}. \quad 3.4.27$$

Eine Verallgemeinerung auf Substanzen mit n titrierbaren Zentren ist durch n Gleichungen der Form

$$x_m = \frac{(a_{\text{H}^+})^{-m} \prod_{i=0}^m K_{a,i-1}}{\sum_{s=0}^n (a_{\text{H}^+})^{-s} \prod_{l=0}^s K_{a,l-1}}, \quad 3.4.28$$

bzw. mit den üblichen logarithmischen Größen

$$x_m = \frac{10^{m \text{ pH}} \prod_{i=0}^m 10^{-\text{p}K_{a,i-1}}}{\sum_{s=0}^n 10^{s \text{ pH}} \prod_{l=0}^s 10^{-\text{p}K_{a,l-1}}} \quad 3.4.29$$

gegeben, wobei wir voraussetzen, dass $\prod_{i=0}^0 f = 1$ für beliebige Funktionen f gilt.

Bei Substanzen mit einer Vielzahl titrierbarer Zentren ist jedoch zu beachten, dass zu den Ionisierungszuständen mehrere Tautomere existieren können, die sich nicht darin unterscheiden, wie viele Zentren mit Protonen besetzt sind, sondern nur darin, welche Zentren protoniert sind. Jedes Tautomer kann wiederum in unterschiedlichen Konformationen vorliegen. Die Gibbs-Energie eines Ionisierungszustands m in Lösung, der alle diese Konformationen c und Tautomere t beinhaltet, wäre^[122–124]

$$\begin{aligned} G_m &= -\beta^{-1} \ln(Z_m) \\ &= -\beta^{-1} \ln\left(\sum_t g_{mt} \exp(-\beta(G_{mt}))\right) \\ &= -\beta^{-1} \ln\left(\sum_t g_{mt} Z_{mt}\right) \\ &= -\beta^{-1} \ln\left(\sum_t g_{mt} \sum_c g_{mtc} \exp(-\beta(w_{mtc}))\right). \end{aligned} \quad 3.4.30$$

Hierbei sind

$$Z_m = \sum_t g_{mt} \exp(-\beta G_{mt}) \quad 3.4.31$$

die kanonische Zustandssumme des Ionisierungszustands m und

$$Z_{mt} = \sum_c g_{mtc} \exp(-\beta w_{mtc}) \quad 3.4.32$$

die kanonische Zustandssumme des Tautomers t im Ionisierungszustand m . g_{mt} und g_{mtc} bezeichnen Entartungsfaktoren.

Wir sehen, dass sich die Gibbs-Energie des Ionisierungszustands in einzelne Beiträge für die Tautomere zerlegen lässt, die sich wiederum aus Einzelbeiträgen für die Konformationen zusammensetzen. Den Tautomeren kann also auch wieder eine Gibbs-Energie

$$G_{mt} = -\beta^{-1} \ln(Z_{mt}) \quad 3.4.33$$

zugeordnet werden. In einer fiktiven Lösung von Tautomeren nur des Ionisierungszustands m wäre dann der Stoffmengenanteil eines Tautomers t

$$x_{t|m} = \exp(-\beta G_{mt}) / Z_m. \quad 3.4.34$$

Damit ist der Stoffmengenanteil des Tautomers t in der Lösung

$$x_{mt} = x_m x_{t|m}, \quad 3.4.35$$

weil die Stoffmengenanteile auch als Wahrscheinlichkeiten für das Auffinden eines bestimmten Zustands interpretiert werden können, und sich eine Gesamtwahrscheinlichkeit multiplikativ aus den Wahrscheinlichkeiten einzelner Teilereignisse zusammensetzt. Die Bestimmung des Stoffmengenanteils einer bestimmten Konformation erfolgt analog als

$$x_{mtc} = x_m x_{t|m} x_{c|t|m}, \quad 3.4.36$$

mit

$$x_{c|t|m} = \exp(-\beta w_{mtc}) / Z_{mt}. \quad 3.4.37$$

Die hier vorgestellte Methodologie wird in den Kapiteln 3.8 und 3.9 verwendet, um Speziesverteilungsdiagramme für Histamin und Pteroinsäure zu berechnen.

3.5 Kraftfeld-Geometrieoptimierung mit AMBER/3D-RISM

Als Vorstufe zu einer 3D-RISM/QM/MM-Geometrieoptimierung wurde ein Programm zur 3D-RISM-Geometrieoptimierung mit dem AMBER-Kraftfeld entwickelt und am Protein Thioredoxin getestet. Ein solches QM/MM-Verfahren würde im Vergleich zu der im Weiteren verwendeten PCM-Methodik einen deutlich konsistenteren Ansatz darstellen. An der Thioredoxin-Struktur aus der Protein Datenbank (1XOA) wurden mit Hilfe des Leap-Programms aus der AMBERTOOLS-Sammlung Wasserstoffe hinzugefügt und Geometrieoptimierungen mit dem Quasi-Newton-Verfahren Broyden-Fletcher-Goldfarb-Shanno (BFGS) durchgeführt. Quasi-Newton-Verfahren bieten sich hier besonders an, da sie im Gegensatz zum Newton-Verfahren die Hesse-Matrix nicht explizit benötigen, sondern Sie auf der Basis von Funktionswerten und Gradienten aus vorhergehenden Iterationen approximieren.

Weiterhin wurden das Reziprokraum-Verfahren, PME mit Lagrange-Polynomen 10. Ordnung, 30 Å Lösungsmittelpuffer verwendet. In jeder Geometrieoptimierungsiteration wurde bis auf 10^{-6} in der Maximumnorm der direkten Korrelationsfunktion konvergiert. Als *Closures* kamen PSE-1 und PSE-2 zum Einsatz, jeweils mit Gittern mit einer Maschenweite von 0.3 Å, 0.45 Å und 0.6 Å. Zusätzlich wurde für PSE-2 eine Optimierung mit einer Maschenweite von 0.6 Å mit 0.45 Å und 0.3 Å fortgesetzt. Diese Optimierung wird im Folgenden mit 0.6→0.45→0.3 bezeichnet.

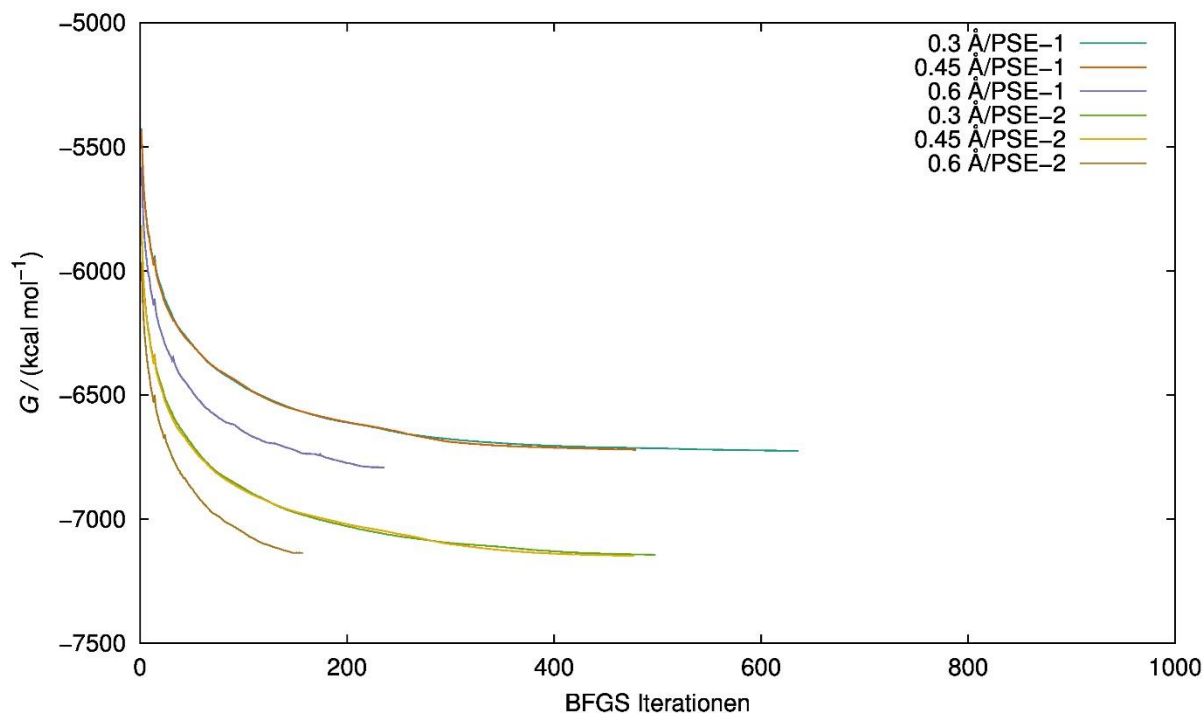


Abbildung 13: Verlauf der Gibbs-Energie während einer AMBER/3D-RISM-Geometrieoptimierung von Thioredoxin mit dem BFGS-Verfahren für verschiedene Gitterauflösungen und PSE-Closure-Ordnungen.

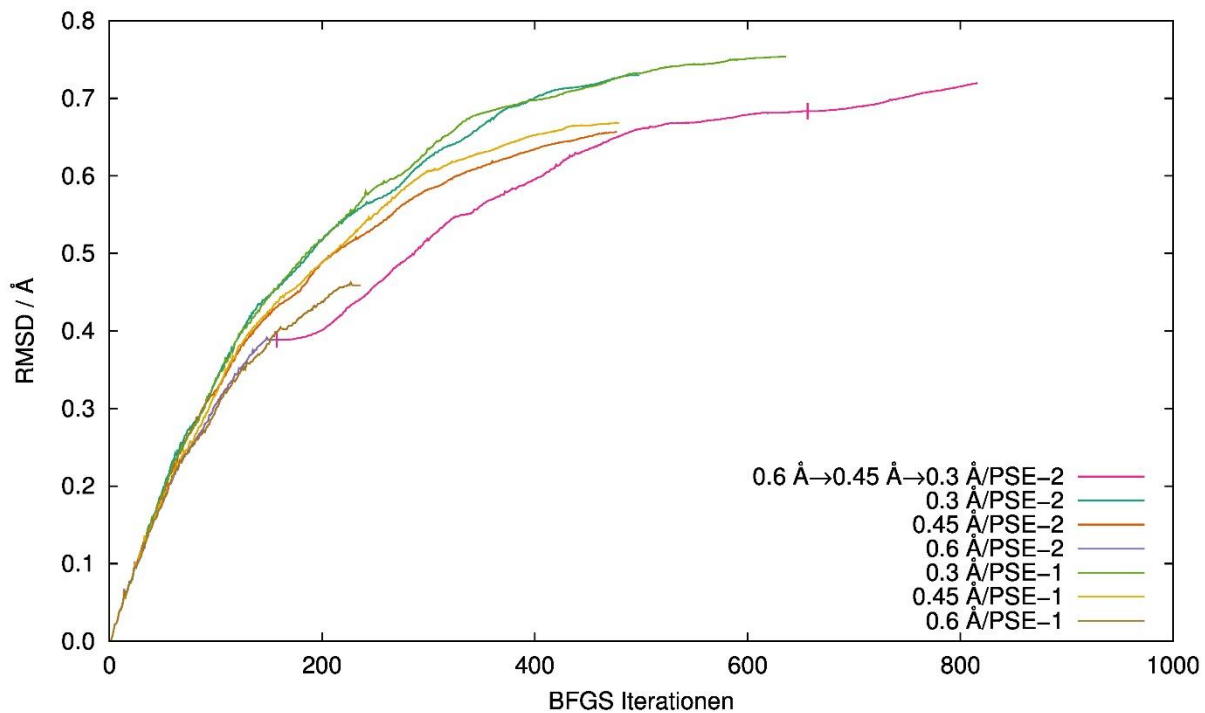


Abbildung 14: Auftragung der quadratisch gemittelten Abweichung von den ursprünglichen Atomkoordinaten im Verlauf von AMBER/3D-RISM-Geometrieoptimierungen von Thioredoxin mit dem BFGS-Verfahren für verschiedene Gitterauflösungen und PSE-Closure-Ordnungen. Die rosafarbige Line zeigt das Ergebnis einer Optimierung, die zunächst mit einem Gitter mit einer Maschenweite von 0.6 Å, dann 0.45 Å und schließlich 0.3 Å konvergiert wurde. Hier zeigen die senkrechten Balken die Stellen, an denen auf das nächstfeinere Gitter umgeschaltet wurde.

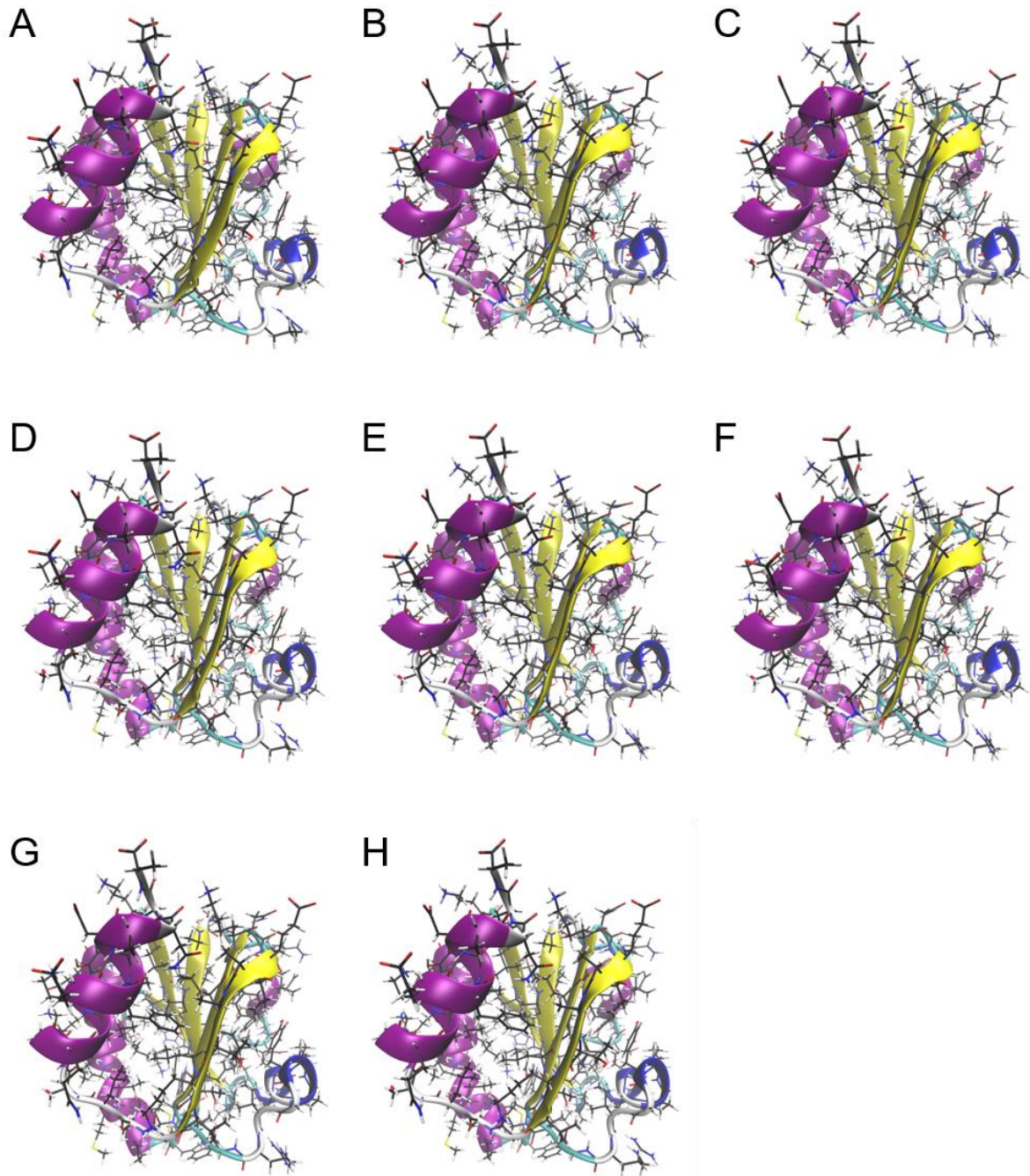


Abbildung 15: Thioredoxin-Strukturen **A**: vor der Optimierung, nach der Optimierung mit **B**: PSE-1/0.3 Å Maschenweite, **C**: PSE-1/0.45 Å, **D**: PSE-1/0.6 Å, **E**: PSE-2/0.6→0.45→0.3 Å, **F**: PSE-2/0.3 Å, **G**: PSE-2/0.45 Å, **H**: PSE-2/0.6 Å.

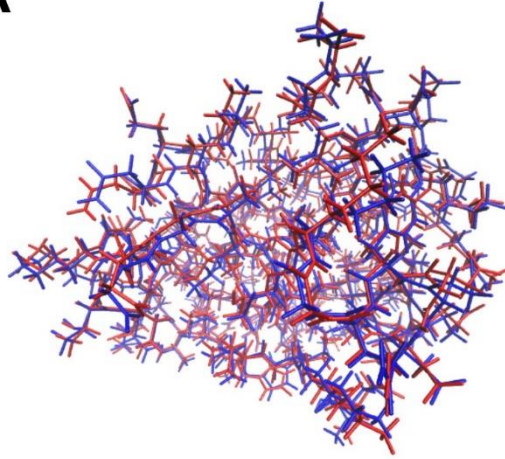
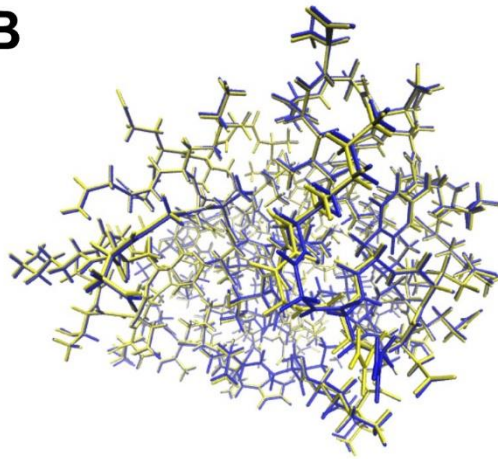
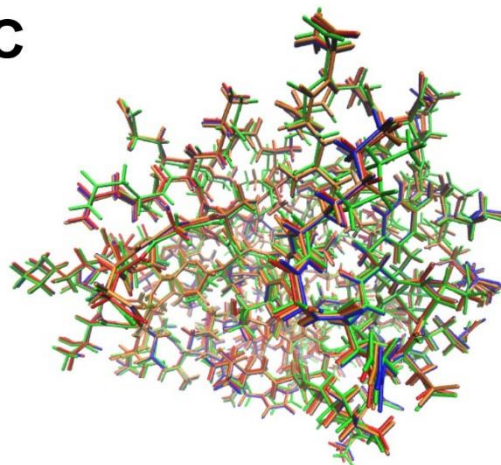
A**B****C**

Abbildung 16: Überlagerte Thioredoxinstrukturen **A**: vor der Optimierung (rot), PSE-2/0.3 Å (blau), **B**: PSE-1/0.3 Å (gelb), PSE-2/0.3 Å (blau), **C**: PSE-2 mit 0.3 Å (blau), 0.45 Å (rot), 0.6 Å (grün) und 0.6→0.45→0.3 Å (orange).

Die Verläufe der Gibbs-Energien und des RMSD-Wertes werden in Abbildung 13 und Abbildung 14 gezeigt. Man erkennt, dass die Gibbs-Energien für 0.3 Å und 0.45 Å sehr ähnlich sind, aber für 0.6 Å recht weit abweichen. Ein ähnliches Bild ergibt sich für die RMSD-Werte, die jedoch allesamt nur geringe Abweichungen von der Ausgangsstruktur zeigen. Dementsprechend sollte man erwarten, dass die Sekundärstrukturelemente des Proteins stabil bleiben. Wie Abbildung 15 zeigt, ist das auch der Fall. Generell lassen sich die Abweichungen zwischen der Ausgangsstruktur und den optimierten Strukturen optisch nicht auf den ersten Blick, sondern nur bei genauem Betrachten ausmachen. Das gilt in noch stärkerem Maße für die optimierten Strukturen untereinander. Dementsprechend sind in Abbildung 16 die Strukturen überlagert dargestellt. Hier wird erkennbar, dass auch die Lage der Seitenketten nur geringen Abweichungen unterliegt.

Die Optimierung mit sukzessive abnehmenden Maschenweiten führt zu deutlich weniger Iterationen mit dem feinsten Gitter, während jedoch die Gesamtzahl an Iterationen deutlich größer wird. Ob der Gesamtaufwand durch die Voroptimierung mit einem gröberem Gitter reduziert wird und der manuelle Mehraufwand gerechtfertigt ist, dürfte vom Einzelfall abhängen.

Tabelle 3 zeigt paarweise RMSD-Werte für die Startstruktur und die optimierten Strukturen. Man erkennt, dass Strukturen, die mit dem groben 0.6 Å-Gitter erhalten wurden, verglichen mit den Ergebnisstrukturen der feineren Gitter deutlich näher an der Ausgangsstruktur bleiben, während diese untereinander recht ähnlich sind, sich jedoch deutlich von der Ausgangsstruktur deutlicher unterscheiden. Die mit PSE-1 und PSE-2 bei gleicher Maschenweite erhaltenen Strukturen sind recht ähnlich. Zusammenfassend lässt sich sagen, dass die hier erhaltenen Ergebnisse darauf hindeuten, dass für qualitative Zwecke eine Optimierung mit PSE-1/0.6 Å hinreichend sein sollte. Für eine strukturelle quantitative Analyse sollte ein Gitter von 0.45 Å verwendet werden, für eine quantitative energetische Analyse kann der Mehraufwand für 0.3 Å gerechtfertigt werden. Die hier durchgeführte Untersuchung müsste jedoch noch zwecks weiterer Verifizierung auf eine Reihe anderer Makromoleküle ausgedehnt werden. Insofern sind die hier gezeigten Ergebnisse als vorläufige Hypothesen zu betrachten.

Tabelle 3: RMSD-Werte der Ergebnis-Strukturen von AMBER/3D-RISM-Geometrieoptimierungen von Thioredoxin untereinander, zur besseren Übersicht farblich hervorgehoben ($0 \text{ \AA} < \text{RMSD} \leq 0.2 \text{ \AA}$: grün, $0.2 \text{ \AA} \leq \text{RMSD} \leq 0.4 \text{ \AA}$ blau, $0.4 \text{ \AA} \leq \text{RMSD} \leq 0.6 \text{ \AA}$ pink, $\text{RMSD} > 0.6 \text{ \AA}$ orange). Die Struktur auf die sich die mit dem Stern markierte Reihe/Spalte bezieht wurde zunächst mit einer Maschenweite von 0.6 \AA , dann 0.45 \AA und schließlich 0.3 \AA optimiert. „Start“ bezeichnet die Ausgangsstruktur aus der Protein Databank mit mittels AMBER hinzugefügten Wasserstoffatomen. Alle Angaben in der Tabelle sind in Ångström.

		Start	PSE-1			PSE-2			
			0.3	0.45	0.6	0.3	0.45	0.6	0.3*
	Start	0.00	0.75	0.67	0.46	0.73	0.66	0.39	0.72
PSE-1	0.3	0.75	0.00	0.23	0.52	0.18	0.26	0.58	0.19
	0.45	0.67	0.23	0.00	0.40	0.23	0.18	0.48	0.21
	0.6	0.46	0.52	0.40	0.00	0.49	0.42	0.20	0.48
PSE-2	0.3	0.73	0.18	0.23	0.49	0.00	0.22	0.55	0.12
	0.45	0.66	0.26	0.18	0.42	0.22	0.00	0.48	0.18
	0.6	0.39	0.58	0.48	0.20	0.55	0.48	0.00	0.54
	0.3*	0.72	0.19	0.21	0.48	0.12	0.18	0.54	0.00

3.6 *ab initio* pK_a-Vorhersage kleiner organischer Moleküle

Seit einiger Zeit ist die Vorhersage der Azidität kleiner organischer Moleküle eine wichtige Anwendung der Quantenchemie.^[120,125] Dabei wird eine Vielzahl unterschiedlicher Solvatationsmodelle aus der Klasse der dielektrischen Kontinuumsmodelle verwendet, die teilweise durch das Hinzufügen expliziter Wassermoleküle erweitert wurden.^[126,127] Typischerweise wird für die Berechnung der pK_a-Werte als QSPR-Ansatz ein lineares Modell nach 3.4.19 und 3.4.20 eingeführt, bei dem die Gibbs-Energie des Protons nach als Anpassungsparameter auftritt. Im Fall einer exakten Berechnung der Gibbs-Energien der Solvatspezies und des Protons sollte die Steigung *A* den Wert eins erreichen. Stattdessen stellt man jedoch fest, dass bei pK_a-Berechnungen mit Kontinuumsmodellen der Faktor in der Regel deutlich kleiner als eins ist. Bei Einbeziehung expliziter Wassermoleküle wird der Faktor deutlich größer,^[126,127] was aber mit einem erhöhten Rechenaufwand erkauft werden muss. Außerdem ist nicht eindeutig, auf welche Art und Weise die Wassermoleküle zu platzieren sind und wie viele eingesetzt werden müssen. Derartige Methoden erscheinen daher sowohl konzeptionell als auch bezogen auf den zusätzlich nötigen Rechenaufwand eher wenig attraktiv.

Da EC-RISM einen Mittelweg zwischen den Kontinuumsmodellen und dem Hinzufügen expliziter Wassermoleküle bietet, stellt sich die Frage, ob mit EC-RISM auch Steigungen nahe eins erhalten werden, und ob mit EC-RISM stärker prädiktive Modelle erstellt werden können. Zusätzlich besteht die Möglichkeit, innerhalb der Datensätze pK_a-Differenzen ΔpK_a zwischen jeweils zwei Substanzen *i* und *j* über

$$\Delta pK_{a,ij} = \frac{\beta(G_{A_j}^0 - G_{AH_j}^0 - (G_{A_i}^0 - G_{AH_i}^0))}{\ln(10)} = \frac{\beta\Delta\Delta_{\text{Diss}} G_{ij}}{\ln(10)} \quad 3.6.1$$

zu berechnen, womit der physikalische Realismus der Modelle ohne Beeinträchtigung durch die Frage des exakten Werts der Gibbs-Energie des Protons direkt überprüft werden kann.

Im Rahmen der Diplomarbeit des Autors^[53] wurden für eine Auswahl von Substanzklassen starrer Moleküle aus Arbeiten von Friesner und Mitarbeitern^[128] für jede Substanz mit MP2/6-311+G** auf Basis von B3LYP/6-311+G**-Vakuum-Geometrien jeweils mit PCM und EC-RISM mit PSE-3-Closure Gibbs-Energien und daraus pK_a -Verschiebungen berechnet. Für die Gruppe der Heterozyklen wurde auch eine aug-cc-pVDZ-Basis getestet. Zusätzlich wurden aus den pK_a -Modellen von Friesner^[128] Gibbs-Energien zurückgerechnet, und damit die Verschiebungen evaluiert. Falls Tautomere existieren wurden diese berücksichtigt und für einige Moleküle mit Hydroxyfunktionalität wurden mehrere Rotamere getestet. Weiterhin wurden für jede Substanzklasse lineare Regressionsmodelle erzeugt und bezüglich ihrer Güte evaluiert. Diese Untersuchungen werden hier auf weitere Kombinationen von MP2 und B3LYP mit unterschiedlichen Basissätzen und unter Verwendung des PCM-Solvatationsmodells berechneten Geometrien ausgedehnt. Die statistische Auswertung der Regressionsmodelle sowie quadratisch gemittelte Abweichungen für die pK_a -Differenzen sind in Tabelle 4 bis Tabelle 8 aufgeführt. Abbildung 17 und Abbildung 18 zeigen die Korrelation zwischen den berechneten pK_a -Werten und -Verschiebungen und den jeweiligen Referenzwerten aus Messungen. Zum Vergleich sind auf der Basis von Ergebnissen aus Ref. ^[128] zurückgerechnete Ergebnisse der Jaguar-SCRF-Methode angegeben. Die Arbeiten wurden bereits im Rahmen einer Diplomarbeit^[53] begonnen und hier fortgesetzt. Während in der Diplomarbeit außer für die Gruppe der Heterozyklen, für die auch die aug-cc-pVDZ-Basis getestet wurde, nur die Methode MP2(EC-RISM)/6-311+G(d,p)//MP2/6-311+G(d,p) in verschiedenen Varianten verwendet wurde, konnte die Anzahl der Kombinationen aus Methode und Basissatz hier deutlich erhöht werden. Desweiteren wurde der Einfluß einer Geometrieoptimierung unter Zuhilfenahme des PCM-Solvatationsmodells untersucht.

Tabelle 4: Statistische Auswertung einer Regressionsanalyse für pK_a -Daten von Pyrrolen. Für alle EC-RISM-Ergebnisse wurde ^a die C2-Protonierung, die nach experimentellen Befunden das globale Minimum-Tautomer darstellt,^[129] berücksichtigt. ^b Es wurde eine Zustandssumme über alle möglichen Protonierungsmuster gebildet. A , B , R^2 und RMSD^s , σ sind Steigung, Abszisse, Regressionskoeffizient und Standardabweichung, RMSD^s ist die quadratisch gemittelte Abweichung der pK_a -Verschiebungen über alle möglichen Substanz-Paare innerhalb einer Substanzklasse. Jaguar-SCRF-Ergebnisse wurden ausgehend von Daten aus Ref. ^[128] berechnet. EC-RISM ist hier durch ECR abgekürzt. Bereits im Rahmen der Diplomarbeit des Autors^[53] vorgestellte Berechnungsmethoden sind grau hervorgehoben.

Method	A	B	R^2	σ	RMSD^s
MP2(ECR)/6-311+G(d,p)//B3LYP/6-311+G(d,p) ^a	0.93	-175.72	0.98	0.29	0.38
MP2(ECR)/6-311+G(d,p)//B3LYP/6-311+G(d,p) ^b	0.94	-176.41	0.98	0.29	0.37
MP2(ECR)/6-311++G(d,p)//B3LYP/6-311++G(d,p) ^a	0.93	-175.92	0.98	0.29	0.39
MP2(ECR)/6-311++G(d,p)//B3LYP/6-311++G(d,p) ^b	0.94	-177.02	0.98	0.30	0.38
MP2(ECR)/aug-cc-pVDZ//B3LYP/6-311+G(d,p) ^a	0.91	-170.74	0.98	0.32	0.43
MP2(ECR)/aug-cc-pVDZ //B3LYP/6-311+G(d,p) ^b	0.92	-171.25	0.98	0.32	0.43
MP2(ECR)/6-311+G(d,p)//B3LYP(PCM)/6-311+G(d,p) ^a	0.92	-173.47	0.98	0.31	0.41
MP2(ECR)/6-311+G(d,p)//B3LYP(PCM)/6-311+G(d,p) ^b	0.92	-174.09	0.98	0.31	0.41
MP2(ECR)/6-311++G(d,p)//B3LYP(PCM)/6-311++G(d,p) ^a	0.92	-174.14	0.97	0.33	0.42
MP2(ECR)/6-311++G(d,p)//B3LYP(PCM)/6-311++G(d,p) ^b	0.93	-174.74	0.97	0.33	0.42
MP2(ECR)/6-311+G(d,p)//MP2(PCM)/6-311+G(d,p) ^a	0.93	-175.07	0.97	0.36	0.45
MP2(ECR)/6-311+G(d,p)//MP2(PCM)/6-311+G(d,p) ^b	0.93	-175.69	0.97	0.35	0.44
B3LYP(ECR)/6-311+G(d,p)//B3LYP(PCM)/6-311+G(d,p) ^a	0.80	-153.34	0.97	0.37	0.72
B3LYP(ECR)/6-311+G(d,p)//B3LYP(PCM)/6-311+G(d,p) ^b	0.80	-153.53	0.97	0.37	0.71
Jaguar-SCRF	0.60	-6.32	0.91	0.59	1.64

Tabelle 5: Statistische Auswertung einer Regressionsanalyse für pK_a -Daten von Indolen. Für alle EC-RISM-Ergebnisse wurde nur die C3-Protonierung berücksichtigt, die nach experimentellen Befunden das globale Minimum-Tautomer darstellt.^[130] A , B , R^2 und RMSD^s , σ sind Steigung, Abszisse, Regressionskoeffizient und Standardabweichung, RMSD^s ist die quadratisch gemittelte Abweichung der pK_a -Verschiebungen über alle möglichen Substanz-Paare innerhalb einer Substanzklasse. Jaguar-SCRF-Ergebnisse wurden ausgehend von Daten aus Ref. ^[128] berechnet. EC-RISM ist hier durch ECR abgekürzt. Bereits im Rahmen der Diplomarbeit des Autors^[53] vorgestellte Berechnungsmethoden sind grau hervorgehoben.

Method	A	B	R^2	σ	RMSD^s
MP2(ECR)/6-311+G(d,p)//B3LYP/6-311+G(d,p)	1.02	-192.90	0.94	0.55	0.64
MP2(ECR)/6-311++G(d,p)//B3LYP/6-311++G(d,p)	1.02	-191.78	0.93	0.58	0.67
MP2(ECR)/aug-cc-pVDZ//B3LYP(PCM)/6-311+G(d,p)	0.97	-181.32	0.93	0.61	0.71
MP2(ECR)/6-311+G(d,p)//B3LYP(PCM)/6-311+G(d,p)	1.00	-188.82	0.94	0.56	0.65
MP2(ECR)/6-311++G(d,p)//B3LYP(PCM)/6-311++G(d,p)	1.00	-187.83	0.93	0.59	0.68
MP2(ECR)/6-311+G(d,p)//MP2(PCM)/6-311+G(d,p)	1.02	-192.39	0.94	0.57	0.66
B3LYP(ECR)/6-311+G(d,p)//B3LYP(PCM)/6-311+G(d,p)	0.87	-165.88	0.98	0.35	0.57
Jaguar-SCRF	0.84	-7.88	0.99	0.23	0.58

Tabelle 6: Statistische Auswertung einer Regressionsanalyse für pK_a -Daten von Anilinen. A , B , R^2 , σ und RMSD^s sind Steigung, Abszisse, Regressionskoeffizient und Standardabweichung, RMSD^s ist die quadratisch gemittelte Abweichung der pK_a -Verschiebungen über alle möglichen Substanz-Paare innerhalb einer Substanzklasse. Jaguar-SCRF-Ergebnisse wurden ausgehend von Daten aus Ref. ^[128] berechnet. EC-RISM ist hier durch ECR abgekürzt. Bereits im Rahmen der Diplomarbeit des Autors^[53] vorgestellte Berechnungsmethoden sind grau hervorgehoben.

Methode	<i>A</i>	<i>B</i>	R^2	σ	RMSD^s
MP2(ECR)/6-311+G(d,p)//B3LYP/6-311+G(d,p)	0.74	-140.05	0.71	1.09	1.51
MP2(ECR)/6-311++G(d,p)//B3LYP/6-311++G(d,p)	0.73	-138.14	0.70	1.10	1.53
MP2(ECR)/aug-cc-pVDZ//B3LYP(PCM)/6-311+G(d,p)	1.00	-188.57	1.00	0.09	0.11
MP2(ECR)/6-311+G(d,p)//B3LYP(PCM)/6-311+G(d,p)	1.44	-275.15	1.00	0.14	0.77
MP2(ECR)/6-311++G(d,p)//B3LYP(PCM)/6-311+G(d,p)	1.45	-276.75	0.99	0.18	0.79
MP2(ECR)/6-311+G(d,p)//MP2(PCM)/6-311+G(d,p)	1.14	-215.76	1.00	0.09	0.32
B3LYP(ECR)/6-311+G(d,p)//B3LYP(PCM)/6-311+G(d,p)	0.68	-125.64	0.99	0.21	1.20
Jaguar-SCRF ^a	0.54	-2.90	0.97	0.35	2.12

Tabelle 7: Statistische Auswertung einer Regressionsanalyse für pK_a -Daten von Heterozyklen. ^a Werte ausgehend von Jaguar-SCRF- pK_a -Daten aus Ref. ^[128] berechnet. A , B , R^2 , σ und RMSD^s sind Steigung, Abszisse, Regressionskoeffizient und Standardabweichung, RMSD^s ist die quadratisch gemittelte Abweichung der pK_a -Verschiebungen über alle möglichen Substanz-Paare innerhalb einer Substanzklasse. Jaguar-SCRF-Ergebnisse wurden ausgehend von Daten aus Ref. ^[128] berechnet. EC-RISM ist hier durch ECR abgekürzt. Bereits im Rahmen der Diplomarbeit des Autors vorgestellte Berechnungsmethoden sind grau hervorgehoben.

Methode	<i>A</i>	<i>B</i>	R^2	σ	RMSD^s
MP2(ECR)/6-311+G(d,p)//B3LYP/6-311+G(d,p)	0.84	-157.34	0.89	0.83	1.29
MP2(ECR)/6-311++G(d,p)//B3LYP/6-311++G(d,p)	0.84	-157.78	0.89	0.83	1.29
MP2(ECR)/aug-cc-pVDZ//B3LYP(PCM)/6-311+G(d,p)	0.84	-157.74	0.91	0.73	1.16
MP2(ECR)/6-311+G(d,p)//B3LYP(PCM)/6-311+G(d,p)	0.85	-160.26	0.90	0.78	1.20
MP2(ECR)/6-311++G(d,p)//B3LYP(PCM)/6-311++G(d,p)	0.85	-160.81	0.90	0.78	1.20
MP2(ECR)/6-311+G(d,p)//MP2(PCM)/6-311+G(d,p)	0.86	-162.27	0.90	0.77	1.17
B3LYP(ECR)/6-311+G(d,p)//B3LYP(PCM)/6-311+G(d,p)	0.80	-150.33	0.90	0.79	1.35
Jaguar-SCRF	0.79	-6.03	0.99	0.28	0.94

Tabelle 8: Statistische Auswertung einer Regressionsanalyse für pK_a -Daten von Phenolen. ^a EC-RISM-Berechnungen mit einem modifizierten Lennard-Jones-Parameter σ von 3.325 Å (PCM-Strukturen) bzw. 3.21 Å (Vakuum-Strukturen) für den basischen Phenol-Sauerstoff. Jaguar-SCRF-Ergebnisse wurden ausgehend von Daten aus Ref. ^[128] berechnet. A , B , R^2 , σ und RMSD^s sind Steigung, Abszisse, Regressionskoeffizient und Standardabweichung, RMSD^s ist die quadratisch gemittelte Abweichung der pK_a -Verschiebungen über alle möglichen Substanz-Paare innerhalb einer Substanzklasse. EC-RISM ist hier durch ECR abgekürzt. Bereits im Rahmen der Diplomarbeit des Autors vorgestellte Berechnungsmethoden sind grau hervorgehoben.

Methode	A	B	R^2	σ	RMSD^s
MP2(ECR)/6-311+G(d,p)//B3LYP/6-311+G(d,p)	0.64	-117.07	0.13	1.28	1.72
MP2(EC-RISM)/6-311++G(d,p)//B3LYP/6-311++G(d,p)	0.66	-120.21	0.13	1.28	1.71
MP2(ECR)/6-311+G(d,p)//B3LYP/6-311+G(d,p) ^a	0.68	-127.64	0.67	0.80	1.26
MP2(ECR)/6-311+G(d,p)//B3LYP(PCM)/6-311+G(d,p) ^a	1.00	-194.99	0.99	0.15	0.20
MP2(ECR)/aug-cc-pVDZ //B3LYP(PCM)/6-311+G(d,p)	-1.19	236.88	0.29	1.16	2.34
MP2(ECR)/6-311+G(d,p)//B3LYP(PCM)/6-311+G(d,p)	-1.20	242.13	0.33	1.13	2.40
MP2(ECR)/6-311+G(d,p)//MP2(PCM)/6-311+G(d,p)	-0.95	192.75	0.07	1.33	1.99
MP2(ECR)/6-311+G(d,p)//B3LYP(PCM)/6-311+G(d,p)	-1.22	246.16	0.35	1.11	2.42
B3LYP(ECR)/6-311+G(d,p)//B3LYP(PCM)/6-311+G(d,p)	1.37	-250.79	0.95	0.31	0.63
Jaguar-SCRF	0.47	0.63	0.98	0.18	2.01

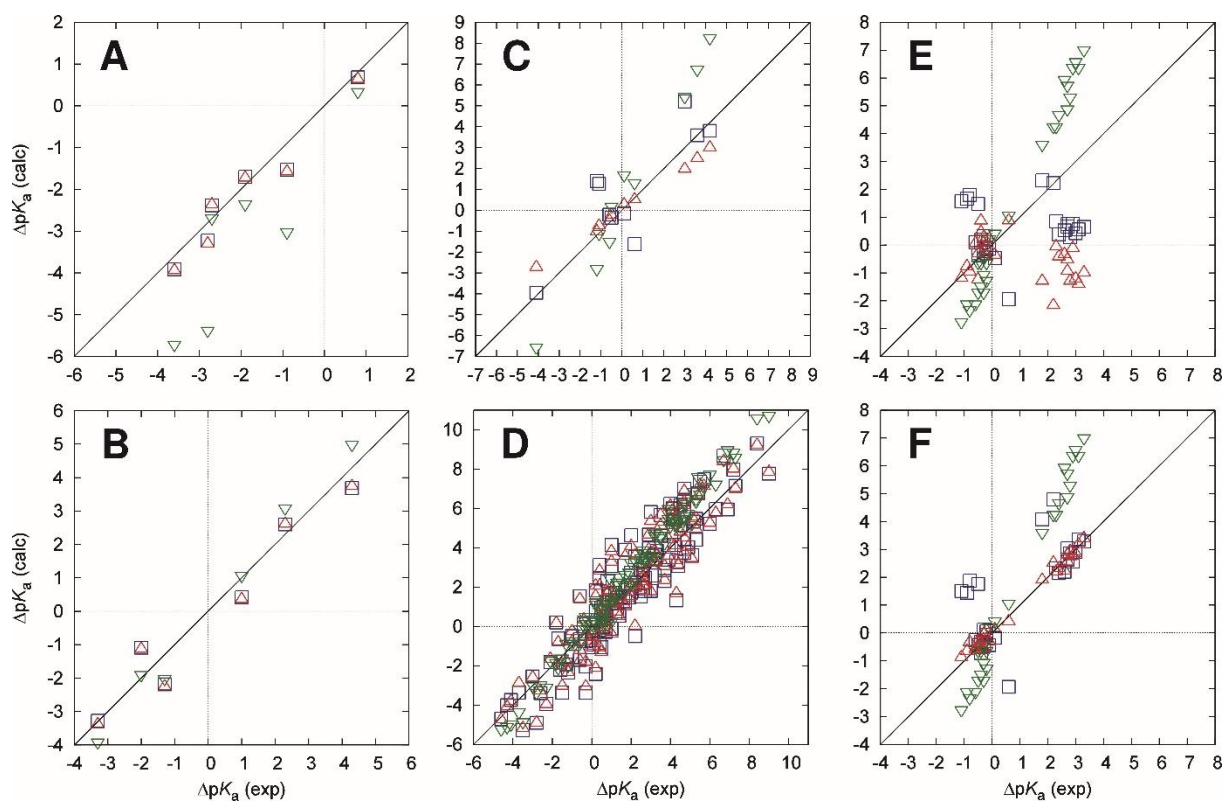


Abbildung 17: Mit MP2(EC-RISM)/6-311+G(d,p)//B3LYP/6-311+G(d,p) berechnete pK_a -Verschiebungen aufgetragen gegen experimentelle Ergebnisse aus Ref. ^[128]. Ergebnisse für Vakuumstrukturen in blau, Ergebnisse für mit PCM optimierte Strukturen in rot, Jaguar-SCRF-Ergebnisse zum Vergleich in grün. **A:** Pyrrole, **B:** Indole, **C:** Aniline, **D:** Heterozyklen, **E:** Phenole (original GAFF LJ-Parameter), **F:** Phenole (modifizierte LJ-Parameters wie im Text beschrieben).

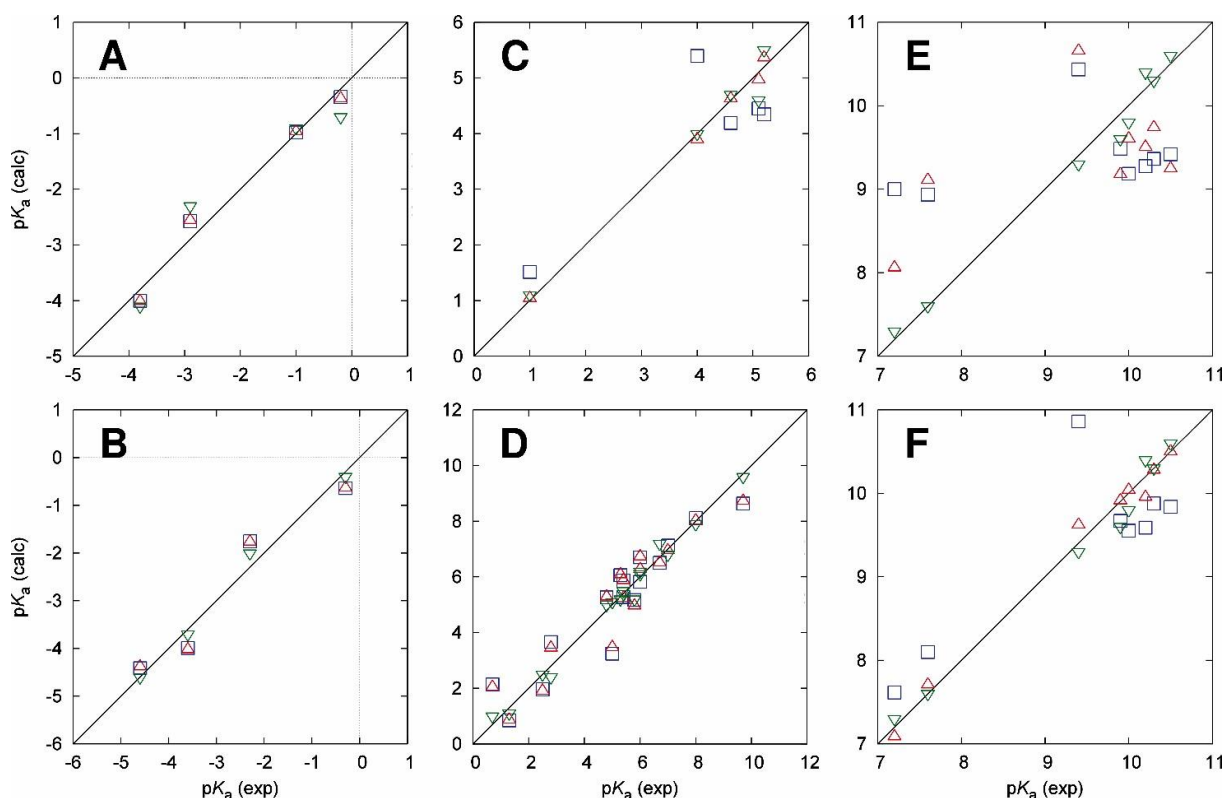


Abbildung 18: pK_a -Werte berechnet mit MP2(EC-RISM)/6-311+G(d,p)//B3LYP/6-311+G(d,p) aufgetragen gegen experimentelle Ergebnisse aus Ref. ^[128]. Ergebnisse für Vakuumstrukturen in blau, Ergebnisse für mit PCM optimierte Strukturen in rot, Jaguar-SCRF-Ergebnisse zum Vergleich in grün. **A:** Pyrrole, **B:** Indole, **C:** Aniline, **D:** Heterozyklen, **E:** Phenole (original GAFF LJ-Parameter), **F:** Phenole (modifizierte LJ-Parameters wie im Text beschrieben).

Grundsätzlich bleiben die im Rahmen der Diplomarbeit gewonnenen qualitativen Schlussfolgerungen erhalten. Insbesondere in denjenigen Fällen, in denen die Substanzen ähnlich sind, sich also wie z.B. bei den Pyrrolen nur bezüglich einer Methylierungsposition unterscheiden, ist eine sehr starke Korrelation zwischen experimentellen und berechneten pK_a -Werten zu beobachten. In diesem Fällen liegt der RMSD^s für die besten Methoden deutlich unterhalb von $0.5 \text{ kcal mol}^{-1}$. Für die wesentlich diverseren Heterozyklen mit Ringen unterschiedlicher Größe mit einer unterschiedlichen Anzahl von Heteroatomen liegt der RMSD^s-Wert deutlich höher. Das GAFF-Kraftfeld differenziert nicht zwischen geladenen und ungeladenen Spezies – Der Phenolat-Sauerstoff erhält die gleichen Lennard-Jones-Parameter zugewiesen wie der Hydroxy-Sauerstoff des Phenols, obwohl durch die zusätzliche Ladung die effektive Größe des Sauerstoffs deutlich zunehmen sollte. In der Tat zeigt sich, dass durch eine Reparametrisierung des LJ- σ -Parameters des Phenolat-Sauerstoffs eine ausgezeichnete Übereinstimmung berechneter und experimenteller Werte hergestellt werden kann. In vielen Fällen ist die Steigung der linearen Modelle nahe bei eins. Das kann als Hinweis interpretiert werden, dass die sehr kleinen Steigungen bei pK_a -Modellen auf der Basis der dielektrischen Kontinuumsmodelle in der

Tat auf intrinsische Limitierungen dieser Methoden zurückzuführen sind. Für Pyrrole und Indole wurden auch individuelle Tautomerenergien auf ihre Kompatibilität mit spektroskopischen Ergebnissen,^[129,130] die für die Pyrrole die C2-Position, für Indole die C3-Position als primäre Protonierungszentren identifizieren, untersucht. Im Rahmen der Diplomarbeit konnte auf der Basis von Vakuumgeometrien jeweils eine C2- bzw. C3-protonierte Pyrrol-/Indol-Spezies als energetisch günstigster Zustand ermittelt werden. Tabelle 9 und Tabelle 10 zeigen die Gibbs-Energien dieser Tautomere.

Tabelle 9: Gibbs-Energien der Pyrroltautomere. Die Strukturen der individuellen Tautomere sind in digitaler Form beigelegt. Gibbs-Energien wurden mittels MP2(EC-RISM)/6-311+G(d,p)//B3LYP(PCM)/6 311+G(d,p) berechnet. Die Differenzen ΔG beziehen sich jeweils auf die energetisch günstigste Struktur der gleichen Substanz im gleichen Ionisierungszustand.

Substanz	Ionisierungszustand	Struktur	$G / (\text{kcal} / \text{mol})$	$\Delta G / (\text{kcal} / \text{mol})$	Protonierungszentrum
	0		-131520.71	0.0	-
Pyrrol	1	a	-131759.28	12.60	N
	1	b	-131771.88	0.0	C-2
	1	c	-131767.12	4.76	C-3
	0		-156107.81	0.0	-
1-Methylpyrrol	1	a	-156349.03	12.04	N
	1	b	-156361.07	0.0	C-2
	1	c	-156357.12	3.95	C-3
	0		-156116.41	0.0	-
2-Methylpyrrol	1	a	-156356.94	15.95	N
	1	b	-156372.89	0.0	C-2
	1	c	-156367.43	5.46	C-2
	1	d	-156367.75	5.14	C-3
	1	e	-156365.31	7.58	C-3
	0		-156114.83	0.0	-
3-Methylpyrrol	1	a	-156356.91	13.51	N
	1	b	-156368.92	1.50	C-2
	1	c	-156370.42	0.0	C-2
	1	d	-156364.64	5.78	C-3
	1	e	-156360.93	9.49	C-3

Tabelle 10: Gibbs-Energien der Indolautomere. Die Strukturen der individuellen Tautomere sind in digitaler Form beigefügt. Gibbs-Energien wurden mittels MP2(EC-RISM)/6-311+G(d,p)//B3LYP(PCM)/6-311+G(d,p) berechnet. Die Differenzen ΔG beziehen sich jeweils auf die energetisch günstigste Struktur der gleichen Substanz im gleichen Ionisierungszustand.

Substanz	Ionisierungszustand	Struktur	$G / (\text{kcal} / \text{mol})$	$\Delta G / (\text{kcal} / \text{mol})$	Protonierungszentrum
Indol	0		-227665.18	0.0	-
	1	a	-227909.48	7.60	N
	1	b	-227903.28	13.80	C-2
	1	c	-227917.08	0.0	C-3
1-Methylindol	0		-252252.96	0.0	-
	1	a	-252499.73	8.14	N
	1	b	-252491.99	15.88	C-2
	1	c	-252507.87	0.0	C-3
2-Methylindol	0		-252262.24	0.0	-
	1	a	-252507.73	9.54	N
	1	b	-252499.25	18.02	C-2
	1	c	-252517.27	0.0	C-3
3-Methylindol	0		-252260.53	0.0	-
	1	a	-252508.16	3.73	N
	1	b	-252502.72	9.17	C-2
	1	c	-252511.89	0.0	C-3

Man erkennt, dass wie erwartet die C2-protonierten Pyrrole und die C3-protonierten Indole jeweils die protonierten Spezies mit der kleinsten Gibbs-Energie sind. Das gilt jedoch bereits auch für alte Rechnungen auf der Basis der Vakuumstrukturen.

Durch die größere methodische Datenbasis konnten aber auch neue Effekte beobachtet werden. Zum Beispiel zeigt sich bei Geometrieoptimierung mit dem PCM-Lösungsmittelmodell bei Anilinen und Phenolen eine entscheidende Verbesserung der Ergebnisse im Vergleich zu Vakuumstrukturen. Dies deckt sich mit bei der „SAMPL2 Tautomer Challenge“ gewonnenen Erkenntnissen. Bei dieser erwies sich die Geometrieoptimierung mit Lösungsmittelmodell ebenfalls als essentiell. Der 6-311++G(d,p)-Basissatz bringt im Vergleich zu einer 6-311+G(d,p)-Basis keinen Genauigkeitsgewinn. Die zusätzlichen diffusen Basisfunktionen an den Wasserstoffen sorgen lediglich für einen höheren Rechenaufwand, sind aber offensichtlich unnötig. MP2-Geometrien sorgen für eine leicht bessere Korrelation mit den experimentalen Daten, was aber die sehr deutlich ansteigenden Rechenzeiten nach

Meinung des Autors kaum aufwiegen dürfte. Wie bei der Geometrieoptimierung stellt man auch für die eigentlichen EC-RISM-Berechnungen fest, dass etwas größere Basissätze hier meist zu keinen Verbesserungen führen. Ergebnisse für MP2-Methoden sind leicht besser als die B3LYP-Ergebnisse. Bei den pK_a -Werten der Aniline wird die Steigung wesentlich durch einen einzelnen etwas abseits liegenden Datenpunkt bestimmt. Dass die pK_a hier speziell für MP2(ECRISM)/aug-cc-pVDZ//B3LYP(PCM)/6-311+G(d,p) fast exakt reproduziert werden können ist somit wahrscheinlich zufällig.

Für kleine Moleküle wie die hier verwendeten ist der Zusatzaufwand für EC-RISM mit MP2 statt B3LYP recht gering, für die Geometrieoptimierung gilt dies aber nicht. Daher bleibt das Fazit, dass hier MP2(ECRISM)/6-311+G(d,p)//B3LYP(PCM)/6-311+G(d,p) einen guten Kompromiss aus Genauigkeit und Geschwindigkeit bietet und bevorzugt für pK_a -Berechnungen verwendet werden sollte.

Bezüglich der schwächer ausfallenden Ergebnisse für die Heterozyklen wurden während der Arbeiten im Rahmen der „SAMPL2-Tautomer Challenge“ mit EC-RISM speziell für heterozyklische Ringsysteme mit 5 Atomen Tautomerenverhältnisse erhalten, die deutlich näher am experimentellen Wert liegen als für Ringsysteme mit 6 Atomen.^[15] Offenbar ist es schwierig, 5- und 6-Ringsysteme wie hier gleichzeitig mit den gleichen Parametern zu beschreiben.

Schlussendlich sind in Tabelle 11 die Gibbs-Energien und pK_a -Werte aller Spezies für die Methode MP2(EC-RISM)/6-311+G(d,p)//B3LYP(PCM)/6-311+G(d,p) dargestellt.

Tabelle 11: pK_a -Werte und mittels MP2(EC-RISM)/6-311+G(d,p)//B3LYP(PCM)/6-311+G(d,p) berechnete Gibbs-Energien unter Anwendung der durch die Daten in Tabelle 4 bis Tabelle 8 gegebenen linearen Modelle. Nur das globale Minimum-Gibbs-Energie-Konformer wurde berücksichtigt. Experimentelle pK_a -Werte wurden aus Ref. [128] übernommen. Für Phenole wurde anstelle des entsprechenden GAFF-Werts (2.91 Å) ein LJ- σ -Parameter von 3.325 Å für den basischen Phenol-Sauerstoff verwendet.

Method/Substanzklasse	G / (kcal / mol) (prot.)	G / (kcal / mol) (deprot.)	Exp. pK_a	Ber. pK_a
Pyrrole				
Pyrrol	-131771.89	-131520.85	-3.8	-4.0
1-Methylpyrrole	-156361.07	-156107.89	-2.9	-2.6
2-Methylpyrrole	-156372.97	-156116.56	-0.2	-0.4
3-Methylpyrrole	-156370.49	-156114.95	-1.0	-1.0
Indole				
Indol	-227917.09	-227665.36	-3.6	-4.0
1-Methylindole	-252507.86	-252253.05	-2.3	-1.8
2-Methylindole	-252518.79	-252262.44	-0.3	-0.6
3-Methylindole	-252511.94	-252260.69	-4.6	-4.4
Aniline				
Anilin	-180235.48	-179970.90	4.6	4.6
4-Chloroanilin	-468296.86	-468032.98	4.0	3.9
4-Methoxyanilin	-251932.68	-251667.41	5.2	5.4
4-Nitroanilin	-308313.65	-308052.46	1.0	1.0
<i>p</i> -Toluidin	-204831.25	-204566.35	5.1	5.0
Heterozyklen				
2-Aminopyridin	-190303.10	-190035.95	6.7	6.5
2-Aminothiazol	-391394.89	-391129.78	5.4	5.2
2-Methylimidazol	-166459.24	-166189.67	8.0	8.0
3-Aminopyridin	-190299.90	-190033.14	6.0	6.3
4-Aminopyridin	-190305.89	-190035.20	9.7	8.7
4-Methylpyridin	-180232.81	-179965.32	6.0	6.7
Benzimidazol	-238002.17	-237737.51	5.8	5.0
Imidazol	-141860.24	-141592.39	7.0	6.9
Isoquinolin	-251777.49	-251511.36	5.4	5.9
Melamin	-279779.83	-279517.54	5.0	3.5
Pyrazin	-165683.65	-165423.66	0.7	2.0
Pyrazol	-141837.77	-141577.99	2.5	1.9
Pyridin	-155635.07	-155368.61	5.3	6.0
Pyrimidin	-165687.48	-165429.37	1.3	0.9
Quinolin	-251777.38	-251512.19	4.8	5.3
Thiazol	-356728.99	-356466.74	2.8	3.5

Phenole				
Phenol	-192427.74	-192153.02	10.0	10.0
4-Aminophenol	-227090.60	-226816.63	9.4	9.2
4-Chlorophenol	-480489.59	-480214.86	9.9	10.1
4-Fluorophenol	-254598.55	-254323.90	10.2	10.0
4-Methoxyphenol	-264123.67	-263848.79	10.3	10.2
4-Methylphenol	-217022.97	-216747.72	10.5	10.7
4-Nitrophenol	-320509.49	-320237.11	7.2	7.3
4-Hydroxybenzaldehyd	-263390.79	-263118.09	7.6	7.7
Thiole				
Methanthiol	-274865.15	-274603.76	10.3	10.5
Ethanthiol	-299459.52	-299199.04	10.6	9.8
2-Thioethanol	-346576.74	-346316.35	9.4	9.7
1,2-Ethandithiol	-548998.19	-548737.30	9.1	9.4
Thiophenol	-394846.94	-394590.64	6.6	6.6

3.7 Tautomerenverhältnisse von Guaninen

Eines der Hauptziele der synthetischen Biologie^[131] ist es, Systeme zu synthetisieren, die emergente Eigenschaften lebendiger Materie wie Vererbung und Evolution nachbilden. In diesem Kontext spielen nichtnatürliche Nucleobasen eine wichtige Rolle. Wie bei den natürlichen Nucleobasenpaaren Guanin-Cytosin (G-C) und Adenin-Thymin (A-T, in DNA) bzw. Adenin-Uracil (A-U, in RNA) bilden auch hier je zwei von ihnen ein Basenpaar mit komplementären Wasserstoffdonoren und -akzeptoren. Dabei ist es entscheidend, dass diejenigen Tautomere dieser Nucleobasen, die zu Watson-Crick-artigen Basenpaaren führen, eine deutlich höhere Stabilität besitzen als andere Tautomere. Diese könnten sonst möglicherweise zu Fehlstellungen führen, die die genetische Funktionalität beeinträchtigen.^[132] Speziell für Isoguanin kann auf der Basis von Experimenten mit 2'-desoxyisoguanosin vermutet werden, dass in wässriger Lösung eine Enolform, die zu Thymin statt zu Isocytosin komplementär ist, in nicht vernachlässigbarer Menge existiert.^[133] Andere Guanin-Derivate, die dieses Problem nicht besitzen sind 7-deaza-Isoguanin,^[134] 8-aza-7-deaza-Isoguanin^[135] und 8-aza-Isoguanin,^[136] die mit Hilfe einer „*Trial and Error*“-Strategie gefunden wurden.

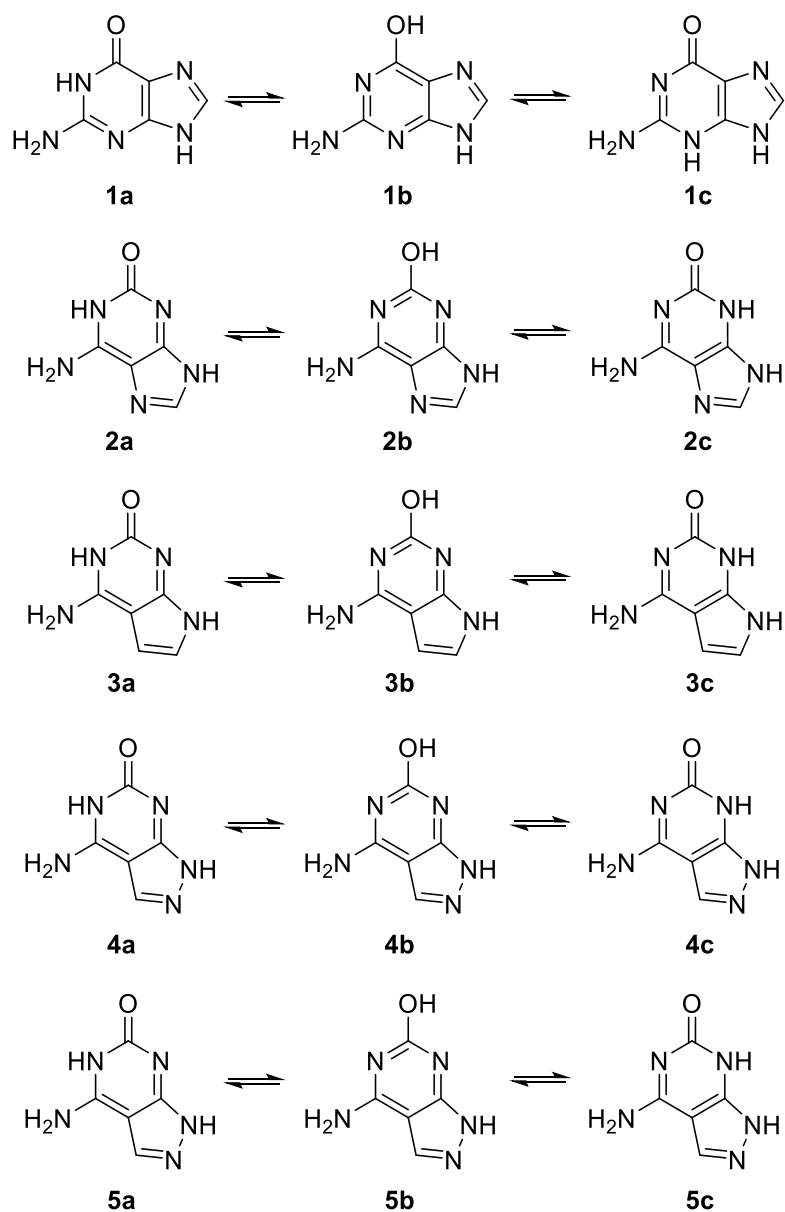


Abbildung 19: Relevante Tautomere von Guanin **1**, Isoguanin **2**, 7-Deaza-9-guanin **3**, 7-Deaza-8-aza-guanin **4** und 8-Azaguanin **5**. N1-H -Tautomere **a**, Enol-Tautomere **b**, N3-H -Tautomere **c**.

Leider ist bisher unklar, welcher Aufwand für theoretische Berechnungen zu betreiben wäre, um den chemischen Raum möglicher Guaninderivate oder anderer Nucleobasen mit *in silico*-Tests zuverlässig einzuschränken, um aufwendige Synthesearbeiten von vorneherein in eine erfolgversprechende Richtung zu lenken. Daher wurden für die Tautomere der Guanin-Derivate in Abbildung 19 EC-RISM-Berechnungen der Gibbs-Energien mit MP2/aug-cc-pVDZ vorgenommen. Mit MP2/cc-pVDZ und IEFPCM optimierte Ausgangsgeometrien (global optimale OH-Rotamere) wurden von Nicolaas van Eikema Hommes (NEH) zur Verfügung gestellt, der auch thermochemische Korrekturen auf B3LYP/6-31(d,p)-Niveau addierte und auf Basis der EC-RISM-Ergebnisse durch Subtrahieren von Gasphasen-MP2/aug-cc-pVDZ-Energien und addieren von Gasphasen-

CCSD/aug-cc-pVDZ- bzw. CCSD(T)/aug-cc-pVTZ-Energien auf die jeweilige *Coupled-Cluster*-Methodik extrapolierte. Das Guanin-N3-H-Tautomer **1c** wurde nicht beachtet, da Gasphasen-CCSD/aug-cc-pVDZ-Rechnungen von NEH bereits eine Differenz von 17.1 kcal/mol zum N1-H-Tautomer **1a** ergaben.

Erhalten wurden die in Tabelle 12 gezeigten Energien. Experimentelle Abschätzungen geben das Ketoform/Enolform-Verhältnis von Isoguanin als 9:1 ($\Delta_{\text{T}}G^{\circ} = 1.3$ kcal/mol) an. Für Guanosin, bei dem Guanin an β -D-Ribose gebunden ist, wird das Verhältnis auf 9999:1 bis 99999:1 ($\Delta_{\text{T}}G^{\circ} = 5.5$ bis 6.8 kcal/mol) geschätzt.^[132]

Tabelle 12: Relative Gibbs-Energien für Guanin-Derivate. a: EC-RISM MP2/aug-cc-pVDZ, b: EC-RISM MP2/aug-cc-pVDZ mit thermochemischen Korrekturen bei 298.15 K, c: EC-RISM MP2/aug-cc-pVDZ mit thermochemischen Korrekturen bei 298.15 K extrapoliert auf CCSD/aug-cc-pVDZ, d: EC-RISM MP2/aug-cc-pVDZ mit thermochemischen Korrekturen bei 298.15 K extrapoliert auf CCSD(T)/aug-cc-pVTZ.

Substanz/Tautomer	$\Delta G/(\text{kcal/mol})^a$	$\Delta G/(\text{kcal/mol})^b$	$\Delta G/(\text{kcal/mol})^c$	$\Delta G/(\text{kcal/mol})^d$
1a	0.0	0.0	0.0	0.0
1b	9.6	9.8	9.8	9.6
1c	-	-	-	-
2a	0.0	0.0	0.0	0.0
2b	6.9	7.3	8.1	8.4
2c	0.6	-0.3	0.1	0.5
3a	0.0	0.0	0.0	0.0
3b	8.0	8.8	9.6	9.9
3c	1.3	1.1	1.3	1.7
4a	0.0	0.0	0.0	0.0
4b	8.7	9.4	10.1	10.3
4c	2.1	2.1	2.2	2.6
5a	0.0	0.0	0.0	0.0
5b	6.1	8.0	8.5	8.9
5c	0.4	1.6	2.0	2.3

Man erkennt, dass in jedem Fall die Enolform deutlich penalisiert ist. Dieses Ergebnis ist also inkompatibel mit Vermutungen bezüglich der Existenz eines nicht vernachlässigbaren Anteils einer Enolform von Isoguanin. Die Gibbs-Energien sind quer über alle Methoden qualitativ ähnlich, auch wenn die Gibbs-Tautomerisierungsenergien für Guanin und Isoguanin leicht oberhalb der Schätzwerte liegen. Thermochemische Korrekturen und *Coupled-Cluster*-Extrapolation verstärken die Unterschiede leicht. Interessant ist die niedrige Energiedifferenz des N3-H-Tautomers speziell des Isoguanins, das in bisherigen Untersuchungen mit Ausnahme von Berechnungen von Switzer und Mitarbeitern,^[137] wo man zu ähnlichen Ergebnissen kam, kaum berücksichtigt wurde.

3.8 pH-abhängige Speziesverteilung von Histamin

Histamin ist ein biogenes Amin, das als Gewebeghormon und Neurotransmitter wirkt und unter anderem im Körper Entzündungsvorgänge reguliert.^[138] Das konformationelle Ensemble und die Tautomerenzpräferenz des Histamins wurden bereits mit Hilfe dielektrischer Kontinuumsmodelle untersucht,^[124,139] so dass sich hier ein Vergleich der Ergebnisse mit denen der EC-RISM-Methode anbietet. Interessant ist auch, dass auf der Basis älterer Rechnungen ein Mechanismus für die Aktivierung von H₂-Rezeptoren vorgeschlagen wurde, der einen Tautomerenzwechsel des Histamins beinhaltet.^[140] Die pK_a-Werte von Histamin sind im physiologisch relevanten Bereich ca. 6.0 und 9.8.^[141]

Histamin besitzt drei titrierbare Stickstoffzentren, die aber in der Literatur nach unterschiedlichen Benennungsschemata gekennzeichnet werden. Im Bereich der Biochemie ist eine Nomenklatur üblich, die sich von der Aminosäure Histidin ableitet.^[114] Dabei beginnt die Nummerierung mit dem der Seitenkette benachbarten Stickstoffatom. In der Organischen Chemie ist es dagegen üblich, dem der Seitenkette entgegengesetzten Stickstoffatom die Zahl 1 zuzuweisen. Um Verwechslungen vorzubeugen, hat die IUPAC ein Benennungsschema eingeführt, bei dem die Ringstickstoffe mit den griechischen Buchstaben π („pros“) und τ („tele“) bezeichnet werden. Damit ist jedoch noch nicht klar, wie die Ring-Kohlenstoffatome zu bezeichnen sind. Im Folgenden wird zusätzlich das Protonierungszentrum am Stickstoffatom der Seitenkette mit dem griechischen Buchstaben α bezeichnet, obwohl dieser sich streng genommen auf das nächstgelegene Kohlenstoffatom bezieht. Die verschiedenen Benennungsschemata sind in Abbildung 19 dargestellt.

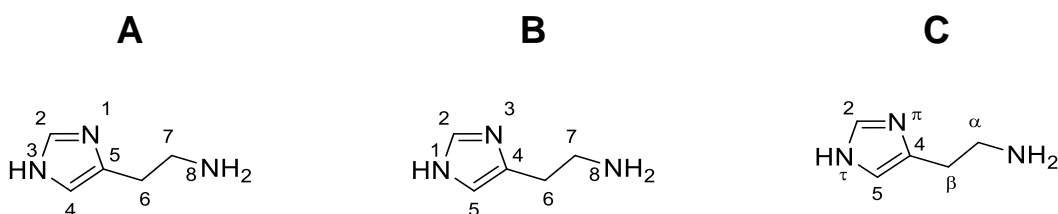


Abbildung 20: Benennungsschemata für das Histamin. **A** Biochemische Nomenklatur, **B** Organische Chemie-Nomenklatur, **C** IUPAC-Benennungsschema für Histamin.

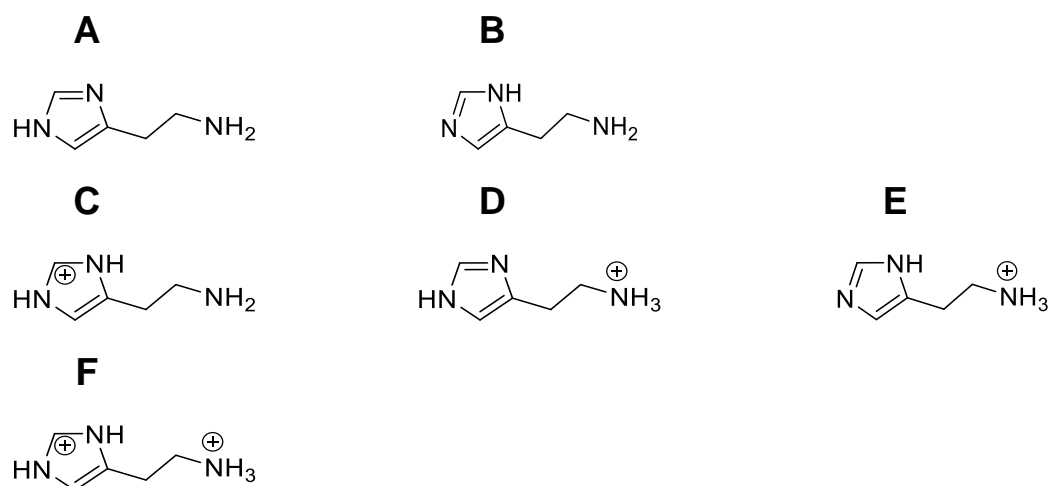


Abbildung 21: Untersuchte Tautomere des neutralen Histamins, des Histaminmonokations und des Histamindikations. **A:** τ Neutralform (0, τ), **B:** π Neutralform (0, π), **C:** π/τ Monokation (+1, π/τ), **D:** α/τ Monokation (+1, α/τ), **E:** α/π Monokation (+1, α/π), **F:** $\alpha/\pi/\tau$ Dikation (+2, $\alpha/\pi/\tau$).

Die Frage nach der Tautomerenpräferenz des Histamins in unterschiedlichen Ionisierungsstufen ist nach Ansicht des Autors immer noch teilweise offen. In der Vergangenheit wurden diesbezüglich einige Studien durchgeführt. Diese Studien kamen aber teilweise zu unterschiedlichen Ergebnissen bezüglich der Präferenz der π - bzw. τ -Tautomere.^[124,139,140,142,143] Aus dem Blickwinkel der Gegenwart betrachtet entspricht jedoch nur die Arbeit von Luque und Mitarbeitern^[124] dem aktuellen Stand der Technik. Hier wurde ein modernes PCM-Modell mit quantenchemischen Rechnungen auf B3LYP/6-31G*-, MP2/aug-cc-pVDZ- und MP2/aug-cc-pVTZ-Niveau und einem komplexen Protokoll zum Konformationssampling kombiniert. Dabei zeigt sich das τ -Tautomere als leicht stabiler bei einem π - zu- τ -Verhältnis von 48:52, was aber einer Gibbs-Energiedifferenz von nur 0.05 kcal/mol entspricht. Ob die angegebene Methode derart kleine Energiedifferenzen auflösen kann erscheint hochgradig fragwürdig. Arbeiten von Ramírez und Mitarbeitern^[139] sind ebenfalls relevant, obwohl dort ein Kontinuumsmodell benutzt wird, das eine Kavität verwendet die nicht der Molekülform folgt und auf Konformationssampling verzichtet wird. Dafür werden Infrarot-Spektroskopie- (IR-) Experimente durchgeführt, und die so erhaltenen Spektren mit mittels Quantenchemie berechneten verglichen. Hierbei zeigte sich, dass sich das berechnete IR-Spektrum des τ -Tautomers in einer *gauche*-Konformation mit intramolekularen Wasserstoffbrücken am besten an das experimentell gemessene Spektrum anpassen ließ. Von Ganellin stammt eine Bestimmung des Tautomerenverhältnisses des Histamin-Monokations aus den Säurekonstanten der *N*-methylierten Histaminderivate.^[144] Dabei wurde die Annahme verwendet, dass diese sich bei der Titration ähnlich verhalten wie sich die

Histamintautomere selbst verhalten würden, wenn sie sich nicht ineinander umlagerten. Dabei ergab sich für das α/τ -Tautomer eine um 0.82 kcal/mol geringere Energie als für das α/π -Tautomer. Nach Einschätzung des Autors erscheint es jedoch keineswegs ausgeschlossen, dass subtile Unterschiede im Solvatationsverhalten zwischen Histamin und seinen *N*-methylierten Derivaten für einen Energieunterschied von mehreren kcal/mol sorgen können. Bei ^{13}C -NMR-Experimenten fanden Reynolds und Tzeng^[145] das τ - und das α/τ -Tautomer als präferierte Spezies für neutrales Histamin und das Monokation. Diese Experimente basieren auf der Annahme, dass sich die ^{13}C chemischen Verschiebungen von Histamin und Histidin bei Titration ähnlich verhalten und dass sich die Histidintautomere und ihre jeweiligen methylierten Derivate im sauren Bereich, in dem sich das Dikation bildet, ähnlich verhalten. Hier könnte das gleiche Gegenargument wie bei den Experimenten von Ganellin angebracht werden. Zusammenfassend lässt sich sagen, dass die experimentellen Daten nicht als endgültige Bestätigung der Präferenz des τ - bzw. α/τ -Tautomers gelten können, da sie im Wesentlichen auf Ähnlichkeitsargumenten beruhen, deren Validität im Einzelfall durchaus in Zweifel gezogen werden kann, so wie bei den spektroskopischen Untersuchungen von Ramírez und Mitarbeitern die Zuordnung der Banden mit Hilfe der durchgeführten Berechnungen. Die bisherigen Arbeiten sind valide Hinweise auf eine Präferenz der τ - und α/τ -Tautomere, weitere Arbeiten mit moderneren theoretischen Ansätzen wie EC-RISM können jedoch weiter zur Klärung beitragen.

Die im Folgenden genannten technischen Arbeiten wurden von Herrn Lukas Eberlein unter Anleitung des Autors durchgeführt. Vom Auto selbst stammt u. a. der in Kapitel 3.4 vorgestellte Formalismus zur Berechnung der pH-abhängigen Speziesverteilung und die dazu gehörende Auswertung. Für die in Abbildung 21 gezeigten Spezies wurde der Konformationsraum durchmustert, indem alle rotierbaren Bindungen in 90° -Abstände eingeteilt, und dann Strukturen auf der Basis aller Permutationen für die Winkelstellung erzeugt wurden. Mit Hilfe von Gaussian 09^[63] wurde unter Vakuumbedingungen eine Geometrieoptimierung mit B3LYP/6-31+G* durchgeführt, gefolgt von einer zweiten mit dem PCM-Solvatationsmodell. Nach Entfernen von Dubletten verblieben pro Tautomer 64 eindeutige Konformationen. Mit diesen Konformationen als Ausgangsstrukturen wurden wieder mit dem PCM-Modell MP2-Geometrieoptimierungen mit den Basissätzen 6-311+G** und aug-cc-pVDZ unternommen. Die mit B3LYP(PCM)/6-31+G*, MP2(PCM)/6-311+G** und MP2(PCM)/aug-cc-pVDZ erhaltenen Strukturen wurden anschließend für Berechnungen mit EC-RISM auf demselben Theorieniveau verwendet. Für B3LYP(PCM)/6-31+G* wurde testweise noch eine D3-Dispersionskorrektur nach Grimme auf die erhaltenen Strukturen angewendet.

Abbildung 22 zeigt die Verteilung der Gibbs-Energien der Konformere aller Tautomere der verschiedenen Ionisierungszustände. Generell unterscheiden sich die Gibbs-Energien der Konformere um bis zu 6 kcal/mol innerhalb eines Tautomers und bis zu ca. 10 kcal/mol innerhalb eines Ionisierungszustands. Für das neutrale Histamin favorisiert EC-RISM das π -Tautomer um 1.2 kcal/mol und 1.4 kcal/mol für B3LYP/6-31+G* und B3LYP-D3/6-31+G* und ca. 0.9 kcal/mol für MP2/6-311+G** und MP2/aug-cc-pVDZ. Für das Monokation ist die Situation weniger eindeutig. Die α/π -Form ist gegenüber der α/τ -Form für die B3LYP-Methoden leicht um 0.5 und 0.8 kcal/mol favorisiert, nicht jedoch für die MP2-Methoden, bei denen die Gibbs-Energien beider Tautomere mit einem Unterschied von nur 0.1 kcal/mol beinahe identisch sind. Die Gibbs-Energien der π/τ -Form spielen mit einem Abstand von ca. 3 kcal/mol für die B3LYP- und ca. 5 kcal/mol für die MP2-Methoden bezüglich des energetisch nächsten Tautomers keine wichtige Rolle. Der RMSD beträgt hier für das MP2/aug-cc-pVDZ-Modell nur 0.57 kcal/mol. Daher erscheint wahrscheinlich, dass in Lösung ein nicht zu vernachlässigender Anteil von π - und insbesondere α/π -Tautomeren existiert. Während des SAMPL2-Methodenvergleichs^[146] konnten mittels EC-RISM speziell für 5-Ring-Heterozyklen gute Ergebnisse erhalten werden,^[15] insofern sind die EC-RISM-Ergebnisse ernst zu nehmen, auch wenn bisherige Arbeiten in eine gegenteilige Richtung weisen.

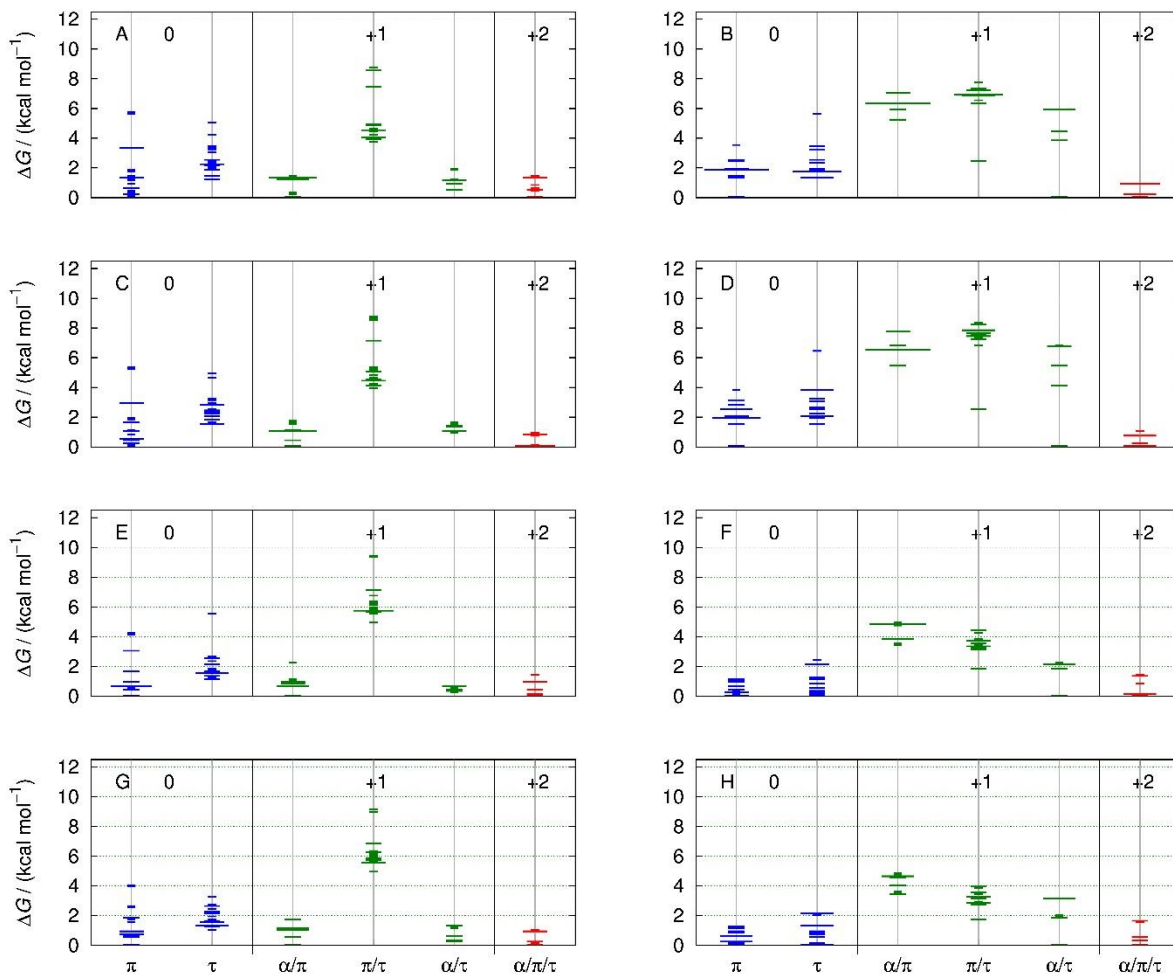


Abbildung 22: Gibbs-Energie-Profile von Histamin für B3LYP/6-31+G* (A, B), B3LYP-D3/6-31+G* (C, D), MP2/6-311+G** (E, F) und MP2/aug-cc-pVDZ (G, H) auf der Basis von PCM-optimierten Strukturen mit jeweils der gleichen Methode und dem gleichen Basissatz, EC-RISM-Ergebnisse links, PCM-Ergebnisse auf der rechten Seite. Gezeigt werden vertikale Histogramme der Gibbs-Energien der Konformere mit einer Korbgröße von 0.25 kcal/mol. B3LYP-D3-Berechnungen wurden auf der Basis von mit ohne D3-Korrektur optimierten Strukturen durchgeführt. Alle Gibbs-Energien sind relativ zum günstigsten Konformer des günstigsten Tautomers des jeweiligen Ionisierungszustands. Die einzelnen Energien sind für alle Methoden und Spezies im Anhang in 5.2 tabelliert.

Die Differenzen der auf der Basis der EC-RISM-Tautomerenergien berechneten Gibbs-Energien der Ionisierungsstufen passen gut zu auf Basis der experimentellen pK_a -Werte berechneten. Daraus resultieren die mittels Gl. 3.4.18 berechneten Gibbs-Energien des Protons von -256.83 kcal/mol für B3LYP/6-31+G*, -257.69 kcal/mol für B3LYP-D3/6-31+G*, -258.79 kcal/mol für MP2/6-311+G** und -256.98 kcal/mol für MP2/aug-cc-pVDZ, sowie die in Abbildung 23 gezeigten pH-abhängigen Verteilungen der Ionisierungsstufen. Wie dort zu sehen ist, ist die Anpassung nach 3.4.18 für die B3LYP-Methoden praktisch perfekt und für die MP2-Methoden immer noch sehr gut. Die zugehörigen pK_a -Werte sind in Tabelle 13 aufgeführt.

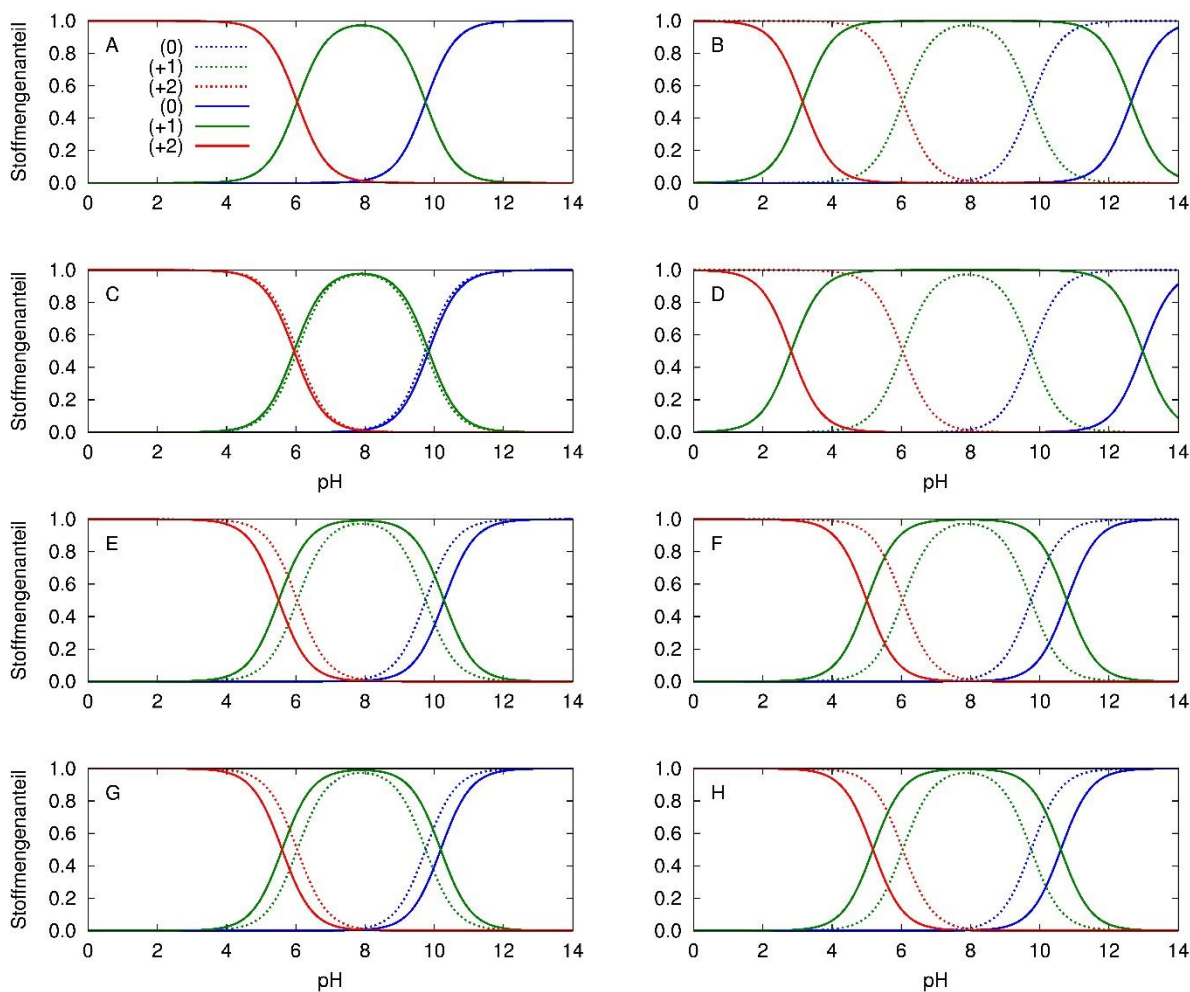


Abbildung 23: pH-abhängige Verteilung der Ionisierungsstufen des Histamins für B3LYP/6-31+G* (A, B), B3LYP-D3/6-31+G* (C, D), MP2/6-311+G** (E, F) und MP2/aug-cc-pVDZ (G, H) berechnet mit den Solvatationsmodellen EC-RISM (linke Seite) und PCM (rechte Seite). Die durchgezogenen Linien zeigen die auf der Basis der berechneten Gibbs-Energien für die Histamin-Spezies und an experimentelle pK_a -Werte angepasster Werte für die Gibbs-Energie des Protons erzeugten Verteilungen. Die gepunkteten Linien zeigen Verteilungen allein auf der Basis der experimentellen pK_a -Werte. Die Farbe der Linien markiert den Ionisierungszustand (Blau 0, grün +1, rot +2). Die zugehörigen pK_a -Werte sind in Tabelle 13 angegeben.

Tabelle 13: pK_a -Werte des Histamins aus der Anpassung berechneter Gibbs-Energien an experimentelle pK_a -Werte für verschiedene quantenchemische Methoden. pK_a 1 bezieht sich auf den Übergang von +2 zu +1, pK_a 2 auf den Übergang von +1 zu 0.

Methodik	EC-RISM		PCM	
	pK_a 1	pK_a 2	pK_a 1	pK_a 2
B3LYP/6-31+G*	6.04	9.75	3.15	12.64
B3LYP-D3/6-31+G*	5.95	9.84	2.82	12.97
MP2/6-311+G**	5.51	10.28	5.01	10.78
MP2/aug-cc-pVDZ	5.60	10.19	5.18	10.61

Die genannten Gibbs-Energien des Protons sind intrinsisch im Sinne von Gl. 3.4.12. Um sie mit experimentellen Werten zu vergleichen, muss demzufolge noch der Galvanianteil addiert werden, womit sie in etwa bei -269 bis -273 kcal/mol und damit nur leicht unterhalb des derzeit präferierten experimentellen Wertes von -265.9 kcal/mol liegen. Hierbei ist jedoch zu beachten, dass die von Misin, Fedorov und Palmer für die Berechnung des Galvanipotentials verwendeten Wassermodelle cSPC/E und cTIP3P^[117] in Details vom hier verwendeten modifizierten SPC/E-Modell abweichen.

Unter Einbeziehung der Tautomerenzverhältnisse ergeben sich dann die in Abbildung 24 gezeigten pH-abhängigen Speziesverteilungsdiagramme.

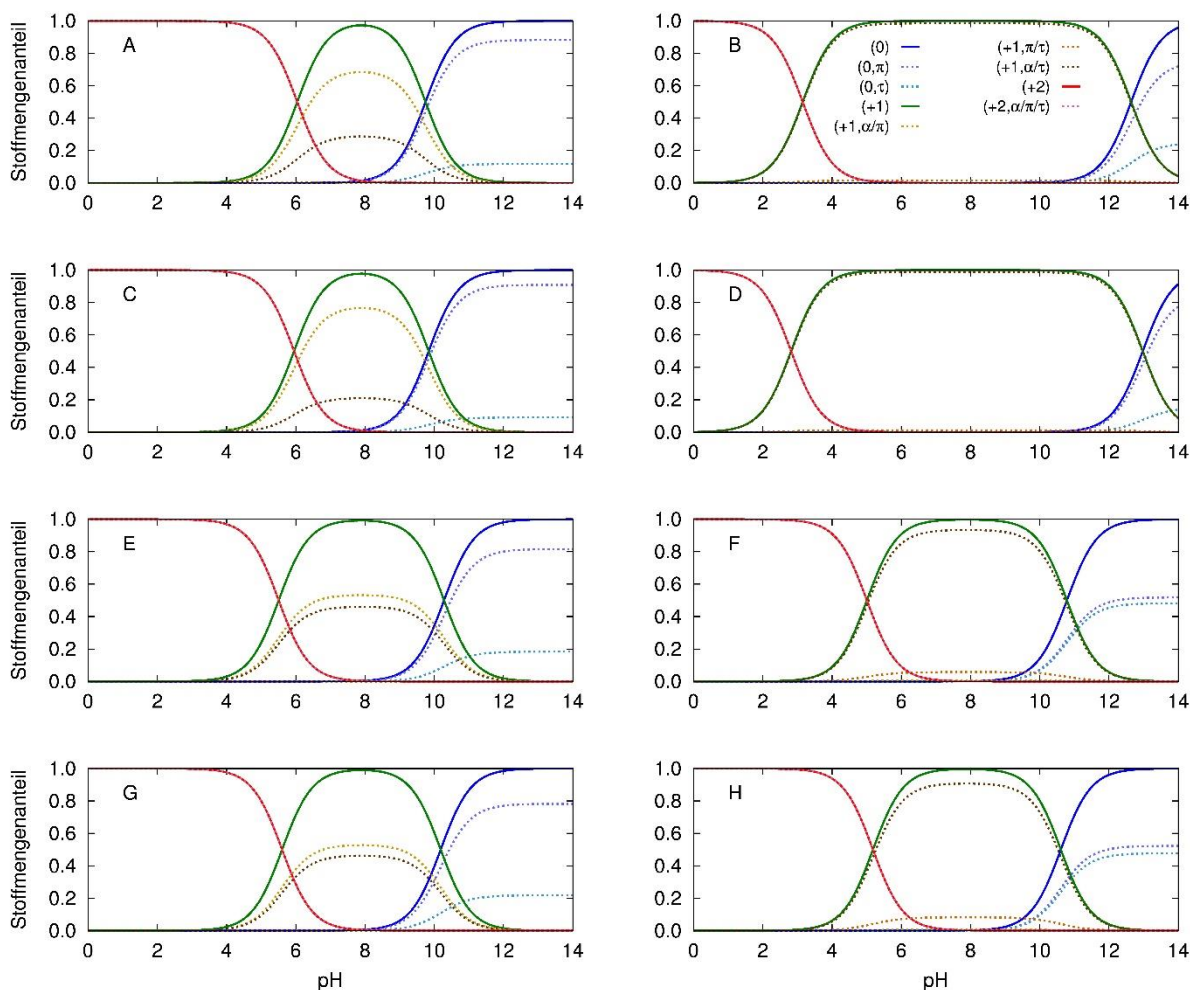


Abbildung 24: pH-abhängige Speziesverteilung des Histamins für B3LYP/6-31+G* (A, B), B3LYP-D3/6-31+G* (C, D), MP2/6-311+G** (C, D) und MP2/aug-cc-pVDZ (G, H) mit dem EC-RISM-Solvatationsmodell (A, C, E, G) und dem PCM-Solvatationsmodell (B, D, F, H). Gestrichelte Linien stellen einzelne Tautomere dar, durchgezogene Linien Ionisationszustände. Die Farbe der Linien markiert den Ionisierungszustand (Blau 0, grün +1, rot +2).

Mit PCM und den B3LYP-Methoden berechnete Ergebnisse für das freie Histamin favorisieren das π -Tautomer gegenüber dem τ -Tautomer um 0.7 (B3LYP/6-31+G*) und 1.0 kcal/mol (B3LYP-D3/6-31+G*), etwas schwächer als EC-RISM, während für die MP2-Methoden die Gibbs-Energien beider Tautomere sehr ähnlich sind. Für das Monokation unterscheiden sich PCM- und EC-RISM-Ergebnisse deutlich. Hier ist das α/τ -Tautomer um beträchtliche 4.0 kcal/mol bzw. 5.1 kcal/mol für die B3LYP-Methoden und 3.0 und 2.9 kcal/mol für die MP2-Methoden präferiert. Interessanterweise sind einige Strukturen des π/τ -Tautomers energetisch günstiger als die des α/π -Tautomers, speziell bei den B3LYP-Methoden. Bei genauerer Betrachtung fällt auf, dass diese eine intramolekulare Wasserstoffbrücke ausbilden. Würde man die Strukturen mit der Wasserstoffbrücke

ignorieren, wäre das π/τ -Tautomer für die B3LYP-Methoden nur wenig bevorzugt und bei den MP2-Methoden sogar energetisch ungünstiger als das α/π -Tautomer.

Betrachtet man die einzelnen Konformere, so zeigt sich eine große strukturelle Variabilität, wie auch bereits von Luque und Mitarbeitern beobachtet. Viele recht unterschiedliche Konformere besitzen eine ähnliche Gibbs-Energie. Im Allgemeinen bevorzugt PCM in starkem Maße die Bildung intramolekularer Wasserstoffbrücken, während diese Tendenz bei EC-RISM weniger ausgeprägt ist, wie man Abbildung 25 entnehmen kann.

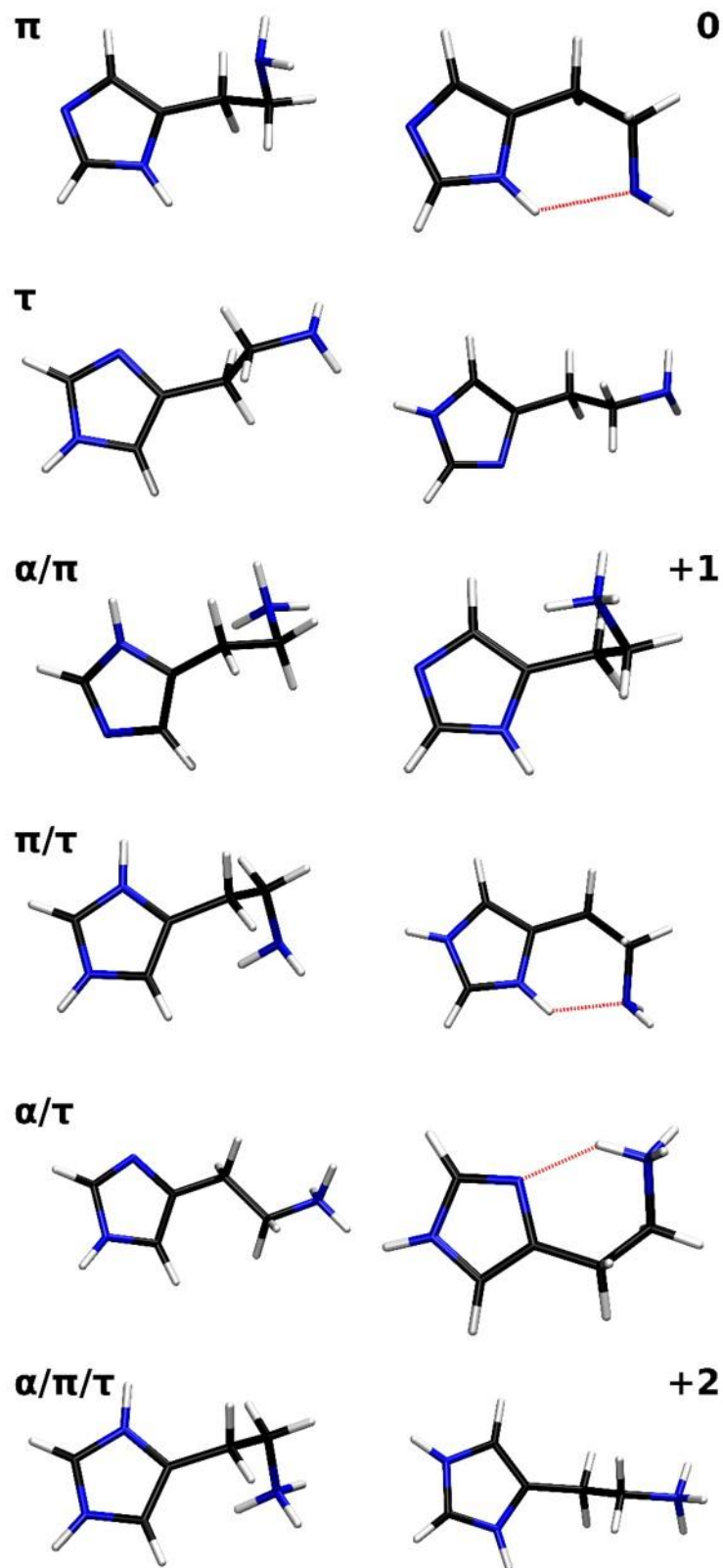


Abbildung 25: Energetisch günstigste Konformere auf Basis von B3LYP/6 31+G* mit EC-RISM (links) und PCM (rechts) für alle untersuchten Histamin-Tautomere. Wasserstoffbrücken sind als gestrichelte rote Linie dargestellt.

Hier zeigt sich, dass die Behandlung subtiler Kompensationseffekte in der Gibbs-Energie der Schlüssel zur korrekten Vorhersage der Tautomererhältnisse ist. Abbildung 26 illustriert dies für ein einzelnes Beispiel, das α/τ -Tautomer des Monokations

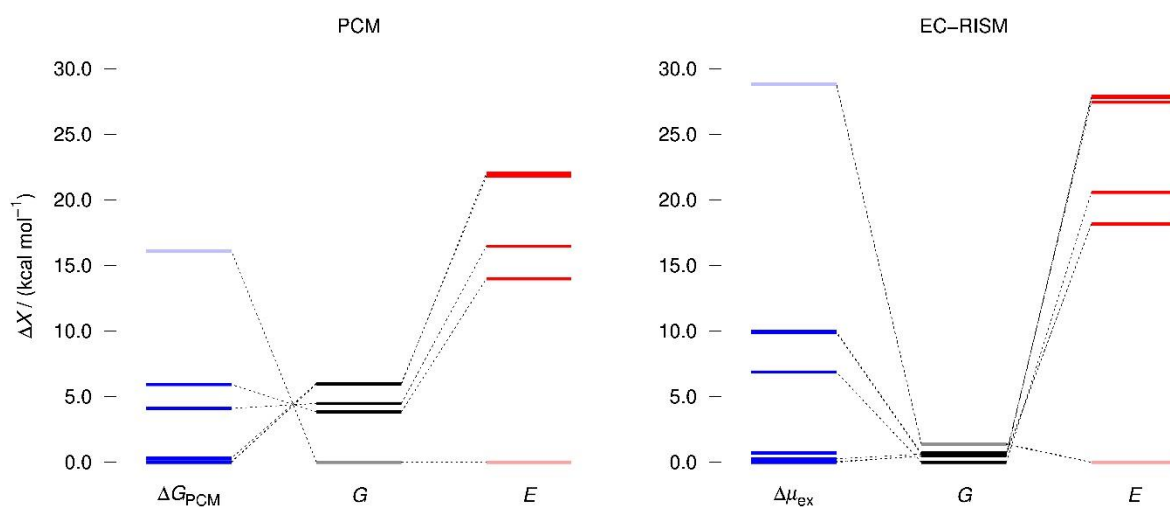


Abbildung 26: Zerlegung der Gibbs-Energie in einen quantenmechanisch-intramolekularen Beitrag und einen Solvatationsbeitrag für das α/τ -Tautomer des Histaminmonokations für PCM (links) und EC-RISM (rechts) auf B3LYP/6-31+G*-Niveau. Alle Energien sind relativ zur kleinsten Energie des jeweiligen Typs aufzufassen. Energien die zum gleichen Konformer gehören sind durch gepunktete Linien miteinander verbunden. Blasse Färbung zeigt Energien von Konformeren mit intramolekularer Wasserstoffbrücke an.

Während für sich genommen der intramolekular-energetische Beitrag und das chemische Exzesspotential für EC-RISM und PCM eine qualitative Übereinstimmung zeigen, führt die Kombination beider Beiträge zur Gibbs-Energie dazu, dass für PCM das Konformer mit der intramolekularen Wasserstoffbrücke das energetisch günstigste, für EC-RISM das energetisch ungünstigste Konformer ist.

Vergleicht man die hier erhaltenen Ergebnisse mit den experimentellen Daten, so fallen einige Diskrepanzen auf. Während die Infrarot-Spektroskopie-Daten von Ramírez und Mitarbeitern für das neutrale Histamin das τ -Tautomer als dominante Spezies hinweisen, zeigen die hier erhaltenen Ergebnisse auf das π -Tautomer. Dass sich das berechnete Spektrum eines Tautomers besser an experimentelle Spektren anpassen lässt schließt jedoch nicht aus, dass auch das andere Tautomer in erheblichen Mengen vorhanden ist. Des Weiteren wurden die Infrarot-Spektren von Ramírez und Mitarbeitern auf der Basis von statischen Strukturen berechnet, ohne dass ein Konformationssampling durchgeführt und Boltzmann-Gewichte berechnet wurden. Obwohl die EC-RISM-Methode prinzipiell das π -Tautomer zu bevorzugen scheint, erlauben die EC-RISM-Fehlerspannen aus dem SAMPL2-

Methodenvergleich auch eine Präferenz des α/τ -Tautomers für das Monokation und möglicherweise eine leichte Präferenz des τ -Tautomers für das neutrale Histamin. In jedem Fall sollten nicht zu vernachlässigende Mengen des π - und α/π -Tautomers im jeweiligen Ionisierungszustand vorhanden sein.

Die Auswertung der Ergebnisse dieses Abschnitts erfolgte in enger Zusammenarbeit mit Lukas Eberlein auf der Basis der vom Autor in Kapitel 3.4 entwickelten Methodologie. Speziell wurden vom Autor entwickelte Auswertungswerkzeuge nach Anleitung des Autors von Lukas Eberlein modifiziert und eingesetzt, um die Speziesverteilungen zu berechnen. Die Schlussfolgerungen sind die des Autors.

3.9 Charakterisierung des Protonierungszustands des RTA/PTA-Komplexes

Das Säure-Base-Verhalten von Protein-Ligand-Komplexen ist von großem Interesse für die medizinische Wirkstoffentwicklung, aber auch für das Verständnis vielfältiger in der Natur beobachtbarer Phänomene, wie am Beispiel des Histamins bereits deutlich gemacht wurde. Leider existieren keine Systeme, bei denen das Säure-Base-Verhalten sowohl des freien Liganden als auch des Protein-Ligand-Komplexes mikroskopisch vollständig charakterisiert ist. Der Grund dafür sind die vielfältigen experimentellen Schwierigkeiten, die bereits für den freien Liganden auftreten. Dazu zählen z.B. unzureichende Löslichkeit in Wasser, Photoinstabilität und Selbstaggregation.^[147] Wie sich auch am Beispiel des Histamins gezeigt hat, ist die NMR-Aufklärung des Protonierungszustands häufig bereits für kleine Moleküle schwierig und die Schwierigkeiten nehmen mit der Größe des Moleküls zu. Ab einer bestimmten Größe wird häufig eine Isotopensubstitution nötig.^[148] Bei röntgenkristallographischen Untersuchungen besteht das Problem, dass die Protonen zu schwach streuen, um gut detektierbare Signale zu erhalten.^[149] In Neutronenstreuexperimenten sind die Wasserstoffe zwar gut sichtbar, derartige Untersuchungen benötigen aber eine Neutronenquelle, d.h. einen Forschungskernreaktor oder eine Spallationsquelle. Weiterhin müssen große Einkristalle vorhanden sein.^[149,150] Für die Strukturanalyse sowohl mit Röntgen- als auch mit Neutronenbeugungsexperimenten muss das entsprechende Protein aufwendig kristallisiert werden, was i.d.R. nur bei einem einzelnen pH-Wert durchgeführt wird. Auch ist nicht sichergestellt, dass beim Übergang von der Lösung in die Festkörperumgebung das ursprüngliche Protonierungsmuster erhalten bleibt. Proteinstrukturmodelle entstehen manuell unter dem Einfluss menschlichen Handelns und sind somit zwangsläufig zum Teil fehlerhaft. Zudem liegt der Fokus häufig nicht auf der Struktur der Liganden.^[151] Zusammenfassend steht und fällt die Qualität einer Vorhersage in diesem Bereich mit der Qualität der Kristall- oder NMR-Struktur aus z.B. der PDB-Datenbank. Neuere Untersuchungen haben

ergeben, dass ein erheblicher Teil der in der PDB abgelegten Strukturen den für die Untersuchung von Protein-Ligand-Wechselwirkungen benötigten Qualitätsmaßstab nicht erfüllt.^[151] Bezüglich der computergestützten Behandlung ist problematisch, dass der Protonierungszustand des Liganden vom Protonierungszustand einer möglicherweise großen Zahl von titrierbaren Aminosäuren des Proteins abhängt, deren jeweilige Protonierungswahrscheinlichkeit wiederum vom Protonierungszustand aller anderen titrierbaren Aminosäuren sowie der des Liganden abhängt. Daher verwenden primär physikbasierte Methoden der pK_a -Berechnung für Proteine verschiedene Näherungen,^[105,152–154] die jedoch zumeist eine Zerlegung der elektrostatischen Energie in Wechselwirkungsenergien zwischen je zwei Zentren erfordern. Im Folgenden wird eine alternative Strategie erprobt, bei der zunächst die Stärke der Kopplung des Protonierungszustands von Protein und Liganden untersucht wird. Da, wie sich zeigen wird, diese Kopplung im für die Kristallisation^[155] verwendeten pH-Bereich um 8.9 nicht ausgeprägt ist, kann der Proteinzustand fixiert werden. Dann ist nur noch der Zustand des Liganden im Protein von Interesse.

Der Rizin-A/Pteroinsäure- (RTA-PTA-) Komplex ist ein Protein-Ligand-Komplex, für den zumindest einige belastbare experimentelle Daten zur Verfügung stehen, die erste Anhaltspunkte für eine Methodvalidierung bieten könnten. Als pK_a -Werte des freien PTA-Liganden wurden mittels druckunterstützter Kapillarelektrophorese (engl. *Pressure-assisted capillary electrophoresis*, PACE) die Werte 2.29, 4.8 und 8.14 gemessen.^[147] Die bei pH 8.9 präparierte Kristallstruktur mit einer Auflösung von 2.3 Å (PDB-Code 1BR6) ist Teil der „Iridium-HT“-Datenbank hochgradig vertrauenswürdiger Strukturen.^[151] Aus der Kristallstruktur lässt sich auf der Basis der Abstände zwischen mutmaßlichen Wasserstoffdonoren und Wasserstoffakzeptoren in Ligand und angrenzenden Aminosäuren des Proteins das in Abbildung 27 gezeigte Protonierungsmuster ableiten.

Neben seiner Eignung zur Methodvalidierung ist der RTA-PTA-Komplex selbst auch von erheblichem Interesse. Das Protein Rizin, das aus den Samen des Wunderbaums (*Ricinus Communis*) extrahiert werden kann, gehört zur Klasse der Lektine. Es besteht aus zwei Teilen, der A- und der B-Kette, die über eine Disulfidbrücke miteinander verbunden sind. Die B-Kette trägt die Lektinfunktionalität und ist für den Transport des Proteins in das Zellinnere verantwortlich. Die A-Kette ist die eigentliche Ursache der Toxizität. Sie katalysiert im endoplasmatischen Retikulum die Abspaltung eines Adenins-Moleküls im RNA-Teil der Ribosomen und negiert so die Fähigkeit des Organismus zur Proteinbiosynthese.^[156,157] Rizin wird von einschlägigen Experten ein erhebliches Gefahrenpotential bezüglich seiner Eignung als Mittel zur Durchführung von Attentaten und

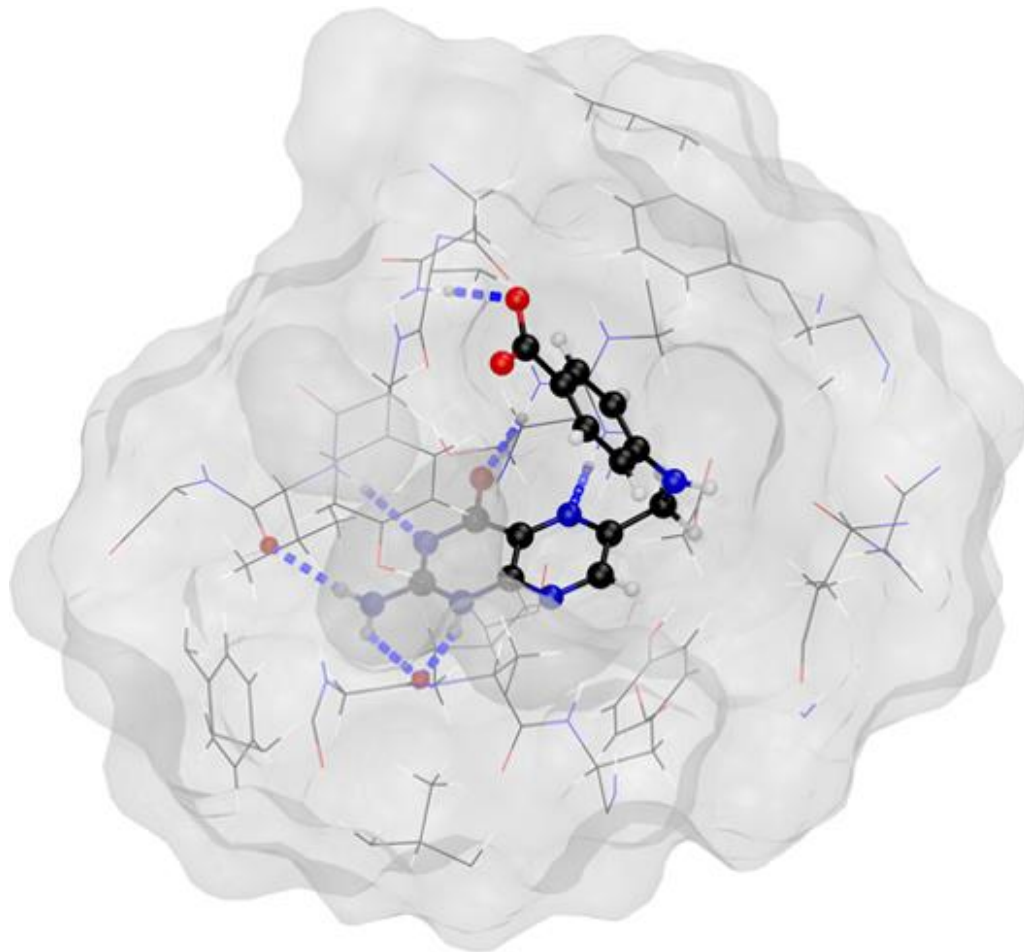
Terroranschlägen nachgesagt.^[158] In der Tat sind in der Vergangenheit mehrere entsprechende Ereignisse in den Fokus der Öffentlichkeit gerückt. Dazu zählen

- das erfolgreiche Attentat auf den bulgarischen Dissidenten Georgi Markov 1978 in London mit einem zur Waffe umgerüsteten Regenschirm (sog. „Bulgarischer Regenschirm“), mit dessen Hilfe dem Opfer eine mit Rizin präparierte Metallkugel in den Unterschenkel injiziert wurde, und
- erfolglose Briefattentate auf unter anderem den US-Präsidenten Barack Obama und den New Yorker Bürgermeister Michael Bloomberg, an die im Jahr 2013 je ein mit Rizin präparierter Brief verschickt wurde. Beide Briefe konnten jedoch von den US-Behörden abgefangen werden.

Vor diesem Hintergrund erscheint es wenig verwunderlich, dass ein erhebliches Interesse an der Entwicklung von RTA-Inhibitoren als wirksame Antidote besteht.

Die Idee, Pterine als RTA-Inhibitoren zu verwenden geht auf Arbeiten von Robertus und Mitarbeitern zurück.^[155] Es besteht die Hoffnung, dass ein besseres Verständnis der Tautomerpräferenz von an Rizin gebundenen Pterinen das rationale Design von Rizin-Inhibitoren voranbringen könnte. Pterinsäure ist ein schwacher RTA-Inhibitor. Abbildung 27 zeigt die Lage des Pterinsäure-Liganden in der Bindetasche der Rizin-A-Kette mit den in Ref. ^[155] postulierten Wasserstoffbrücken. In Abbildung 28 sind Pterinsäure und verwandte Moleküle dargestellt. Folsäure oder Pteroylglutaminsäure (Vitamin B₉, **B**) ist ein Amid aus *L*-Glutaminsäure und Pterinsäure. Die Pterinsäure selbst (**A**) besteht wiederum aus der Pterineinheit (**D**) und der über eine Methylengruppe mit ihr verbundenen *p*-Aminobenzoesäure (**C**).

A



B

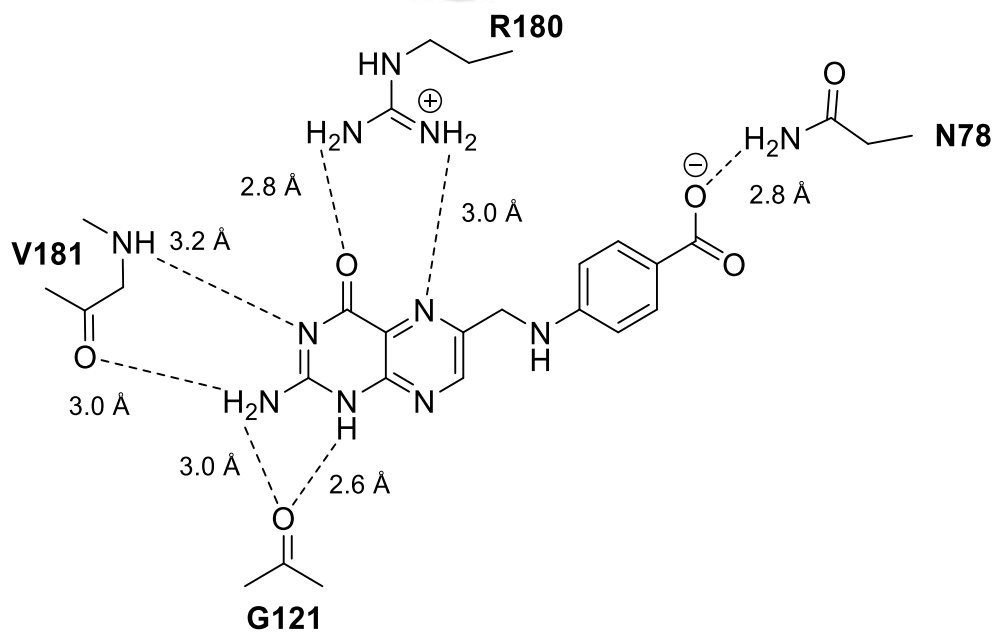


Abbildung 27: Pteroinsäure-Ligand in der Bindetasche des Rizin-A-Proteins mit den in Ref. ^[155] postulierten Wasserstoffbrücken zwischen Protein und Ligand als 3D-Darstellung (A) und Schemazeichnung (B).

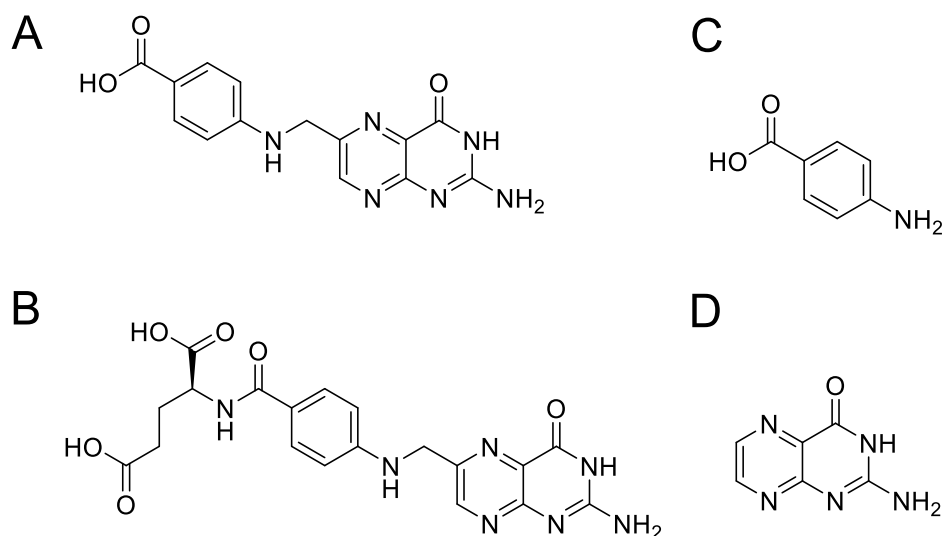


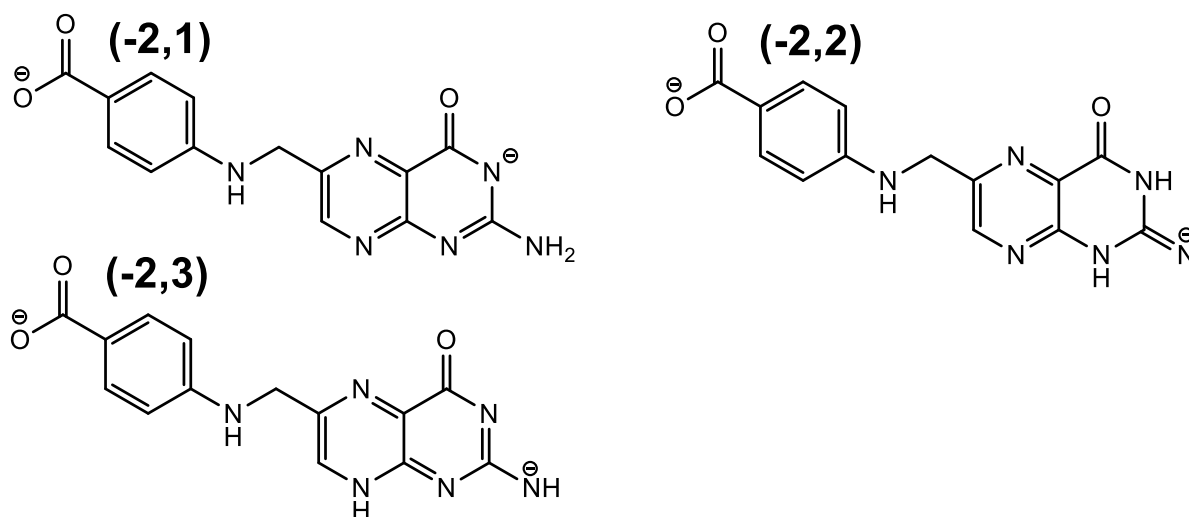
Abbildung 28: Strukturen von Pterinsäure (A), Folsäure (B), *p*-Aminobenzoesäure (C) und Pterin (D).

3.9.1 Ligand

3.9.1.1 Berechnung von Gibbs-Energien für Konformere, Tautomere und Ionisierungsstufen

Strukturen von Tautomeren des PTA-Liganden in den Ionisierungszuständen -2, -1, 0 und +1 wurden von S. Güssregen mit dem Programm MoKa 2.6^[159] erstellt. Erhalten wurden die in Abbildung 29 bis Abbildung 32 dargestellten 31 Strukturen. Die 31 Strukturen wurden in Anlehnung an vorherige Arbeiten^[160] behandelt. Zunächst wurde für jedes Pterinsäure-Tautomer ein GAFF-Kraftfeld^[95] mit AM1-BCC-Ladungen^[161,162] erzeugt. Mit diesen Kraftfeldern wurden Moleküldynamik-Simulationen über einen Zeitraum von je 0.2 μ s bei einer Temperatur von 500 K mit einer Zeitschrittlänge von 1 fs durchgeführt. Die Temperatur wurde mittels eines Andersen-Thermostats^[163] näherungsweise konstant gehalten. Um eine stabile Simulation zu gewährleisten, war es nötig, die Geschwindigkeiten der Atome alle 80 Schritte zu reinitialisieren. Lösungsmittelleffekte wurden durch das implizite Solvatationsmodell ALPB^[164,165] (ohne unpolare Beiträge) berücksichtigt. Über die Simulationsdauer wurden in gleichmäßigen Abständen von 0.1 ns je 2000 und in Abständen von 0.2 ns testweise 1000 Strukturen extrahiert und bis zum Unterschreiten eines Kriteriums von 10^{-5} kcal/(mol Å) im *RMS* des Energiegradienten aller Atome mit der *Limited-Memory*-Broyden-Fletcher-Goldfarb-Shanno- (LBFGS-) Methode^[166,167] geometrieoptimiert. Mit diesen Strukturen wurden *Clustering*-Analysen durchgeführt, bei denen die Konformere eines Tautomers mittels des Jarvis-Patrick-Algorithmus^[168] gruppiert wurden. Hierbei kam ein RMSD-Kriterium von 0.7 Å zum Einsatz. Bezüglich der Anzahl der extrahierten

Konformere zeigt Tabelle 14, dass die Anzahl der mit einem RMSD-Kriterium von 0.7 Å erhaltenen *Cluster* bei 1000 und 2000 Strukturen meist identisch ist. Für die Tautomere 1, 2 und 7 der Ionisierungsstufe +1 ist die *Cluster*-Anzahl sogar geringer als bei 1000 Strukturen. Anscheinend koaleszieren hier einige *Cluster*, neue Konformere entstehen bei Entnahme von mehr Strukturen jedoch offensichtlich nicht. Am Beispiel von Tautomer (+1,2) wurde überprüft, welchen Einfluss das RMSD-Kriterium auf die Anzahl der *Cluster* hat. Dazu wurde das *Clustering* mit RMSD-Werten von 0.6 Å, 0.7 Å und 0.8 Å durchgeführt. Für 0.6 Å wurden 150 *Cluster* erhalten, für 0.7 Å 14 und für 0.8 Å 12 *Cluster*. Je 150 *Cluster* verursachen bei quantenchemischen Berechnungen mit 31 Tautomeren einen dramatischen Rechenzeitbedarf. Insofern erscheint der Wert 0.7 Å sinnvoll. Die weitere Reduktion von 14 auf 12 *Cluster* verringert die benötigte Rechenzeit dann nicht mehr signifikant, erhöht aber die Gefahr, dass möglicherweise wichtige Konformere unberücksichtigt bleiben.



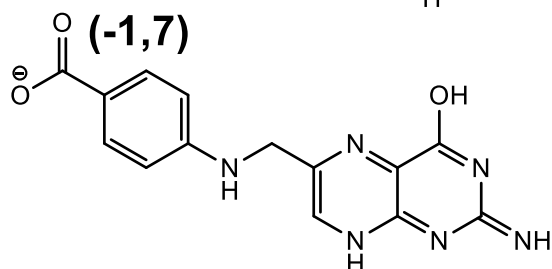
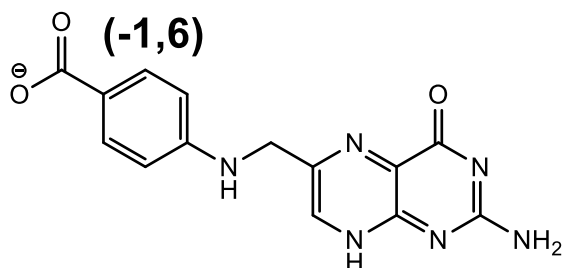
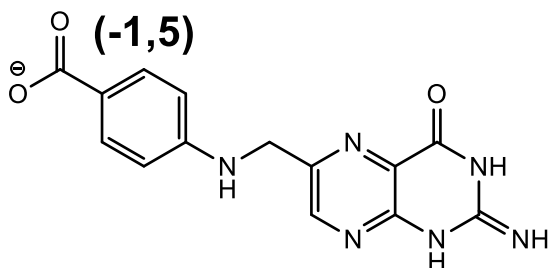
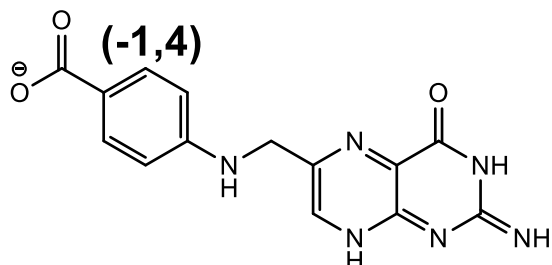
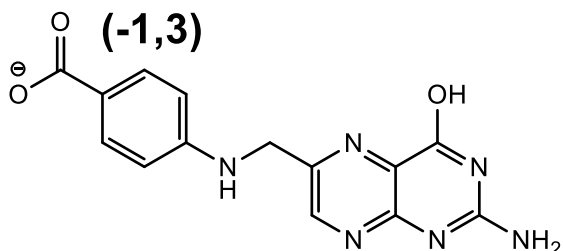
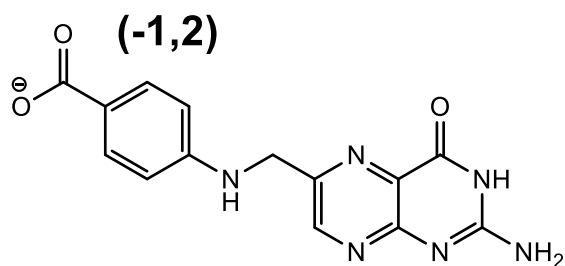
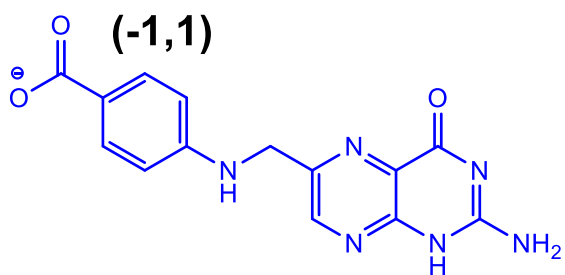
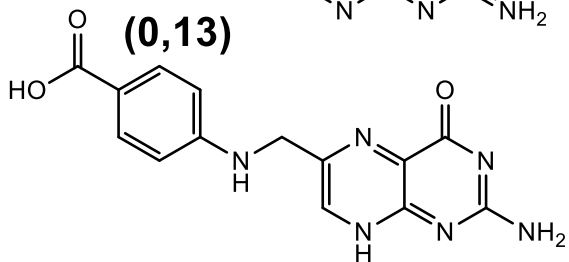
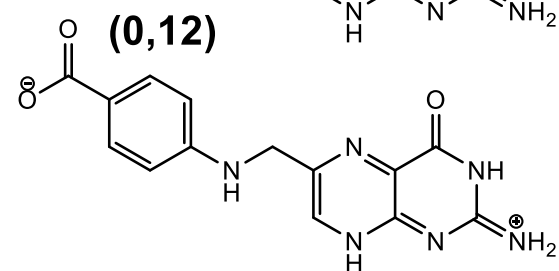
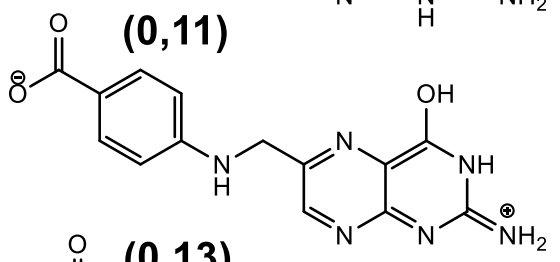
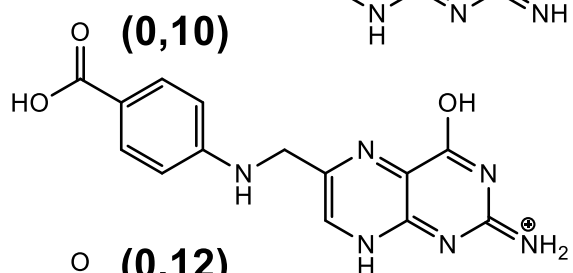
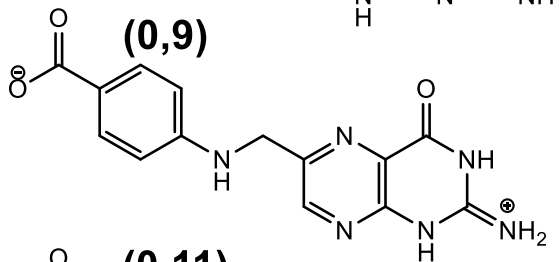
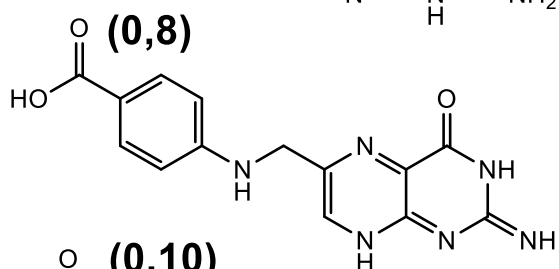
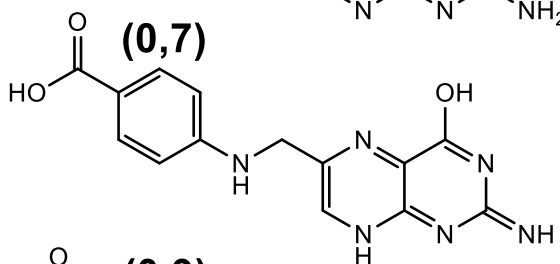
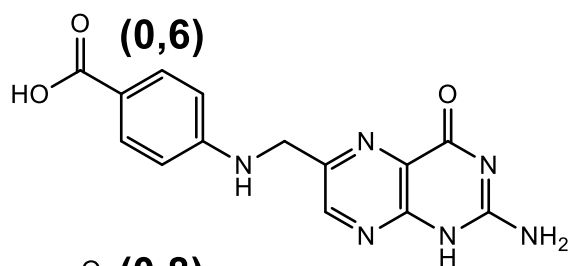
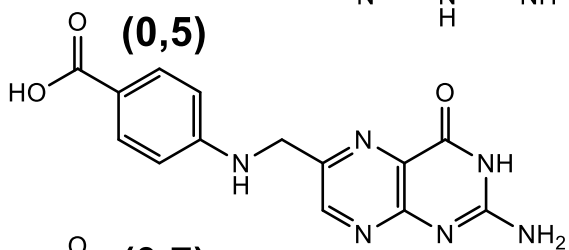
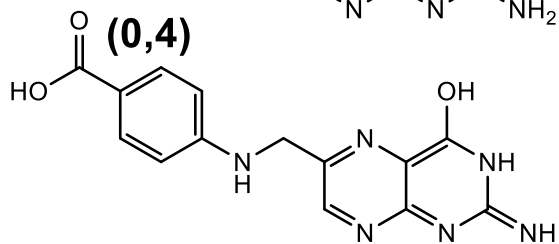
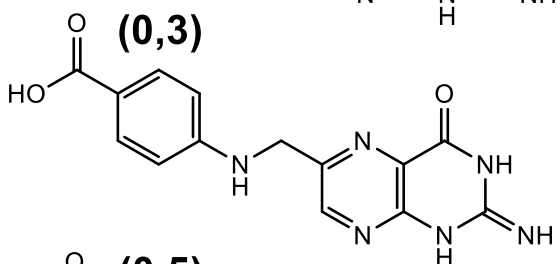
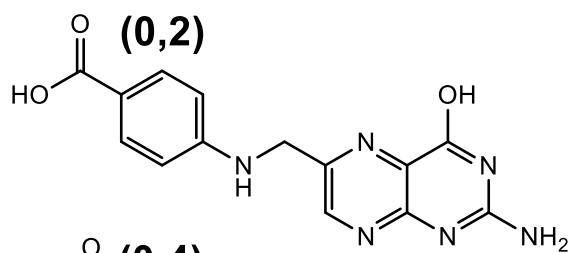
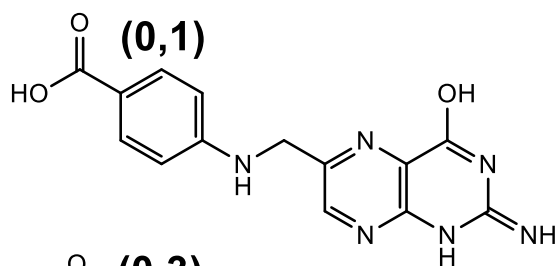
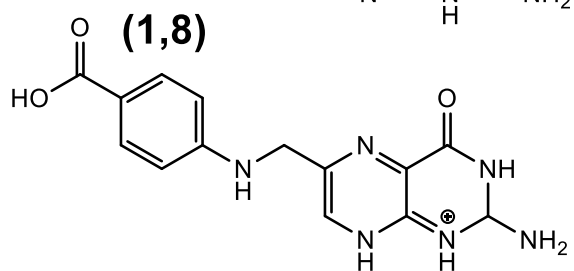
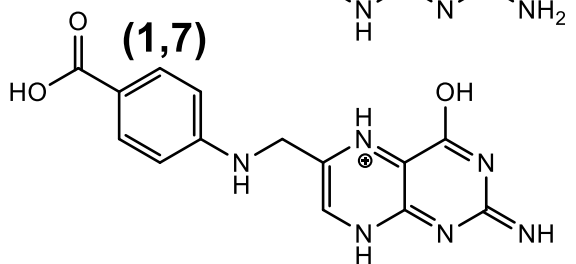
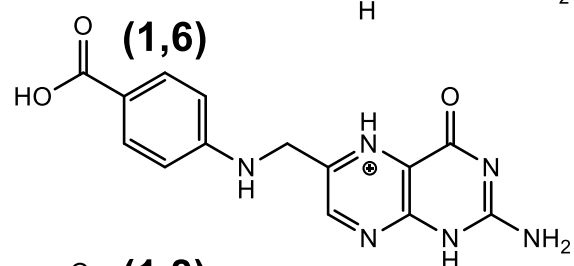
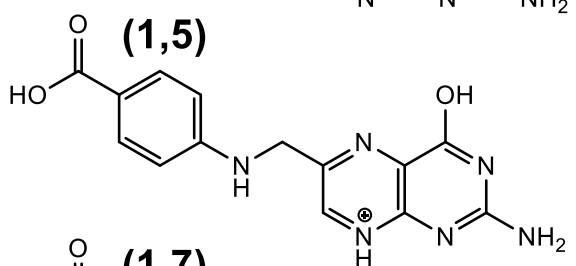
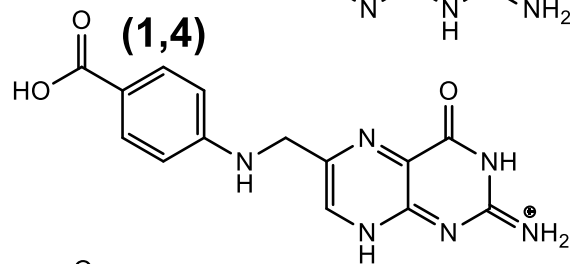
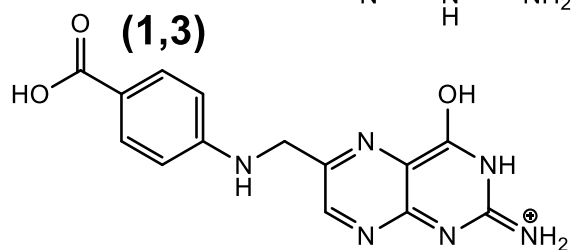
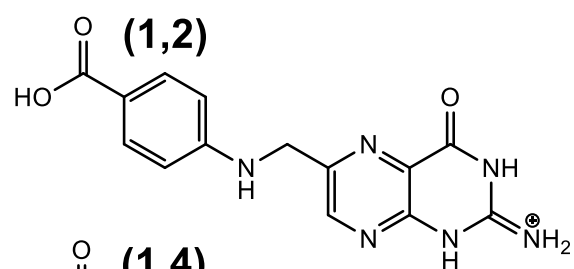
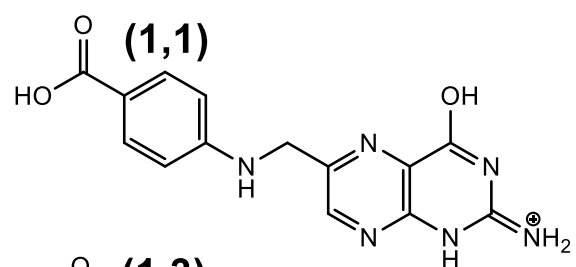


Abbildung 30: Tautomere der Pteroinsäure in der Ionisierungsstufe -1. Die von Robertus und Mitarbeitern im Kristall mit Rizin bei pH 8.9 aufgrund einer Analyse der Wasserstoffbrücken angenommene Struktur^[155] ist in blau hervorgehoben.





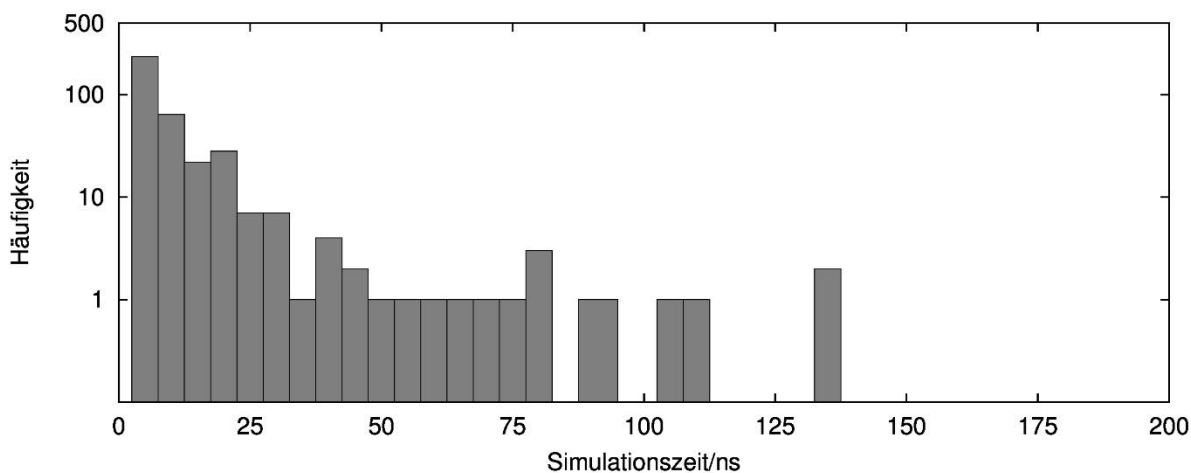


Abbildung 33: Häufigkeitsverteilung des ersten Auftretens der *Cluster*-Minima von Pteroisäure während der Simulation, akkumuliert über alle Tautomere aller Ionisierungszustände.

Für Tautomere mit deprotonierter Carboxylgruppe wurde aus jedem *Cluster* die Struktur mit der niedrigsten GAFF/ALPB-Energie extrahiert. Da von GAFF bekannt ist, dass die Carboxylgruppen bezüglich der OH-Rotamere häufig fälschlicherweise als *trans*-Konformere vorliegen^[169], wurden für Tautomere mit protonierter Carboxylgruppe sowohl die *cis*- als auch die *trans*-Form mit der niedrigsten Energie extrahiert, sofern der *Cluster* beide beinhaltet.

Ob die Länge der Simulation ausreicht wurde wie folgt überprüft: Für jedes *Cluster*-Minimum jedes Tautomers jedes Ionisierungszustands wurde ermittelt, an welcher Stelle in der Simulation die jeweilige Struktur erhalten wurde. Das resultierende Histogramm ist in Abbildung 33 dargestellt. Offensichtlich ist die Simulationslänge hinreichend, nach 150 ns treten keine neuen Konformationen mehr auf.

Mit jeder der erhaltenen Strukturen wurden eine quantenmechanische Geometrieoptimierung mit B3LYP/6-31+G*, B3LYP/6-31G* und B3LYP/6-311+G** und B3LYP/6-31+G*-D3 mit dem PCM Lösungsmittelmodell durchgeführt. Testweise wurde für B3LYP/6-31+G* zusätzlich die klassische D3-Dispersionskorrektur verwendet. Während der quantenmechanischen Berechnungen wurde für die Evaluation des Austausch-Korrelationsfunktionals mittels numerischer Integration das „UltraFine“-Gitter verwendet, bei dem der Radialanteil mit 99 Gitterpunkten, der Winkelraum mit 590 Gitterpunkten diskretisiert wird. Mit B3LYP/6-311+G** und B3LYP/6-31+G*-D3 konnten für viele Strukturen die von Gaussian09 vorgegebenen Konvergenz-Schwellenwerte (Standardeinstellungen) nicht erreicht werden. Teilweise wurden auch Konvergenzschwierigkeiten im SCF-Zyklus beobachtet, insbesondere bei zwitterionischen Spezies. Durch die Verwendung des „UltraFine“-Gitters für alle Optimierungen sowie mittels initialer Kraftkonstanten aus einer Frequenzrechnung zu Beginn des

Optimierungszyklus und in einigen Fällen in jedem Schritt konnte die Anzahl nicht konvergierter Strukturen signifikant gesenkt werden, trotzdem verblieb eine erhebliche Menge nicht konvergierter Rechnungen, so dass ein Vergleich mit B3LYP/6-31+G* nicht sinnvoll erscheint. Daher werden B3LYP/6-311+G** und B3LYP/6-31+G*-D3 im Folgenden nicht mehr betrachtet. Nach den Optimierungen wurden alle Doubletten entfernt.

Tabelle 14: Besetzungszahlen der Jarvis-Patrick-Cluster bei Entnahme von 1000 und 2000 Strukturen aus der Trajektorie.

Ionisierungszustand	Tautomer	$N_{\text{Cluster},1000}$	$N_{\text{Cluster},2000}$
-2	1	12	12
	2	12	12
	3	12	12
-1	1	16	16
	2	10	10
	3	12	12
	4	14	14
	5	16	16
	6	14	14
	7	12	12
0	1	10	10
	2	10	10
	3	10	10
	4	10	10
	5	10	10
	6	10	10
	7	10	10
	8	14	14
	9	10	10
	10	10	10
	11	13	13
	12	14	14
	13	14	14
+1	1	20	14
	2	15	14
	3	14	14
	4	14	14
	5	10	10
	6	12	12
	7	22	17
	8	14	14

Mit den optimierten Strukturen der *Cluster* wurden EC-RISM-Berechnungen entsprechend der für die Geometrieoptimierungen verwendeten Methodik auf B3LYP/6-31+G*- oder B3LYP/6-31G*-Niveau durchgeführt und aus jedem der *Cluster* mittels der Gleichungen 3.4.30 bis 3.4.37 in verschiedenen Varianten Gibbs-Energien für die Konformationsbassins, und aus diesen die Gibbs-Energien für Tautomere und die Ionisierungsstufen berechnet. Für B3LYP/6-31+G* sind die Gibbs-Energien für die Konformationsbassins in Tabelle 15 und Tabelle 16 aufgeführt, die Gibbs-Energien für Tautomere und die Ionisierungsstufen in Tabelle 17. Die in Kapitel 3.4 dargestellte Methodik der Berechnung der Gibbs-Energien mittels Zustandssummen lässt hier Spielraum bezüglich der Wahl der Entartungsfaktoren in Gl. 3.4.30. Prinzipiell ist denkbar, die *Cluster*-Populationen als Entartungsfaktoren zu verwenden oder die Entartungsfaktoren zu vernachlässigen. Wie aus Tabelle 15 und Tabelle 16 leicht ersichtlich, ist der Einfluss auf die relative Lage der Gibbs-Energien der Tautomere jedoch meist gering und die Trends sind in beiden Fällen identisch. Eine Ausnahme bildet das Verhältnis von (0,9) zu (0,12) für EC-RISM aus Tabelle 16. Der Unterschied von 0.49 kcal/mol entspricht einem Verhältnis von 0.4, 0.97 kcal/mol einem Verhältnis von 0.2.

Für B3LYP/6-31G* sind die Gibbs-Energien für die Konformationsbassins Tabelle 18 und Tabelle 19 entnehmbar, die Gibbs-Energien für Tautomere und die Ionisierungsstufen Tabelle 20. Im Gegensatz zu den Berechnungen mit dem 6-31G*-Basissatz sind hier für die neutrale Pteroinsäure und sowohl EC-RISM als auch PCM die Tautomere (0,5) und (0,6) energetisch günstiger als (0,9) und (0,12). Ansonsten sind die Trends ähnlich.

Tabelle 15: Gibbs-Energien von Pteroinsäure-Tautomeren mit B3LYP/6-31+G* und **a/b**: EC-RISM, **c/d**: PCM. Die Gibbs-Energien wurden mittels einer Zustandssumme über Konformercluster aus MD-Simulationen erhalten und die *Cluster*-Besetzungszahlen aus Tabelle 14 wurden als Entartungsfaktoren verwendet. Für **a** und **c** läuft die Zustandssumme über alle *Cluster*, für **b** und **d** sind die Gibbs-Energien der Minimum-Gibbs-Energie-*Cluster* jedes Tautomers angegeben.

Ionisierungs- zustand	Tautomer	G/(kcal mol ⁻¹) ^a	G/(kcal mol ⁻¹) ^b	G/(kcal mol ⁻¹) ^c	G/(kcal mol ⁻¹) ^d
-2	1	-686622.23	-686618.11	-686561.26	-686556.96
	2	-686618.23	-686614.02	-686548.54	-686544.20
	3	-686613.36	-686609.08	-686535.79	-686531.78
-1	1	-686883.14	-686879.13	-686849.95	-686845.58
	2	-686884.00	-686879.99	-686852.31	-686848.21
	3	-686872.39	-686868.74	-686844.70	-686840.43
	4	-686874.64	-686870.58	-686837.75	-686833.55
	5	-686877.49	-686873.40	-686845.86	-686841.45
	6	-686880.65	-686876.57	-686843.46	-686839.27
	7	-686860.52	-686856.75	-686821.47	-686817.29
0	1	-687117.84	-687113.72	-687121.38	-687116.93
	2	-687124.81	-687120.63	-687134.72	-687130.39
	3	-687130.14	-687126.06	-687135.82	-687131.54
	4	-687112.29	-687108.17	-687113.30	-687108.96
	5	-687136.54	-687132.52	-687142.64	-687138.41
	6	-687135.96	-687131.87	-687140.01	-687135.72
	7	-687112.38	-687108.25	-687110.88	-687106.78
	8	-687126.91	-687123.18	-687127.13	-687123.07
	9	-687139.56	-687135.39	-687124.63	-687120.37
	10	-687130.95	-687126.88	-687119.51	-687115.29
	11	-687122.91	-687118.78	-687110.92	-687106.74
	12	-687140.04	-687135.93	-687124.94	-687121.03
	13	-687133.34	-687129.42	-687132.86	-687128.78
+1	1	-687382.52	-687378.44	-687408.67	-687404.31
	2	-687391.86	-687387.68	-687414.03	-687409.69
	3	-687377.00	-687372.92	-687400.42	-687396.07
	4	-687392.81	-687388.68	-687414.02	-687409.85
	5	-687381.47	-687377.13	-687405.45	-687401.03
	6	-687390.40	-687386.02	-687411.73	-687407.26
	7	-687360.18	-687356.25	-687373.56	-687369.10
	8	-687386.84	-687382.65	-687400.94	-687396.62

Tabelle 16: Vergleich der relativen Lage der Gibbs-Energien der Tautomere von Pteroinsäure mit **(a, c)** und ohne **(b, d)** *Cluster*-Populationen als Entartungsfaktoren in Gl. 3.4.30 für B3LYP/6-31+G* mit EC-RISM **(a, b)** und PCM **(c, d)** und. Gibbs-Energien sind relativ zum energetisch günstigsten Tautomer des jeweiligen Ionisierungszustands.

Ion.-Zust.	Tautomer	$\Delta G/(\text{kcal mol}^{-1})^a$	$\Delta G/(\text{kcal mol}^{-1})^b$	$\Delta G/(\text{kcal mol}^{-1})^c$	$\Delta G/(\text{kcal mol}^{-1})^d$
-2	1	0.00	0.00	0.00	0.00
	2	4.00	3.98	12.72	12.64
	3	8.88	8.85	25.48	25.02
-1	1	0.87	0.75	2.36	2.12
	2	0.00	0.00	0.00	0.00
	3	11.61	11.63	7.61	7.48
	4	9.37	9.13	14.56	14.27
	5	6.52	6.39	6.45	6.20
	6	3.36	3.42	8.85	8.57
	7	23.49	23.26	30.84	30.68
0	1	22.20	22.76	21.26	20.98
	2	15.23	15.79	7.92	7.80
	3	9.90	10.00	6.82	6.74
	4	27.74	28.40	29.34	29.21
	5	3.50	3.62	0.00	0.00
	6	4.08	4.22	2.63	2.57
	7	27.66	27.65	31.76	31.28
	8	13.12	12.95	15.51	14.90
	9	0.48	0.97	18.01	17.70
	10	9.09	9.75	23.13	22.75
	11	17.13	17.23	31.72	31.22
	12	0.00	0.00	17.70	17.08
	13	6.70	6.60	9.78	9.20
+1	1	10.28	10.78	5.36	5.36
	2	0.95	0.86	0.00	0.15
	3	15.81	16.37	13.61	13.77
	4	0.00	0.00	0.01	0.00
	5	11.33	11.62	8.58	8.89
	6	2.41	2.68	2.30	2.64
	7	32.62	32.39	40.46	40.78
	8	5.96	5.99	13.08	13.12

Tabelle 17: Gibbs-Energien von Pteroinsäure-Ionisierungszuständen mit B3LYP/6-31+G* und **a/b**: EC-RISM, **c/d**: PCM. Die Gibbs-Energien wurden mittels einer Zustandssumme über Tautomere und Konformerencluster aus MD-Simulationen erhalten und die *Cluster*-Besetzungszahlen wurden als Entartungsfaktoren verwendet. Für **a** und **c** läuft die Zustandssumme über alle Tautomere aller *Cluster*, für **b** und **d** sind die Gibbs-Energien der Minimum-Gibbs-Energie-Tautomere jedes Ionisierungszustands angegeben.

Ionisierungszustand	G/(kcal mol⁻¹)^a	G/(kcal mol⁻¹)^b	G/(kcal mol⁻¹)^c	G/(kcal mol⁻¹)^d
-2	-686622.23	-686622.23	-686561.26	-686561.26
-1	-686884.13	-686884.00	-686852.32	-686852.31
0	-687140.26	-687140.04	-687142.65	-687142.64
+1	-687392.92	-687392.81	-687414.44	-687414.03

Tabelle 18: Gibbs-Energien von Pteroinsäure-Tautomeren mit B3LYP/6-31G* und **a/b**: EC-RISM, **c/d**: PCM. Die Gibbs-Energien wurden mittels einer Zustandssumme über Konformercluster aus MD-Simulationen erhalten und die Cluster-Besetzungszahlen aus Tabelle 14 wurden als Entartungsfaktoren verwendet. Für **a** und **c** läuft die Zustandssumme über alle Cluster, für **b** und **d** sind die Gibbs-Energien der Minimum-Gibbs-Energie-Cluster jedes Tautomers angegeben.

Ionisierungs- zustand	Tautomer	G/(kcal mol ⁻¹) ^a	G/(kcal mol ⁻¹) ^b	G/(kcal mol ⁻¹) ^c	G/(kcal mol ⁻¹) ^d
-2	1	-686555.19	-686550.99	-686504.03	-686499.77
	2	-686548.51	-686544.39	-686490.39	-686486.11
	3	-686541.08	-686536.84	-686476.24	-686471.82
-1	1	-686829.21	-686825.09	-686802.33	-686798.05
	2	-686831.40	-686827.35	-686805.79	-686801.55
	3	-686820.93	-686816.92	-686798.79	-686794.36
	4	-686820.03	-686815.91	-686790.29	-686786.17
	5	-686824.59	-686820.49	-686799.46	-686795.11
	6	-686824.84	-686820.72	-686794.64	-686790.50
	7	-686805.38	-686801.62	-686774.08	-686770.06
0	1	-687083.23	-687079.05	-687089.51	-687085.08
	2	-687091.49	-687087.32	-687103.05	-687098.73
	3	-687095.56	-687091.44	-687103.89	-687099.66
	4	-687076.87	-687072.68	-687081.17	-687076.87
	5	-687102.29	-687098.13	-687110.43	-687106.23
	6	-687100.30	-687096.11	-687106.90	-687102.69
	7	-687076.21	-687071.89	-687077.88	-687073.79
	8	-687090.93	-687086.98	-687094.16	-687089.96
	9	-687095.66	-687091.52	-687084.13	-687079.88
	10	-687087.94	-687083.90	-687079.79	-687075.68
	11	-687079.48	-687075.40	-687070.94	-687066.95
	12	-687095.73	-687091.61	-687084.17	-687080.43
	13	-687096.10	-687092.09	-687098.57	-687094.27
+1	1	-687357.35	-687353.23	-687383.10	-687378.83
	2	-687365.53	-687361.34	-687387.83	-687383.56
	3	-687351.23	-687347.17	-687374.65	-687370.25
	4	-687366.03	-687361.95	-687387.58	-687383.39
	5	-687355.49	-687351.27	-687379.20	-687374.95
	6	-687363.67	-687359.30	-687385.61	-687381.14
	7	-687331.17	-687327.03	-687346.64	-687342.17
	8	-687358.15	-687354.03	-687372.81	-687368.60

Tabelle 19: Vergleich der relativen Lage der Gibbs-Energien der Tautomere von Pteroinsäure mit (a, c) und ohne (b, d) *Cluster*-Populationen als Entartungsfaktoren in Gl. 3.4.30 für B3LYP/6-31G* mit EC-RISM (a, b) und PCM (c, d). Gibbs-Energien sind relativ zum energetisch günstigsten Tautomer des jeweiligen Ionisierungszustands.

Ion.-Zust.	Tautomer	$\Delta G/(\text{kcal mol}^{-1})^a$	$\Delta G/(\text{kcal mol}^{-1})^b$	$\Delta G/(\text{kcal mol}^{-1})^c$	$\Delta G/(\text{kcal mol}^{-1})^d$
-2	1	0.00	0.00	0.00	0.00
	2	6.68	6.60	13.64	13.66
	3	14.11	14.15	27.79	27.95
-1	1	2.19	2.26	3.46	3.50
	2	0.00	0.00	0.00	0.00
	3	10.47	10.43	7.00	7.19
	4	11.37	11.44	15.50	15.38
	5	6.81	6.86	6.33	6.44
	6	6.56	6.63	11.15	11.05
	7	26.02	25.73	31.71	31.49
0	1	19.06	19.08	20.92	21.15
	2	10.80	10.81	7.38	7.50
	3	6.73	6.69	6.54	6.57
	4	25.42	25.45	29.26	29.36
	5	0.00	0.00	0.00	0.00
	6	1.99	2.02	3.53	3.54
	7	26.08	26.24	32.55	32.44
	8	11.36	11.15	16.27	16.27
	9	6.63	6.61	26.30	26.35
	10	14.35	14.23	30.64	30.55
	11	22.81	22.73	39.49	39.28
	12	6.56	6.52	26.26	25.80
	13	6.19	6.04	11.86	11.96
+1	1	8.68	8.72	4.73	4.73
	2	0.50	0.61	0.00	0.00
	3	14.80	14.78	13.18	13.31
	4	0.00	0.00	0.25	0.17
	5	10.54	10.68	8.63	8.61
	6	2.36	2.65	2.22	2.42
	7	34.86	34.92	41.19	41.39
	8	7.88	7.92	15.02	14.96

Tabelle 20: Gibbs-Energien von Pteroinsäure-Ionisierungszuständen mit B3LYP/6-31G* und **a/b**: EC-RISM, **c/d**: PCM. Die Gibbs-Energien wurden mittels einer Zustandssumme über Tautomere und Konformercluster aus MD-Simulationen erhalten und die Cluster-Besetzungszahlen wurden als Entartungsfaktoren verwendet. Für **a** und **c** läuft die Zustandssumme über alle Tautomere aller Cluster, für **b** und **d** sind die Gibbs-Energien der Minimum-Gibbs-Energie-Tautomere jedes Ionisierungszustands angegeben.

Ionisierungszustand	G/(kcal mol ⁻¹) ^a	G/(kcal mol ⁻¹) ^b	G/(kcal mol ⁻¹) ^c	G/(kcal mol ⁻¹) ^d
-2	-686555.19	-686555.19	-686504.03	-686504.03
-1	-686831.42	-686831.40	-686805.79	-686805.79
0	-687102.31	-687102.29	-687110.43	-687110.43
+1	-687366.25	-687366.03	-687388.14	-687387.83

3.9.1.2 pH-abhängige Speziesverteilung

Auf der Basis der Gibbs-Energien der Pteroinsäure in den verschiedenen Ionisierungsstufen und Tautomeren sowie der Gibbs-Energie des Protons lässt sich wie für das Histamin die pH-abhängige Speziesverteilung der Pteroinsäure berechnen. Mit Hilfe von Gl. 3.4.18 wurden für B3LYP/6-31+G* mit EC-RISM Werte für die Gibbs-Energie des Protons durch Anpassen an die experimentellen pK_a -Werte 2.29, 4.8 und 8.14 ermittelt. Diese pK_a -Werte wurden zwar bei einer Ionenstärke von 0.05 M gemessen, der Vergleich der pK_a -Werte für Pteroinsäure in den Tabellen 1 und 3 in Ref. ^[147] zeigt jedoch, dass die Abhängigkeit der pK_a -Werte von der Ionenstärke recht schwach ist. Die intrinsische Gibbs-Energie des Protons beträgt hier -249.97 kcal/mol, damit berechnet man nach Gl. 3.4.16 die pK_a -Werte 1.97, 4.51 und 8.74. Wie man Abbildung 34 entnehmen kann, passen dann die berechneten Verteilungen der Ionisierungszustände sehr gut zu auf der Basis der experimentellen pK_a -Werte berechneten Verteilungen. Mit dem Galvani-Potential von ca. -13 kcal/mol kommt man somit mit ca. -263 kcal/mol noch etwas näher als für das Histamin an die experimentell bestimmte Gibbs-Energie des Protons von -265.9 kcal/mol. Die mit dem PCM-Solvatationsmodell berechneten Verteilungen erscheinen dagegen nicht sinnvoll und bereits die intrinsische Gibbs-Energie des Protons ist mit -277.56 kcal/mol deutlich zu klein, die pK_a -Werte aus der Anpassung der Gibbs-Energien von -4.14, 9.40 und 9.96 weichen stark vom Experiment ab. Die Speziesverteilung von Pteroinsäure ist in Abbildung 35 dargestellt.

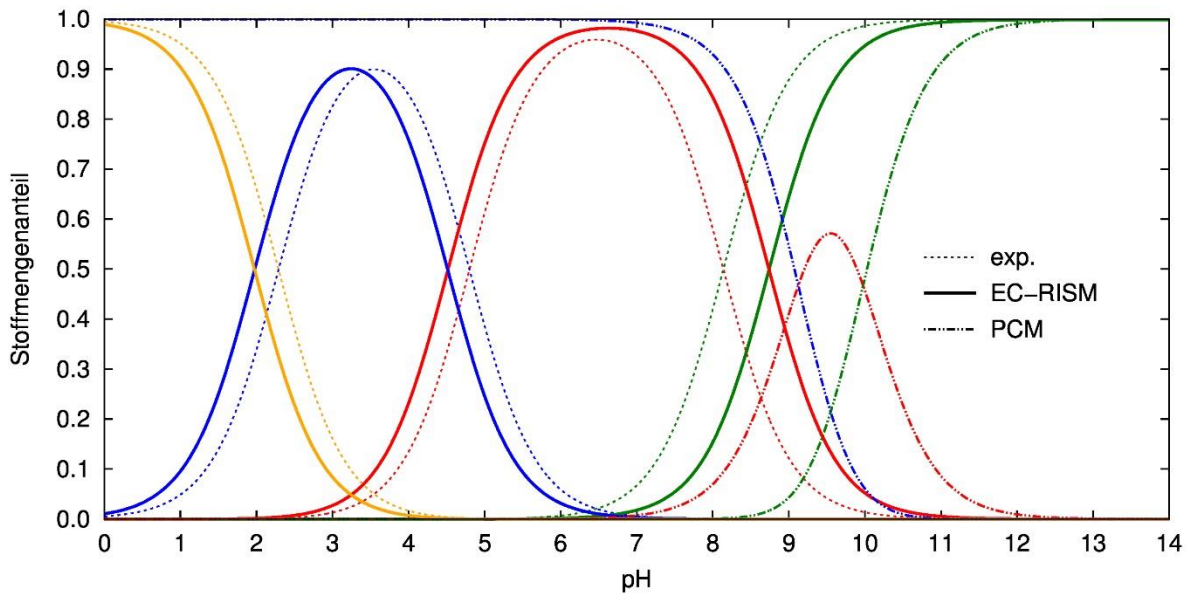


Abbildung 34: Stoffmengenanteile der einzelnen Ionisierungsstufen für experimentelle pK_a -Werte und mit B3LYP/6-31+G* und EC-RISM bzw. PCM berechnete. Die Farben kennzeichnen die Ionisierungsstufen +1 (gelb), 0 (blau), -1 (rot) und -2 (grün).

Der Grund für das merkwürdige Verhalten von PCM wird bei einem Blick auf Abbildung 36 und Abbildung 37 deutlich. Das PCM-Modell präferiert für neutrales PTA Tautomere mit protonierter Carboxylgruppe wie (0,5/6), EC-RISM zwitterionische Tautomere mit deprotonierter Carboxylgruppe und Protonierung am Pterinrest, wie (0,9/12).

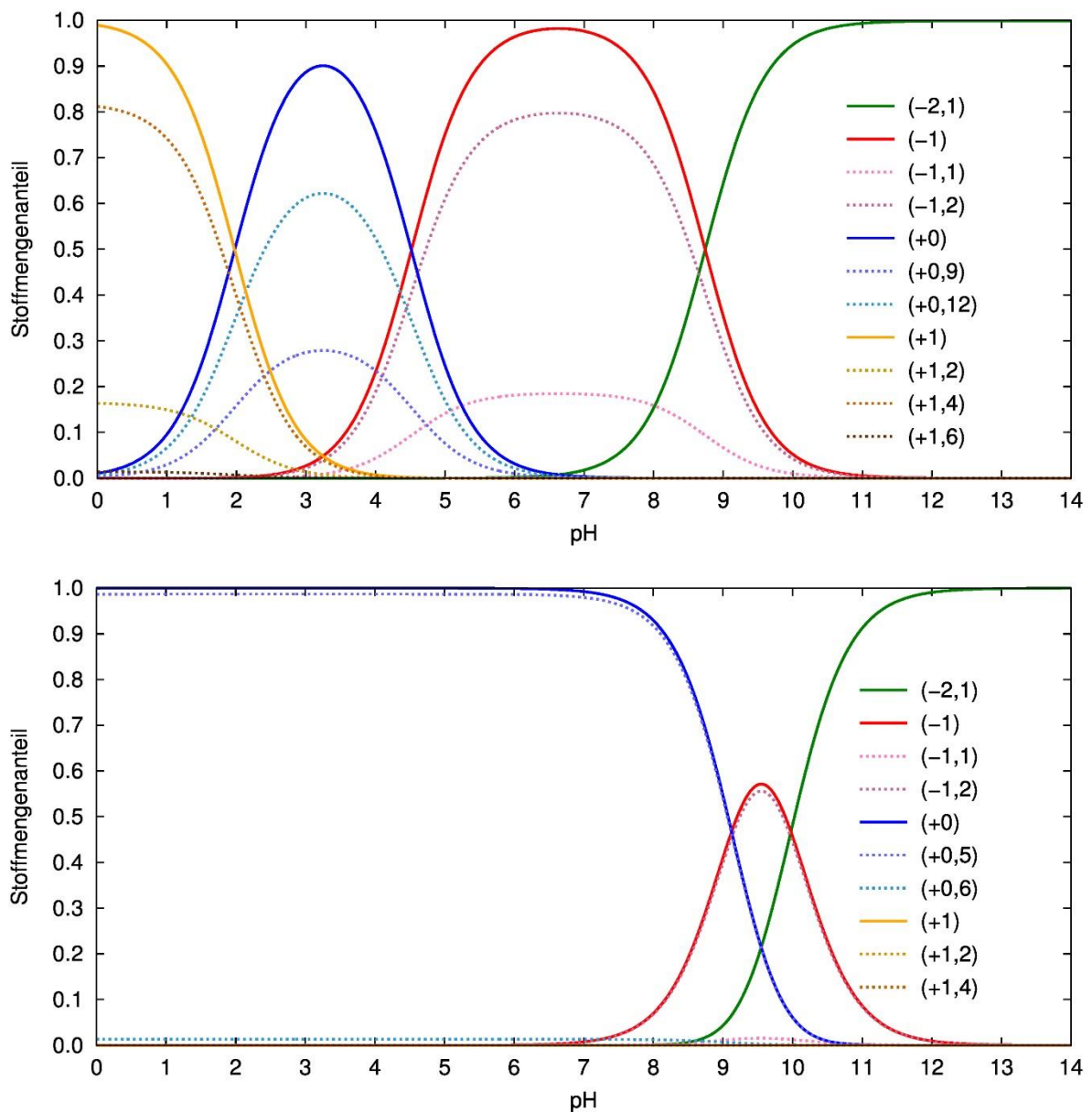


Abbildung 35: Die Speziesverteilungsdiagramme für EC-RISM/PSE-2 (oben) und PCM (unten), jeweils mit B3LYP/6-31+G* gerechnet, zeigen Stoffmengenanteile der verschiedenen Ionisierungsstufen bzw. Tautomere von PTA.

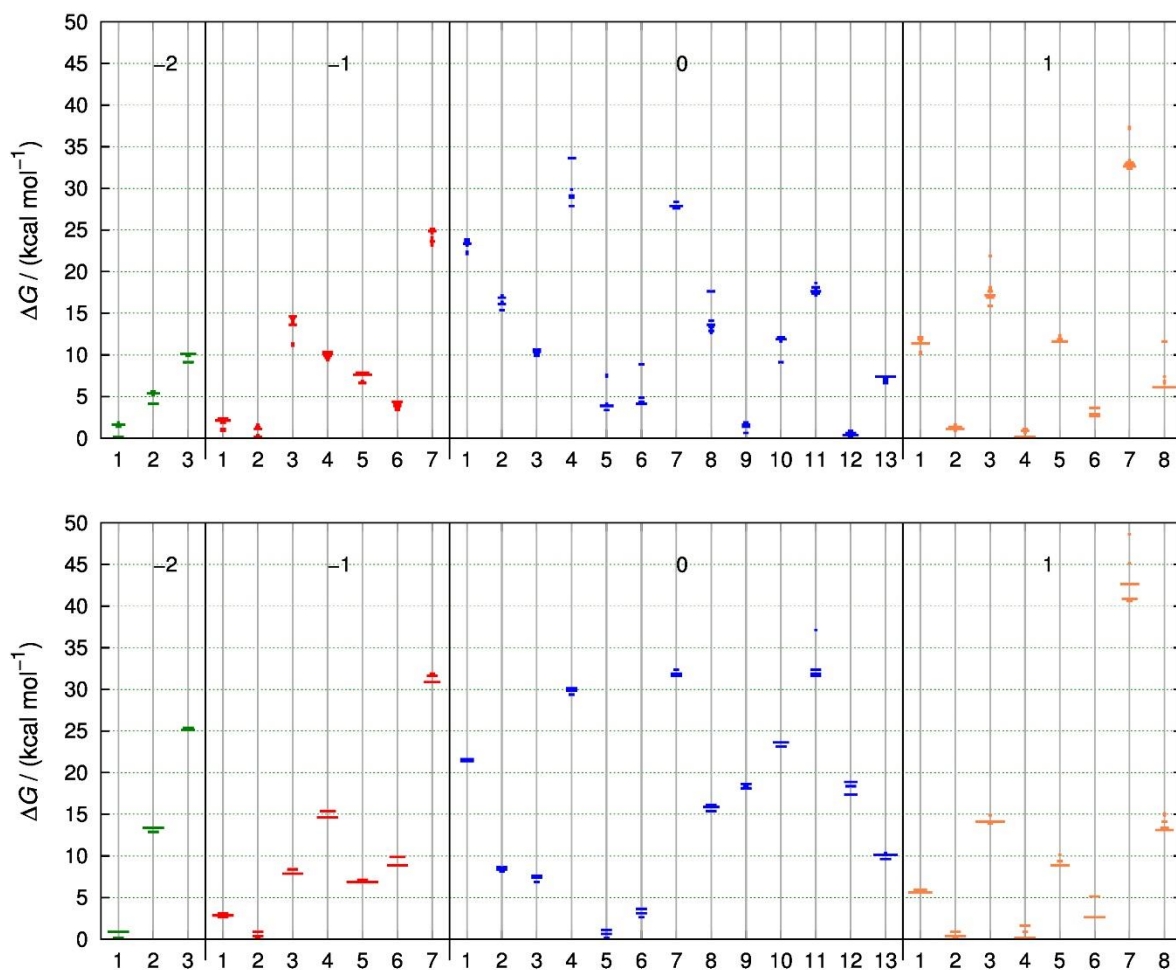


Abbildung 36: Vertikale Histogramme der Energie der einzelnen Tautomere von Pteroinsäure in den Ionisierungsstufen -2 bis +1 für B3LYP/6-31+G* und EC-RISM (oben) und PCM (unten). Die Gibbs-Energien sind als relativ zur günstigsten Konformation des günstigsten Tautomers der jeweiligen Ionisierungsstufe zu verstehen. Die Breite der Histogrammbehälter beträgt $0.25 \text{ kcal mol}^{-1}$. Auf der X-Achse sind die Tautomeren aufnummeriert, d.h. zum Beispiel das erste Histogramm auf der linken Seite gehört zur Spezies (-2,1). Die zugehörigen 2D-Strukturen sind Abbildung 29 bis Abbildung 32 zu entnehmen.

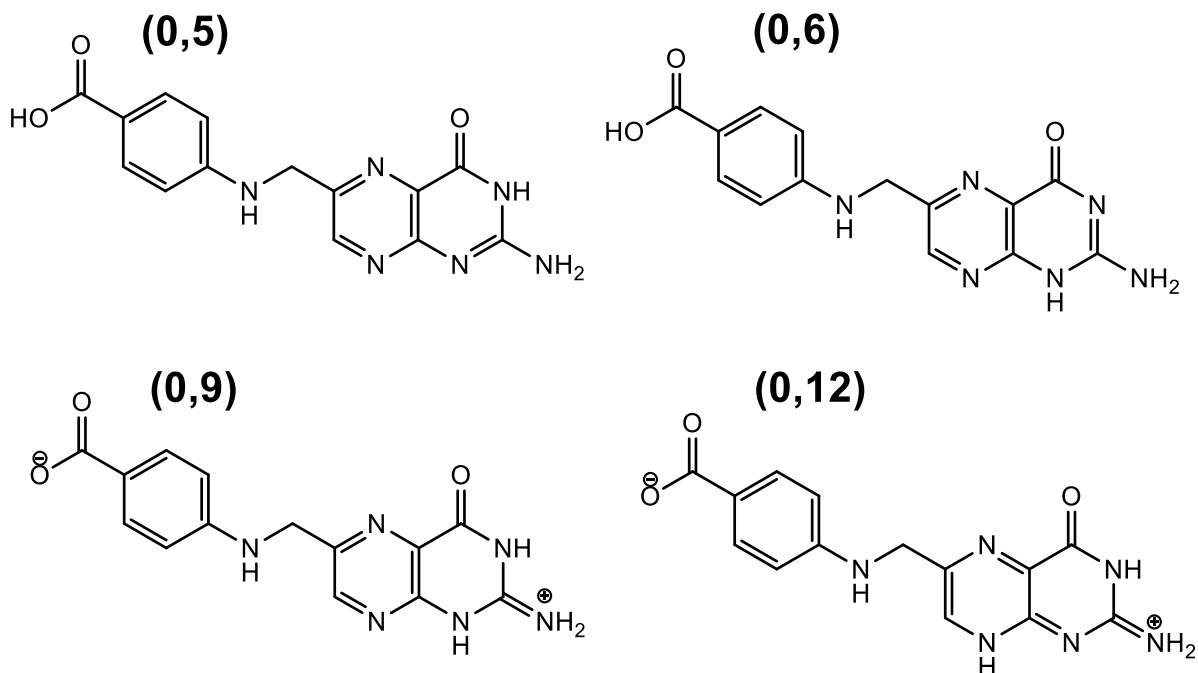
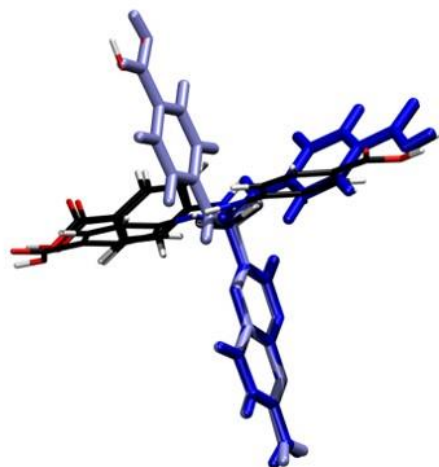


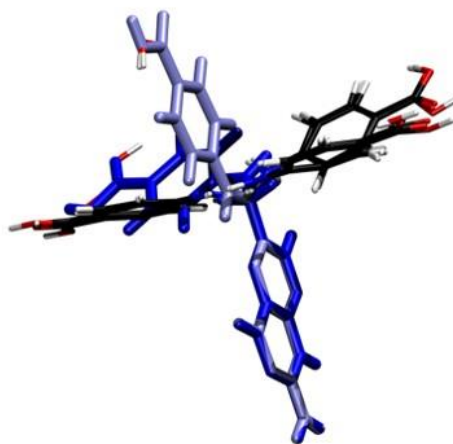
Abbildung 37: Jeweils günstigste Tautomere der neutralen Pteroinsäure für B3LYP/6-31+G* mit PCM (0,5/6) und EC-RISM (0,9/12).

Eine Analyse der konformationellen Bassins der für die pK_a -Berechnung relevanten Tautomere des Ionisierungszustands 0 zeigt, dass EC-RISM und PCM unterschiedliche Konformationen bevorzugen. Diese Konformationen sind in Abbildung 38 dargestellt. Für EC-RISM sind bei (0,9) und (0,12), gestreckte Konformationen, bei (0,5) und (0,6), geknickte Konformationen favorisiert. Bei PCM ist es genau umgekehrt.

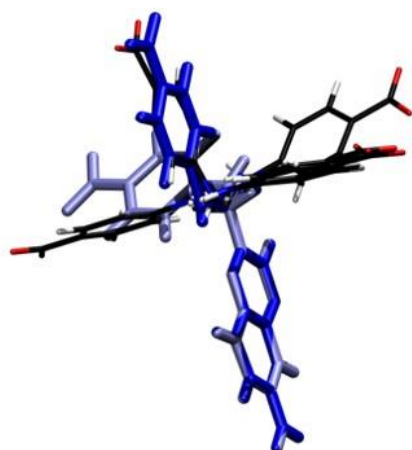
(0,5)



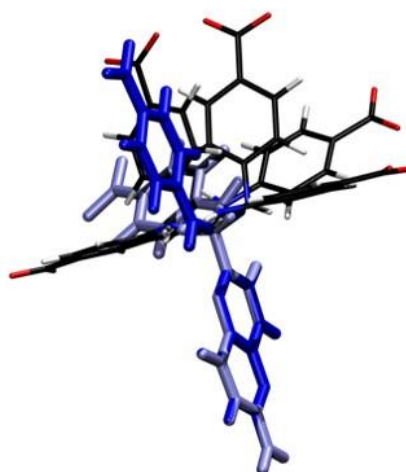
(0,6)



(0,9)



(0,12)



Der kleinere Basissatz 6-31G*, der keine diffusen Basisfunktionen besitzt, ermöglicht deutlich schnellere Berechnungen. Hier erhält man für EC-RISM eine intrinsische Gibbs-Energie des Protons von -263.43 kcal/mol, die damit nach Gl. 3.4.16 berechneten pK_a -Werte sind 0.37, 5.47 und 9.38. Die

mit dem PCM-Solvatationsmodell berechneten Verteilungen erscheinen dagegen nicht sinnvoll und bereits die intrinsische Gibbs-Energie des Protons ist mit -287.78 kcal/mol deutlich zu klein. Die pK_a -Werte aus der Anpassung der Gibbs-Energien von -7.38 , 10.25 und 12.37 weichen stark vom Experiment ab. Die resultierenden Speziesverteilungsdiagramme sind in Abbildung 39 gezeigt. Überlagert man, wie in Abbildung 40 dargestellt, die Lage der Ionisierungszustände mit experimentellen Werten, und vergleicht das Ergebnis mit Abbildung 34, so erkennt man, dass die Ergebnisse mit EC-RISM und B3LYP/6-31G* verglichen mit B3LYP/6-31+G* erheblich schlechter mit experimentellen Befunden übereinstimmen. Diffuse Basisfunktionen sind offensichtlich essentiell. Ergebnisse für PCM und B3LYP/6-31G* sind nicht sinnvoll.

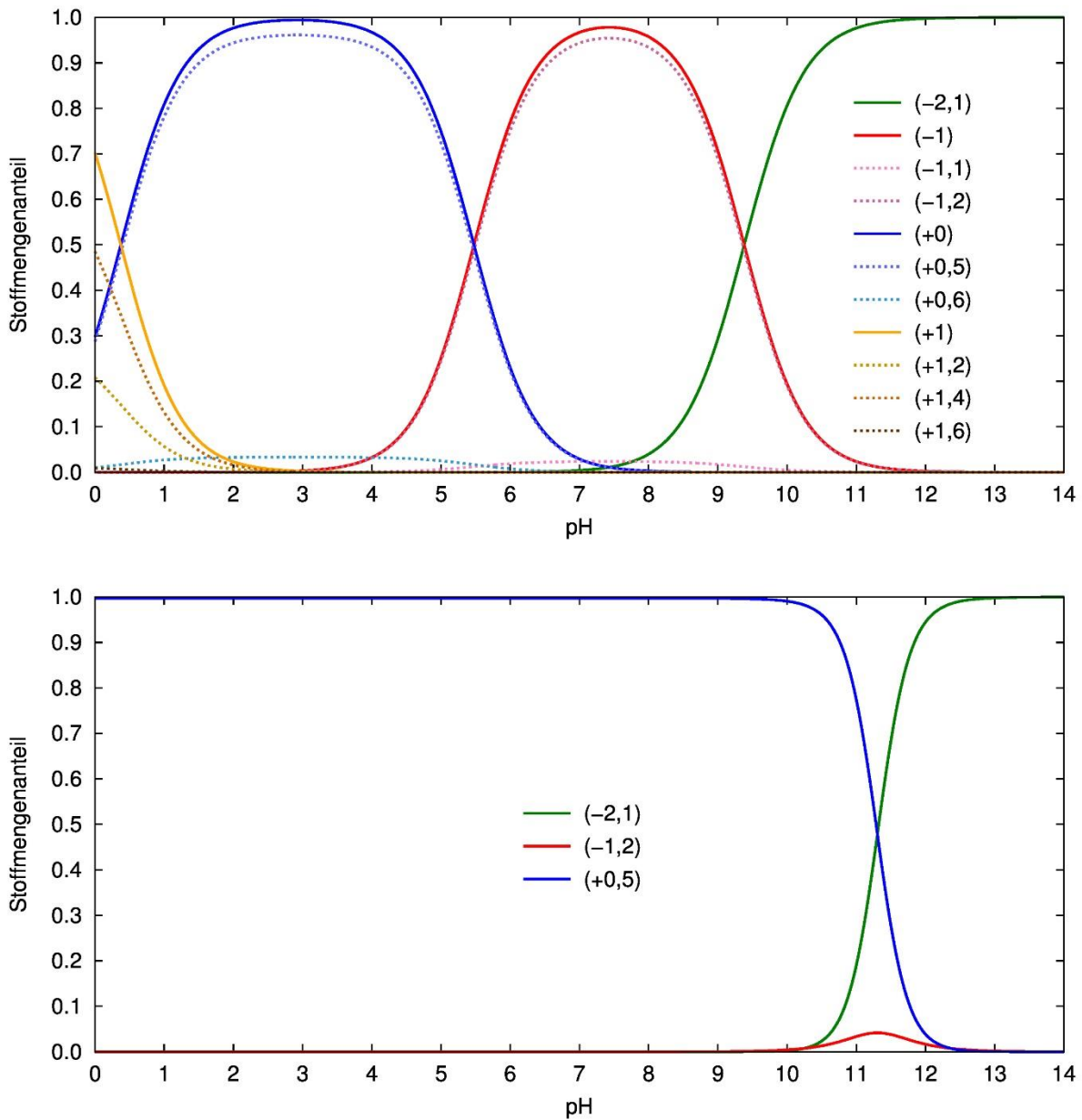


Abbildung 39: Die Speziesverteilungsdiagramme für B3LYP/6-31G* mit EC-RISM/PSE-2 (oben) und PCM (unten), zeigen Stoffmengenanteile der verschiedenen Ionisierungsstufen bzw. Tautomere von Pteroinsäure.

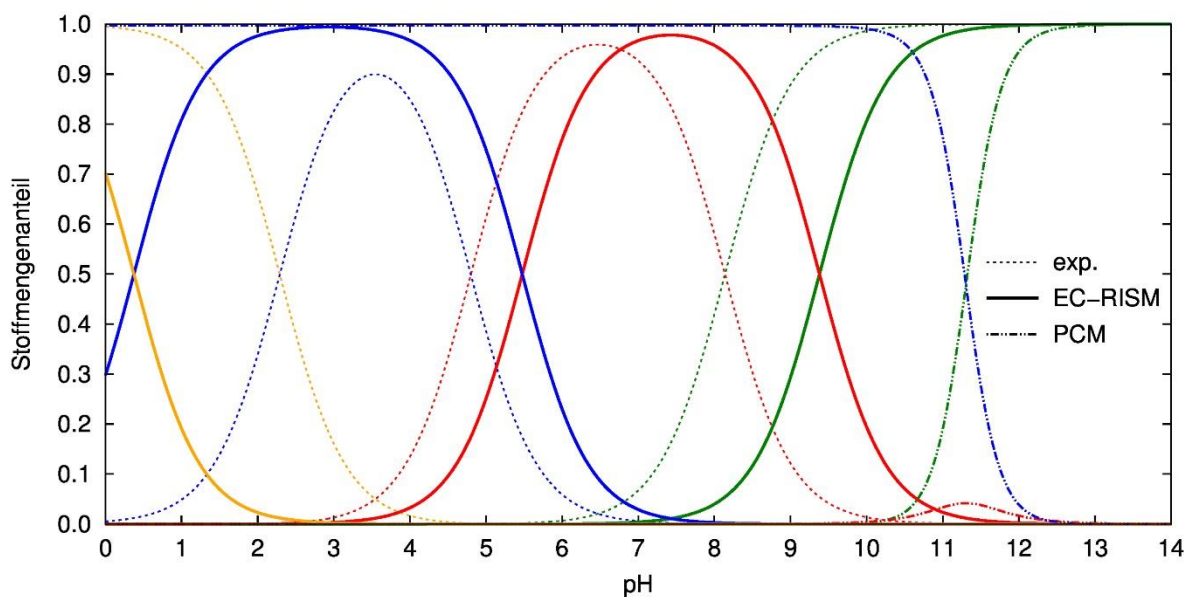


Abbildung 40: Stoffmengenanteile der einzelnen Ionisierungsstufen für experimentelle pK_a -Werte und mit B3LYP/6-31G* und EC-RISM bzw. PCM berechnete. Die Farben kennzeichnen die Ionisierungsstufen +1 (gelb), 0 (blau), -1 (rot) und -2 (grün).

3.9.1.3 Zuordnung der pK_a -Werte zu Ionisierungsstufen

Die Zuordnung der experimentellen pK_a -Werte zu den einzelnen Ionisierungsstufen wird z.B. von Szakács und Noszál ausgelassen, jedoch sollte die im Folgenden skizzierte Zuordnung plausibel erscheinen. Bezüglich saurer funktioneller Gruppen existieren in PTA eine Carbonsäurefunktionalität und eine gegebenenfalls durch Tautomerisierung erzeugbare phenolische Hydroxygruppe am Pterinrest. Der den phenolischen Sauerstoff tragende Ring ist analog zu Isocytosin, von dem die pK_a -Werte 4.0 (Übergang 0/+1) und 9.6 (Übergang -1/0) aus Messungen auf Basis von UV-Spektroskopie bekannt sind.^[170] Unter Betrachtung des Einflusses typischer Substituenteneffekte sind dramatische Änderungen dieser prinzipiellen Perspektive nicht zu erwarten, und unter stark basischen wässrigen Bedingungen sollten somit, ausgehend von der Neutralform von PTA, mindestens zwei Protonen abspaltbar sein. Alle anderen Protonen sind an Kohlenstoffatome gebunden, die kaum C-H-Azidität aufweisen, oder aber an die verschiedenen Stickstoffatome in PTA. Betrachtet man die 2D-Strukturen der jeweils energetisch günstigsten Spezies der Ionisierungszustände -1 und -2 in Abbildung 36, so zeigt sich sowohl für EC-RISM als auch PCM, dass beim Übergang von Ionisierungszustand -2 zu -1 die Erhaltung der exozyklischen primären Aminofunktionalität energetisch bevorzugt ist. Zur Verdeutlichung sind die beteiligten Spezies in Abbildung 41 noch einmal dargestellt. Um die dort lokalisierten Protonen oder alternativ die der aromatischen sekundären Aminogruppe im *p*-Aminobenzoessäurefragment abzuspalten, sollten metallorganische Reagenzien nötig sein, in wässriger Umgebung ist hiermit nicht

zu rechnen. Dies illustriert auch der pK_a -Wert von (neutralem) Ammoniak als Säure von 32.5.^[171] Dazu kommt, dass die beiden in PTA vorliegenden konjugierten π -Systeme im Ionisierungszustand -2 offensichtlich bereits je eine negative Elementarladung tragen und eine weitere Deprotonierung damit auch aus elektrostatischer Sicht ungünstig ist. Damit ergeben sich mit den 3 pK_a -Werten von Szakács und Noszál die Ionisierungsstufen -2, -1, 0 und +1. Tabelle 21 stellt noch einmal alle hier verwendeten pK_a -Werte des freien Pteroinsäureliganden übersichtlich dar.

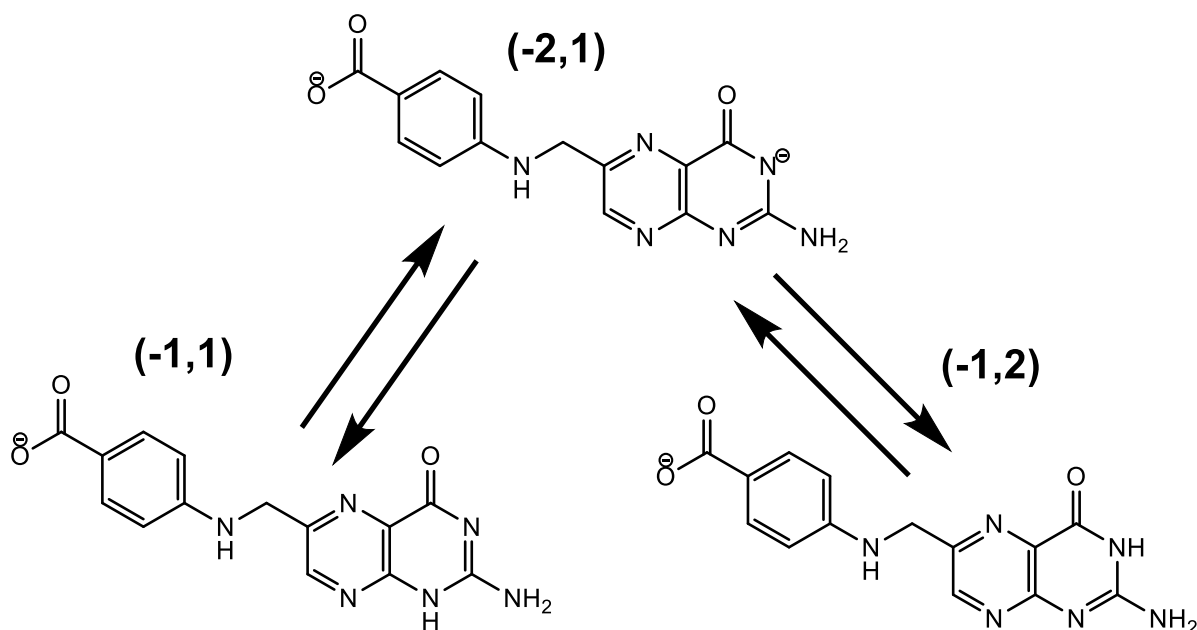


Abbildung 41: Die energetisch günstigsten Tautomere von Pteroinsäure in den Ionisierungszuständen -1 und -2 für B3LYP/6-31+G* mit sowohl dem EC-RISM- als auch dem PCM-Solvatationsmodell. Die exozyklische primäre Amino-Funktionalität bleibt bei Protonierungs- bzw. Deprotonierungsreaktionen zwischen zwischen den Ionisierungszuständen -1 und -2 unverändert.

Tabelle 21: Experimentelle pK_a -Werte des Pteroinsäureliganden sowie pK_a -Werte aus der Anpassung der mit EC-RISM und PCM berechneten Gibbs-Energien.

Ionisierungszustand	Exp.	EC-RISM 6-31+G*	EC-RISM 6-31G*	PCM 6-31+G*	PCM 6-31G*
-2 \rightarrow -1	2.29	1.97	0.37	-4.14	-7.38
-1 \rightarrow 0	4.80	4.51	5.47	9.40	10.25
0 \rightarrow 1	8.14	8.74	9.38	9.96	12.37

3.9.2 Komplex

Mit den anhand der Rechnungen für den Pteroinsäure-Liganden gewonnenen Erfahrungen konnte jetzt die Berechnung der Tautomeren- und Ionisationszustandspräferenz im Protein-Ligand-Komplex angegangen werden. Dazu ist es zunächst wichtig zu wissen, ob titrierbare Aminosäuren des Proteins den Ionisationszustand des Liganden merklich beeinflussen. Dazu wurden mit Hilfe des H⁺⁺-Webservers,^[172] der Protein-pK_a-Berechnungen auf der Basis der Poisson-Boltzmann-Theorie ermöglicht, alle in Abbildung 29 bis Abbildung 32 aufgeführten PTA-Spezies auf ihren Einfluss auf die pK_a-Werte der titrierbaren Proteinseitenketten hin untersucht. Auf diese Weise ist es möglich, einen Eindruck von der Stärke der Kopplung zwischen den titrierbaren Zentren des Proteins und des PTA-Liganden zu erhalten. Die Annahme ist hier, dass wenn die Änderung des Ionisierungszustands eines Zentrums einen großen Einfluss auf den Ionisierungszustand eines zweiten Zentrums hat, der Einfluss des Ionisierungszustands dieses zweiten Zentrums auf den des ersten Zentrums ebenfalls groß ist.^[105] Die nötigen 3D-Strukturen aller Spezies, die auch im Folgenden Verwendung fanden, wurden erzeugt indem ausgehend von der Struktur der Schweratome des Liganden innerhalb der Kristallstruktur des Komplexes je nach Zustand mit Avogadro^[1] Wasserstoffe hinzugefügt und unter Konstanthaltung der Abstände der Schweratome durch Geometrieoptimierung mit dem MMFF94s-Kraftfeld^[173] an ihre endgültige Position bewegt wurden. Abbildung 42 zeigt, dass der Ligand bzgl. seines Ionisierungszustands nur mit TYR-80 und GLU-208 stark koppelt. Der Bereich der pK_a-Werte von TYR-80 liegt jedoch deutlich oberhalb des pH-Wertes von 8.9, bei dem der RTA-PTA-Komplex kristallisiert wurde. Der pK_a der Seitenketten-Carboxy-Funktionalität von Glutaminsäure in Proteinen liegt mit typischerweise 3.5-4.5^[174] normalerweise so weit im sauren Bereich, dass hier bei pH 8.9 klar die deprotonierte Form vorliegen sollte. Wie aus Abbildung 43 ersichtlich, interagiert GLU-208 nicht direkt mit dem Liganden, könnte aber eine Wasserstoffbrücke zum Amid-Sauerstoff von GLN-173 bilden, wodurch der protonierte Zustand stabilisiert würde. Trotzdem erscheint der Abstand von 6.2 Å von Sauerstoff- zu Sauerstoffatom recht weit für eine Wasserstoffbrücke, da bei einer OH-Bindung mit ca. 1 Å noch 5.2 Å für die Wasserstoffbrückenlänge verbleiben, während Wasserstoffbrücken typischerweise nicht länger als 4 Å sind.^[175] Damit ist höchstens eine schwache Stabilisierung des protonierten Zustands für GLU-208 vorstellbar. Da der pK_a von 8.9 deutlich am äußeren Rand des in Abbildung 42 dargestellten Bereiches für GLU-208 liegt, erscheint der protonierte Zustand bei pH 8.9 wenig wahrscheinlich. In der Abbildung wird deutlich, dass ansonsten lediglich für Tyrosin 194 der Ionisierungszustand angepasst werden könnte. Für das deprotonierte Tyrosin sind aber keine AMBER-

Parameter verfügbar und die Wechselwirkung mit dem Liganden erscheint schwach. Daher wurde für alle Aminosäuren der Rizin-A-Kette der Standardionisierungszustand angenommen.

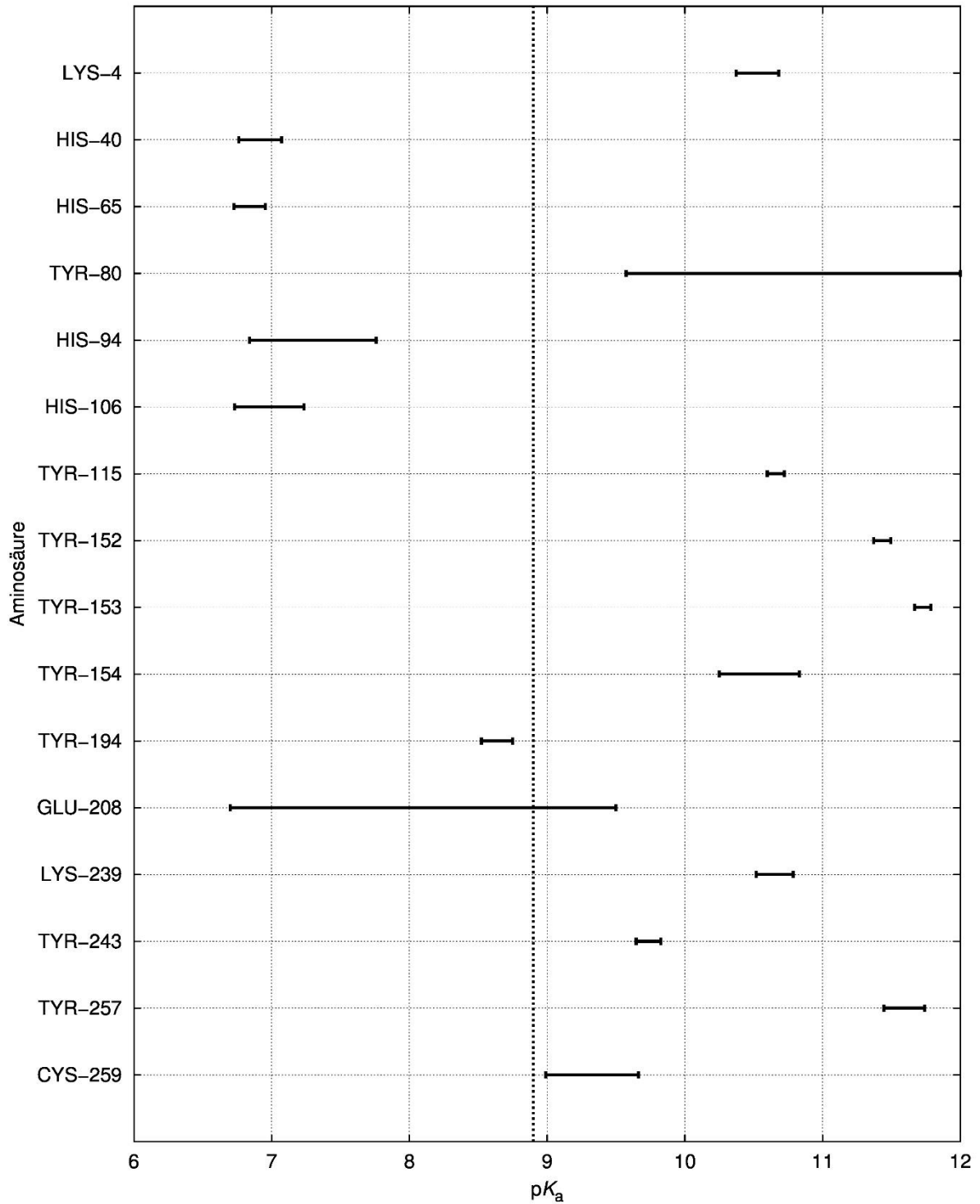


Abbildung 42: Gezeigt wird der Bereich, in dem die pK_a -Werte titrierbarer Aminosäuren der Rizin-A-Kette des RTA-PTA-Komplexes schwanken, wenn unterschiedliche Ionisierungszustände und Tautomere des Pteroinsäure-Liganden vorliegen, fokussiert auf die Aminosäuren der Rizin-A-Kette mit pK_a im pH-Bereich 6-12. Die Bereiche geben die Spanne der mit Hilfe des H⁺⁺-Servers^[172] ermittelten pK_a -Werte aller 31 hier betrachteten Ionisierungs- und Tautomerenzustände an. Die gepunktete Linie zeigt den pH-Wert von 8.9 an, bei dem der RTA-PTA-Komplex kristallisiert wurde.^[155]

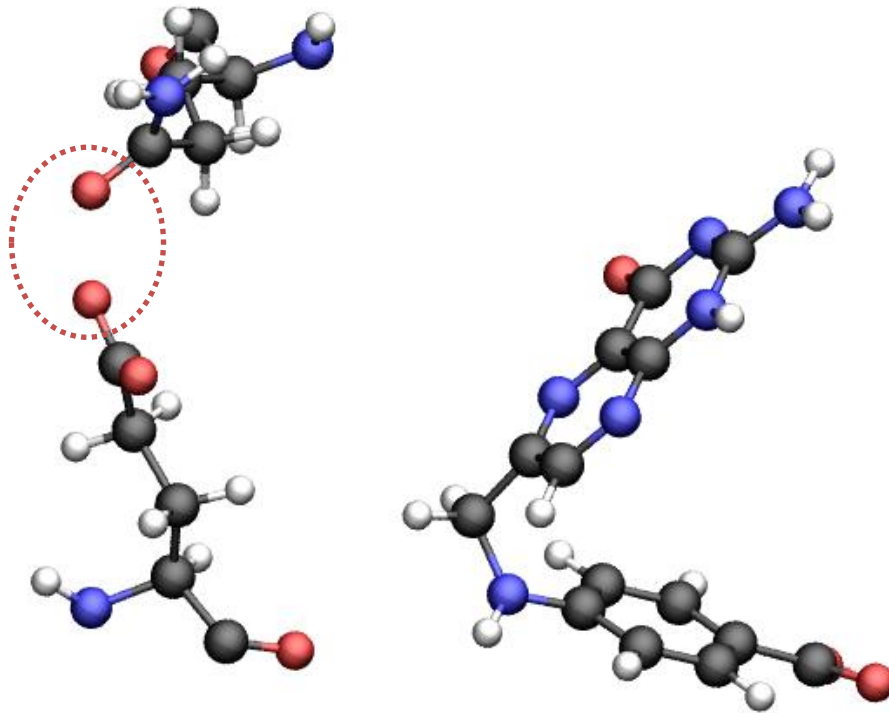


Abbildung 43: Ausschnitt aus dem RTA-PTA-Komplex (PDB-Code 1BR6). Pteroinsäure-Ligand (rechts), GLU-208 (links unten) und GLN-173 (links oben). Zwischen der Carboxygruppe von GLU-208 und dem Amid-Sauerstoff von GLN-173 (Markierung) könnte ein Proton stabilisiert werden.

Um dem Liganden die Möglichkeit zur Relaxation unter gleichzeitiger Beibehaltung seiner Position in der Bindetasche zu geben wurde als nächstes die Geometrie des Liganden in der Proteinbindetasche optimiert. Die Erfahrungen aus den Computereperimenten mit kleinen Molekülen aus den vorangegangenen Kapiteln zeigen, dass der Einsatz eines Lösungsmittelmodells essentiell ist, und dass bereits kleine Abweichungen in der Struktur erhebliche Auswirkungen auf die Gibbs-Energie haben können. Daher wurde versucht, Fehlerkompensationseffekte auszunutzen, indem die Lage möglichst aller Proteinatome außer den direkten Wasserstoffbrückenpartnern des Liganden unverändert gehalten wurde. Erste Versuche einer Geometrieoptimierung mit Gaussian09 rev. A02 zeigten, dass der

Zeitbedarf einer einzelnen PCM-Berechnung mit dem B3LYP-Dichtefunktional und einem kleinen Basissatz mit der QM/MM-ONIOM-PCM-Methode weit jenseits des Akzeptablen liegt (die Berechnung wurde nach einigen Tagen abgebrochen, ohne dass auch nur ein SCF-Zyklus abgeschlossen war), und dass bezüglich der Koordinaten der festzuhaltenden Proteinatome während der Berechnung eine erhebliche Drift festzustellen war. Um diese Probleme zu umgehen wurde ein auf dem LBFGS-Verfahren^[166,167] basierendes eigenes Optimierungsprogramm entwickelt, das Gaussian09^[63] zur Berechnung der Energie und der Kräfte auf die einzelnen Atome aufruft, und die Einhaltung der Nebenbedingungen dadurch sicherstellt, dass Kräfte auf festzuhaltende Atome ignoriert werden. Neben den Atomen des Liganden wurden auch die Atome einiger funktioneller Gruppen des Proteins, die als essentiell erachtet wurden um die Lage des Liganden im Protein zu erhalten, quantenmechanisch behandelt. Dabei musste die Anzahl der Protein-Atome stark beschränkt werden, um den Rechenzeitbedarf im Rahmen zu halten. Konkret wurden folgende Protein-Atome quantenmechanisch behandelt:

- Die die Wasserstoffbrücken zwischen ARG-180 und dem Liganden bildenden Atome der Guanidingruppe von ARG-180 und das Kohlenstoffatom dazwischen, das mit einem Wasserstoff abgesättigt wurde.
- Die Amidfunktionalitäten von ASN-78, VAL-81 und TYR-123 wurden durch Ammoniak ersetzt.
- Die Carbonylgruppe innerhalb der Amidgruppe von GLY-121 wurde durch Formaldehyd ersetzt.
- Die der Exozyklischen Aminogruppe des Pteringerüsts des Liganden am nächsten gelegene Methylgruppe von ILE-172 wurde durch Methan ersetzt. Dies erwies sich als notwendig, um die gegenseitige Durchdringung von Protein und Ligand zu verhindern.

Durch den Verzicht auf den Großteil der Atome des Proteins besitzen die übrigen Proteinatome teilweise freie Valenzen, die durch Wasserstoffatome abgesättigt wurden. Diese Wasserstoffatome bekamen die Positionen der Atome des Proteins zugewiesen, die durch eben diese Wasserstoffatome ersetzt wurden. Zur Verdeutlichung ist die Lage der genannten Atome noch einmal in Abbildung 44 gezeigt. Alle quantenmechanisch behandelten Protein-Atome wurden fixiert. Um zu verhindern, dass Ligand und Protein sich gegenseitig durchdringen, wurden zusätzliche apolare Wechselwirkungen zwischen Protein und Ligand hinzugefügt. Hierfür wurden die üblichen AMBER/GAFF-Lennard-Jones-Terme und

AMBER/GAFF-Lennard-Jones-Parameter verwendet. Zunächst wurde versucht, elektrostatische Wechselwirkungen zwischen Protein und Ligand durch zusätzliche Punktladungen in der QM-Berechnung an den Orten der Proteinatome zu behandeln. Es zeigte sich jedoch, dass unter diesen Bedingungen das SCF-Verfahren nicht zur Konvergenz zu bringen war. Daher wurde an dieser Stelle auf die Modellierung der Proteinelektrostatik verzichtet. Die Geometrieoptimierung wurde dann wie beschrieben bis zu einem Energieschwellenwert von 0.01 kJ/mol bei zwei aufeinanderfolgenden Iterationen durchgeführt.

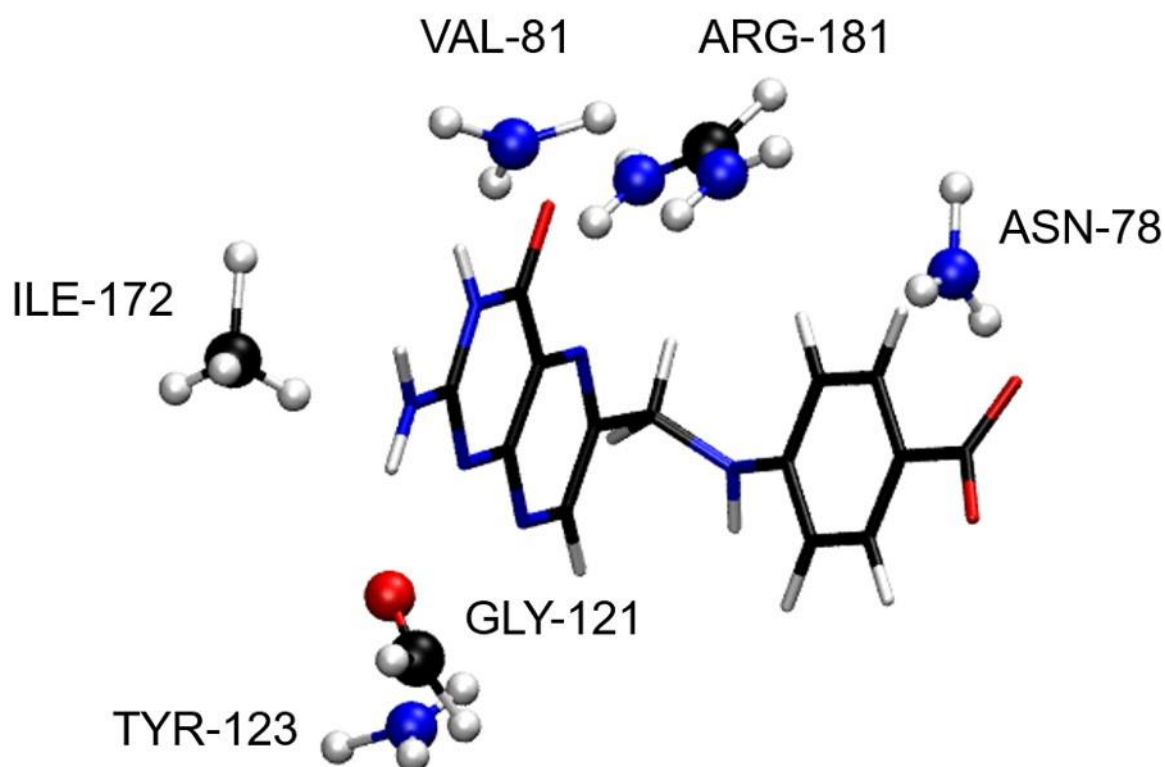


Abbildung 44: Das für die Geometrieoptimierung des Pteroinsäure-Rizin-Komplexes verwendete reduzierte QM-Modell-Subsystem. Der frei bewegliche Ligand ist als Stäbchenmodell dargestellt, die kleinen Moleküle, die das Verhalten des Proteins imitieren sollen, als „Ball-and-Stick“-Modelle. Alle nicht an Wasserstoffbrücken zum Ligand beteiligten Atome, die nicht zum Liganden gehören, wurden fixiert. Startkoordinaten sind in Kapitel 5.3 in Tabelle 29 und Tabelle 30 angegeben.

Mit den nach der Optimierung erhaltenen Geometrien wurden nun wie in Kapitel 3.2 beschrieben nach Gl. 2.2.2-2.2.8 und 3.2.1-3.2.4 EC-RISM-QM/MM-Rechnungen mit B3LYP/6-31+G* und der PSE-2-Closure durchgeführt. Berechnungen mit der kleineren 6-31G*-Basis wurden nicht mehr durchgeführt, da hiermit wie bereits ausgeführt für den Liganden keine glaubhaften

und zuverlässigen Ergebnisse erhalten werden konnten. Die QM-Zone wurde zunächst auf den Liganden beschränkt. In einem zweiten Schritt wurde dann das bereits für die Geometrieoptimierung aufgebaute und in Abbildung 44 gezeigte Subsystem verwendet, um die Wasserstoffbrücken zwischen Protein und Ligand in die quantenchemischen Berechnungen miteinzubeziehen. Alle Protein-Atome, die nicht mit den gezeigten Atomen des QM-Systems zusammenfallen, wurden als MM-Atome behandelt. Das Starrhalten des Proteins bietet hier den Vorteil, dass Intramolekulare Kraftfeldterme des Proteins für die Berechnungen mit allen betrachteten Tautomeren des Pteroinsäureliganden identisch sind. Daher ist eine erweiterte Betrachtung des *Link*-Bereichs, d.h. der QM/MM-Übergangszone, nicht erforderlich, weil die Anteile der Tautomeren und Ionisierungszustände durch den immer gleichen additiven Energiebeitrag nicht verändert werden. Dazu kommen in beiden Fällen ein Gitter mit 320 Gitterpunkten in X-, Y- und Z-Richtung und einer Maschenweite von 0.3 Å und die in den Kapiteln 3.1 und 3.3 entwickelten Methoden zur Beschleunigung von 3D-RISM und EC-RISM-Berechnungen an großen Systemen zum Einsatz: Chemische Exzesspotentiale wurden mit der Reziprokraumtechnik evaluiert, die langreichweitigen Reziprokraumpotentiale mit der PME-Technik mit Lagrange-Polynomen 10. Ordnung und die Solvens-Ladungsdichte wurde auf einem adaptiven Gitter mit den Parametern $\tau_1 = 0.3$, $\tau_2 = 0.005$ und $\tau_3 = 0.35$ diskretisiert. Zusätzlich wurde die Evaluation des für die Berechnung der Solvat-Solvens-Wechselwirkungsenergie benötigten vollen elektrostatischen Potentials des Solvats auf dem Gitter auf mehrere Prozesskerne aufgeteilt. Die weiteren Parameter entsprechen bis auf die Anzahl der MDIIS-Krylov-Vektoren und den Relaxationsfaktor, die beide keinen Einfluss auf das Ergebnis haben sollten, denen der EC-RISM-Berechnungen für den Liganden. Die erhaltenen Gibbs-Energien sind in Tabelle 22 aufgeführt.

Tabelle 22: Gibbs-Energien des Pteroinsäure-Rizin-Komplexes für unterschiedliche Pteroinsäurespezies aus Berechnungen mit B3LYP(EC-RISM)/6-31+G*. ^a Nur der Ligand wird quantenmechanisch behandelt, ^b QM-Zone wie für die Geometrieoptimierung. Die Gibbs-Energie-Differenzen sind relativ zum energetisch günstigsten Tautomer des jeweiligen Ionisierungszustands angegeben.

Ion.-Zust.	Tauto-mer	$G/(\text{kcal mol}^{-1})^a$	$G/(\text{kcal mol}^{-1})^b$	$\Delta G/(\text{kcal mol}^{-1})^a$	$\Delta G/(\text{kcal mol}^{-1})^b$
-2	1	-686462.32	-984410.40	0.00	0.00
	2	-686459.99	-984404.47	2.33	5.93
	3	-686435.03	-984385.39	27.29	25.01
-1	1	-686731.25	-984679.89	0.00	0.00
	2	-686708.72	-984663.77	22.53	16.12
	3	-686706.69	-984659.40	24.56	20.49
	4	-686685.59	-984642.76	45.66	37.13
	5	-686706.98	-984660.11	24.27	19.78
	6	-686709.34	-984669.66	21.91	10.23
	7	-686675.33	-984630.83	55.92	49.06
0	1	-686944.14	-984894.19	30.67	32.37
	2	-686950.37	-984905.50	24.44	21.06
	3	-686954.73	-984907.05	20.08	19.51
	4	-686902.42	-984858.19	72.39	68.37
	5	-686951.73	-984908.21	23.08	18.35
	6	-686974.81	-984926.56	0.00	0.00
	7	-686914.91	-984874.93	59.90	51.63
	8	-686931.88	-984890.02	42.93	36.54
	9	-686953.40	-984906.49	21.41	20.07
	10	-686956.60	-984906.58	18.21	19.98
	11	-686905.28	-984864.67	69.53	61.89
	12	-686935.60	-984896.94	39.21	29.62
	13	-686952.02	-984912.34	22.79	14.22
+1	1	-687199.30	-985150.97	0.00	2.43
	2	-687197.24	-985152.07	2.06	1.33
	3	-687152.01	-985113.06	47.29	40.34
	4	-687186.67	-985146.82	12.63	6.58
	5	-687180.96	-985145.40	18.34	8.00
	6	-687182.58	-985135.25	16.72	18.15
	7	-687121.46	-985084.56	77.84	68.84
	8	-687194.84	-985153.40	4.46	0.00

Mit den so berechneten Energien und dem Wert für die Gibbs-Energie des Protons aus den Berechnungen am freien Liganden konnten nun die in Abbildung 45 und Abbildung 46 gezeigten Speziesverteilungsdiagramme generiert werden.

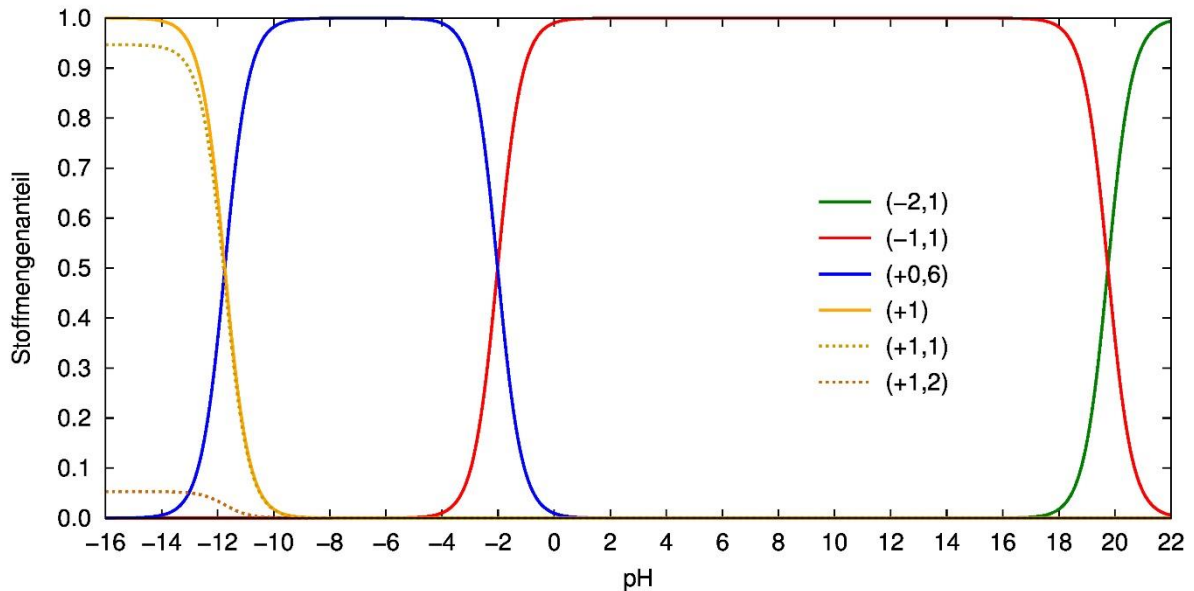


Abbildung 45: pH-abhängiges Speziesverteilungsdiagramm von Pteroinsäure im Komplex mit Rizin für ein QM/MM-System mit QM-Ligand und MM-Protein.

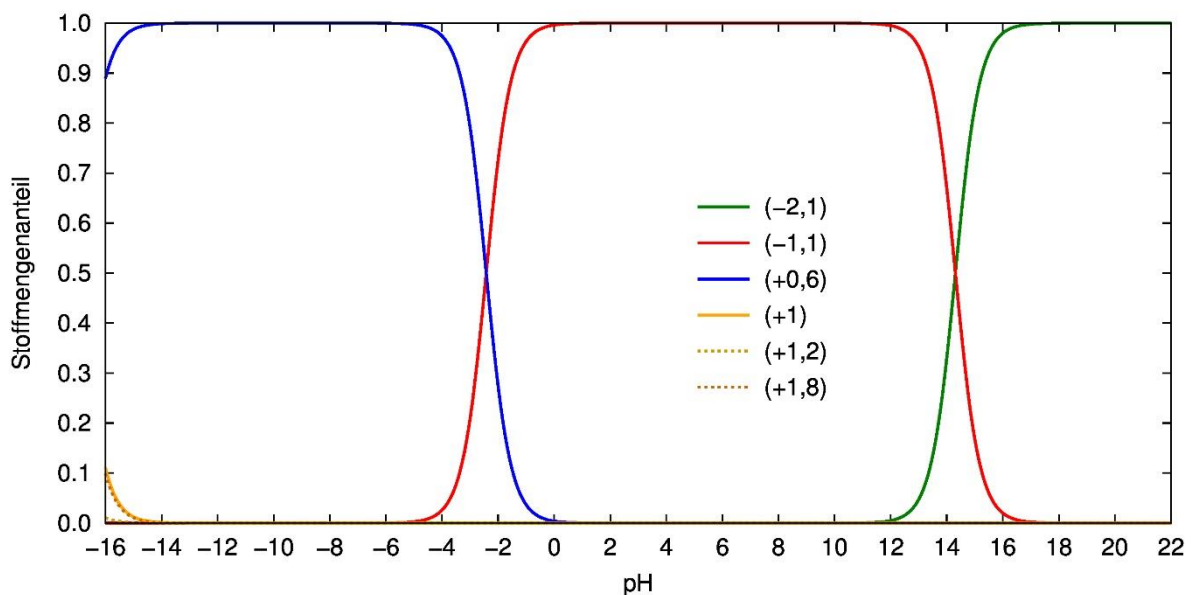


Abbildung 46: pH-abhängiges Speziesverteilungsdiagramm von Pteroinsäure im Komplex mit Rizin für ein QM/MM-System mit dem in Abbildung 44 gezeigten QM-System. Alle Protein-Atome, die nicht mit den gezeigten Atomen des QM-Systems zusammenfallen, wurden als MM-Atome behandelt.

Das bei pH 8.9 vorhergesagte Tautomer (-1,1) entspricht der Vorhersage von Robertus und Mitarbeitern auf Basis der Kristallstruktur. Wie man sieht wird der gesamte praxisrelevante pH-Bereich vom diesem

Tautomer dominiert. Das dürfte allerdings auch daran liegen, dass hier mit einem entweder komplett oder zumindest größtenteils starren Protein mit fixiertem Protonierungszustand gearbeitet wurde. Das bedeutet, das Protein ist für einen bestimmten Protonierungszustand des Liganden vorangepasst, wodurch dieser artifiziell stabilisiert wird. Die Differenz der Gibbs-Energien zwischen dem dominanten Tautomer in freier Lösung, (-1,2), und dem dominanten Tautomer im Komplex, (-1,1) kann Tabelle 16 bzw. Tabelle 22 entnommen werden und beträgt in freier Lösung unter Einbeziehung der *Clusterpopulationen* in der Zustandssumme 0.75, und ohne *Clusterpopulationen* 0.87 kcal/mol in entgegengesetzter Richtung. Im Komplex wurden 22.53 kcal/mol mit rein klassisch behandeltem Protein und 16.12 kcal/mol mit zum Teil quantenmechanisch, zum Teil klassisch behandeltem Protein erhalten, da hier kein Konformationssampling und kein *Clustering* durchgeführt wurden, entfällt diese Betrachtung hier. Abbildung 47 und Tabelle 23 zeigen noch einmal die Ionisierungsstufen des freien Liganden und der beiden Berechnungsvarianten des Komplexes im Vergleich, sowie die berechneten pK_a -Werte der Pteroinsäure im Komplex.

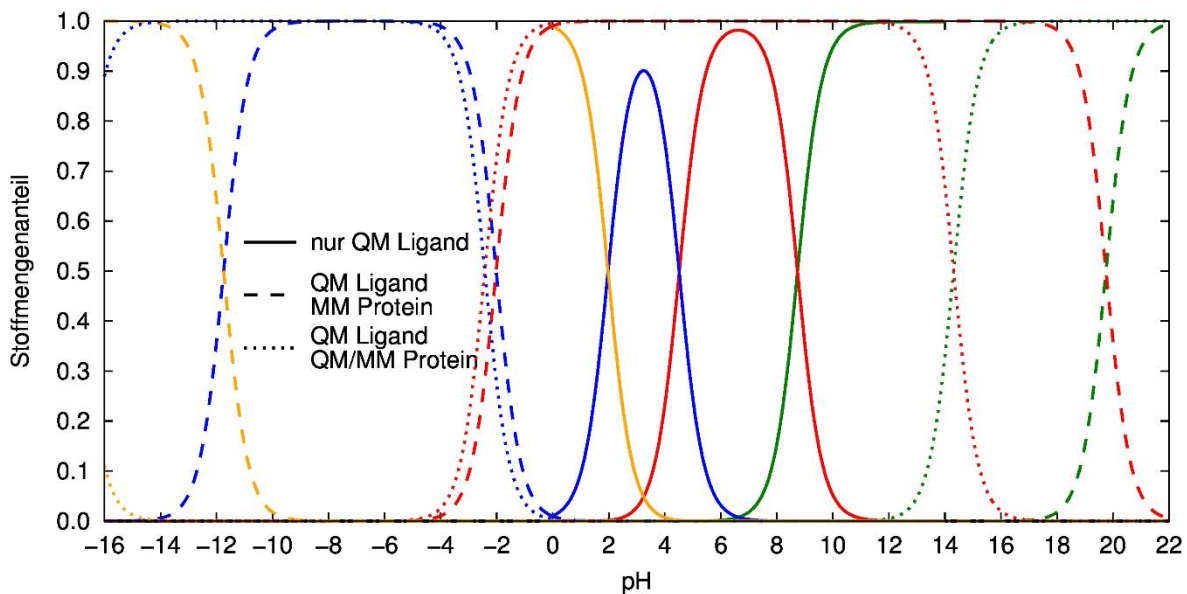


Abbildung 47: Stoffmengenanteile der einzelnen Ionisierungsstufen in Abhängigkeit des pH-Werts von Pteroinsäure, Pteroinsäure im Komplex mit Rizin für ein QM/MM-System mit quantenmechanisch behandeltem Liganden und klassisch behandeltem Protein, sowie Pteroinsäure im Komplex mit Rizin für ein QM/MM-System mit quantenmechanisch behandeltem Liganden und zum Teil quantenmechanisch, zum Teil klassisch behandeltem Protein.

Tabelle 23: Berechnete pK_a -Werte des Pteroinsäureliganden im Komplex mit Rizin.

Ionisierungszustand	Exp. freier Ligand	EC-RISM QM Ligand, MM Protein	EC-RISM QM Ligand, QM/MM Protein
-2 → -1	2.29	-11.77	-16.91
-1 → 0	4.80	-2.02	2.42
0 → 1	8.14	19.75	14.31

4 Zusammenfassung und Ausblick

In der vorliegenden Arbeit konnte gezeigt werden, dass die Kombination aus 3D-RISM und Quantenchemie, EC-RISM, ein vielversprechendes Mittel zur Untersuchung von Dissoziations- und Tautomerisierungsvorgängen ist.

Für eine Reihe von unterschiedlichen Stoffklassen wurden lineare pK_a -Modelle erstellt, die zumeist ohne weitere spezielle Parametrisierung gute Ergebnisse liefern. Im Gegensatz zu Modellen auf der Basis dielektrischer Kontinuumsmodelle zeigen viele der hier entwickelten Modelle Steigungen nahe dem idealen Wert eins, vor allem solche, die experimentell gemessene pK_a -Werte besonders gut reproduzieren. Das lässt darauf schließen, dass EC-RISM-basierte Modelle die Energetik des Dissoziationsvorgangs grundsätzlich realistischer beschreiben.

Für einen Satz von Guanin und Guanin-Derivaten, die für artifizielle genetische Systeme genutzt werden könnten, wurden in Zusammenarbeit mit weiteren Beteiligten Tautomerisierungs-Gibbs-Energien berechnet, die gut zu experimentellen Beobachtungen passen. Das hier verwendete EC-RISM/*Coupled-Cluster*-Verfahren eignet sich offenbar gut für diesen Zweck. Weiterhin wurde für das bisher wenig beachtete N3-H-Tautomer des Isoguanins eine verglichen mit dem dominanten N1-H-Tautomer nur wenig höhere Gibbs-Energie gefunden. Dementsprechend sollte dieses Tautomer in wässriger Lösung zumindest in geringer Menge vorliegen.

Für Histamin wurden Speziesverteilungsdiagramme erstellt, die die pH-abhängigen Stoffmengenanteile aller relevanten Spezies in übersichtlicher Form darstellen. Dazu wurde ausgehend von grundlegender statistischer Mechanik ein konsistenter Formalismus entwickelt. Im Gegensatz zu den meisten anderen experimentellen und theoretischen Arbeiten weisen die EC-RISM-Berechnungen auf eine leichte Präferenz der π -protonierten Spezies hin. Ein Vergleich der Zusammensetzung der Gibbs-Energien aus intramolekularem Beitrag und Solvensbeitrag für PCM und EC-RISM zeigt, dass trotz qualitativer Übereinstimmung der Einzelbeiträge im einen Fall Konformationen mit intramolekularen Wasserstoffbrücken bevorzugt werden während sie im anderen aber energetisch ungünstig sind. Im Allgemeinen folgen die günstigsten Konformation keinem einfach nachvollziehbaren Schema, ihre korrekte Identifizierung ist aber dennoch entscheidend für den Erfolg. Daher erweist es sich ein Konformationssampling als notwendig. Die berechneten Gibbs-Energien passen sehr gut zu den experimentellen pK_a -Werten. Während für EC-RISM das B3LYP-Dichtefunktional mit und ohne D3-Dispersionskorrektur und MP2, sowie die Basissätze 6-31+G*, 6-311+G** und aug-cc-pVDZ die

Speziesverteilung nur wenig beeinflussen, weichen die mit PCM und der 6-31+G*-Basis erzeugten pK_a -Werte deutlich vom Experiment ab.

Die Speziesverteilung von Pteroinsäure wurde sowohl in freier Form als auch im Komplex mit Rizin untersucht. Dazu wurde ein EC-RISM-QM/MM-Verfahren entwickelt, das in starkem Maße von einer Reihe weiterer technischer Innovationen profitiert. Es zeigt sich, dass für den freien Pteroinsäure-Liganden mit Hilfe von EC-RISM, Konformerensampling, und dem B3LYP-Dichtefunktional mit dem 6-31+G*-Basissatz eine glaubhafte Speziesverteilung erstellt werden konnte. Mit PCM konnten dagegen keine sinnvollen Speziesverteilungen erhalten werden. Eine Untersuchung der Energetik der einzelnen Tautomere und Konformationen zeigt, dass EC-RISM und PCM bereits hier unterschiedliche Präferenzen zeigen. Mit dem kleineren 6-31G*-Basissatz ohne Polarisationsfunktionen erhaltene Speziesverteilungen müssen als weniger glaubhaft erachtet werden, da sich die berechneten pK_a -Werte verglichen mit einer 6-31+G*-Basis deutlich schlechter an experimentelle anpassen lassen.

Mit Hilfe der aus der Anpassung der Ionisationsstufen an die experimentellen pK_a -Werte gewonnenen Gibbs-Energie des Protons wurde nun auch für den Pteroinsäure-Rizin-Komplex eine Speziesverteilung berechnet. Das Ergebnis deckt sich auch hier mit Erwartungen auf der Basis einer qualitativ hochwertigen Kristallstruktur. Die hier gewonnenen Erkenntnisse über die Energetik der unterschiedlichen Tautomere könnten in Zukunft bei der Optimierung des Inhibitors nützlich sein.

Die genannten technischen Innovationen sind ein Verfahren zur Verlegung aller langreichweitigen Potentialberechnungen für 3D-RISM in den reziproken Raum, die schnelle Berechnung der langreichweitigen Reziprokraumpotentiale mittels einer *Particle-Mesh-Ewald*-Methode und die Reduktion der Anzahl der für die EC-RISM-Kopplung zwischen Quantenchemie und 3D-RISM nötigen Punktladungen durch Verwendung eines adaptiven Gitters. Durch alle drei Methoden können mit minimalem Verlust an Genauigkeit erhebliche Mengen Rechenzeit eingespart werden.

Die in Kapitel 3.1.1 und 3.1.2 vorgestellten Methoden zur Beschleunigung des 3D-RISM-Lösers durch Verlagerung eines Großteils der Elektrostatikberechnungen in den reziproken Raum und die schnelle Berechnung des Reziprokraumpotentials mittels PME lassen sich prinzipiell auch auf die Berechnung der Lennard-Jones-Wechselwirkungen erweitern.^[47,70] Das ermöglicht die Verwendung sehr kleiner Abschneideradien für die verbliebenen Realraum-Potentialberechnungen. Sollte dies umgesetzt werden, bestimmt die 3D-RISM-Iteration praktisch alleine den Rechenaufwand für 3D-RISM und weitere Anstrengungen um das 3D-RISM-Verfahren zu beschleunigen sollten sich auf diesen Teilaspekt konzentrieren. Zusätzlich besteht die Hoffnung, die in Abbildung 5 noch sichtbare residuale

Zellengrößenabhängigkeit weiter zu reduzieren und trotz stark beschleunigter Berechnungsverfahren im Gegensatz zur Verwendung von Abschneideradien auch den Absolutwert des chemischen Exzesspotentials im Wesentlichen unverändert zu lassen. Eventuell könnte damit auch eine Verkleinerung der Simulationszelle ohne signifikante Genauigkeitsverluste möglich werden.

Für die in Kapitel 3.3 angesprochene schnelle Approximation der Solvens-Elektrostatik könnten für einen Testdatensatz mit Hilfe von Optimierungstechniken auf bestimmte Zielgenauigkeiten abgestimmte optimale Kombinationen der drei Parameter im Akzeptanzkriterium Gl. 7.4.2 gefunden werden, wodurch der routinemäßige Einsatz möglich würde. Damit könnte der für Berechnungen mit EC-RISM nötige Aufwand substantiell reduziert werden.

Das für die Berechnungen am Rizin-Komplex entwickelte QM/MM-Verfahren besitzt eine Reihe von Schwächen, die in Zukunft vermieden werden sollten. Die Hauptprobleme sind aus Sicht des Autors das fehlende Konformerensampling und die fehlende Möglichkeit der Relaxation des Großteils des Proteins. Die Durchführung des Konformerensampling scheiterte zum einen am Rechenzeitbedarf, der bereits für die freie Pteroinsäure beträchtlich ist, zum anderen an den vielen manuellen Schritten, die für die Geometrieoptimierungen und die EC-RISM-QM/MM-Berechnungen nötig sind. Einiges könnte sicher automatisiert werden, nachdem erst einmal ein *modus operandi* festgelegt ist und die während der Implementierungs- und Testphase nötige Flexibilität nicht mehr benötigt wird. Wenn schnelle PCM-basierte QM/MM-Lösungsmittelmodelle in Gaussian allgemein verfügbar werden, sollten zumindest die Geometrieoptimierungen in Zukunft erheblich einfacher, genauer und effizienter durchführbar werden. Kann das EC-RISM-QM/MM-Verfahren z.B. durch die bereits angesprochenen Verbesserungen des 3D-RISM-Lösers ebenfalls beschleunigt werden, rückt die Möglichkeit des Konformerensamplings erkennbar näher. Ein großes Problem stellt die Behandlung der Wasserstoffbrücken zwischen Protein und Ligand dar. Idealerweise würden alle an der Wasserstoffbrücke beteiligten Atome bereits während der Geometrieoptimierung quantenmechanisch behandelt und wären frei beweglich. Dadurch würde aber die Komplexität des Verfahrens deutlich erhöht, weil dann die QM/MM-Grenze durch kovalente Bindungen im Protein verlaufen und die intramolekulare Energieänderung des Proteins berücksichtigt werden müsste.^[87,88]

Bezüglich der Methodik für die Tautomeren- und Ionisationsgleichgewichte sollten sogenannte indirekte Verfahren getestet werden, bei denen die Standard-Reaktions-Gibbs-Energien über einen Kreisprozess mit Gasphasenreaktionen berechnet werden. Dazu sind Lösungsmittelmodelle nötig, die auf die Reproduktion der Standard-Solvatations-Gibbs-Energien ausgelegt sind. Derartige Modelle auf

der Basis von EC-RISM wurden kürzlich von Daniel Tomazic und Stefan M. Kast entwickelt^[176] und von Kast und Mitarbeitern im Rahmen des SAMPL5-Wettbewerbs eingesetzt.^[177] Sie könnten für den hier skizzierten Zweck verwendet werden. Bei indirekten Verfahren ist jedoch auch entscheidend, dass die Tautomerisierungs- bzw. Dissoziationsreaktion in der Gasphase mit hochgenauen Methoden behandelt werden kann. Dazu würden sich in neuerer Zeit verfügbar gewordene schnelle lokale *Post-Hartree-Fock*-Methoden wie die DLPNO-Verfahren von Neese und Mitarbeitern,^[178] insbesondere in der explizit korrelierten F12-Variante^[179] und in Kombination mit einer Korrektur für die Nullpunktschwingungsenergie anbieten. Die Maximierung der Genauigkeit der Elektronenstrukturberechnungen würde auch verbesserte Möglichkeiten bieten, 3D-RISM und EC-RISM besser zu parametrisieren. Von Tomazic und Kast begonnene Parametrisierungsarbeiten könnten dann auf diese Methodologie ausgedehnt werden. Weiter sollte geprüft werden, ob Möglichkeiten bestehen, speziell auf geladene Spezies angepasste LJ-Kraftfeldparameter zu erzeugen. In jüngerer Zeit werden zunehmend effiziente quantenchemische Methoden zur Berechnung intermolekularer Wechselwirkungsenergien entwickelt und verfeinert,^[180–184] die zusammen mit einer globalen Parameteroptimierungsprozedur^[185] als Basis hierfür dienen könnten, wenn die Berechnungen hinreichend effizient und genau wären.

5 Anhang

5.1 Details der pK_a-Modelle

5.1.1 Gibbs-Energien aller Spezies

Tabelle 24: Gibbs-Energien aller Spezies für EC-RISM mit MP2(EC-RISM)/6-311+G(d,p)//B3LYP(PCM)/6-311+G(d,p), jeweils für die Minimum-Gibbs-Energie-Struktur. Der Suffix „K“ kennzeichnet die Protonierungsstelle, „R“ die Rotamere, siehe die zugehörigen digital beigefügten Strukturen. Experimentelle pK_a-Werte entstammen Ref. ^[128].

Substanz/ Gruppe	Suffix prot./deprot.	G/(kcal mol ⁻¹)		pK _a (exp.)	pK _a (ber.)
		protoniert	deprotoniert		
Pyrrole					
Pyrrol	Kc2_R0/K0_R0	-131771.89	-131520.85	-3.8	-4.0
1-Methylpyrrol	Kc2_R0/K0_R0	-156361.07	-156107.89	-2.9	-2.6
2-Methylpyrrol	Kc2a_R0/K0_R0	-156372.97	-156116.56	-0.2	-0.4
3-Methylpyrrol	Kc2b_R0/K0_R0	-156370.49	-156114.95	-1.0	-1.0
Indole					
Indol	K0_R0/K0_R0	-227917.09	-227665.36	-3.6	-4.0
1-Methylindol	K1_R0/K0_R0	-252507.86	-252253.05	-2.3	-1.8
2-Methylindol	K1_R0/K0_R0	-252518.79	-252262.44	-0.3	-0.6
3-Methylindol	K1_R0/K0_R0	-252511.94	-252260.69	-4.6	-4.4
Aniline					
Anilin	K0_R0/K0_R0	-180235.48	-179970.90	4.6	4.6
4-Chloroanilin	K0_R0/K0_R0	-468296.86	-468032.98	4.0	3.9
4-Methoxyanilin	K0_R0/K0_R0	-251932.68	-251667.41	5.2	5.4
4-Nitroanilin	K0_R0/K0_R0	-308313.65	-308052.46	1.0	1.0
<i>p</i> -Toluidin	K0_R0/K0_R0	-204831.25	-204566.35	5.1	5.0
Heterozyklen					
2-Aminopyridin	K0_R0/K0_R0	-190303.10	-190035.95	6.7	6.5
2-Aminothiazol	K0_R0/K0_R0	-391394.89	-391129.78	5.4	5.2
2-Methylimidazol	K0_R0/K0_R0	-166459.24	-166189.67	8.0	8.0
3-Aminopyridin	K0_R0/K0_R0	-190299.90	-190033.14	6.0	6.3
4-Aminopyridin	K0_R0/K0_R0	-190305.89	-190035.20	9.7	8.7
4-Methylpyridin	K0_R0/K0_R0	-180232.81	-179965.32	6.0	6.7
Benzimidazol	K0_R0/K0_R0	-238002.17	-237737.51	5.8	5.0

Substanz/ Gruppe	Suffix prot./deprot.	$G/(\text{kcal mol}^{-1})$		$\text{p}K_{\text{a}}(\text{exp.})$	$\text{p}K_{\text{a}}(\text{ber.})$
		protoniert	deprotoniert		
Imidazol	K0_R0/K0_R0	-141860.24	-141592.39	7.0	6.9
Isochinolin	K0_R0/K0_R0	-251777.49	-251511.36	5.4	5.9
Melamin	K0_R0/K0_R0	-279779.83	-279517.54	5.0	3.5
Pyrazin	K0_R0/K0_R0	-165683.65	-165423.66	0.7	2.0
Pyrazol	K0_R0/K0_R0	-141837.77	-141577.99	2.5	1.9
Pyridin	K0_R0/K0_R0	-155635.07	-155368.61	5.3	6.1
Pyrimidin	K0_R0/K0_R0	-165687.48	-165429.37	1.3	0.9
Chinolin	K0_R0/K0_R0	-251777.38	-251512.19	4.8	5.3
Thiazol	K0_R0/K0_R0	-356728.99	-356466.74	2.8	3.5
Phenole					
Phenol	K0_R0/K0_R0	-192427.74	-192149.37	10.0	10.0
4-Chlorophenol	K0_R0/K0_R0	-480489.59	-480211.40	9.9	9.9
4-Aminophenol	K0_R0/K0_R0	-227090.60	-226812.80	9.4	9.6
4-Fluorophenol	K0_R0/K0_R0	-254598.55	-254320.29	10.2	10.0
4-Methoxyphenol	K0_R0/K0_R0	-264123.67	-263844.98	10.3	10.3
4-Methylphenol	K0_R0/K0_R0	-217022.97	-216743.98	10.5	10.5
4-Nitrophenol	K0_R0/K0_R0	-320509.49	-320235.13	7.2	7.1
4-Hydroxybenzaldehyd	K0_R1/K0_R0	-263390.79	-263118.09	7.6	7.7

5.2 Gibbs-Energien aller Histaminkonformere

Die Atomkoordinaten der Konformere sind der Masterarbeit von Lukas Eberlein^[2] zu entnehmen.

Tabelle 25: Energien der Histaminkonformere für B3LYP/6-31+G**/B3LYP/6-31+G*. Für die PCM-Energien wurde die Gaussian-Option „*ExternalIteration*“ verwendet, da die Standard-PCM-Methode keine Zerlegung der Gibbs-Energie in Solvat-Energie und chemisches Exzesspotential erlaubt.

Tautom.	Konform.	PCM			EC-RISM		
		$G/(\text{kcal mol}^{-1})$	$E/(\text{kcal mol}^{-1})$	$\mu^{\text{ex}}/(\text{kcal mol}^{-1})$	$G/(\text{kcal mol}^{-1})$	$E/(\text{kcal mol}^{-1})$	$\mu^{\text{ex}}/(\text{kcal mol}^{-1})$
π		-226046.54			-226044.26		
	1	-226044.05	-226031.27	-12.78	-226041.84	-226020.70	-21.14
	2	-226044.53	-226031.46	-13.07	-226042.80	-226020.74	-22.07
	3	-226044.10	-226031.99	-12.11	-226042.98	-226021.58	-21.41
	10	-226044.50	-226032.17	-12.33	-226042.60	-226021.79	-20.81
	11	-226045.91	-226035.99	-9.92	-226037.53	-226029.55	-7.98
	14	-226043.93	-226033.95	-9.98	-226039.87	-226025.88	-13.99
	16	-226043.49	-226029.45	-14.04	-226043.05	-226017.10	-25.96
	18	-226044.50	-226032.17	-12.33	-226042.61	-226021.70	-20.91
	21	-226044.11	-226031.84	-12.27	-226041.81	-226021.05	-20.76
	22	-226045.92	-226035.99	-9.93	-226037.47	-226029.64	-7.83
	23	-226043.94	-226033.94	-10.00	-226039.88	-226025.94	-13.94
	24	-226044.12	-226031.84	-12.28	-226041.90	-226021.03	-20.87
	29	-226044.03	-226031.27	-12.76	-226041.82	-226020.70	-21.12
	30	-226044.09	-226031.99	-12.10	-226042.87	-226021.67	-21.20
	34	-226044.52	-226031.48	-13.04	-226042.96	-226020.67	-22.29
	36	-226043.37	-226031.10	-12.27	-226041.50	-226021.00	-20.50
	38	-226043.93	-226033.90	-10.03	-226039.91	-226025.89	-14.02
	48	-226042.38	-226028.81	-13.57	-226042.07	-226016.64	-25.43
	50	-226044.01	-226031.03	-12.98	-226042.24	-226019.84	-22.40
	54	-226043.49	-226029.45	-14.04	-226043.22	-226017.06	-26.16
	58	-226044.02	-226031.03	-12.99	-226041.96	-226020.11	-21.85
	62	-226043.36	-226031.10	-12.26	-226041.41	-226021.02	-20.40
τ		-226045.89			-226043.07		
	1	-226044.11	-226032.91	-11.20	-226040.70	-226024.30	-16.41
	2	-226044.58	-226032.77	-11.81	-226041.95	-226023.28	-18.67
	3	-226044.17	-226032.14	-12.03	-226041.79	-226021.98	-19.81
	10	-226044.14	-226033.60	-10.54	-226040.14	-226025.83	-14.30
	11	-226044.16	-226033.51	-10.65	-226041.08	-226024.73	-16.35

12	-226042.68	-226029.51	-13.17	-226040.94	-226016.77	-24.16
14	-226043.93	-226032.88	-11.05	-226039.75	-226023.31	-16.44
17	-226043.51	-226031.31	-12.20	-226041.15	-226020.80	-20.35
18	-226044.55	-226032.84	-11.71	-226041.41	-226023.56	-17.85
20	-226040.27	-226025.99	-14.28	-226038.94	-226010.55	-28.39
21	-226044.17	-226033.49	-10.68	-226040.96	-226024.86	-16.10
23	-226043.93	-226032.88	-11.05	-226039.85	-226023.21	-16.64
26	-226044.14	-226033.61	-10.53	-226040.00	-226025.97	-14.03
28	-226044.10	-226032.91	-11.19	-226040.73	-226024.24	-16.49
30	-226044.16	-226032.13	-12.03	-226041.73	-226022.04	-19.69
34	-226044.58	-226032.78	-11.80	-226041.99	-226023.27	-18.72
38	-226042.69	-226029.51	-13.18	-226040.90	-226016.92	-23.98
48	-226042.44	-226032.37	-10.07	-226038.15	-226025.48	-12.68
52	-226042.42	-226031.48	-10.94	-226040.63	-226022.31	-18.32
53	-226044.54	-226032.84	-11.70	-226041.39	-226023.61	-17.77
54	-226043.51	-226031.31	-12.20	-226041.01	-226020.75	-20.27
61	-226043.36	-226031.23	-12.13	-226041.09	-226021.11	-19.99
62	-226043.36	-226031.24	-12.12	-226041.23	-226021.10	-20.13
α/π	-226328.65			-226314.25		
1	-226327.11	-226249.15	-77.96	-226313.38	-226240.28	-73.10
10	-226326.75	-226252.31	-74.44	-226312.12	-226243.81	-68.31
11	-226327.80	-226255.13	-72.67	-226313.13	-226247.27	-65.87
12	-226326.80	-226253.35	-73.45	-226311.98	-226245.30	-66.69
14	-226326.77	-226254.25	-72.52	-226312.05	-226246.34	-65.71
16	-226326.74	-226253.85	-72.89	-226312.06	-226245.85	-66.21
17	-226327.82	-226255.13	-72.69	-226313.04	-226247.39	-65.65
21	-226326.74	-226252.42	-74.32	-226312.08	-226244.07	-68.01
22	-226326.76	-226254.25	-72.51	-226312.05	-226246.23	-65.82
30	-226327.08	-226249.15	-77.93	-226313.36	-226240.27	-73.09
37	-226326.80	-226253.34	-73.46	-226312.04	-226245.28	-66.76
59	-226326.71	-226252.92	-73.79	-226312.08	-226244.58	-67.51
61	-226326.11	-226249.29	-76.82	-226312.12	-226241.04	-71.08
62	-226326.10	-226249.28	-76.82	-226311.97	-226241.13	-70.84
63	-226326.06	-226249.35	-76.71	-226311.98	-226241.07	-70.92
α/τ	-226333.52			-226313.74		
1	-226333.11	-226279.74	-53.37	-226311.51	-226279.50	-32.01
10	-226327.14	-226257.70	-69.44	-226312.39	-226251.58	-60.81
11	-226333.11	-226279.73	-53.38	-226311.45	-226279.49	-31.96
12	-226328.62	-226263.25	-65.37	-226312.84	-226258.92	-53.92
14	-226328.63	-226263.26	-65.37	-226312.86	-226258.92	-53.93
18	-226327.09	-226257.93	-69.16	-226312.15	-226252.05	-60.10

23	-226329.28	-226265.75	-63.53	-226312.18	-226261.35	-50.84
29	-226329.27	-226265.73	-63.54	-226312.26	-226261.33	-50.93
46	-226327.15	-226257.68	-69.47	-226312.39	-226251.64	-60.75
50	-226327.15	-226257.69	-69.46	-226312.24	-226251.69	-60.55
π/τ	-226331.06			-226310.74		
1	-226326.20	-226268.68	-57.52	-226308.77	-226263.71	-45.06
2	-226326.75	-226267.69	-59.06	-226309.38	-226262.57	-46.81
3	-226326.26	-226269.13	-57.13	-226309.18	-226264.47	-44.71
10	-226326.63	-226266.86	-59.77	-226309.33	-226261.39	-47.94
11	-226330.64	-226282.07	-48.57	-226305.88	-226282.24	-23.64
17	-226326.23	-226271.83	-54.40	-226309.34	-226267.76	-41.58
18	-226326.18	-226269.02	-57.16	-226308.82	-226263.94	-44.88
22	-226326.19	-226271.83	-54.36	-226309.41	-226267.54	-41.87
23	-226330.65	-226282.06	-48.59	-226305.92	-226282.25	-23.68
26	-226326.18	-226266.28	-59.90	-226308.80	-226260.47	-48.32
29	-226326.24	-226269.13	-57.11	-226309.40	-226264.26	-45.14
31	-226326.23	-226268.67	-57.56	-226308.86	-226263.64	-45.21
34	-226326.72	-226267.69	-59.03	-226309.60	-226262.36	-47.24
35	-226325.74	-226268.65	-57.09	-226308.47	-226263.64	-44.83
45	-226325.91	-226273.77	-52.14	-226304.65	-226273.34	-31.31
49	-226325.42	-226271.12	-54.30	-226308.49	-226266.83	-41.67
51	-226325.89	-226273.67	-52.22	-226304.87	-226273.10	-31.77
54	-226325.89	-226273.67	-52.22	-226304.79	-226273.12	-31.67
55	-226326.14	-226269.02	-57.12	-226308.97	-226263.83	-45.14
56	-226326.18	-226269.15	-57.03	-226308.55	-226264.44	-44.11
61	-226325.71	-226268.66	-57.05	-226308.38	-226263.72	-44.66
$\alpha/\pi/\tau$	-226607.36			-226579.53		
10	-226606.12	-226419.90	-186.22	-226578.11	-226419.10	-159.01
11	-226605.42	-226418.25	-187.17	-226577.36	-226417.37	-159.99
13	-226605.37	-226418.64	-186.73	-226577.25	-226417.79	-159.46
16	-226606.25	-226424.25	-182.00	-226578.68	-226423.83	-154.85
17	-226606.12	-226419.85	-186.27	-226578.21	-226419.00	-159.21
22	-226606.30	-226424.26	-182.04	-226578.68	-226423.81	-154.87
24	-226606.14	-226419.83	-186.31	-226578.16	-226418.98	-159.17
37	-226605.38	-226418.66	-186.72	-226577.33	-226417.81	-159.52
38	-226605.43	-226418.25	-187.18	-226577.35	-226417.38	-159.97
41	-226606.12	-226419.91	-186.21	-226578.07	-226419.10	-158.97
61	-226605.46	-226424.34	-181.12	-226577.80	-226424.01	-153.79

Tabelle 26: Energien der Histaminkonformere für B3LYP-D3/6-31+G**/B3LYP/6-31+G*. Die D3-Dispersionskorrektur wurde *ex post* mit Hilfe des D3-Programms von Grimme und Mitarbeitern durchgeführt.

Tautomer	Konformer	PCM G/(kcal/mol)	EC-RISM G/(kcal/mol)
π		-226062.62	-226059.99
	1	-226059.51	-226057.31
	2	-226060.07	-226058.35
	3	-226059.56	-226058.45
	10	-226060.54	-226058.66
	11	-226062.06	-226053.68
	14	-226059.98	-226055.92
	16	-226059.19	-226058.76
	18	-226060.54	-226058.66
	21	-226060.09	-226057.79
	22	-226062.06	-226053.62
	23	-226059.98	-226055.93
	24	-226060.10	-226057.88
	29	-226059.51	-226057.29
	30	-226059.56	-226058.34
	34	-226060.07	-226058.51
	36	-226058.93	-226057.07
	38	-226059.97	-226055.96
	48	-226058.19	-226057.88
	50	-226060.11	-226058.33
54	-226059.19	-226058.92	
58	-226060.11	-226058.05	
62	-226058.93	-226056.98	
τ		-226061.60	-226058.63
	1	-226059.43	-226056.03
	2	-226059.99	-226057.36
	3	-226059.49	-226057.12
	10	-226060.10	-226056.10
	11	-226060.02	-226056.92
	12	-226058.25	-226056.50
	14	-226059.81	-226055.64
	17	-226059.06	-226056.69
	18	-226060.47	-226057.34
	20	-226055.65	-226054.30
	21	-226060.03	-226056.83
	23	-226059.81	-226055.73

26	-226060.10	-226055.97
28	-226059.43	-226056.06
30	-226059.49	-226057.06
34	-226059.99	-226057.40
38	-226058.25	-226056.45
48	-226058.23	-226053.93
52	-226058.18	-226056.39
53	-226060.47	-226057.31
54	-226059.06	-226056.55
61	-226058.86	-226056.58
62	-226058.86	-226056.72
<hr/>		
α/π	-226345.60	-226331.01
1	-226343.48	-226329.73
10	-226343.75	-226329.09
11	-226344.84	-226330.15
12	-226343.77	-226328.99
14	-226343.80	-226329.10
16	-226343.80	-226329.10
17	-226344.84	-226330.05
21	-226343.73	-226329.06
22	-226343.80	-226329.10
30	-226343.48	-226329.71
37	-226343.78	-226329.05
59	-226343.77	-226329.07
61	-226342.59	-226328.60
62	-226342.59	-226328.45
63	-226342.54	-226328.45
<hr/>		
α/τ	-226350.74	-226330.25
1	-226350.33	-226328.71
10	-226343.55	-226328.77
11	-226350.33	-226328.65
12	-226344.87	-226329.08
14	-226344.87	-226329.09
18	-226343.52	-226328.50
23	-226346.19	-226329.08
29	-226346.19	-226329.15
46	-226343.57	-226328.76
50	-226343.57	-226328.62
<hr/>		
π/τ	-226348.15	-226327.27
1	-226342.50	-226325.04
2	-226343.09	-226325.73

3	-226342.51	-226325.46
10	-226343.43	-226326.17
11	-226347.73	-226322.96
17	-226342.80	-226325.90
18	-226342.91	-226325.57
22	-226342.80	-226325.98
23	-226347.73	-226323.00
26	-226343.03	-226325.65
29	-226342.51	-226325.67
31	-226342.50	-226325.13
34	-226343.09	-226325.96
35	-226342.11	-226324.85
45	-226342.64	-226321.39
49	-226342.00	-226325.06
51	-226342.66	-226321.62
54	-226342.66	-226321.55
55	-226342.92	-226325.72
56	-226342.88	-226325.29
61	-226342.11	-226324.76
$\alpha/\pi/\tau$	-226624.92	-226596.97
10	-226623.81	-226595.76
11	-226623.07	-226594.95
13	-226623.06	-226594.88
16	-226623.57	-226595.83
17	-226623.78	-226595.83
22	-226623.57	-226595.83
24	-226623.78	-226595.79
37	-226623.06	-226594.97
38	-226623.07	-226594.95
41	-226623.81	-226595.72
61	-226622.78	-226595.07

Tabelle 27: Energien der Histaminkonformere für MP2/6-311+G**//MP2/6-311+G**.

Tautomer	Konformer	PCM	EC-RISM		
		G/(kcal/mol)	G/(kcal/mol)	E/(kcal/mol)	μ^{ex} /(kcal/mol)
π		-224663.19	-225444.48		
	1	-224661.63	-225441.78	-225419.90	-21.88
	2	-224661.82	-225442.91	-225419.35	-23.56
	3	-224661.68	-225442.84	-225420.52	-22.31
	10	-224661.48	-225443.47	-225421.49	-21.98
	11	-224661.83	-225439.34	-225430.01	-9.33
	14	-224660.89	-225440.44	-225426.62	-13.82
	16	-224660.95	-225443.00	-225416.98	-26.02
	18	-224661.50	-225443.42	-225421.40	-22.03
	21	-224661.23	-225442.48	-225421.20	-21.29
	22	-224661.85	-225439.27	-225430.20	-9.07
	23	-224660.90	-225440.41	-225426.66	-13.75
	24	-224661.23	-225442.53	-225421.20	-21.34
	29	-224661.60	-225441.87	-225419.78	-22.09
	30	-224661.68	-225442.83	-225420.60	-22.23
	34	-224661.78	-225442.85	-225419.56	-23.29
	50	-224660.72	-225442.82	-225419.62	-23.20
	54	-224660.96	-225443.04	-225416.97	-26.08
58	-224660.74	-225442.82	-225419.53	-23.28	
τ		-224663.15	-225443.60		
	1	-224661.71	-225440.89	-225424.48	-16.41
	2	-224661.91	-225442.11	-225422.78	-19.34
	3	-224661.79	-225441.79	-225421.85	-19.94
	10	-224660.80	-225441.34	-225426.28	-15.06
	11	-224661.32	-225441.89	-225426.11	-15.77
	12	-224659.72	-225441.94	-225417.51	-24.43
	14	-224660.70	-225441.91	-225423.91	-18.01
	17	-224661.02	-225441.10	-225421.33	-19.77
	18	-224661.60	-225442.25	-225424.25	-18.00
	20	-224659.72	-225442.31	-225416.90	-25.42
	21	-224661.35	-225441.91	-225426.07	-15.84
	23	-224660.70	-225441.65	-225424.06	-17.58
	26	-224660.80	-225441.27	-225426.41	-14.86
	28	-224661.70	-225440.90	-225424.40	-16.51
	30	-224661.78	-225441.85	-225421.80	-20.05
	34	-224661.89	-225442.15	-225422.74	-19.41
	38	-224659.72	-225441.76	-225417.62	-24.14

	52	-224659.42	-225437.88	-225426.27	-11.61
	53	-224661.60	-225442.28	-225424.33	-17.95
	54	-224661.01	-225440.85	-225421.68	-19.17
α/π		-224951.08	-225717.04		
	1	-224949.78	-225715.46	-225640.75	-74.71
	10	-224948.75	-225715.25	-225647.48	-67.77
	11	-224950.12	-225716.08	-225650.09	-65.98
	12	-224948.81	-225715.12	-225648.32	-66.80
	14	-224948.78	-225715.03	-225648.91	-66.12
	16	-224948.75	-225715.12	-225648.06	-67.06
	17	-224950.05	-225716.11	-225650.21	-65.90
	21	-224948.79	-225715.24	-225648.31	-66.93
	22	-224948.79	-225715.24	-225648.06	-67.19
	30	-224949.74	-225715.44	-225640.85	-74.59
	37	-224948.81	-225715.18	-225648.20	-66.98
	59	-224948.67	-225714.92	-225649.09	-65.83
	61	-224949.79	-225715.43	-225640.72	-74.72
	62	-224948.90	-225713.90	-225641.72	-72.19
	63	-224949.78	-225715.45	-225640.89	-74.55
α/τ		-224954.06	-225716.96		
	1	-224953.57	-225715.77	-225682.94	-32.83
	10	-224951.43	-225715.63	-225660.37	-55.26
	11	-224953.61	-225715.70	-225682.93	-32.78
	12	-224951.43	-225715.44	-225660.36	-55.08
	14	-224951.44	-225715.50	-225660.77	-54.73
	18	-224951.42	-225715.56	-225660.65	-54.91
	23	-224951.72	-225715.77	-225665.65	-50.12
	29	-224951.79	-225715.90	-225665.19	-50.71
	46	-224951.39	-225715.47	-225660.50	-54.97
π/τ		-224952.41	-225712.00		
	1	-224950.25	-225709.77	-225664.60	-45.17
	2	-224950.45	-225710.38	-225662.88	-47.50
	3	-224950.35	-225710.19	-225665.29	-44.90
	10	-224950.20	-225711.16	-225662.41	-48.75
	11	-224951.79	-225708.93	-225684.81	-24.12
	17	-224950.09	-225710.51	-225669.88	-40.63
	18	-224949.91	-225710.38	-225666.30	-44.08
	22	-224950.05	-225710.36	-225670.01	-40.36
	23	-224951.72	-225708.95	-225684.83	-24.12
	26	-224949.40	-225710.31	-225661.70	-48.61
	29	-224950.34	-225710.38	-225665.08	-45.30

31	-224950.31	-225709.85	-225664.58	-45.27
34	-224950.46	-225710.47	-225662.92	-47.55
35	-224949.73	-225709.38	-225664.27	-45.11
45	-224949.18	-225706.66	-225675.26	-31.40
54	-224949.16	-225706.71	-225675.30	-31.41
55	-224949.84	-225710.44	-225666.11	-44.32
56	-224949.83	-225710.30	-225666.46	-43.84
61	-224950.23	-225709.94	-225664.28	-45.65
$\alpha/\pi/\tau$	-225236.75	-225983.71		
10	-225235.62	-225982.66	-225822.87	-159.79
11	-225234.34	-225981.81	-225821.53	-160.28
16	-225235.67	-225982.31	-225827.20	-155.11
17	-225235.55	-225982.76	-225822.82	-159.94
22	-225235.72	-225982.26	-225827.21	-155.05
24	-225235.58	-225982.63	-225822.78	-159.85
37	-225234.32	-225981.83	-225821.51	-160.32
38	-225234.37	-225981.78	-225821.47	-160.31
41	-225235.59	-225982.61	-225822.76	-159.85
61	-225234.90	-225981.26	-225827.23	-154.02

Tabelle 28: Energien der Histaminkonformere für MP2/aug-cc-pVDZ//MP2/aug-cc-pVDZ.

Tautomer	Konformer	PCM		EC-RISM	
		$G/(\text{kcal mol}^{-1})$	$G/(\text{kcal mol}^{-1})$	$E/(\text{kcal mol}^{-1})$	$\mu^{\text{ex}}/(\text{kcal mol}^{-1})$
π		-224635.04	-225402.85		
	1	-224633.54	-225400.93	-225379.05	-21.88
	2	-224633.70	-225401.86	-225378.28	-23.59
	3	-224633.60	-225402.00	-225379.16	-22.85
	10	-224633.07	-225402.65	-225380.17	-22.48
	11	-224633.52	-225398.69	-225389.92	-8.76
	14	-224632.62	-225400.04	-225385.57	-14.47
	16	-224632.77	-225401.97	-225375.34	-26.63
	18	-224633.09	-225402.59	-225380.26	-22.33
	21	-224632.86	-225401.71	-225380.57	-21.13
	22	-224633.54	-225398.62	-225390.07	-8.55
	23	-224632.64	-225400.05	-225385.73	-14.32
	24	-224632.87	-225401.73	-225380.61	-21.12
	29	-224633.50	-225401.06	-225378.85	-22.21
	30	-224633.59	-225402.07	-225379.13	-22.93
	34	-224633.67	-225401.89	-225378.37	-23.52
	36	-224633.11	-225400.79	-225379.25	-21.54
	50	-224632.50	-225401.94	-225378.65	-23.30
	54	-224632.78	-225402.08	-225375.13	-26.95
58	-224632.51	-225401.74	-225378.70	-23.04	
62	-224633.10	-225400.85	-225379.14	-21.71	
τ		-224634.98	-225404.85		
	1	-224633.58	-225399.98	-225383.48	-16.50
	2	-224633.75	-225401.31	-225381.09	-20.22
	3	-224633.68	-225401.00	-225380.34	-20.66
	10	-224632.42	-225400.71	-225384.59	-16.12
	11	-224632.92	-225401.26	-225384.75	-16.51
	12	-224631.57	-225401.25	-225375.91	-25.35
	14	-224632.40	-225401.09	-225382.20	-18.89
	17	-224632.81	-225400.17	-225379.73	-20.44
	18	-224633.16	-225401.41	-225382.41	-19.00
	20	-224631.59	-225401.61	-225375.30	-26.31
	21	-224632.95	-225401.14	-225384.60	-16.54
	23	-224632.41	-225401.01	-225382.11	-18.90
	26	-224632.42	-225400.47	-225384.97	-15.50
	28	-224633.57	-225400.01	-225383.31	-16.70
30	-224633.67	-225401.05	-225380.28	-20.77	

34	-224633.73	-225401.29	-225381.23	-20.06
38	-224631.57	-225400.92	-225376.33	-24.59
48	-224631.61	-225399.36	-225383.00	-16.36
52	-224631.68	-225400.43	-225380.32	-20.11
53	-224633.16	-225401.42	-225382.73	-18.68
54	-224632.80	-225399.92	-225380.32	-19.59
61	-224633.01	-225400.41	-225379.66	-20.75
62	-224633.02	-225400.51	-225379.55	-20.97
α/π	-224922.37	-225674.25		
1	-224921.36	-225672.92	-225599.07	-73.85
10	-224920.01	-225672.40	-225607.19	-65.21
11	-22492123	-225673.43	-225609.52	-63.91
12	-224920.11	-225672.31	-225607.76	-64.55
14	-224920.21	-225672.35	-225608.00	-64.35
16	-224920.28	-225672.25	-225608.17	-64.08
17	-224921.34	-225673.34	-225609.36	-63.98
21	-224920.33	-225672.31	-225607.94	-64.37
22	-224920.19	-225672.36	-225607.90	-64.46
30	-224921.32	-225672.84	-225599.26	-73.57
37	-224920.14	-225672.39	-225607.29	-65.10
59	-224920.16	-225672.25	-225608.12	-64.12
61	-224920.76	-225671.67	-225600.43	-71.23
62	-224920.75	-225671.70	-225600.43	-71.27
α/τ	-224925.28	-225674.17		
1	-224924.80	-225673.17	-225640.84	-32.33
10	-224921.71	-225672.04	-225611.04	-61.01
11	-224924.83	-225673.11	-225640.84	-32.27
12	-224922.97	-225672.77	-225618.09	-54.68
14	-224922.97	-225672.76	-225618.52	-54.24
18	-224921.68	-225672.13	-225610.89	-61.24
23	-224922.85	-225673.13	-225622.60	-50.53
29	-224922.82	-225673.11	-225622.77	-50.34
46	-224921.70	-225672.24	-225610.85	-61.39
50	-224921.72	-225672.10	-225611.02	-61.08
π/τ	-224923.86	-225669.34		
1	-224921.90	-225667.28	-225621.71	-45.57
2	-224922.09	-225667.69	-225619.90	-47.80
3	-224922.00	-225667.56	-225622.45	-45.12
10	-224921.61	-225668.50	-225619.53	-48.97
11	-224923.13	-225666.58	-225642.80	-23.78
17	-224921.58	-225667.41	-225627.01	-40.40

18	-224921.38	-225667.92	-225623.17	-44.76
22	-224921.58	-225667.47	-225627.03	-40.44
23	-224923.07	-225666.58	-225642.82	-23.77
26	-224920.92	-225667.58	-225618.80	-48.78
29	-224921.99	-225667.84	-225621.98	-45.86
31	-224921.95	-225667.22	-225621.77	-45.45
34	-224922.05	-225667.71	-225619.91	-47.80
35	-224921.68	-225667.16	-225621.44	-45.71
45	-224920.91	-225664.32	-225633.40	-30.93
54	-224920.98	-225664.50	-225633.00	-31.50
55	-224921.31	-225667.86	-225623.07	-44.78
56	-224921.28	-225667.75	-225623.56	-44.19
61	-224921.64	-225667.11	-225621.62	-45.49
$\alpha/\pi/\tau$	-225207.84	-225939.25		
10	-225206.68	-225938.21	-225778.84	-159.37
11	-225205.38	-225937.39	-225777.44	-159.95
16	-225206.96	-225938.07	-225783.40	-154.67
17	-225206.64	-225938.32	-225778.89	-159.4
22	-225207.03	-225938.07	-225783.39	-154.67
24	-225206.51	-225938.24	-225778.78	-159.46
37	-225205.40	-225937.34	-225777.50	-159.84
38	-225205.44	-225937.31	-225777.36	-159.95
61	-225206.44	-225937.35	-225783.76	-153.59

5.3 Aufbau des QM/MM-Subsystems des RTA-PTA-Komplexes

Tabelle 29: Atome des QM/MM-Subsystems, die zum Protein gehören oder Wasserstoffatome, die Proteinatome ersetzen.

Element	x / Å	y / Å	z / Å
O	5.777	1.205	9.349
C	6.659	2.054	9.449
H	6.775	3.065	8.609
H	7.671	1.936	10.553
H	4.161	2.305	5.775
N	5.080	1.925	5.603
H	5.216	0.960	4.494
H	6.100	2.286	6.401
H	1.103	-0.030	13.489
N	1.030	0.060	14.492
H	1.809	-0.856	15.307
H	0.280	1.032	15.009
H	-0.673	7.329	13.383
N	-0.637	6.676	14.150
H	0.233	6.478	14.623
H	-1.740	6.116	14.600
H	-2.891	3.062	11.779
N	-2.619	2.433	11.044
H	-1.654	2.159	10.952
H	-3.824	0.798	8.544
N	-3.151	1.121	9.229
H	-2.181	0.874	9.120
C	-3.529	1.959	10.189
H	-4.819	2.309	10.328
H	3.209	-1.618	9.299
H	4.188	-2.523	10.463
H	3.779	-3.249	8.895
C	3.415	-2.610	9.699

H	2.132	-3.213	10.334
---	-------	--------	--------

Tabelle 30: Schweratome und nicht zu titrierbaren Gruppen gehörende Wasserstoffatome des Pteroinsäure-Liganden.

Element	x / Å	y / Å	z / Å
C	0.580	0.687	10.717
C	0.828	1.307	9.400
C	2.108	1.528	8.916
N	-0.247	1.640	8.653
C	3.032	0.606	10.890
N	3.203	1.167	9.673
C	-0.042	2.204	7.456
N	4.087	0.327	11.624
C	1.277	2.426	6.992
C	-0.132	7.740	10.901
O	0.988	8.337	10.999
O	-0.497	0.487	11.175
N	1.777	0.360	11.395
N	2.340	2.087	7.715
C	-1.224	2.724	6.672
N	-1.104	4.168	6.509
C	-0.854	4.996	7.620
C	0.197	5.920	7.565
C	-1.648	4.938	8.778
C	0.438	6.784	8.622
C	-1.402	5.812	9.831
C	-0.362	6.754	9.769
O	-1.106	7.931	11.719
H	1.439	2.909	6.031
H	-1.242	2.260	5.684
H	-2.153	2.485	7.192
H	0.818	5.968	6.681
H	-2.471	4.239	8.841

H	1.245	7.503	8.556
H	-2.03	5.779	10.711

6 Literaturverzeichnis

- [1] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, G. R. Hutchison, *J. Cheminformatics* **2012**, *4*, 17.
- [2] L. Eberlein, *Konformationsaufklärung von pH-abhängigen Histamin-Spezies*, TU Dortmund, **2016**.
- [3] T. Kloss, J. Heil, S. M. Kast, *J. Phys. Chem. B* **2008**, *112*, 4337–4343.
- [4] R. Ramirez, D. Borgis, *J. Phys. Chem. B* **2005**, *109*, 6754–6763.
- [5] R. Ramirez, M. Mareschal, D. Borgis, *Chem. Phys.* **2005**, *319*, 261–272.
- [6] L. Gendre, R. Ramirez, D. Borgis, *Chem. Phys. Lett.* **2009**, *474*, 366–370.
- [7] S. Zhao, R. Ramirez, R. Vuilleumier, D. Borgis, *J. Chem. Phys.* **2011**, *134*, 194102.
- [8] R. Ramirez, R. Gebauer, M. Mareschal, D. Borgis, *Phys. Rev. E* **2002**, *66*, 31206.
- [9] D. Beglov, B. Roux, *J. Phys. Chem. B* **1997**, *101*, 7821–7826.
- [10] A. Kovalenko, F. Hirata, *Chem. Phys. Lett.* **1998**, *290*, 237–244.
- [11] E. L. Ratkova, D. S. Palmer, M. V. Fedorov, *Chem. Rev.* **2015**, *115*, 6312–6356.
- [12] J. Tomasi, B. Mennucci, R. Cammi, *Chem. Rev.* **2005**, *105*, 2999–3093.
- [13] C. J. Cramer, D. G. Truhlar, *Chem. Rev.* **1999**, *99*, 2161–2200.
- [14] B. Roux, T. Simonson, *Biophys. Chem.* **1999**, *78*, 1–20.
- [15] S. M. Kast, J. Heil, S. Güssregen, K. F. Schmidt, *J. Comput.-Aided. Mol. Des.* **2010**, *24*, 343–353.
- [16] D. Chandler, H. C. Andersen, *J. Chem. Phys.* **1972**, *57*, 1930–1937.
- [17] T. Morita, K. Hiroike, *Prog. Theor. Phys.* **1960**, *23*, 1003–1027.
- [18] S. J. Singer, D. Chandler, *Mol. Phys.* **1985**, *55*, 621–625.
- [19] F. Hirata, P. J. Rossky, *Chem. Phys. Lett.* **1981**, *83*, 329–334.
- [20] L. Blum, A. J. Torruella, *J. Chem. Phys.* **1972**, *56*, 303.
- [21] J. Percus, *Phys. Rev. Lett.* **1962**, *8*, 462–463.
- [22] J. S. Perkyns, B. M. Pettitt, *J. Chem. Phys.* **1992**, *97*, 7656–7666.
- [23] J. S. Perkyns, B. M. Pettitt, *Chem. Phys. Lett.* **1992**, *190*, 626–630.
- [24] Q. Du, D. Beglov, B. Roux, *J. Phys. Chem. B* **2000**, *104*, 796–805.
- [25] S. M. Kast, J. Heil, F. Hoffgaard, in *Computational Trends in Solvation and Transport in Liquids* (Hrsg.: G. Sutmann, J. Grotendorst, G. Gompper, D. Marx), FZ Jülich, Jülich, **2015**, pp. 419–434.
- [26] J.-P. Hansen, I. R. MacDonald, *Theory of Simple Liquids*, Academic Press, London, **2006**.
- [27] F. Hirata, Ed., *Molecular Theory of Solvation*, Kluwer Academic Publishers, Dordrecht, **2003**.
- [28] P. Hohenberg, W. Kohn, *Phys. Rev.* **1964**, *136*, B864–B871.
- [29] D. N. Mermin, *Phys. Rev.* **1965**, *137*, A1441.
- [30] R. Evans, *Adv. Phys.* **1979**, *28*, 143–200.
- [31] S. M. Kast, T. Kloss, *J. Chem. Phys.* **2008**, *129*, 236101.
- [32] A. Kovalenko, F. Hirata, *J. Chem. Phys.* **1999**, *110*, 10095.
- [33] R. Ishizuka, N. Yoshida, *J. Chem. Phys.* **2013**, *139*, 84119.

- [34] K. M. Dyer, J. S. Perkyns, G. Stell, B. M. Pettitt, *J. Chem. Phys.* **2008**, *129*, 104512.
- [35] J. S. Høye, G. Stell, *J. Chem. Phys.* **1976**, *65*, 18.
- [36] P. T. Cummings, G. Stell, *Mol. Phys.* **1981**, *44*, 529–531.
- [37] S. M. Kast, K. F. Schmidt, B. Schilling, *Chem. Phys. Lett.* **2003**, *367*, 398–404.
- [38] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- [39] J. G. Kirkwood, *J. Chem. Phys.* **1935**, *3*, 300–313.
- [40] R. A. Chiles, P. J. Rossky, *J. Am. Chem. Soc.* **1984**, *106*, 6867–6868.
- [41] R. W. Zwanzig, *J. Chem. Phys.* **1954**, *22*, 1420–1426.
- [42] T. Miyata, F. Hirata, *J. Comput. Chem.* **2008**, *29*, 871–882.
- [43] L. L. Lee, *Molecular Thermodynamics of Nonideal Fluids*, Butterworth Publishers, Stoneham, MA, **1988**.
- [44] J. W. Cooley, J. W. Tukey, *Math. Comput.* **1965**, *19*, 297–301.
- [45] K.-C. Ng, *J. Chem. Phys.* **1974**, *61*, 2680–2689.
- [46] P. P. Ewald, *Ann. Phys.* **1921**, *369*, 253–287.
- [47] B. A. Wells, A. L. Chaffee, *J. Chem. Theory Comput.* **2015**, *11*, 3684–3695.
- [48] A. Kovalenko, F. Hirata, *J. Chem. Phys.* **2000**, *112*, 10391.
- [49] T. Kloss, S. M. Kast, *J. Chem. Phys.* **2008**, *128*, 134505.
- [50] J. D. Talman, *J. Comput. Phys.* **1978**, *29*, 35–48.
- [51] S. Gusarov, B. S. Pujari, A. Kovalenko, *J. Comput. Chem.* **2012**, *33*, 1478–1494.
- [52] J. S. Perkyns, G. C. Lynch, J. J. Howard, B. M. Pettitt, *J. Chem. Phys.* **2010**, *132*, 64106.
- [53] J. Heil, *pK_a-Vorhersagemodelle auf der Basis von dreidimensionaler Integralgleichungstheorie*, TU Darmstadt, **2008**.
- [54] S. A. Petrosyan, A. A. Rigos, T. A. Arias, *J. Phys. Chem. B* **2005**, *109*, 15436–15444.
- [55] S. Ten-no, F. Hirata, S. Kato, *Chem. Phys. Lett.* **1993**, *214*, 391–396.
- [56] H. Sato, F. Hirata, S. Kato, *J. Chem. Phys.* **1996**, *105*, 1546–1551.
- [57] A. Kovalenko, F. Hirata, *J. Chem. Phys.* **1999**, *110*, 10095–10112.
- [58] H. Sato, A. Kovalenko, F. Hirata, *J. Chem. Phys.* **2000**, *112*, 9463–9468.
- [59] N. Yoshida, S. Kato, *J. Chem. Phys.* **2000**, *113*, 4974–4984.
- [60] K. Kido, K. Kasahara, D. Yokogawa, H. Sato, *J. Chem. Phys.* **2015**, *143*, 14103.
- [61] G. N. Chuev, M. Valiev, M. V Fedotova, *J. Chem. Theory Comput.* **2012**, *8*, 1246–1254.
- [62] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. Montgomery, J. A., T. Vreven, K. N. Kudin, J. C. Burant, et al., **2003**.
- [63] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, et al., **2009**.
- [64] L. Greengard, V. Rokhlin, *J. Comput. Phys.* **1987**, *73*, 325–348.
- [65] C. Møller, M. S. Plesset, *Phys. Rev.* **1934**, *46*, 618–622.
- [66] *OpenMP Application Program Interface*, Version 3.0, OpenMP Architecture Review Board, **2008**.
- [67] R. Hockney, J. W. Eastwood, *Computer Simulation using Particles*, Taylor & Francis, New York, **1988**.
- [68] M. Griebel, S. Knapek, G. Zumbusch, A. Caglar, *Numerische Simulation in der Moleküldynamik*, Springer, **2004**.

- [69] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- [70] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **1995**, *103*, 8577.
- [71] M. Pippig, D. Potts, *SIAM J. Sci. Comput.* **2013**, *35*, C411–C437.
- [72] Y.-L. Wang, F. Hedman, M. Porcu, F. Mocci, A. Laaksonen, *Appl. Math.* **2014**, *5*, 520–541.
- [73] V. Ballenegger, J. J. Cerdà, C. Holm, *J. Chem. Theory Comput.* **2012**, *8*, 936–947.
- [74] A. Arnold, F. Fahrenberger, C. Holm, O. Lenz, M. Bolten, H. Dachsels, R. Halver, I. Kabadshow, F. Gähler, F. Heber, et al., *Phys. Rev. E* **2013**, *88*, 63308.
- [75] C. Runge, *Z. Math. Phys.* **1901**, *46*, 224.
- [76] M. Heinen, E. Allahyarov, H. Löwen, *J. Comput. Chem.* **2014**, *35*, 275–289.
- [77] R. D. Blumofe, C. F. Joerg, B. C. Kuszmaul, C. E. Leiserson, K. H. Randall, Y. Zhou, *SIGPLAN Not.* **1995**, *30*, 207–216.
- [78] G. Hager, G. Wellein, *Introduction to High Performance Computing for Scientists and Engineers*, CRC Press, **2011**.
- [79] J. W. Neidigh, R. M. Fesinmeyer, N. H. Andersen, *Nat. Struct. Biol.* **2002**, *9*, 425–430.
- [80] *MPI-2: Extensions to the Message-Passing Interface*, Message Passing Interface Forum, **2003**.
- [81] F. Hoffgaard, J. Heil, S. M. Kast, unveröffentlichte Ergebnisse.
- [82] A. Ruf, G. De Murcia, G. E. Schulz, *Biochemistry* **1998**, *37*, 3893–3900.
- [83] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33–38.
- [84] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
- [85] D. Porezag, T. Frauenheim, T. Köhler, *Phys. Rev. B* **1995**, *51*, 12947–12957.
- [86] G. Seifert, D. Porezag, T. Frauenheim, *Int. J. Quantum Chem.* **1996**, *58*, 185–192.
- [87] H. Lin, D. G. Truhlar, *Theor. Chem. Acc.* **2007**, *117*, 185–199.
- [88] H. M. Senn, W. Thiel, *Angew. Chemie - Int. Ed.* **2009**, *48*, 1198–1229.
- [89] T. Vreven, B. Mennucci, C. O. Da Silva, K. Morokuma, J. Tomasi, *J. Chem. Phys.* **2001**, *115*, 62–72.
- [90] F. Lipparini, B. Stamm, E. Cancès, Y. Maday, B. Mennucci, *J. Chem. Theory Comput.* **2013**, *9*, 3637–3648.
- [91] F. Lipparini, G. Scalmani, L. Lagardère, B. Stamm, E. Cancès, Y. Maday, J.-P. Piquemal, M. J. Frisch, B. Mennucci, *J. Chem. Phys.* **2014**, *141*, 184108.
- [92] F. Lipparini, L. Lagardère, G. Scalmani, B. Stamm, E. Cancès, Y. Maday, J.-P. Piquemal, M. J. Frisch, B. Mennucci, *J. Phys. Chem. Lett.* **2014**, *5*, 953–958.
- [93] S. Caprasecca, S. Jurinovich, L. Lagardère, B. Stamm, F. Lipparini, *J. Chem. Theory Comput.* **2015**, *11*, 694–704.
- [94] E. Cancès, Y. Maday, B. Stamm, *J. Chem. Phys.* **2013**, *139*, 54111.
- [95] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- [96] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- [97] N. Yoshida, Y. Kiyota, F. Hirata, *J. Mol. Liq.* **2011**, *159*, 83–92.
- [98] J. Barnes, P. Hut, *Nature* **1986**, *324*, 446–449.
- [99] S. Fortune, *Algorithmica* **1987**, *2*, 153–174.

- [100] D. G. Kirkpatrick, R. Seidel, *SIAM J. Comput.* **1986**, *15*, 287–299.
- [101] C. H. Rycroft, *Multiscale Modeling in Granular Flow*, Massachusetts Institute of Technology, **2007**.
- [102] C. H. Rycroft, *Chaos* **2009**, *19*, 41111.
- [103] D. M. Greaves, A. G. L. Borthwick, *Int. J. Numer. Methods Eng.* **1999**, *45*, 447–471.
- [104] C. Tanford, J. G. Kirkwood, *J. Am. Chem. Soc.* **1957**, *79*, 5333–5339.
- [105] D. Bashford, M. Karplus, *J. Phys. Chem.* **1991**, *95*, 9556–9561.
- [106] G. M. Ullmann, E.-W. Knapp, *Eur. Biophys. J.* **1999**, *28*, 533–51.
- [107] A. Onufriev, D. A. Case, G. M. Ullmann, *Biochemistry* **2001**, *40*, 3413–3419.
- [108] M. K. Gilson, *Proteins Struct. Funct. Genet.* **1993**, *15*, 266–282.
- [109] M. K. Gilson, J. A. Given, B. L. Bush, J. A. McCammon, *Biophys. J.* **1997**, *72*, 1047–1069.
- [110] S.-H. Chong, S. Ham, *J. Chem. Theory Comput.* **2015**, *11*, 378–380.
- [111] V. Sergiievskiy, G. Jeanmairet, M. Levesque, D. Borgis, *J. Chem. Phys.* **2015**, *143*, 184116.
- [112] M. D. Tissandier, K. A. Cowen, W. Y. Feng, W. Y. Feng, E. Gundlach, M. H. Cohen, A. D. Earhart, J. V. Coe, T. R. Tuttle Jr., *J. Phys. Chem. A* **1998**, *102*, 7787–7794.
- [113] C.-G. Zhan, D. A. Dixon, *J. Phys. Chem. A* **2001**, *105*, 11534–11540.
- [114] A. D. McNaught, A. Wilkinson, Eds., *IUPAC Compendium of Chemical Terminology*, Blackwell Scientific Publications, Oxford, **1997**.
- [115] G. S. Kell, *J. Chem. Eng. Data* **1975**, *20*, 97–105.
- [116] A. Ben-Naim, *Molecular Theory of Solutions*, Oxford University Press, New York, **2006**.
- [117] T. Luchko, S. Gusarov, D. R. Roe, C. Simmerling, D. A. Case, J. Tuszynski, A. Kovalenko, *J. Chem. Theory Comput.* **2010**, *6*, 607–624.
- [118] M. Misin, M. V. Fedorov, D. S. Palmer, *J. Chem. Theory Comput.* **2016**, *120*, 975–983.
- [119] D. M. Camaioni, C. A. Schwerdtfeger, *J. Phys. Chem. A* **2005**, *109*, 10795–10797.
- [120] J. Ho, M. L. Coote, *Theor. Chem. Acc.* **2010**, *125*, 3–21.
- [121] R. de Levie, *Aqueous Acid-Base Equilibria and Titrations*, Oxford University Press, Oxford, **1999**.
- [122] E. E. Dahlke, C. J. Cramer, *J. Phys. Org. Chem.* **2003**, *16*, 336–347.
- [123] A. Jinich, D. Rappoport, I. Dunn, B. Sanchez-Lengeling, R. Olivares-Amaya, E. Noor, A. B. Even, A. Aspuru-Guzik, *Sci. Rep.* **2014**, *4*, 7022.
- [124] F. Forti, C. N. Cavasotto, M. Orozco, X. Barril, F. J. Luque, *J. Chem. Theory Comput.* **2012**, *8*, 1808–1819.
- [125] P. G. Seybold, G. C. Shields, *WIREs Comput. Mol. Sci.* **2015**, *5*, 290–297.
- [126] R. A. Abramson, K. K. Baldridge, *J. Chem. Theory Comput.* **2013**, *9*, 1027–1035.
- [127] F. Eckert, M. Diedenhofen, A. Klamt, *Mol. Phys.* **2010**, *108*, 229–241.
- [128] J. J. Klicic, R. A. Friesner, S.-Y. Liu, W. C. Guida, *J. Phys. Chem. A* **2002**, *106*, 1327–1335.
- [129] Y. Chiang, E. B. Whipple, *J. Am. Chem. Soc.* **1963**, *85*, 2763–2767.
- [130] R. L. Hinman, E. B. Whipple, *J. Am. Chem. Soc.* **1962**, *84*, 2534–2539.
- [131] S. A. Benner, M. A. Sismour, *Nat. Rev. Genet.* **2005**, *6*, 533–543.
- [132] M. D. Topal, J. R. Fresco, *Nature* **1976**, *263*, 285–289.
- [133] T. A. Martinot, S. A. Benner, *J. Org. Chem.* **2004**, *69*, 3972–3975.
- [134] F. Seela, R. Kröschel, *Nucleic Acids Res.* **2003**, *31*, 7150–7158.

- [135] D. Jiang, F. Seela, *J. Am. Chem. Soc.* **2010**, *132*, 4016–4024.
- [136] D. Loakes, P. Holliger, *Chem. Commun.* **2009**, 4619–31.
- [137] C. Roberts, R. Bandaru, C. Switzer, *J. Am. Chem. Soc.* **1997**, *119*, 4640–4649.
- [138] H. L. Haas, O. A. Sergeeva, O. Selbach, *Physiol. Rev.* **2008**, *88*, 1183–1241.
- [139] F. J. Ramírez, I. Tuñón, J. A. Collado, E. Silla, *J. Am. Chem. Soc.* **2003**, *125*, 2328–2340.
- [140] H. Weinstein, D. Chou, C. L. Johnson, S. Kang, J. P. Grenn, *Mol. Pharmacol.* **1976**, *12*, 738–745.
- [141] T. B. Paiva, M. Tominaga, A. C. M. Paiva, *J. Med. Chem.* **1970**, *13*, 689–692.
- [142] G. A. Worth, P. M. King, W. G. Richards, *Biochim. Biophys. Acta, Gen. Subj.* **1990**, *1036*, 158–161.
- [143] G. Karpińska, J. C. Dobrowolski, A. P. Mazurek, *J. Mol. Struct. THEOCHEM* **1996**, *369*, 137–144.
- [144] C. R. Ganellin, *J. Pharm. Pharmacol.* **1973**, *25*, 787–792.
- [145] W. F. Reynolds, C. W. Tzeng, *Can. J. Biochem.* **1976**, *55*, 576–578.
- [146] M. T. Geballe, A. G. Skillman, A. Nicholls, J. P. Guthrie, P. J. Taylor, *J. Comput.-Aided. Mol. Des.* **2010**, *24*, 259–279.
- [147] Z. Szakács, B. Noszál, *Electrophoresis* **2006**, *27*, 3399–3409.
- [148] K. Wüthrich, *J. Biol. Chem.* **1990**, *265*, 22059–22062.
- [149] F. Shu, V. Ramakrishnan, B. P. Schoenborn, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 3872–3877.
- [150] N. Shimizu, S. Sugiyama, M. Maruyama, Y. Takahashi, M. Adachi, T. Tamada, K. Hidaka, Y. Hayashi, T. Kimura, Y. Kiso, et al., *Cryst. Growth Des.* **2010**, *10*, 2990–2994.
- [151] G. L. Warren, T. D. Do, B. P. Kelley, A. Nicholls, S. D. Warren, *Drug Discov. Today* **2012**, *17*, 1270–1281.
- [152] P. Beroza, D. R. Fredkin, M. Y. Okamura, G. Feher, *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 5804–8.
- [153] C. Tanford, R. Roxby, *Biochemistry* **1972**, *11*, 2192–2198.
- [154] J. Myers, G. Grothaus, S. Narayanan, A. V Onufriev, *Proteins* **2006**, *63*, 928–938.
- [155] X. Yan, T. Hollis, M. Svinth, P. Day, A. F. Monzingo, G. W. Milne, J. D. Robertus, *J. Mol. Biol.* **1997**, *266*, 1043–1049.
- [156] Y. Endo, K. Tsurugi, *J. Biol. Chem.* **1987**, *262*, 8128–8130.
- [157] Y. Endo, K. Tsurugi, *J. Biol. Chem.* **1988**, *263*, 8735–8739.
- [158] M. Karwa, B. Currie, V. Kvetan, *Crit. Care Med.* **2005**, *33*, S75–S95.
- [159] F. Milletti, L. Storchi, G. Sforza, S. Cross, G. Cruciani, *J. Chem. Inf. Model.* **2009**, *49*, 68–75.
- [160] J. Engel, A. Richters, M. Getlik, S. Tomassi, M. Keul, M. Termathe, J. Lategahn, C. Becker, S. Mayer-Wrangowski, C. Grütter, et al., *J. Med. Chem.* **2015**, *58*, 6844–6863.
- [161] A. Jakalian, D. B. Jack, C. I. Bayly, *J. Comput. Chem.* **2002**, *23*, 1623–1641.
- [162] A. Jakalian, B. L. Bush, D. B. Jack, C. I. Bayly, *J. Comput. Chem.* **1999**, *21*, 132–146.
- [163] H. C. Andersen, *J. Chem. Phys.* **1980**, *72*, 2384–2393.
- [164] G. Sigalov, A. Fenley, A. Onufriev, *J. Chem. Phys.* **2006**, *124*, 124902.
- [165] G. Sigalov, P. Scheffel, A. Onufriev, *J. Chem. Phys.* **2005**, *122*, 94511.
- [166] D. C. Liu, J. Nocedal, *Math. Program.* **1989**, *45*, 503–528.
- [167] J. Nocedal, *Math. Comput.* **1980**, *35*, 773–782.

- [168] R. A. Jarvis, E. A. Patrick, *IEEE Trans. Comput.* **1973**, C-22, 1025–1034.
- [169] J. P. M. Jämbeck, A. P. Lyubartsev, *J. Phys. Chem. B* **2014**, 118, 3793–3804.
- [170] D. J. Brown, T. Teitei, *Aust. J. Chem.* **1965**, 18, 559–568.
- [171] D. D. Perrin, *Ionisation Constants of Inorganic Acids and Bases in Aqueous Solution*, Pergamon Press, New York, **1982**.
- [172] R. Anandakrishnan, B. Aguilar, A. V Onufriev, *Nucleic Acids Res.* **2012**, 40, W537-41.
- [173] T. A. Halgren, *J. Comput. Chem.* **1999**, 20, 720–729.
- [174] P. M. Dewick, *Essentials of Organic Chemistry*, John Wiley & Sons, Chichester, **2006**.
- [175] G. A. Jeffrey, *An Introduction to Hydrogen Bonding*, Oxford University Press, Oxford, **1997**.
- [176] D. Tomazic, *Optimizing Free Energy Functionals in Integral Equation Theories*, TU Dortmund, **2016**.
- [177] N. Tielker, D. Tomazic, J. Heil, T. Kloss, S. Ehrhart, S. Güssregen, K. F. Schmidt, S. M. Kast, *J. Comput.-Aided. Mol. Des.* **2016**, DOI 10.1007/s10822-016-9939-7.
- [178] C. Riplinger, F. Neese, *J. Chem. Phys.* **2013**, 138, 34106.
- [179] F. Pavošević, P. Pinski, C. Riplinger, F. Neese, E. F. Valeev, *J. Chem. Phys.* **2016**, 144, 144109.
- [180] T. H. Rasmussen, Y. M. Wang, T. Kjærgaard, K. Kristensen, *J. Chem. Phys.* **2016**, 144, 204102.
- [181] A. Heßelmann, G. Jansen, *Chem. Phys. Lett.* **2003**, 367, 778–784.
- [182] A. J. Misquitta, B. Jeziorski, K. Szalewicz, *Phys. Rev. Lett.* **2003**, 91, 33201.
- [183] B. Jeziorski, R. Moszynski, K. Szalewicz, *Chem. Rev.* **1994**, 94, 1887–1930.
- [184] K. U. Lao, J. M. Herbert, *J. Phys. Chem. Lett.* **2012**, 3, 3241–3248.
- [185] R. M. Betz, R. C. Walker, *J. Comput. Chem.* **2015**, 36, 79–87.