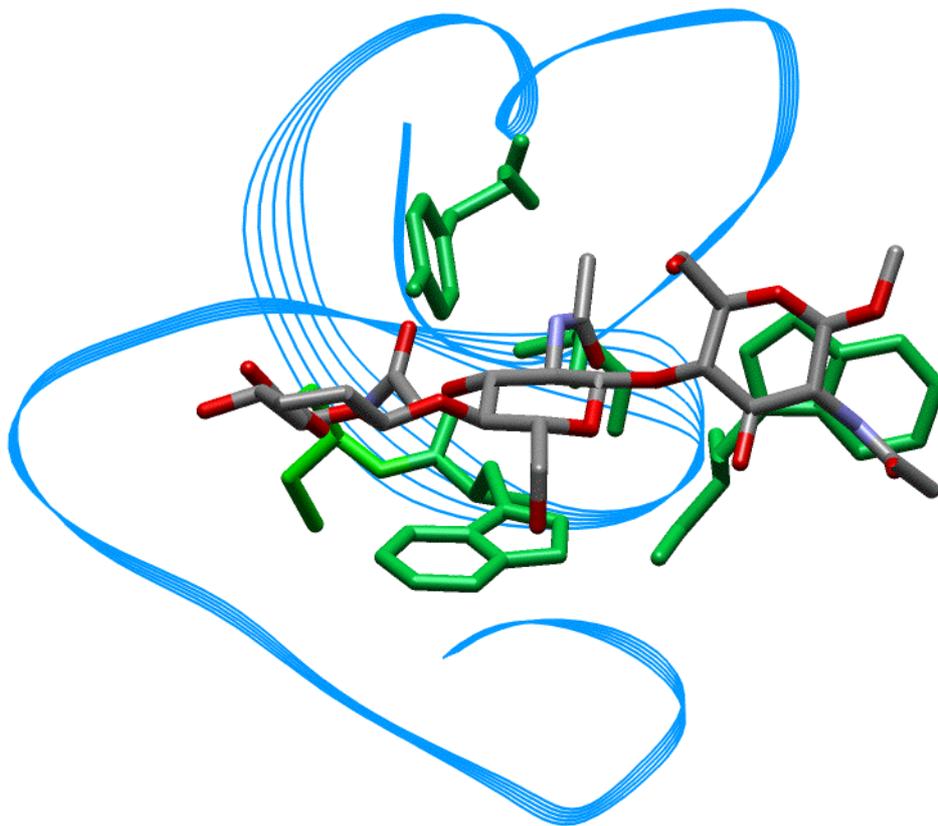


# Konformationsanalyse von Oligosacchariden

---

im freien und gebundenen Zustand



---

Martin Frank

# INAUGURAL-DISSERTATION

zur  
Erlangung der Doktorwürde  
der  
Naturwissenschaftlich-Mathematischen Gesamtfakultät  
der  
Ruprecht-Karls-Universität  
Heidelberg

vorgelegt von  
Diplom-Chemiker Martin Frank  
aus Schwäbisch Hall

Tag der mündlichen Prüfung: 15. März 2000

# **Konformationsanalyse von Oligosacchariden im freien und gebundenen Zustand**

Entwicklung effizienter Simulations-Strategien und Analyse-Methoden,  
experimentelle Untersuchungen zur Bindungsaffinität mittels NMR-Methoden

Gutachter: Prof. Dr. Peter Comba

Prof. Dr. Manfred Wießler

Für Sylvia

*Carpe Diem*

# Danksagung

Prof. Dr. Peter Comba (Uni Heidelberg) danke ich für die Betreuung dieser Arbeit, die anregenden wissenschaftlichen Diskussionen, die Bewegung an der frischen Luft beim Skiseminar und die tolle Atmosphäre im AK Comba.

Mein besonderer Dank gilt Dr. Claus-Wilhelm von der Lieth (DKFZ Heidelberg, Zentrale Spektroskopie) für die Überlassung des faszinierenden Themas wie auch für die fachliche Betreuung dieser Arbeit. Dank auch dafür, daß er mir die Möglichkeit gegeben hat, mich kreativ frei zu entfalten und für sein stetes Bemühen mich mit allem zu versorgen, was ich zum effektiven Arbeiten benötigte.

Dem Deutschen Krebsforschungszentrum (DKFZ) Heidelberg danke ich für die finanzielle und technische Unterstützung dieser Arbeit.

Namentlich gedankt sei weiterhin:

Prof. Dr. Manfred Wießler (DKFZ, Molekulare Toxikologie) für die Begutachtung der Arbeit und die anregenden wissenschaftlichen Diskussionen

Dr. William E. Hull (DKFZ, Zentrale Spektroskopie) für die Einarbeitung am NMR-Spektrometer und die tiefgreifenden Fachdiskussionen

Andreas Bohne (DKFZ) für die Hilfe bei UNIX und C Problemen

Gabi Schwebel-Schilling (DKFZ) für die Unterstützung bei den alltäglichen Formalitäten

Dr. Jesús Jiménez-Barbero, Dr. Francisco Javier Canada und Dr. Juan-Félix Espinosa (CSIC) für die Unterstützung bei den NMR-Messungen in Madrid

Dr. Emadeddin Tajkhorshid (DKFZ) für die Durchführung der *ab initio* Rechnungen

Priv. Doz. Dr. Hans-Christian Siebert (LMU München) für die Koordination des DMSO-Projekts

Prof. Dr. Hans-Joachim Gabius (LMU München) für das Interesse an meiner Arbeit und für Bereitstellung der Galectine

Prof. Dr. Janus Dabrowski (MPI) für die langen Diskussionen zu Sialyl-Lewis<sup>X</sup>

Dr. Michael Pawlita (DKFZ, ATV) für die Züchtung des Polyoma-Virus

Prof. Dr. Reinhard Brossmer (Uni Heidelberg) für die Bereitstellung der Sialinsäurederivate

Tobias Reber (DKFZ, ZDV) für die Administration der SP2

den Mitarbeitern des AK Comba und der Zentralen Spektroskopie (DKFZ), daß sie mich in ihren Kreis aufgenommen haben

Besondere Dankbarkeit gilt meiner Lebenspartnerin Sylvia, die oft auf mich verzichten mußte, weil ich wieder nächtelang am Computer saß. Danke auch für die konstruktiven Ermahnungen, mich nicht in Details zu verlieren.

Meinen Eltern und meiner Familie danke ich für ihre Unterstützung und dafür, daß sie immer zu mir gehalten und nicht die Geduld verloren haben.

Veröffentlichungen, bei denen bereits Ergebnisse und Methoden dieser Arbeit eingeflossen sind:

Frank, M. and C.-W. von der Lieth (1997). "Comparison of the conformational behavior of sialyllactose complexed with the two viral attachment proteins influenza a hemagglutinin and the murine poliovirus." *J. Mol. Model.* 3: 408-414.

Siebert, H.-C.; André, S.; Dong, X.; Espinosa, J.-F.; Frank, M.; Gilleron, M.; Kaltner, H.; Kozár, T.; Beintema, J. J.; Bovin, N. V.; von der Lieth, C.-W.; Vliegthart, J. F. G.; Jiménez-Barbero, J.; Gabius, H.-J. (2000). "NMR investigation of carbohydrate-protein (lectin and immunoglobulin G) interactions in dimethyl sulfoxide: strategy for identification of new conformational constraints of the bound ligand by maintenance of receptor activity in the aprotic solvent." *to be submitted*.

# Abkürzungen

CAT	Conformational Analysis Tools
CC	Conformational Clustering
CIDNP	Chemically Induced Dynamic Nuclear Polarization
CPU	Central Processing Unit
CSL	CAT Script Language
CVFF	Consistent Valence Force Field
DMSO	Dimethylsulfoxid
FGE	Freie Gridgruppen Energie
Fuc	Fucose
Gal	Galactose
GCMD	Grid Search Conformational Clustering Molecular Dynamics
Glc	Glucose
HSEA	Hard Sphere Exo Anomeric
HTMD	High Temperature Molecular Dynamics
iProp	iso-Propyl
MD	Molekül Dynamik
Neu5Ac	N-Acetylneuraminsäure, Sialinsäure (Sia)
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
PAVDO	Programm zu Analyse von Discover Outputfiles
ROESY	Rotating Frame Nuclear Overhauser Effect Spectroscopy
$t_0$	Equilibrierungs Phase
$t_s$	Sampling Phase (Data Collection Period)
$\Delta t_s$	Sampling Intervall
vdW	van-der-Waals

Ich versichere an Eides Statt,  
daß ich die vorliegende Arbeit  
selbständig und nur mit den  
erlaubten Hilfsmitteln durch-  
geführt habe.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung .....</b>	<b>1</b>
1.1	Kohlenhydrate - natürliche Schlüsseltechnologie des Informationstransfers .....	1
1.2	Auf dem Weg zur Entwicklung hochaffiner Therapeutika .....	3
1.2.1	Experimentelle Methoden zur Untersuchung von Protein-Kohlenhydrat-Komplexen .....	3
1.2.2	Welche Faktoren beeinflussen die Komplexbildung? .....	4
1.2.3	Thermodynamik des Bindungsprozesses .....	5
1.2.4	Semiquantitative Vorhersage von Bindungskonstanten durch Kraftfeldmethoden .....	7
1.3	Konformationsanalyse von Oligosacchariden .....	9
1.4	MD Simulationen - Theorie .....	11
<b>2</b>	<b>Allgemeine Definitionen und Grundlagen .....</b>	<b>14</b>
2.1	Wichtige Strukturparameter eines Oligosaccharids .....	14
2.1.1	Übersicht .....	14
2.1.2	Stellung der Hydroxymethylgruppe - Definitionen .....	14
2.1.3	Systematische Aufteilung eines Kohlenhydrats in funktionelle Untereinheiten .....	15
2.2	Conformational Analysis Tools (CAT2000) .....	16
2.3	Gridbasierte statistische Analyse von molekularen Größen .....	18
2.3.1	Definition eines Grids .....	18
2.3.2	Häufigkeitsverteilungen .....	19
2.3.3	Energie-Grids .....	21
2.3.4	Überführung von Populations-Grids in Energie-Grids (Freie Gridgruppen-Energie) .....	22
2.4	Konformationsgleichgewichte .....	25
2.4.1	Definition eines Parameters zur Beschreibung der Einstellung von Konformationsgleichgewichten .....	25
2.4.2	Bewertung der verschiedenen Gleichgewichtsparameter anhand der Einstellung eines Hydroxymethyl-Rotamerengleichgewichts .....	31
<b>3</b>	<b>Moleküldynamik-Simulationen als Methode zur effizienten Konformationsanalyse von Oligosacchariden .....</b>	<b>32</b>
3.1	Einleitung und Zielsetzung .....	32
3.2	Allgemeine Angaben zu den Simulationen .....	32
3.3	MD-basierte Methoden zur Population des Konformationsraums von Oligosacchariden .....	34
3.3.1	Hochtemperatur-MD-Simulationen (HTMD) .....	34
3.3.2	Kombination von Grid-Search- und MD-Methoden zur Konformationsanalyse von Oligosacchariden (GCMD-Methode) .....	42
3.3.3	Analyse von MD-Ensembles zur Unterstützung der Strukturaufklärung mittels NMR-Methoden .....	43

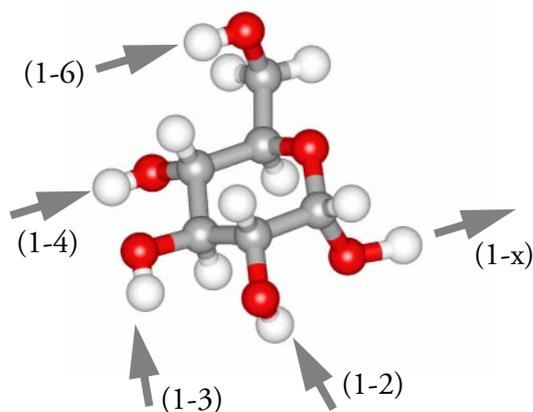
3.4	Analyse der Rotamerengleichgewichte .....	44
3.4.1	Einleitung .....	44
3.4.2	Rotationsprofile, Rotamerenpopulation und Gleichgewichtseinstellung der OH-Gruppen .....	44
3.4.3	„Nachbargruppeneffekte“ .....	46
3.4.4	Rotationsprofile, Rotamerenpopulation und Gleichgewichtseinstellung der Hydroxymethylgruppen .....	47
3.4.5	Rotationsprofile, Rotamerenpopulation und Gleichgewichtseinstellung der glycosidischen Torsionswinkel .....	48
3.4.6	Reproduzierbarkeit von Rotamerengleichgewichten .....	49
3.4.7	Einfluß der Simulationstemperatur auf die Einstellung der Rotamerengleichgewichte .....	49
3.5	Analyse von Rotamerenübergängen .....	51
3.5.1	Einleitung .....	51
3.5.2	Bestimmung der mittleren Lebensdauern von Rotameren .....	53
3.5.3	Die mittlere Lebensdauer von $\Phi/\Psi$ -Konformationen bei 300 K .....	58
3.6	Arrhenius Analyse zur Bestimmung von Übergangsenergien .....	60
3.6.1	Bestimmung von Übergangsbarrieren ausgewählter Modellsysteme und Vergleich mit experimentellen Werten .....	60
3.6.2	Bestimmung der effektiven Übergangsbarrieren von Kohlenhydraten aus der mittleren Lebensdauer einer Konformation .....	62
3.6.3	Bestimmung der Übergangsbarrieren definierter Konformationsübergänge .....	64
3.6.4	Bestimmung bevorzugter Übergangswege .....	65
3.7	Einfluß der Dielektrizitätskonstanten und Solvenseffekte .....	68
3.7.1	Radial-Histogramme zur Analyse der Solvensstruktur .....	68
3.7.2	Minimierungen von solvatisierten Molekülen .....	71
3.7.3	Einfluß der Dielektrizitätskonstanten und des Solvens auf die Freie Gridgruppen-Energie .....	72
3.7.4	Analyse spezifischer Solvens-Effekte .....	73
3.8	Zusammenfassung .....	75
<b>4</b>	<b>Berechnung von Rotationsprofilen der Freien Energie basierend auf Grid-Search-Methoden .....</b>	<b>76</b>
4.1	Einleitung und Problemstellung .....	76
4.2	Methoden .....	77
4.2.1	Beschreibung der Methode .....	77
4.2.2	Berechnung von Rotationsprofilen der Freien Energie aus multidimensionalen Grids .....	78
4.3	Ergebnisse aus dem Grid-Search .....	79
4.3.1	Zweidimensionale Energiekarten .....	79
4.3.2	FGE-Profile .....	81
4.3.3	Abhängigkeit der Rotamerenverteilung vom Kraftfeld und der Dielektrizitätskonstanten .....	82
4.3.4	Vergleich und Bewertung der Ergebnisse für Glc $\alpha$ -O-iProp und Gal $\beta$ -O-iProp .....	84
4.3.5	Einfluß der Orientierung der 4OH-Gruppe auf die Rotamerenverteilung .....	86
4.4	Vergleich mit Ab-Initio-Rechnungen .....	88

<b>5</b>	<b>Untersuchungen zur Bindungsaffinität von Protein-Kohlenhydrat-Komplexen .....</b>	<b>90</b>
5.1	Einleitung .....	90
5.2	Modellierung von Hevein-Chitotriose-Komplexen .....	90
5.2.1	Problemstellung und Zielsetzung .....	90
5.2.2	Methoden .....	91
5.2.3	Überprüfung der Stabilität der Proteinstruktur bei MD-Simulationen .....	92
5.2.4	Interaktions-Analyse des Hevein/Chitobiose-Komplexes .....	93
5.2.5	Modellierung des Hevein-Chitotriose-Komplexes .....	95
5.2.6	Interaktionsanalyse des Hevein-Chitotriose-Komplexes in DMSO .....	98
5.3	Experimentelle Bestimmung der Bindungsaffinität von Hevein-Chitotriose-Komplexen in Wasser/DMSO-Mischungen mittels NMR-Methoden .....	99
5.3.1	Problemstellung und Zielsetzung .....	99
5.3.2	Beschreibung der Methode .....	99
5.3.3	Vorbereitung und Reinigung der Hevein-Probe .....	101
5.3.4	Änderung der chemischen Verschiebung der Protonensignale von Hevein durch Zugabe von DMSO .....	101
5.3.5	Hevein/Chitotriose: Bestimmung der Bindungskonstanten in 20% DMSO .....	102
5.3.6	Hevein/Chitotriose: Van't Hoff Analyse (20% DMSO) .....	103
5.3.7	Hevein-Chitotriose: Bestimmung der Bindungskonstante in 50% DMSO (309 K) .....	105
5.3.8	Änderung der Bindungskonstante mit zunehmender DMSO-Konzentration .....	107
5.3.9	Interpretation der Ergebnisse .....	108
5.4	Zusammenfassung und Ausblick .....	109
<b>6</b>	<b>Zusammenfassung .....</b>	<b>111</b>
<b>7</b>	<b>Anhang .....</b>	<b>114</b>
7.1	Gruppierungs- und Clustering-Methoden von CAT2000 .....	114
7.2	Experimentelle Bestimmung des Rotamerengleichgewichts .....	115
7.3	Tabellen .....	115
<b>8</b>	<b>Abbildungen (X) .....</b>	<b>117</b>
<b>9</b>	<b>Literatur .....</b>	<b>257</b>

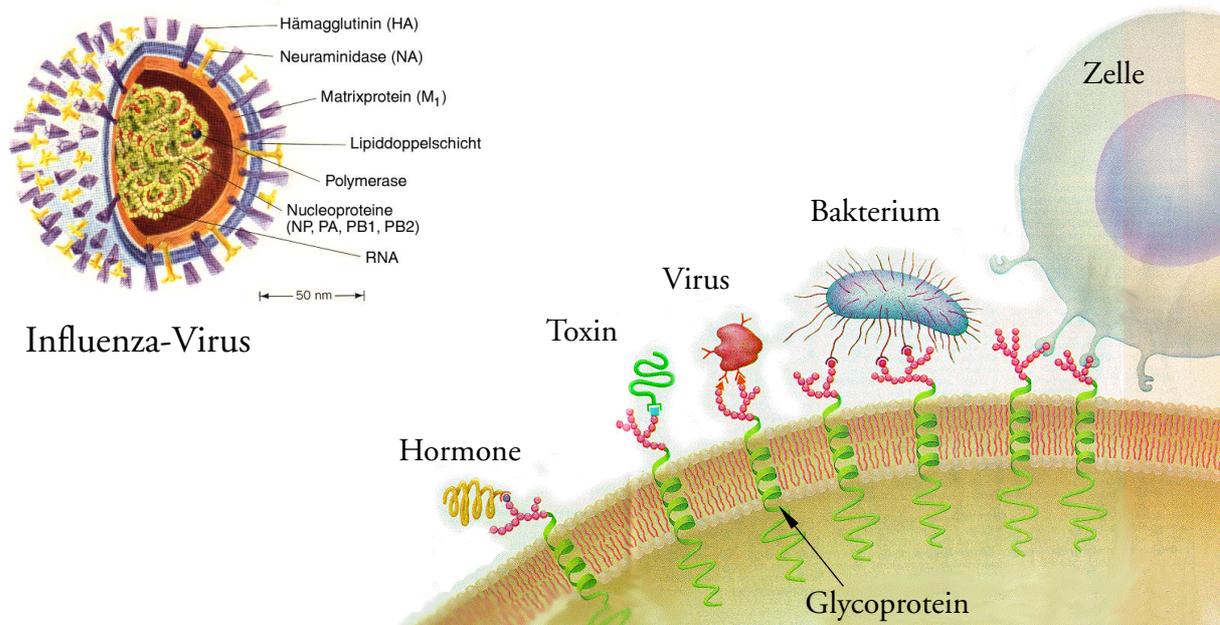
# 1 Einleitung

## 1.1 Kohlenhydrate - natürliche Schlüsseltechnologie des Informationstransfers

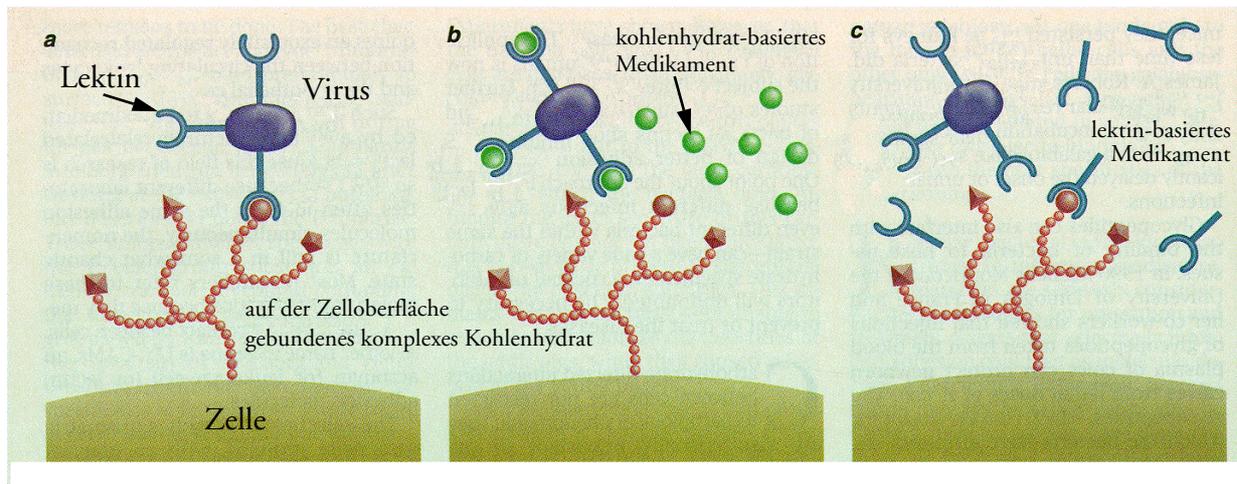
Kohlenhydrate („Zucker“) [1,2], allein oder kovalent gebunden in Glycolipiden [3] und Glycoproteinen [4] sind allgegenwärtig in der Natur. Die Möglichkeit, die Einheiten (Monosaccharide) über mehrere verschiedene Verknüpfungspositionen (vgl. Abbildung rechts), durch Ausbildung von glycosidischen Bindungen, zu Ketten (mit und ohne Verzweigung) zu verknüpfen, ermöglicht eine Informationskodierung auf kleinstem Raum, welche die von anderen Biomolekülen wie Aminosäuren oder Nukleinsäuren bei weitem übersteigt. So läßt sich z.B. aus drei Hexosen bereits die unvorstellbar große Anzahl von 38000 isomeren Permutationen generieren. Vergleicht man dies mit den 27 Permutationen, die durch Kombination von drei Aminosäuren oder Nukleinsäuren erzeugt werden können, so wird das Potential von Kohlenhydraten als Träger biologischer Information offensichtlich [5].



Das Potential der strukturellen Vielfalt auf kleinstem Raum wird von der Natur benutzt, um Informationen zwischen Zellen zu transportieren. Im Wechselspiel mit einem Lektin [7-9], einem Antikörper [10,11] oder einem Enzym [12] wird die auf dem Glycan gespeicherte Information dekodiert und führt zu einer zellulären Reaktion wie Zell-Adhäsion oder Wachstumsregulation [6,13-18].



**Abb. 1:** Beispiele für Protein-Kohlenhydrat-Wechselwirkungen [1,6].

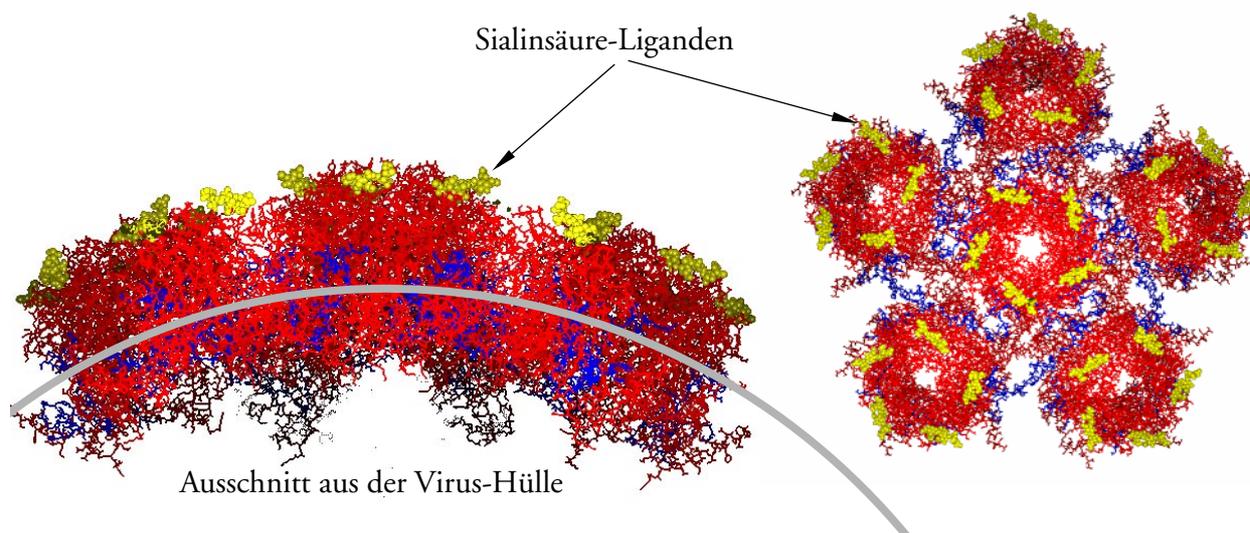


**Abb. 2:** Strategien zur Bekämpfung von Virusinfektionen [6].

Desweiteren wird die spezifische Wechselwirkung einer bestimmten Oligosaccharid-Struktur mit einem Lektin auch von Viren oder Bakterien benutzt, um ihre Wirtszelle zu finden und zu infizieren (vgl. Abb. 1). Der jährliche Ausbruch einer mehr oder weniger großen Grippe-Epidemie ist sicher eines der populärsten Beispiele für die, im negativen Sinne, Allgegenwärtigkeit von Protein-Kohlenhydrat-Interaktionen. Die Erkrankung findet ihren Anfang im Andocken des Influenza-A-Virus mit seinem Oberflächenlektin Hämagglutinin (vgl. Abb. 1) an ein sialinsäurehaltiges Glycoprotein auf der Oberfläche bronchialer Epithelzellen [19,20]. Der naheliegendste Ansatz zur Entwicklung eines „Anti-Grippe-Mittels“, liegt in der Blockierung der Bindungstasche des Hämagglutinins durch ein Kohlenhydrat, welches eine höhere Bindungsaffinität besitzt, als die auf der Wirtszelle gebundene Sialinsäure (N-Acetylneuraminsäure) (vgl. Abb. 2). Wegen der Spezifität der Bindung ist es dazu jedoch notwendig, die wesentlichen Wechselwirkungen auf atomarer Ebene zu identifizieren und zu optimieren. Da von Hämagglutinin eine Kristallstruktur mit komplex gebundener Sialinsäure vorliegt [21] sind die Voraussetzungen für ein „rationales Drug Design“<sup>1</sup> hochaffiner „Hämagglutinin-Blocker“ sehr günstig. Daher wird dieser Ansatz auch von verschiedenen Forschergruppen verfolgt, leider bis jetzt ohne Erfolg [20,23-29]. Mittlerweile ist es jedoch, u.a. durch Einsatz von Computer-Methoden [12,30], gelungen, einen Wirkstoff auf Kohlenhydratbasis zu entwickeln, der selektiv die Bindungstasche der Neuraminidase, ein ebenfalls sialinsäurebindendes Enzym auf der Virusoberfläche, blockiert und damit eine Verbreitung der Viren einschränkt [31]. Ausgangsbasis für das rationale Drug Design war auch hier die Aufklärung der dreidimensionalen Struktur der Neuraminidase und die Entschlüsselung der Wechselwirkung durch Molecular Modelling Methoden [32].

Auch im Bereich der Krebsentstehung spielen die Wechselwirkungen zwischen Kohlenhydraten und Proteinen eine wichtige Rolle (vgl. Abb. 3), daher ist die Fragestellung, wie Oligosaccharide von Lektinen, Antikörpern oder Enzymen erkannt werden, Gegenstand aktueller Forschung [9,13,34-42]. Wie schon beim Influenza-A-Virus beschrieben, kann auch hier die Kenntnis der dreidimensionalen Struktur des Protein-Kohlenhydrat-Komplexes und die Entschlüsselung des Informationstransfers durch Molecular Modelling Studien als Grundlage für die Entwicklung neuer kohlenhydratbasierter Therapeutika im Rahmen eines rationalen Drug Designs dienen [6,29,31,43-48](vgl. Abb. 2).

1. Zielgerichteter Entwurf von Wirkstoffen [22].



**Abb. 3:** Sialinsäurehaltiger Zucker (Neu5NAcα(2-3)Galβ(1-3)[Neu5NAcα(2-6)]GalNAc) komplexiert an Murine Polyoma-Virus VP1 [33]. Dieser Virus verursacht Tumore bei Mäusen.

## 1.2 Auf dem Weg zur Entwicklung hochaffiner Therapeutika

### 1.2.1 Experimentelle Methoden zur Untersuchung von Protein-Kohlenhydrat-Komplexen

Detaillierte Informationen über die dreidimensionale Struktur von Protein-Kohlenhydrat-Komplexen erhält man in erster Linie aus Kristallstrukturanalysen [21,33,49,50]. Die Zahl der gelösten Proteinstrukturen nahm in den letzten Jahren stark zu [51]. Für Oligosaccharide, kokristallisiert mit einem Proteinrezeptor, ist oft nur eine teilweise hochauflösende Bestimmung der Struktur möglich, da wegen der Flexibilität der glycosidischen Bindungen eine lokale Unordnung im Kristall vorliegt und so Teile des Oligosaccharids nicht sicher lokalisiert werden können [50,52].

In den letzten Jahren hat die NMR-Spektroskopie zunehmend wichtige Beiträge zum Verständnis der Komplexbildung geleistet [53-57]. Die direkte strukturelle Untersuchung der zuckerbindenden Proteine mit NMR-Methoden ist wegen ihrer großen molekularen Masse zwar oft nicht möglich [54], durch Transfer-NOE-Messungen [58-61] lassen sich jedoch Informationen über die Konformation des Liganden im gebundenen Zustand erhalten.

Die wichtigsten experimentellen Methoden zur Strukturaufklärung von Oligosacchariden in Lösung sind NMR (Nuclear Magnetic Resonance), FET (Fluorescence Energie Transfer), OR (Optical Rotation) und ROA (Raman Optical Activity) [50,62]. Durch  $^1\text{H}$ -NMR-Messungen in Wasser als Lösungsmittel erhält man im wesentlichen Informationen über räumliche Abstände von CH-Protonen. Die OH-Protonen des Oligosaccharids tauschen unter diesen (physiologischen) Bedingungen zu schnell mit dem Wasser aus und sind daher in einem in wässriger Lösung aufgenommenen  $^1\text{H}$ -NMR-Spektrum nicht sichtbar. Dadurch gehen jedoch wichtige Strukturinformationen verloren und man ist daher bemüht, dieses Problem zu umgehen: Da die Geschwindigkeit des Protonenaustauschs von der Temperatur abhängig ist, besteht eine Methode in einer Temperaturerniedrigung unter  $0\text{ }^\circ\text{C}$  nach Beimischung eines gefrierpunktserniedrigenden Lösungsmittels [63] oder durch Verwendung unterkühlter wässriger Lösungen [64]. Eine weitere Methode um Strukturinformationen basierend auf transfer-NOEs von OH-Protonen zu erhalten, ist die Verwendung von aprotischen Lösungsmitteln

wie Dimethylsulfoxid (DMSO) [65,66]. Es hat sich gezeigt, daß sowohl die Struktur der Oligosaccharide als auch die Aktivität der Proteine in DMSO zum großen Teil erhalten bleiben [67]. Daher besitzen NMR-Messungen in DMSO ein großes Potential zur Strukturaufklärung von komplexen Oligosacchariden oder gar von Protein-Kohlenhydrat-Komplexen [68,69].

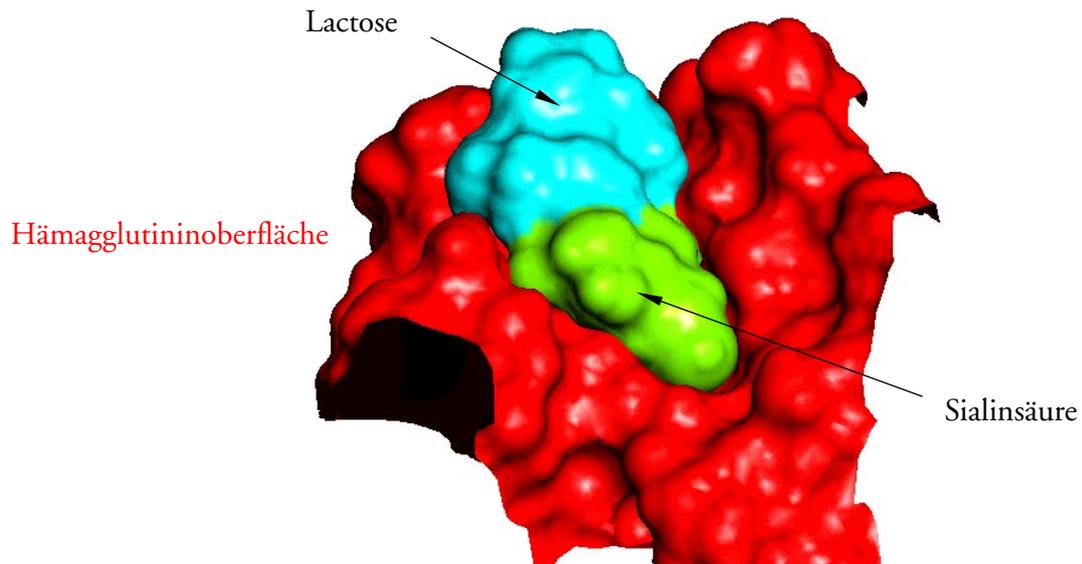
Mit Hilfe von NMR-Titrationsen ist es weiterhin möglich, Informationen über die Bindungsspezifität und -affinität zu erhalten, durch NOESY, ROESY und CIDNP Experimente lassen sich Rückschlüsse auf die an der Bindung beteiligten Aminosäuren ziehen [70-74]. Im günstigsten Fall ist sogar eine komplette Strukturaufklärung eines Protein-Kohlenhydrat-Komplexes durch NMR-Methoden möglich [75,76].

### 1.2.2 Welche Faktoren beeinflussen die Komplexbildung?

Kombiniert man die aus den Kristallstrukturen erhaltenen Daten mit den mittels Titrations-Mikrokalorimetrie [77-79] oder NMR-Methoden [75,80] bestimmten thermodynamischen Daten der Komplex-Bindung, so läßt sich daraus auf die wesentlichen Faktoren schließen, die eine Protein-Kohlenhydrat-Wechselwirkung bestimmen [81].

Es hat sich gezeigt, daß die treibende Kraft bei der Bindung eines Kohlenhydrats an eine Proteinoberfläche die Ausbildung von van-der-Waals-Kontakten (vdW-Wechselwirkungen) zu Aminosäuren mit aromatischen und aliphatischen Seitenketten ist. Kommt es darüber hinausgehend noch zur Ausbildung von Wasserstoffbrückenbindungen (H-Brücken) [82], so sind die Voraussetzungen für eine starke Komplexbindung aus struktureller (enthalpischer) Sicht günstig [54,83]. Die Rolle des Wassers bei der Komplexbildung wird nach wie vor kontrovers diskutiert [84-86], ebenso wie auch der Einfluß der Entropie [86-88]. Man kann jedoch davon ausgehen, daß im Bezug auf das Wasser zwei Einflüsse wichtig sind: Desolvatisierung von Protein und Kohlenhydrat und die Fähigkeit der Wassermoleküle sich aufgrund ihrer geringen Größe in Lücken zwischen der Protein- und Zuckeroberfläche zu setzen und so eventuell zusätzliche H-Brücken auszubilden [81,89]. Es konnte bewiesen werden, daß die Desolvatisierung im wesentlichen enthalpisch begünstigt ist, da Wassermoleküle im freien Wasser („bulk water“) eine energetisch günstigere Anordnung einnehmen können als auf einer Protein- oder Zuckeroberfläche [90,91]. Die Reduzierung der Protein- und Zuckeroberfläche durch die Komplexbildung ist also ein Faktor, der mit einem starken Energiegewinn verbunden ist und daher wesentlich zu Stabilisierung eines Komplexes beiträgt. Die Reduzierung der Oberfläche sollte um so größer sein, je komplementärer die Topologien von Ligand und Bindungstasche sind. Dies wurde bereits von Emil Fischer gegen Ende des 19. Jahrhunderts entdeckt und als „Schlüssel-Schloß-Prinzip“ (1897) bezeichnet (vgl. Abb. 4).

Ein weiterer wichtiger Faktor, der zur Erhöhung der Stärke der Komplexbindung in biologischen Systemen beiträgt, ist die Existenz polyvalenter Wechselwirkungen [92]. Darunter versteht man die gleichzeitige Bindung mehrerer Liganden einer biologischen Einheit (eines Moleküls, einer Oberfläche) an mehrere Rezeptoren einer zweiten biologischen Einheit. Durch Ausbildung von multivalenten Wechselwirkungen kann es zu einer erheblichen Steigerung der Bindungsaffinität kommen, deren wichtigste Ursache in der entropischen Stabilisierung des Komplexes liegt [93]. Multivalente Wechselwirkungen spielen z.B. auch bei der Bindung eines Virus an die Zelloberfläche eine zentrale Rolle [94]. Es besteht daher großes Interesse sich dieses Prinzip im Drug Design durch die Entwicklung von Medikamenten mit multivalenten aktiven Zentren zunutze zu machen [95]. Mit der kürzlich gelungenen Synthese von Kohlenhydratclustern auf Inositol-Basis stehen neue vielversprechende Grundbausteine zum Design hochaffiner multivalenter Therapeutika zur Verfügung [96].



**Abb. 4:** Sialyllactose gebunden an Influenza-A-Hämagglutinin als Beispiel für Emil Fischers „Schlüssel-Schloß-Prinzip“

Nicht alle Informationen, die zum Verständnis der Komplexbildung zwischen einem Kohlenhydrat und einem Protein notwendig sind, lassen sich experimentell gewinnen, so daß, ergänzend zum Experiment, Molecular Modelling Methoden in zunehmendem Maße eingesetzt werden, um zu einem Verständnis der Komplexbindung auf atomarer Ebene zu gelangen [28,51,97-102]. Hierzu werden bevorzugt Kraftfeld-Methoden eingesetzt, da eine Behandlung des Problems mit Hilfe von semiempirischen oder gar quantenchemischen (*ab initio*) Methoden aufgrund der Größe des Systems meistens (noch) nicht möglich ist. Reviews über die zum Modelling von Kohlenhydraten eingesetzten Methoden und Kraftfelder finden sich in [97,101,103-106].

### 1.2.3 Thermodynamik des Bindungsprozesses

Die Vorhersage der (relativen) Bindungsaffinität stellt eines der wichtigsten Ziele der Anwendung von Computer-Methoden im strukturbasierten Wirkstoffdesign dar. Entsprechend umfangreich ist die Anzahl der Methoden und deren Varianten, die es ermöglichen sollen, Aussagen über die Bindungsaffinität eines bestimmten Liganden zu machen [22,102,107-116].

Voraussetzung zur Abschätzung der Bindungsaffinität eines Liganden mit Computer-Methoden ist ein detailliertes thermodynamisches Modell der Komplexbildung, welches geeignet ist, berechnete Daten mit experimentellen Daten, z.B. aus der Mikrokalorimetrie [78] oder der NMR-Spektroskopie [54] zu korrelieren.

Das Maß für die Stärke einer Komplexbindung ist die Komplexbildungskonstante  $K_a$  (Gleichung 1)<sup>1</sup>.



1. Häufig wird auch die Dissoziationskonstante  $K_i = 1/K_a$  (Einheit: [mol/l]) verwendet.

Sie gibt an, welcher Anteil des Liganden im Mittel an das Protein gebunden ist und beschreibt damit auch die Triebkraft der Komplexbildung zwischen Protein und Ligand.

Für den Zusammenhang zwischen der Komplexbildungskonstanten  $K_a$  und der Freien Enthalpie der Komplexbildung  $\Delta G$  („Bindungsenergie“) gilt Gleichung 2. Die Temperaturabhängigkeit der Komplexbildung wird durch Gleichung 3 und Gleichung 4 beschrieben.

$$\Delta G^0 = \Delta H^0 - T\Delta S^0 = -RT\ln K_a \quad (2)$$

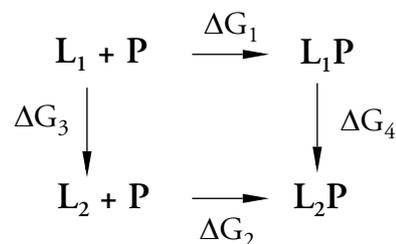
$\Delta G^0$ : Freie Standard-Bindungsenthalpie [ $\text{J mol}^{-1}$ ]  
 $\Delta H^0$ : Standard-Bindungsenthalpie [ $\text{J mol}^{-1}$ ]  
 $\Delta S^0$ : Standard-Bindungsentropie [ $\text{J mol}^{-1}\text{K}^{-1}$ ]  
 $T$ : Temperatur [K]  
 $K_a$ : Komplexbildungskonstante [ $\text{mol}^{-1}$ ]  
 $R$ : Gaskonstante ( $= 8.31441 \text{ J mol}^{-1}\text{K}^{-1}$ )

$$\ln K_a = (-\Delta H^0 / R) \cdot T^{-1} + \Delta S^0 / R \quad \text{Temperaturabhängigkeit der Komplexbildung} \quad (3)$$

$$\frac{d\ln K_a}{d(1/T)} = -\frac{\Delta H^0}{R} \quad \text{van 't Hoffsche Reaktionsisochore} \quad (4)$$

Während Komplexbildungskonstanten ausschließlich experimentell direkt zugänglich sind, lassen sich die Freie Enthalpie oder zumindest enthalpische oder entropische Einzelbeiträge zur Freien Enthalpie durch Berechnungen erhalten.

Exemplarisch für eine Methode zur Berechnung der Freien Enthalpie sei die sogenannte Free Energy Perturbation (FEP) Methode erwähnt, die auf einem thermodynamischen Zyklus beruht und im Prinzip einen Zugang zur relativen Freien Bindungsenthalpie  $\Delta\Delta G$  ermöglicht [109,111,117-119].



$$\Delta\Delta G = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3$$

Kennt man aus dem Experiment die Freie Bindungsenthalpie  $\Delta G_1$  eines Liganden  $L_1$ , so kann die Freie Bindungsenthalpie eines zweiten Liganden  $\Delta G_2$  erhalten werden, indem man  $\Delta G_3$  und  $\Delta G_4$  berechnet. Dazu überführt

(„mutiert“) man den Liganden  $L_1$  einmal in Lösung und einmal im Komplex gebunden während einer Moleküldynamik Simulation in den Liganden  $L_2$  und berechnet die dabei auftretende Änderung der Freien Enthalpie. Dies funktioniert praktisch jedoch nur, wenn die Veränderungen des Liganden nicht allzu gravierend sind. So ist es zum Beispiel problemlos möglich einen Fluor-Substituenten in einen Chlor-Substituenten oder eine Methylgruppe in eine OH-Gruppe zu mutieren. Der Austausch größerer Gruppen oder gar Konformationsübergänge lassen sich damit nicht handhaben.

### 1.2.4 Semiquantitative Vorhersage von Bindungskonstanten durch Kraftfeldmethoden

Da die direkte Berechnung der Freien Bindungsenthalpie nur in speziellen Fällen möglich ist, behilft man sich, indem man die Freie Enthalpie in Einzelbeiträge aufsplittet und versucht diese unter Zuhilfenahme verschiedener Methoden zu berechnen oder durch Inkrementenregeln abzuschätzen [108,120-123].

Eine Möglichkeit der Aufteilung der Freien Bindungsenthalpie besteht darin, sie in für die Komplexbildung vorteilhafte Anteile, wie polare Wechselwirkungen, van-der-Waals-Wechselwirkungen und hydrophobe Effekte, und für die Komplexbildung ungünstige Terme wie Entropieverluste und Konformationsenergie aufzuteilen (Gleichung 5) [88,120,122]. Der „hydrophobe Effekt“ wird in diesem Zusammenhang mit der Entfernung von Kohlenwasserstoffoberflächen aus dem Solvens durch Komplexbildung gleichgesetzt. Die Energiestabilisierung wird mit 0.125 kJ/mol für jedes Å<sup>2</sup> aus dem Solvens entfernter hydrophober Oberfläche bewertet.

$$\Delta G = \Delta G_{trans+rot} + \Delta G_{rotors} + \Delta G_{conform} + \sum \Delta G_i + \sum \Delta G_{vdW} + \sum \Delta G_H \quad (5)$$

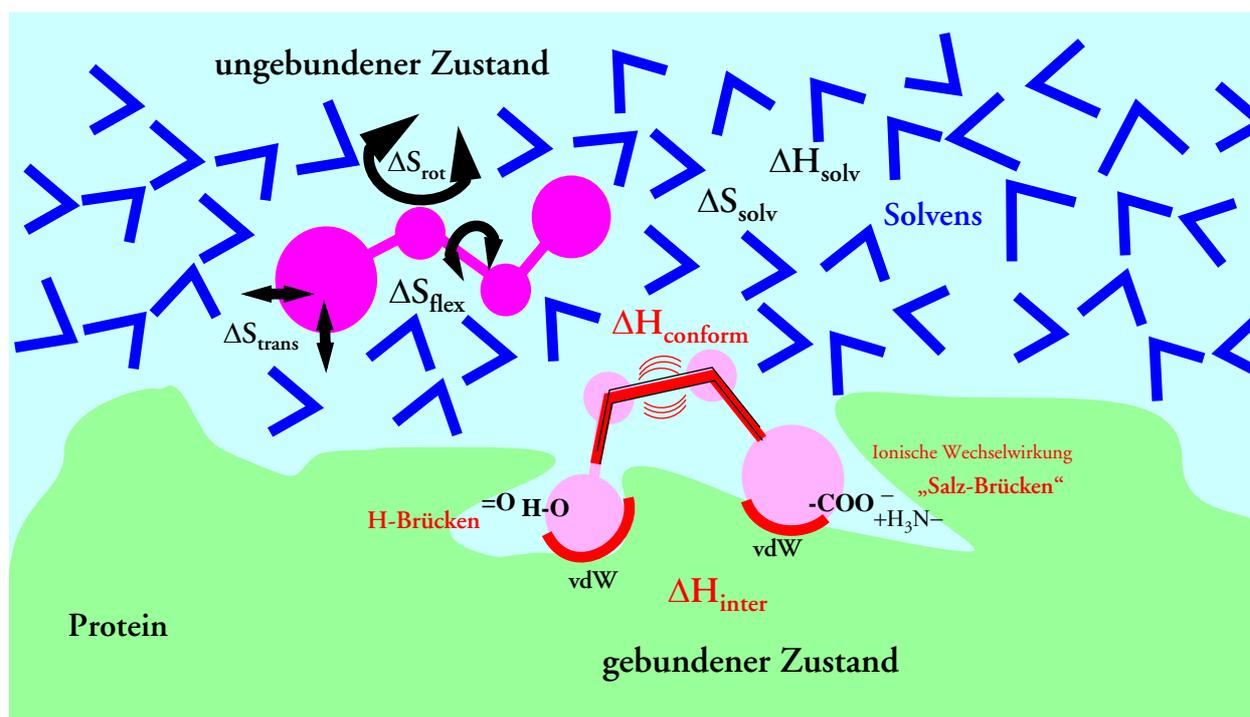
$\Delta G_{trans+rot}$ :	Verlust an Translations- und Rotationsentropie
$\Delta G_{rotors}$ :	Entropieverlust durch Einschränkung der intramolekularen Flexibilität im gebundenen Zustand
$\Delta G_{conform}$ :	Konformationsenergie
$\Delta G_i$ :	polare Wechselwirkungen
$\Delta G_{vdW}$ :	Van-der-Waals-Wechselwirkungen
$\Delta G_H$ :	hydrophober Effekt

Eine ähnliche Aufteilung der Freien Enthalpie stellt Gleichung 6 dar [28]. Im Gegensatz zu Gleichung 5 sind hier die Solvenseffekte explizit ausgewiesen (vgl. Abb. 5). Da Solvenseffekte sehr schwer mit Kraftfeld-Methoden zu simulieren sind, kann man mit diesem Ansatz diese zuerst einmal von der expliziten Berechnung ausklammern und durch empirische Regeln abschätzen.

$$\Delta G = \Delta H_{inter} + \Delta H_{conform} + \Delta H_{solv} - T\Delta S_{trans} - T\Delta S_{rot} - T\Delta S_{flex} - T\Delta S_{solv} \quad (6)$$

$\Delta H_{inter}$ :	Wechselwirkungsenergie zwischen Protein und Ligand (Van-der-Waals, Coulomb)
$\Delta H_{conform}$ :	Konformationsenergie
$\Delta H_{solv}$ :	Solvatisierungsenergie
$\Delta S_{trans}, \Delta S_{rot}$ :	Verlust an Translations- und Rotationsentropie
$\Delta S_{flex}$ :	Entropieverlust durch Einschränkung der intramolekularen Flexibilität im gebundenen Zustand
$\Delta S_{solv}$ :	Entropieänderungen im Solvens

Die Wechselwirkungsenergie  $\Delta H_{inter}$  läßt sich direkt mit Hilfe von Kraftfeldmethoden berechnen, da Terme zur Beschreibung der van-der-Waals- und Coulomb-Wechselwirkungen in jedem Kraftfeld explizit vorhanden sind [28,110,124-126]. Einziges Problem hierbei ist, daß minimale Änderungen der Position von Liganden- oder Proteinatomen große Änderungen der Wechselwirkungsenergie her-



**Abb. 5:** Schematische Illustration der Einzelbeiträge zur Freien Bindungsenthalpie

vorrufen können. Es ist daher günstiger die Wechselwirkungsenergie nicht anhand einer einzigen Struktur zu berechnen, sondern über eine Ensemble ähnlicher Strukturen, z.B. aus einer Moleküldynamik-Simulation, zu mitteln.

Theoretische Voraussetzung zur Berechnung von  $\Delta H_{\text{conform}}$  und  $\Delta S_{\text{flex}}$  ist eine vollständige Konformationsanalyse des Liganden und des Proteins. Ziel ist es dabei alle möglichen Konformationen und deren relative Energie zu ermitteln. In der Praxis beschränkt man sich meist auf eine Konformationsanalyse des Liganden, da eine vollständige Konformationsanalyse des Proteins wegen seiner vielen Freiheitsgrade praktisch nicht durchführbar ist.

$\Delta H_{\text{conform}}$  berechnet sich für den Liganden aus der Energiedifferenz zwischen der Energie der im Komplex gebundenen Konformation und der Energie des globalen Minimums im freien Zustand [127]. Der Entropieverlust des Liganden durch Einfrieren von Torsionsfreiheitsgraden ( $\Delta S_{\text{flex}}$ ) kann theoretisch erhalten werden, indem man die Ensemble-Entropie des Liganden im freien Zustand und im gebundenen Zustands nach Gleichung 7 berechnet [86].

$$S = -R \sum_i p_i \ln p_i \quad \text{Entropie eines Ensembles} \quad (7)$$

$$p_i = \frac{e^{-\frac{E_i}{kT}}}{Q} \quad Q = \sum_i e^{-\frac{E_i}{kT}} \quad (8)$$

$p_i$  : statistisches Gewicht der Konformation  $i$   
 $E_i$  : Energie der Konformation  $i$   
 $Q$  : molekulare Zustandssumme  
 $k$  : Boltzmannkonstante ( $= 1.38066 \cdot 10^{-23}$ )

Der Entropieverlust durch die Reduktion von Translations- und Rotationsfreiheitsgraden bei der Komplexbildung beträgt 10 kJ/mol [81].

Durch Analyse der Interaktion von Protein und Ligand mit dem Solvens im gebundenen und freien Zustand gelangt man zu einer Abschätzung der Solvatisierungsenergie  $\Delta H_{\text{solv}}$ . Deutlich schwieriger zu berechnen ist hingegen der Entropieterm  $\Delta S_{\text{solv}}$ .

### 1.3 Konformationsanalyse von Oligosacchariden

Wie im Abschnitt 1.2.4 bereits erläutert wurde, ist zur Berechnung von  $\Delta H_{\text{conform}}$  und  $\Delta S_{\text{flex}}$  eine Konformationsanalyse des Oligosaccharid-Liganden durchzuführen. In diesem Abschnitt sollen kurz die wesentlichen Methoden vorgestellt und ihre Grenzen aufgezeigt werden. Reviews zu diesem Thema finden sich in [97,128]. Die Grundlagen der zur Konformationsanalyse verwendeten Kraftfeld-Methoden sind ausführlich beschrieben in [129-131].

Jede Monosaccharid-Einheit besitzt normalerweise sechs frei drehbare Torsionswinkel. Geht man von drei stabilen Einstellungen für jeden Torsionswinkel aus ( $-60^\circ$ ,  $+60^\circ$  und  $180^\circ$ ), so ergeben sich theoretisch für ein Monosaccharid bereits  $3^6 = 729$  stabile Konformationen. Für ein Disaccharid sind bereits 531441 und für ein Trisaccharid die unvorstellbar große Anzahl von  $4 \cdot 10^8$  stabile Konformationen denkbar. Diese Zahlen machen deutlich, daß eine vollständige Konformationsanalyse durch systematisches Absuchen des Konformationsraumes für alle exocyclischen Torsionswinkel bereits beim Trisaccharid technisch nicht mehr durchführbar ist.

Die wichtigsten, die Topologie eines Oligosaccharids bestimmenden, Torsionswinkel sind die glycosidischen Winkel  $\Phi$  und  $\Psi$ , für (1-6)-Verknüpfungen noch zusätzlich  $\omega$  [103] (vgl. Abb. 6). Daher beschränkt man sich bei der systematischen Konformationsanalyse von Oligosacchariden meistens auf diese drei Torsionswinkel.

Zum Aufspüren lokaler Minima im  $\Phi/\Psi$ -Koordinatensystem ( $\Phi/\Psi$ -Raum) werden traditionell meist Grid-Search-Methoden verwendet. Bei der Generierung sogenannter „Rigid-Maps“ werden die  $\Phi/\Psi$ -Winkel systematisch variiert und die Energie der erhaltenen Konformation als Funktion von  $\Phi/\Psi$  aufgetragen. Durch Energieminimierung aller Freiheitsgrade der erzeugten Strukturen außer  $\Phi$  und  $\Psi$  erhält man „Relaxed Maps“. Oft werden für jeden  $\Phi/\Psi$ -Gridpunkt noch weitere Freiheitsgrade (z.B. die Hydroxymethylgruppen) systematisch variiert und minimiert und man erhält dadurch mehrere Relaxed  $\Phi/\Psi$ -Maps. Ermittelt man nun für jeden  $\Phi/\Psi$ -Gridpunkt die Konformation mit der niedrigsten Energie, so erhält man eine sogenannte „Adiabatic Map“ [132-134]. Eine spezielle Art der Grid-Optimierung stellt die Optimierung der „Nichtgrid-Freiheitsgrade“ (z.B. OH-Gruppen) durch „random walk“ dar (RAMM) [135].

Das Problem bei Grid-Search-Methoden ist, daß die Zahl der zu generierenden Strukturen exponentiell mit der Zahl der zu rotierenden Torsionswinkel ansteigt. Um im  $\Phi/\Psi$ -Raum lokale Minima mit einer ausreichenden Auflösung zu identifizieren, müssen die  $\Phi/\Psi$ -Winkel maximal in  $20^\circ$ -Schritten inkrementiert werden. Für eine Relaxed  $\Phi/\Psi$ -Map eines Disaccharids sind daher  $18 \times 18 = 324$  Strukturen zu erzeugen und zu minimieren. Rotiert man noch die beiden Hydroxymethylgruppen in  $120^\circ$ -Schritten (gg, gt, tg; vgl. Abb. 7) um eine adiabatische Karte zu erhalten, so erhöht sich die Zahl der zu minimierenden Strukturen auf  $324 \times 3 \times 3 = 2916$ . Dies ist noch problemlos innerhalb weniger Stunden (CPU-Zeit) durchzuführen. Bei einem Trisaccharid wird dies mit  $2\,834\,352$  zu erzeugenden und zu minimierenden Strukturen schon schwierig. Bis zum Konvergenzkriterium von  $0.001$  kcal/Å sind für ein Trisaccharid im Mittel ca. 3000 Minimierungsschritte notwendig (Sialyllactose, CVFF,  $\varepsilon=4$ ), d.h. zur Berechnung der Adiabatic Map eines Trisaccharids sind ca.  $10^{10}$  Energieberechnungen durchzuführen. Eine moderne Workstation würde hierzu ca. 1 Jahr CPU-Zeit benötigen. Schon allein der benötigte Festplattenspeicher macht die Berechnung einer Adiabatic Map in der beschriebenen Weise für ein Trisaccharid praktisch nicht durchführbar.

Zur Konformationsanalyse von Oligo- oder Polysacchariden werden diese daher in Disaccharid-Einheiten zerteilt und diese einzeln analysiert [97,136]. Aus den globalen  $\Phi/\Psi$ -Minima der einzelnen Disaccharid-Einheiten wird anschließend das globale Minimum des Oligosaccharids zusammengesetzt. Der Nachteil der Disaccharid-Methode zur Analyse von Oligosacchariden ist, daß man aus den erhaltenen  $\Phi/\Psi$ -Maps der Disaccharid-Einheiten keinerlei Informationen über Wechselwirkungen nicht direkt miteinander verbundener Monosaccharid-Einheiten erhält. Aus der Kombination der globalen Minima der Disaccharid- $\Phi/\Psi$ -Maps zu einer einzigen Oligosaccharidkonformation muß sich nicht zwangsweise die energieärmste Struktur ergeben, da durch interresiduale Wechselwirkungen eventuell ganz andere  $\Phi/\Psi$ -Kombinationen stabilisiert werden können. Weiterhin ist die Konformationsanalyse von komplexeren Oligosacchariden mit Verzweigungen aufgrund der vielen sich ergebenden Disaccharid-Einheiten sehr unübersichtlich.

Eine spezielle Version der Konformationssuche stellt der CICADA<sup>1</sup>-Ansatz (Heuristic Search) dar [103,137,138]. Dabei werden ausgewählte Torsionswinkel unabhängig voneinander systematisch inkrementiert und die neu erzeugte Struktur anschließend minimiert, wobei der zuvor inkrementierte Torsionswinkel während der Minimierung festgehalten wird. Wird ein lokales Minimum gefunden, so wird dieses, sofern es vorher noch nicht gefunden wurde<sup>2</sup>, gespeichert und dient als Startstruktur für die weitere Konformationssuche. Für eine CICADA-Suche mit 5-8 Torsionswinkeln, die inkrementiert werden, sind durchschnittlich 20000 bis 30000 Strukturen zu minimieren, bis keine neue Struktur mehr gefunden wird.

Metropolis-Monte-Carlo-Methoden (MMC) [139] werden zur Konformationsanalyse von Oligosacchariden selten angewandt [97]. Bei MMC werden die  $\Phi$ ,  $\Psi$  und  $\omega$ -Winkel einer Ausgangsstruktur in einem bestimmten Inkrement-Bereich ( $\pm 20^\circ$ ) zufällig variiert. Ist die Energie der neu generierten Struktur  $E_{\text{neu}}$  kleiner oder gleich der Energie der Ausgangsstruktur  $E_{\text{alt}}$ , so wird die neu generierte Struktur als Ausgangsstruktur für den nächsten Schritt akzeptiert. Ansonsten wird aus der Energiedifferenz ( $\Delta E = E_{\text{neu}} - E_{\text{alt}}$ ) der Boltzmann-Faktor  $\exp(-\Delta E/RT)$  berechnet und mit einer Zufallszahl zwischen 0 und 1 verglichen. Ist der Boltzmann-Faktor größer als die Zufallszahl, dann wird die Struktur ebenfalls als neue Startstruktur akzeptiert. Wiederholt man diese Prozedur mehrere hunderttausend mal, so erhält man ein Ensemble, das einer Boltzmann-Verteilung der Konformationen für die ver-

1. CICADA = Channels in Conformational Space Analysed by Driver Approach

2. Zwei Strukturen werden als verschieden voneinander betrachtet, wenn sie sich in mindestens einem Torsionswinkel um  $\geq 30^\circ$  unterscheiden.

wendete Temperatur entspricht. Das für MCC-Simulationen von Kohlenhydraten meistens verwendete Programmpaket GEGOP [140-142] basiert auf einem HSEA-Kraftfeld<sup>1</sup> und liefert daher nur einen groben Ansatz zur Berechnung der Energie.

Mixed Mode Monte Carlo / Stochastic Dynamics (MC/SD) [143] wurde ebenfalls schon für Disaccharide getestet [144]. Die CPU-Zeiten betragen 28 h für 1 ns MC/SC mit GB/SA-Solventmodell (MACROMODEL [145]).

Eine weitere Methode zur Konformationsanalyse von Oligosacchariden ist die Durchführung von Moleküldynamik-Simulationen (MD-Simulationen) [146,147]. Der wesentliche Vorteil der Methode liegt darin, daß sich zeitabhängige molekulare Eigenschaften analysieren lassen und die Konformationsanalyse unter Einbeziehung von Solvensmolekülen erfolgen kann. Dies ist bei Grid-Search-Methoden nicht und bei MMC-Methoden nur bedingt möglich.<sup>2</sup>

Die momentan üblichen Simulationszeiten bzw. -perioden bei MD-Simulationen bewegen sich im Bereich von maximal einigen Nanosekunden (ns) [148,149]. Bei den in der Literatur beschriebenen MD-Simulationen von Kohlenhydraten handelt es sich meist um Simulationen, die bei einer Simulationstemperatur von 300 K durchgeführt wurden [58,70,149-155]. Der Nachteil von MD-Simulationen bei Raumtemperatur liegt darin, daß extrem lange Simulationszeiten von mehr als 100 ns [152] erforderlich sind um die Rotationsbarrieren zu überwinden, wie sie bei den  $\Phi$ ,  $\Psi$  und Hydroxymethylgruppen ( $\omega$ ) vorliegen ( $>4$  kcal/mol)<sup>3</sup>. MD-Simulationen bei 300 K sind daher zur Konformationsuche im  $\Phi$ - $\Psi$ - $\omega$ -Raum nicht geeignet. Sie lassen sich jedoch zum Studium des dynamischen Verhaltens innerhalb einzelner lokaler  $\Phi$ - $\Psi$ - $\omega$ -Minima sehr gut einsetzen [156].

Es gibt zwei Ansätze um das Problem der hohen Energiebarrieren zu umgehen: erstens man erniedrigt die Energiebarrieren durch Anlegen eines negativen Zusatzpotentials (Umbrella Sampling) [157] oder man erhöht die Simulationstemperatur. Beispiele für den erfolgreichen Einsatz von Hochtemperatur-MD-Simulationen zur Konformationsanalyse finden sich in [158-161].

## 1.4 MD Simulationen - Theorie

Die moderne Kristallstrukturanalyse liefert nur ein statisches Bild der Moleküle. Für das Verständnis vieler biologischen Funktionen ist jedoch die Kenntnis des dynamischen Verhaltens der Moleküle (Konformationsänderungen, intermolekulare Stöße) wichtige Voraussetzung. Diese Informationen lassen sich durch NMR-Methoden oder Moleküldynamik-Simulationen (MD-Simulationen) erhalten. MD-Simulationen finden daher zur Simulation von Biomolekülen eine breite Anwendung [146,162-169]. Zur Optimierung von Kristallstrukturen oder NMR-Strukturen werden sie standardmäßig in Form einer als „Restrained Molecular Dynamics“ bezeichneten Variante eingesetzt [170-172].

Theoretische Grundlage [129,131] einer MD-Simulation bildet die Newtonsche Bewegungsgleichung (Gleichung 9). Danach erfährt ein Atom  $i$  der Masse  $m_i$ , auf das zum Zeitpunkt  $t$  eine Kraft  $F_i(t)$  wirkt, eine Beschleunigung  $a_i(t)$ .

1. HSEA = Hard Sphere Exo Anomeric

2. Der Grund hierfür sind Probleme bei der Minimierung des Solvens. Durch die abrupte systematische bzw. zufällige Änderung der Atomkoordinaten des Systems wird die Solvensstruktur so stark gestört, daß sie durch eine einfache Minimierung nicht mehr regeneriert werden kann. Die erhaltene Energie des Systems hat daher keine Aussagekraft. Weiterhin kann es zu Atomüberlappungen kommen, die teilweise zu irreversiblen Artefakten bei der Minimierung führen können.

3. Die Wahrscheinlichkeit bzw. Geschwindigkeit, mit der ein Übergang über eine Energiebarriere von 4 kcal/mol bei 300K erfolgt, liegt nach Arrhenius (Gleichung 36) bei ca. 1 Übergang pro Nanosekunde ( $A = 10^{12} \text{ s}^{-1}$ ).

$$F_i(t) = m_i a_i(t) = -\frac{\partial V}{\partial r_i} = m_i \frac{\partial^2 r_i}{\partial t_i^2} \quad (9)$$

$F_i(t)$ :	Kraft auf das Atom $i$ zum Zeitpunkt $t$
$m_i$ :	Masse des Atoms $i$
$a_i(t)$ :	Beschleunigung, die das Atom $i$ zum Zeitpunkt $t$ erfährt
$r_i$ :	Koordinaten des Atoms $i$
$V$ :	Potentielle Energie

Die Abhängigkeit der potentiellen Energie  $V$  von den Koordinaten  $r$  eines Systems wird durch ein sogenanntes Kraftfeld beschrieben. Das in dieser Arbeit am häufigsten verwendete Consistent-Valence Forcefield (CVFF) [173,174] (Gleichung 10) berechnet die Valenzenergie aus Diagonal-Termen für Bindungslänge ( $b$ ), Bindungswinkel ( $\theta$ ), Torsionswinkel ( $\varphi$ ) und Out-of-Plane-Energie ( $\chi$ ). Weiterhin enthält es noch Kreuz-Terme ( $X$ )<sup>1</sup> um die Abhängigkeiten zwischen Bindungswinkel, Bindungslängen und Torsionswinkel zu berücksichtigen. Die nichtbindenden Wechselwirkungen werden in van-der-Waals-Wechselwirkungen (Lennard-Jones-Potential) und elektrostatische Wechselwirkungen (Coulomb-Term) aufgeteilt. Es gibt keinen speziellen Term für Wasserstoffbrückenbindungen (H-Brücken), da diese bereits durch die Coulomb- und vdW-Terme beschrieben werden. Alternativ zum Morse-Potential (Gleichung 10, erster Term) kann man auch ein harmonisches Potential der Form  $\frac{1}{2}k(b-b_0)^2$  zur Beschreibung der Streckschwingungen zwischen zwei Atomen wählen.

$$V = \sum_b D_b [1 - e^{-\alpha(b-b_0)}] + \sum_\theta H_\theta (\theta - \theta_0)^2 + \sum_\varphi H_\varphi [1 + s \cos(n\varphi)] + \sum_\chi H_\chi \chi^2 \quad (10)$$

$$+ X_{bb'} + X_{\theta\theta'} + X_{b\theta} + X_{\varphi\Theta\Theta'} + X_{\chi\chi'} + \sum_r D_r \left( \left( \frac{r'}{r} \right)^{12} - 2 \left( \frac{r'}{r} \right)^6 \right) + \sum \frac{q_i q_j}{\epsilon r_{ij}}$$

$D_b, D_r$ :	Tiefe des Potentialminimums für das Morse-Potential (Dissoziationsenergie) bzw. Lennard-Jones Potential
$\alpha$ :	Konstante $\alpha = (k/2D_b)^{1/2}$ [129]
$b, \theta$ :	Bindungslänge, -winkel
$\varphi$ :	Torsionswinkel
$r$ :	Abstand zweier Atome
$k, H$ :	Kraftkonstanten
$q_i$ :	(Partial-) Ladung des Atoms $i$
$\epsilon$ :	Dielektrizitätskonstante

Die Newtonschen Bewegungsgleichungen sind deterministisch, d.h. kennt man die Position der Atome und ihre Geschwindigkeit zum Zeitpunkt  $t$ , so kann man ihre Position und Geschwindigkeit zum Zeitpunkt  $(t+\Delta t)$  berechnen. Einer der am häufigsten hierzu verwendeten Algorithmen ist der Verlet Velocity Integrator (Gleichungen 11 bis 13). Ein Schlüsselparameter hierbei ist die Wahl des Integrationszeitschritts  $\Delta t$ . Um die Länge einer MD-Simulation möglichst groß, die dafür benötigte Rechenzeit jedoch kurz zu halten, ist ein möglichst großer Zeitschritt  $\Delta t$  notwendig. Andererseits verursacht ein zu großer Zeitschritt Instabilität und Ungenauigkeit bei der Integration.

1. Zur genauen Definition der Kreuzterme siehe DISCOVER 2.9.7 User Guide, Part 1 (1995) S. 3-7

Um einen Schwingungsvorgang ausreichend gut zu beschreiben, ist es notwendig 8-10 Datenpunkte innerhalb einer Schwingungsperiode zu erfassen. Die CH-Streckschwingung hat eine Frequenz von ca.  $10^{14} \text{ s}^{-1}$  und ist die schnellste Schwingung, die bei MD-Simulationen zu berücksichtigen ist. Die Größe des Zeitschritts  $\Delta t$  ist daher auf maximal  $10^{-15} \text{ s}$  (1 Femtosekunde) beschränkt. Weiterhin hängt die Wahl des angemessenen Zeitschritts vom verwendeten Integrator ab.

$$r_i(t + \Delta t) = r_i(t) + \Delta t \cdot v_i(t) + \frac{\Delta t^2 \cdot a_i(t)}{2} \quad (11)$$

$$a_i(t + \Delta t) = F_i(t + \Delta t) / m_i \quad (12)$$

$$v_i(t + \Delta t) = v_i(t) + \frac{\Delta t \cdot [a_i(t) + a_i(t + \Delta t)]}{2} \quad (13)$$

$r_i(t)$ : Atomkoordinaten von Atom  $i$  zum Zeitpunkt  $t$   
 $v_i(t)$ : Geschwindigkeit von Atom  $i$  zum Zeitpunkt  $t$   
 $a_i(t)$ : Beschleunigung von Atom  $i$  zum Zeitpunkt  $t$   
 $F_i(t)$ : Kraft auf Atom  $i$  zum Zeitpunkt  $t$   
 $m_i$ : Masse von Atom  $i$

Ein zweiter wichtiger Parameter bei der Durchführung einer MD-Simulation ist die Festlegung einer geeigneten Simulationstemperatur. Die Anfangsgeschwindigkeit eines Atoms  $v_i(t_0)$  wird zufällig aus der Maxwell-Boltzmannschen Geschwindigkeitsverteilung für die eingestellte Simulationstemperatur ausgewählt. In der ersten Phase der MD-Simulation (Equilibrierungsphase) müssen die einzelnen Geschwindigkeiten der Atome noch so aufeinander abgestimmt werden (Velocity Scaling), daß die einzelnen Geschwindigkeiten aller Atome des Systems zusammen einer Gleichgewichtsverteilung entsprechen. Im Gleichgewicht ist die mittlere kinetische Energie des Systems  $3/2 NkT$ , wobei  $N$  die Zahl der Atome,  $k$  die Boltzmannkonstante und  $T$  die thermodynamische Temperatur ist.

In der, sich der Equilibrierungsphase anschließenden, Sampling-Phase (Data Collection Period) wird die Temperatur durch Kopplung des Systems mit einem Temperaturbad konstant gehalten [175].

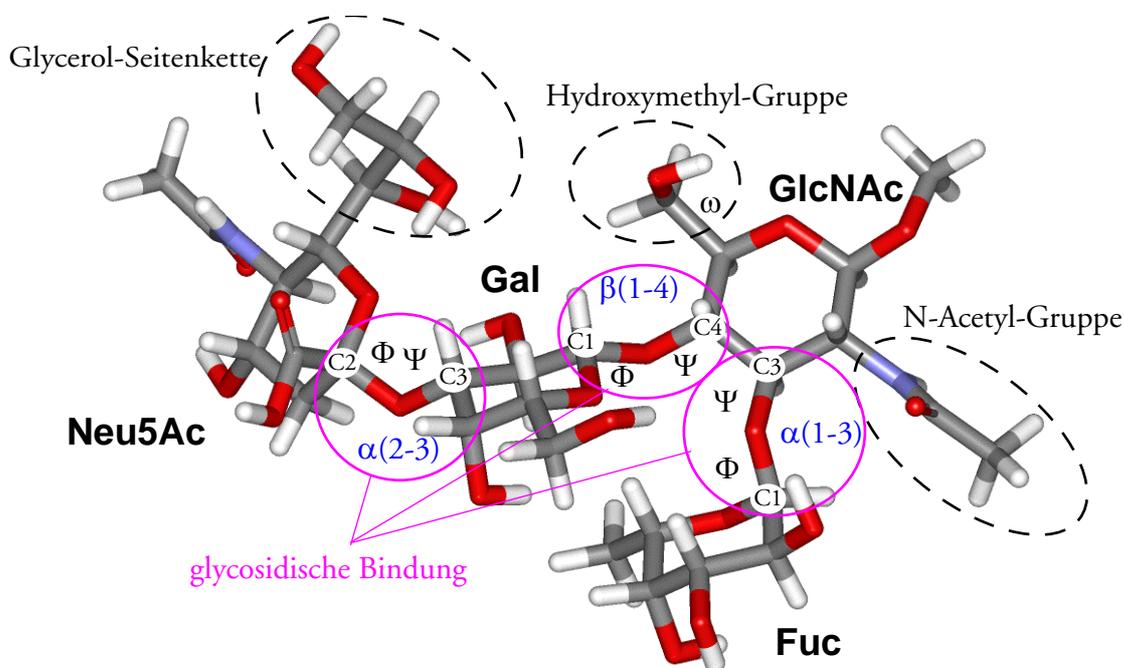
Die zeitliche Abfolge der Koordinaten und der Geschwindigkeiten der Atome während einer MD-Simulation nennt man Trajektorie.

Üblicherweise werden MD-Simulationen bei konstanter Temperatur und konstantem Volumen durchgeführt (NVT-Ensemble, kanonisches Ensemble). Bei MD-Simulationen im Vakuum, d.h. ohne Anwendung von Periodic Boundary Conditions (PBC), ist der Druck und das Volumen des Systems nicht definiert.

## 2 Allgemeine Definitionen und Grundlagen

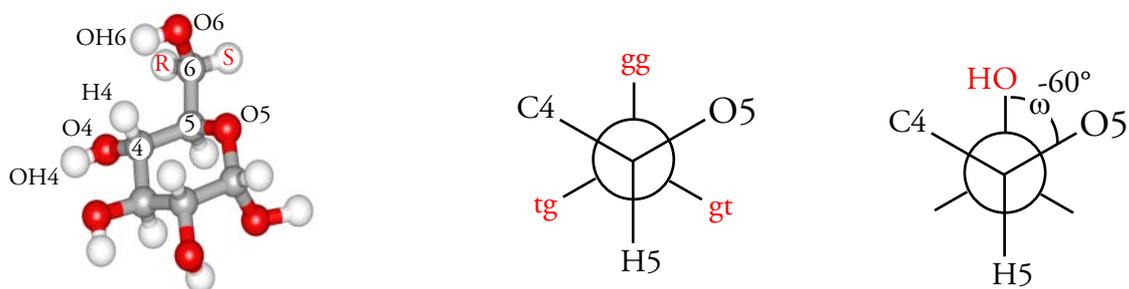
### 2.1 Wichtige Strukturparameter eines Oligosaccharids

#### 2.1.1 Übersicht



**Abb. 6:** Wichtige Strukturparameter eines Oligosaccharids am Beispiel von Sialyl-Lewis<sup>X</sup>.  
Fuc = L-Fucopyranose; Gal = D-Galactopyranose, GlcNAc = 2-Acetamido-2-Deoxy-D-Glucopyranose, Neu5Ac = N-Acetyl-D-Neuraminsäure (Sialinsäure)

#### 2.1.2 Stellung der Hydroxymethylgruppe - Definitionen



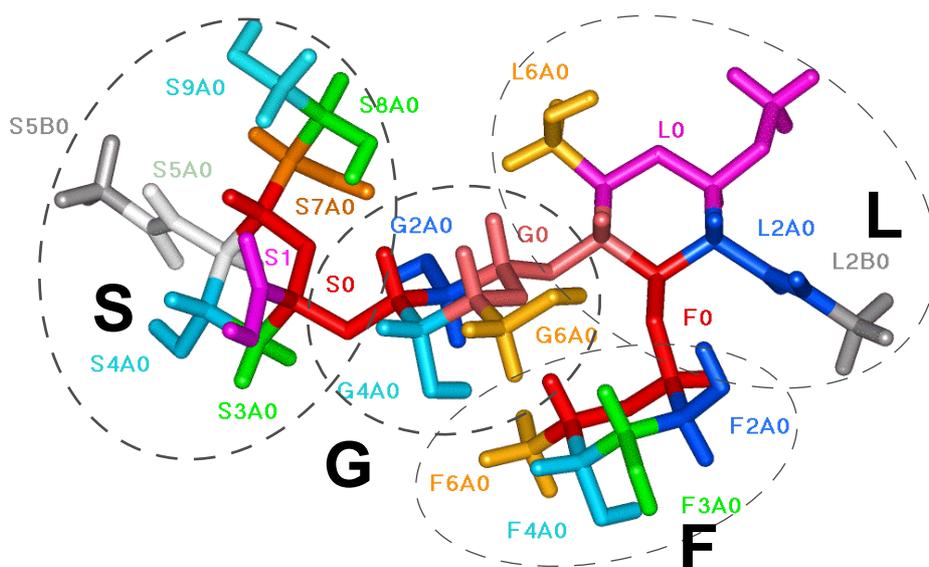
Bezeichnung der H-Atome am C6    Orientierung der OH-Gruppe    Torsionswinkel  $\omega$

**Abb. 7:** Stellung der Hydroxymethylgruppe (Blickrichtung entlang der C5-C6-Bindungsachse).

Die Stellung der Hydroxymethylgruppe wird durch die zwei Buchstaben „g“ und „t“ gekennzeichnet (g = gauche, t = trans). Dabei wird zuerst der Torsionswinkel O5-C5-C6-O6 und anschließend C4-C5-C6-O6 betrachtet. Definiert man den Torsionswinkel  $\omega$  durch O5-C5-C6-O6, so ergibt sich „tg“  $\approx 180^\circ$ , „gg“  $\approx -60^\circ$  und „gt“  $\approx +60^\circ$

### 2.1.3 Systematische Aufteilung eines Kohlenhydrats in funktionelle Untereinheiten

Zur Identifizierung bestimmter funktioneller Atomgruppierungen, vor allem im Hinblick auf die Interaktionsanalyse (vgl. Abschnitt 5.2.4), ist es notwendig, das Kohlenhydrat systematisch durch einen vierziffrigen Code in Untereinheiten (Residues) aufzuteilen. Hierzu wurde eine spezielle Nomenklatur entwickelt, die es erlaubt die Position eines Residues anhand seiner Benennung zu lokalisieren. Dazu wird den Monosacchariden zuerst willkürlich ein Einbuchstaben-Code zugewiesen (vgl. Abb. 8), der sich am Namen der Monosaccharid-Einheit orientiert (Sia  $\rightarrow$  „S“, Gal  $\rightarrow$  „G“, Glc  $\rightarrow$  „L“, Fuc  $\rightarrow$  „F“). Im zweiten Schritt erfolgt eine Einteilung des Monosaccharids in funktionelle Untereinheiten derart, daß der zentrale Baustein (Core) die Kennung „0“ erhält. Als zentraler Baustein bei Zuckern wird „H5-O5-C1-O-CX-HX“ definiert, wobei „X“ die Verknüpfungsposition der glycosidischen Bindung ist. Die Bezifferung der restlichen Residues richtet sich nach der Numerierung des Basis-C-Atoms (Bsp.: Fucose: F2A0 = H2-C2-O2-OH2). Besteht ein Substituent aus mehreren funktionellen Untereinheiten, so wird die sequentielle Abfolge durch die Buchstabenfolge A,B,C,... codiert, beginnend mit dem Residue, der das Basis-C-Atom enthält [Bsp. N-Acetyl=Amidgruppierung (L2A0 = H2-C2-NH-C=O) + Methylgruppe (L2B0 = CH<sub>3</sub>)]. Die vierte Ziffer ist für weitere Unterverzweigungen des Substituenten reserviert. Ist keine weitere Verzweigung vorhanden, so erhält der Residue als vierte Ziffer eine Null zugewiesen.



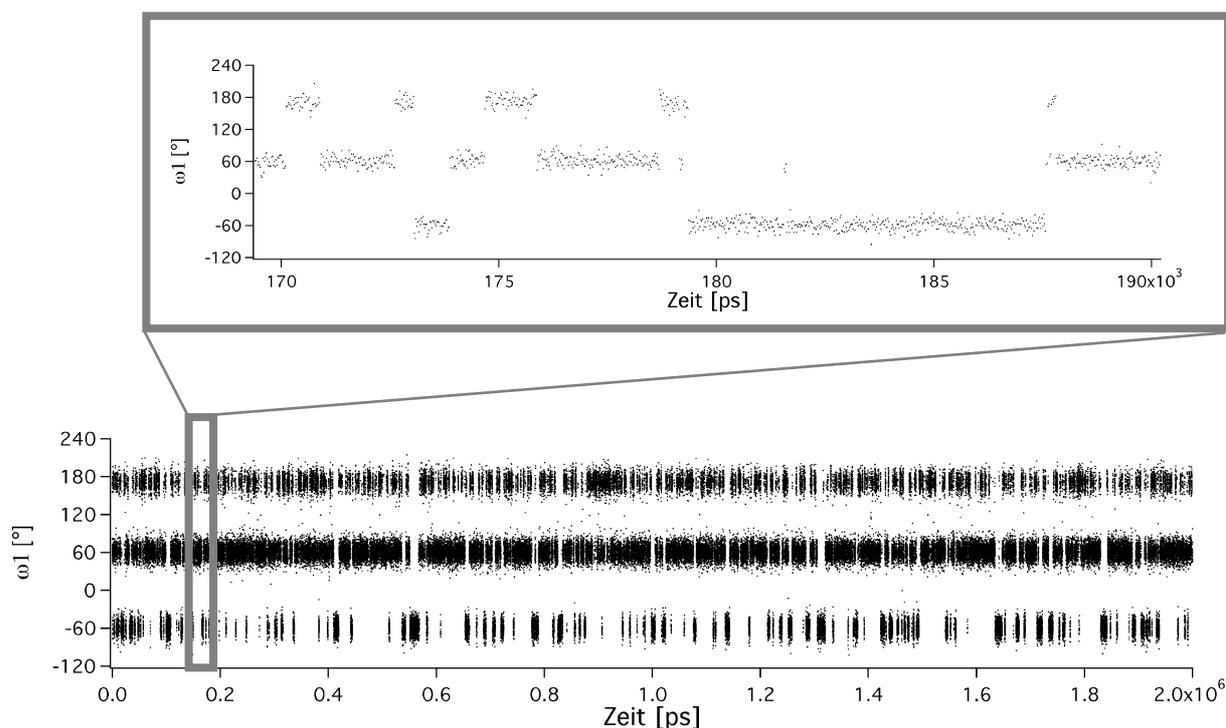
**Abb. 8:** Systematische Einteilung eines Oligosaccharids in Residues am Beispiel von Sialyl-Lewis<sup>X</sup>

## 2.2 Conformational Analysis Tools (CAT2000)

Programme zur Analyse von MD-Simulationen wie z.B. das Analysis-Modul von INSIGHT II<sup>1</sup> beschränken sich meist auf die Visualisierung der während der MD-Simulation gespeicherten Strukturen (Frames, Snapshots) sowie auf die Darstellung der Trajektorien von Systemenergien und Strukturparametern wie Atomabstände und Torsionswinkel (vgl. Abb. 9). Traditionell werden daher die Ergebnisse einer MD-Simulation wie Anzahl der auftretenden Konformationen, Lebensdauern von  $\Phi$ -,  $\Psi$ - oder  $\omega$ -Rotameren, Flexibilitäten, Lebensdauern von H-Brücken, usw. direkt anhand der entsprechenden Trajektorien diskutiert [28,149,150,152,154,176-178]. Trajektorien sind jedoch nur solange manuell (z.B. durch „Auszählen“) analysierbar, wie nur wenige Ereignisse (z.B. Konformationsübergänge) innerhalb der Simulationsperiode stattfinden.

Die in der Literatur bis jetzt publizierten Trajektorien von Oligosacchariden überspannen eine Simulationsperiode von maximal 10 ns. Innerhalb dieser Zeitperiode erfolgen z.B. nur wenige Hydroxymethyl-Rotamerenübergänge (vgl. Abb. 9, oben), so daß diese Torsionswinkel-Trajektorie noch gut überschaubar ist. Trajektorien mit nur wenigen Ereignissen sind jedoch in ihrer Aussagekraft äußerst limitiert, da keine statistische Signifikanz der Ergebnisse aufgrund der wenigen Beobachtungen zu erwarten ist. Mittlerweile sind durch den Einsatz von Hochleistungsrechnern MD-Simulationen mit Simulationsperioden bis in den Mikrosekundenbereich möglich (vgl. Abb. 9, unten). Die zugehörigen Trajektorien lassen sich nicht mehr durch einfaches Betrachten oder Auszählen analysieren, die Zuhilfenahme einer leistungsfähigen Analysesoftware ist unumgänglich.

Da zum Zeitpunkt des Beginns dieser Arbeit (1996) keine kommerziell erhältliche Analysesoftware<sup>2</sup>

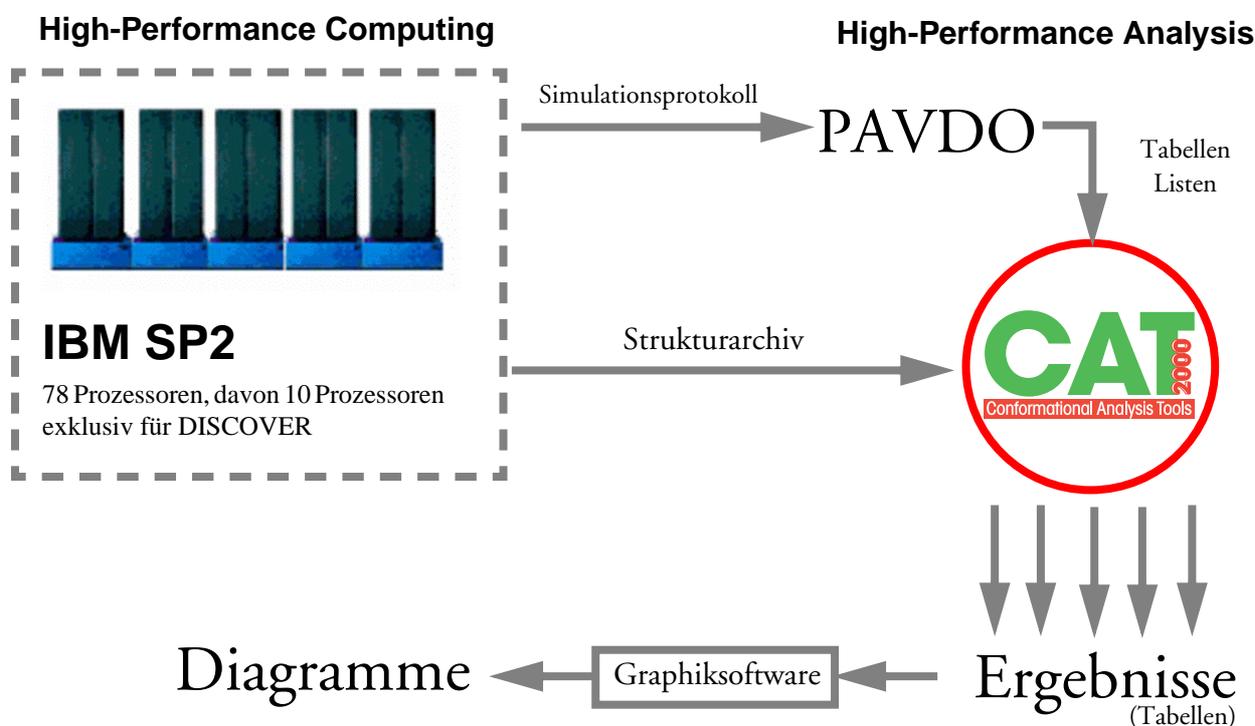


**Abb. 9:** Beispiel einer Trajektorie eines Torsionswinkels (Hydroxymethylgruppe) ( $\beta$ -Maltose, G56O,  $T = 300$  K,  $t_s = 2$   $\mu$ s, Ausschnitt: 20 ns)

1. BIOSYM/MSI, San Diego
2. Nichtkommerzielle oder sogenannte „In House Software“ konnte nicht vollständig überprüft werden, da über deren Funktionalitäten oft nichts publiziert ist.

bekannt war, die z.B. Konformationsübergänge oder Wechselwirkungsenergien basierend auf einer großen Anzahl von Konformationen ( $> 10^4$ ) analysieren konnte, bestand die erste Aufgabe darin, eine entsprechend leistungsfähige Software zur Analyse von MD-Trajektorien selbst zu programmieren. Das Ergebnis einer mehr als dreijährigen Entwicklungszeit ist das Programm CAT2000 („Conformational Analysis Tools“). Es handelt sich hierbei um eine Sammlung von mehr als 200 Funktionen zur Konformationsanalyse. Komplexe Analysesequenzen lassen sich mit Hilfe der integrierten CAT Script Language (CSL) entweder im Batchmodus (automatisch) oder über Menüs (interaktiv) durchlaufen. CAT2000 besitzt kein GUI (Graphical User Interface) und läuft momentan ausschließlich unter UNIX (IRIX, AIX, LINUX).

Viele der Funktionen von CAT2000 sind auch in anderen allgemeinen Datenanalyse-Programmen wie SPSS, SAS, MATLAB, usw. vorhanden, im Gegensatz zu CAT2000 bieten diese Programme jedoch keine Möglichkeit Struktur-Archive (xyz-Koordinaten) direkt zu analysieren. CAT2000 bietet neben einfachen Funktionen wie deskriptive Statistik (`$ANALYSE_STATISTIC`), Populationsanalyse (`$ANALYSE_POPULATION`) oder Datentransformation (`$TRANSFORM_COLUMNS`) auch komplexe Funktionen zur Gruppierung (`$CREATE_GROUPS`) bzw. zum Clustern der Daten (`$FIND_CLUSTERS`). Hinzu kommen Funktionen zur Analyse von Konformationsübergängen (`$ANALYSE_TRANSITIONS`) und zur Interaktionsanalyse (`$ANALYSE_INTERACTION`). Weiterhin sind viele praktische Funktionen, wie die Anpassung der Torsionswinkelskala (`$OPTIMIZE_TORSION_SCALE`) oder die Berechnung von relativen Werten (`$CALC_DELTA_REF`, `$CALC_DELTA_MIN`) vorhanden. Die komplette Aufzählung und Beschreibung der Funktionen würde den Umfang dieser Arbeit sprengen, daher sei auf das CAT-Referenzhandbuch verwiesen<sup>1</sup>.



**Abb. 10:** Einbindung von CAT2000 in die technische Abfolge (Workflow) einer Konformationsanalyse basierend auf großen Datenmengen (Geschätzter Durchsatz: bis zu  $10^6$  Konformationen / Tag).

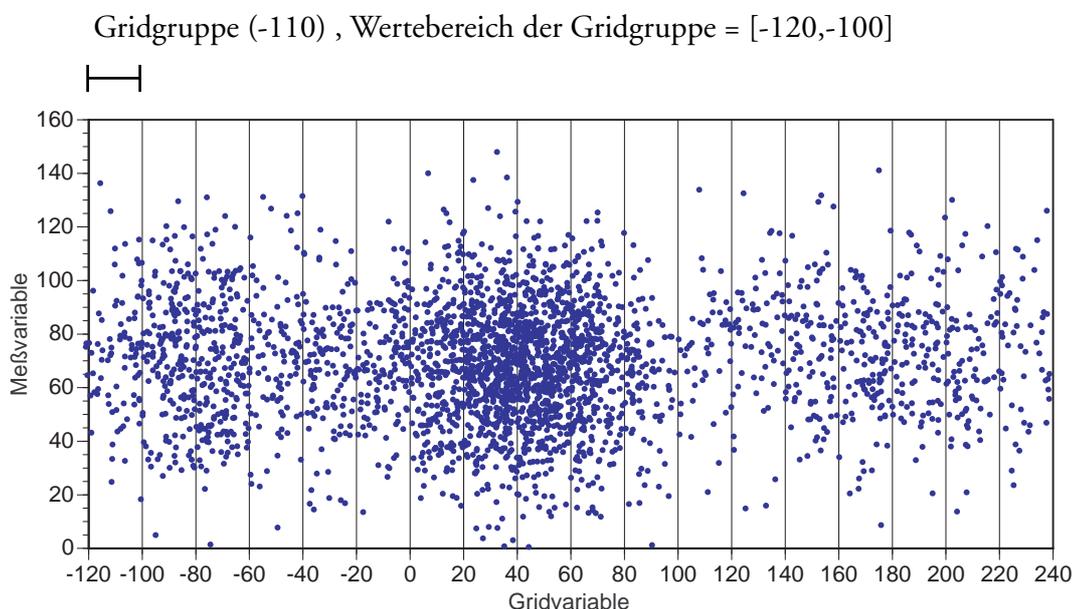
1. Aktuelle Informationen zu CAT2000 sind unter <http://www.martin-frank.de> abrufbar.

## 2.3 Gridbasierte statistische Analyse von molekularen Größen

### 2.3.1 Definition eines Grids

Unabhängig davon, welche Methode man zur Konformationssuche (Conformational Search) anwendet, das Resultat ist immer ein Satz von Strukturen (Molecular Ensemble), deren xyz-Atomkoordinaten in einem (Struktur-)Archiv gespeichert sind und/oder Tabellen (Datensätze) in denen jede Zeile eine Struktur (Beobachtung, Frame) repräsentiert und in den Spalten die einzelnen zugehörigen Variablen (Energiewerte, Torsionswinkel, Abstände, xyz-Koordinaten, usw.) gespeichert sind. Die Werte der einzelnen Variablen werden entweder direkt während der Konformationssuche berechnet oder nachträglich aus dem Struktur-Archiv.

Bei fast allen anfallenden Daten handelt es sich um stetige Daten, d.h. sie können jeden beliebigen Zahlenwert innerhalb eines gegebenen Intervalls annehmen. Zur statistischen Analyse der Daten ist es oft notwendig die Frames hinsichtlich einer oder mehrerer Variablen zu gruppieren. Dazu werden die stetigen Daten der Variablen, die gruppiert werden sollen (Gridvariablen) in ein Raster (Grid) eingeordnet (vgl. Abb. 11). Jeder Rasterbereich (Gridgruppe) wird durch eine eindeutige Bezeichnung (Grid-Gruppen-Code, GG-Code) indiziert<sup>1</sup>. Zur Bezeichnung der einzelnen Intervalle werden entweder die Intervallgrenzen angegeben (Bsp.: [15, 25]) oder, sofern wie im vorliegenden Fall, die Intervallbreite über den gesamten Wertebereich der Variablen konstant ist, der Mittelwert des Intervalls (Gridkoordinaten, 1D-GG-Code). Im Fall des eindimensionalen Grids ist die Gridkoordinate auch gleichzeitig der GG-Code.<sup>2</sup>



**Abb. 11:** Definition eines eindimensionalen, regelmäßigen Grids. Bsp. Gridvariable = Torsionswinkel, Wertebereich des Grids =  $[-120^\circ, 240^\circ]$ , Intervallbreite =  $20^\circ$ , Meßvariable = Energie. Jeder Punkt repräsentiert eine Beobachtung (Frame). Alle Punkte, die zwischen zwei senkrechten Linien liegen, gehören zu derselben Gridgruppe.

1. Daten, die eine Gruppe identifizieren zu der ein Frame gehört, nennt man nominale Daten
2. Im Programm CAT 2000 wird der GG-Code in speziellen Gruppierungsspalten eingetragen.

Ein Grid kann im Prinzip beliebig viele Dimensionen besitzen. Ein mehrdimensionales Grid setzt sich aus den eindimensionalen (1D-) Grids der einzelnen Gridvariablen zusammen. Eine Gridgruppe in einem mehrdimensionalen Grid ist durch die Gridkoordinaten (1D-GG-Codes) der einzelnen 1D-Grids der Gridvariablen definiert. Die Bezeichnung der Gridgruppen (GG-Codes) eines mehrdimensionalen Grids hat keine anschauliche Bedeutung; daher werden die Gridgruppen normalerweise sequentiell durchnummeriert (siehe Abb. 13). Ausgehend von der Gruppierung der Daten kann nun für jede Gridgruppe ein statistischer Wert (Population, Mittelwert und Standardabweichung einer Meßvariable, usw.) berechnet werden, der als Ergebnis der Analyse stellvertretend für die ganze Gridgruppe steht (Gridgruppen-Ergebniswert, Eigenschaft der Gruppe). Mit der Gruppierung der Daten in Gridgruppen werden stetige Daten in Pseudo-Zustände<sup>1</sup>, für die mittels der statistischen Analyse mittlere „molekulare Eigenschaften“ berechnet werden können, überführt.

Ein spezielles Struktur-Ensemble wird durch einen Grid-Search erzeugt. Hierbei haben die Gridvariablen von vornerein nominale Werte. Man unterscheidet:

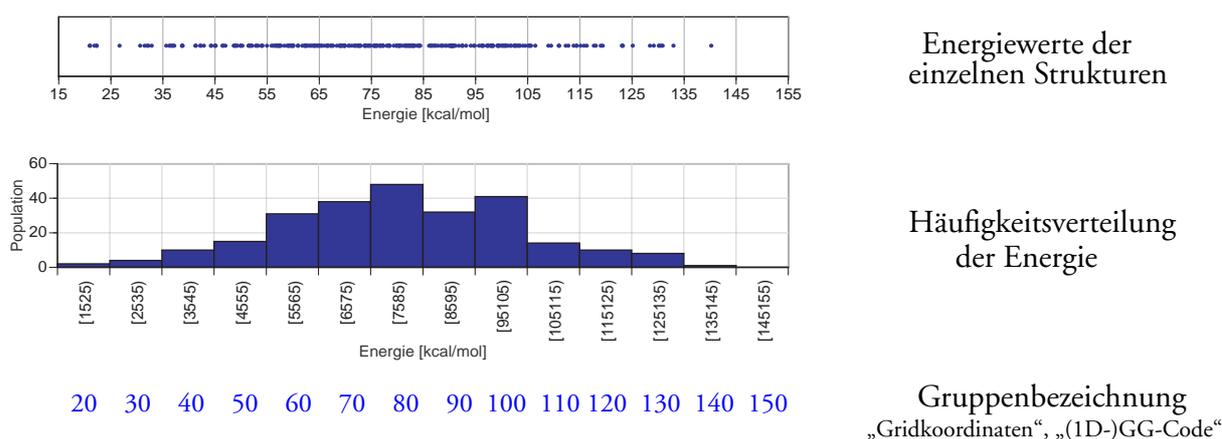
„*Rigid Grid*“: Ein Satz von Strukturen, der durch systematische Variation der Gridvariablen erzeugt wurde.

„*Relaxed Grid*“: Ein Satz von Strukturen, der durch Energieminimierung eines „Rigid Grid“ mit fixierten Werten für die Gridvariablen erhalten wurde.

„*Adiabatic Grid*“: Für jeden Gridpunkt wird nur die Struktur mit der geringsten Energie betrachtet.

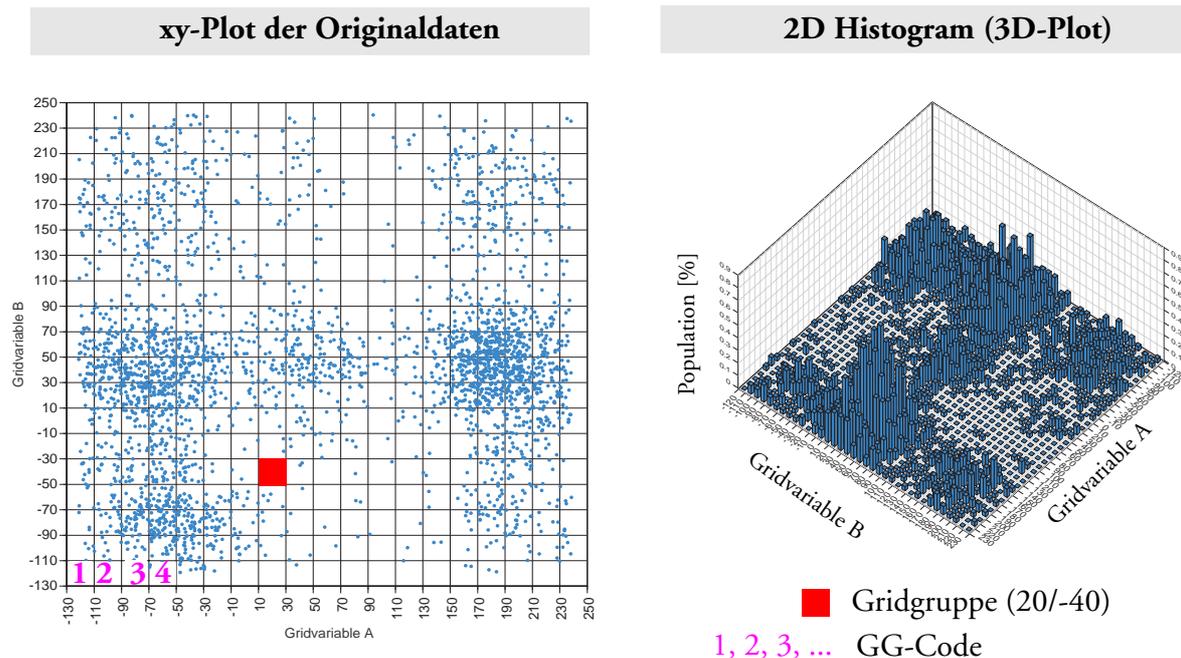
### 2.3.2 Häufigkeitsverteilungen

Eine wichtige Anwendung grid-basierter statistischer Analysen ist die Berechnung von Häufigkeitsverteilungen. Um z.B. die Energieverteilung innerhalb eines Ensembles zu analysieren, ist es notwendig, den gesamten Wertebereich (z.B. 150 kcal/mol) über den sich die Energie erstreckt in regelmäßige, kleinere Intervalle (z.B. 10 kcal/mol) aufzuteilen und die Anzahl der Strukturen zu



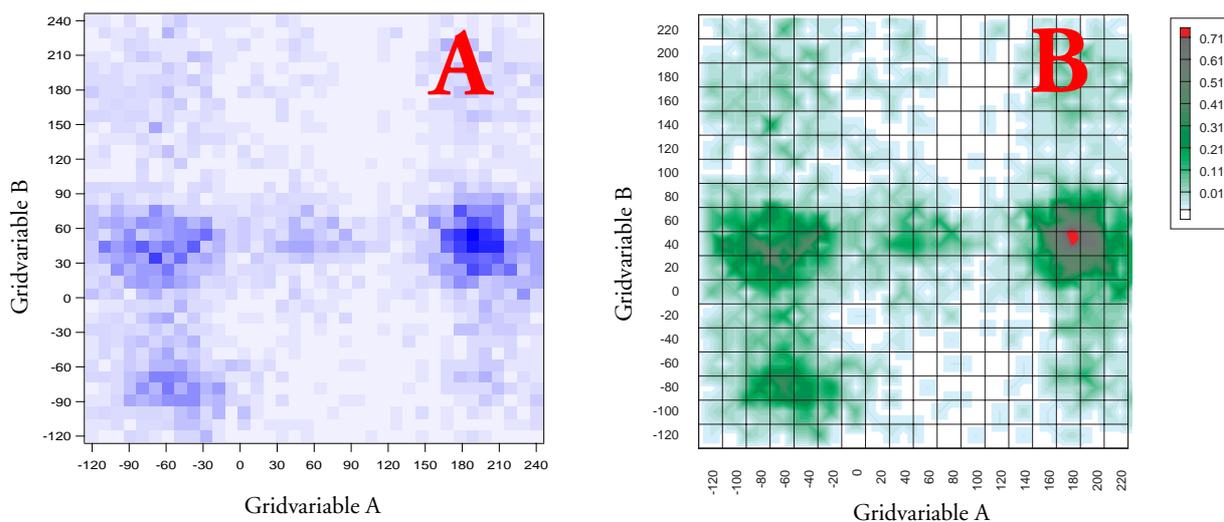
**Abb. 12:** Berechnung einer Häufigkeitsverteilung (1D Histogramm) der Systemenergie als Beispiel für die Anwendung eines eindimensionalen, regelmäßigen Grids.

1. Verwendet man interne Koordinaten als Gridvariablen, so repräsentieren die Gridgruppen Konformationen. Die Bezeichnung „Gridgruppen“ wird im Rahmen dieser Arbeit für alle Konformationen, deren Definition (Gruppierung) basierend auf Torsionswinkeln erfolgt ist, verwendet.



**Abb. 13:** Berechnung einer korrelierten Häufigkeitsverteilung (2D Histogramm) zweier Torsionswinkel als Beispiel für die Anwendung eines zweidimensionalen, regelmäßigen Grids.

bestimmen, deren Energie in den Wertebereich der einzelnen Intervalle fällt (Abb. 12). Da hier nur eine Gridvariable vorliegt, bildet ein eindimensionales, regelmäßiges Grid die Basis der Analyse. Das Ergebnis der statistischen Analyse (Gridgruppen-Ergebniswert) ist die Population eines Gridpunkts.



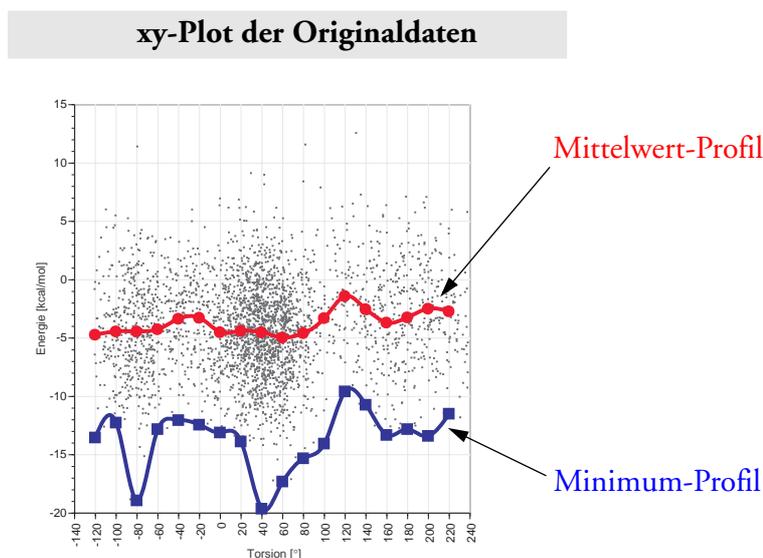
**Abb. 14:** 2D Diagramme einer korrelierten Häufigkeitsverteilung (2D Histogramm) zweier Torsionswinkel als Beispiel für die Anwendung eines zweidimensionalen, regelmäßigen Grids.  
 A: Image Plot (Igor Pro, Wavemetrics); B: gefüllter Contour-Plot (DeltaGraph Pro, SPSS)

Um z.B. zu prüfen, ob es bestimmte Kombinationen zweier Gridvariablen A und B (z.B. zweier Torsionswinkel) gibt, die besonders häufig im Ensemble vorhanden sind, werden die Daten der Gridvariablen in ein zweidimensionales Grid eingeordnet (siehe Abb. 13) und für jede (zweidimensionale) Gridgruppe die Population bestimmt.

Die graphische Darstellung gridbasierter Analysen beschränkt sich auf maximal dreidimensionale Grids (Volumenplots, Isosurface-Plots → 4D-Plots: drei Gridvariablen, ein Ergebniswert), wobei in der Praxis normalerweise nur maximal zweidimensionale Grids graphisch in Form sogenannter Contour- oder Höhenlinien-Diagramme Verwendung finden. Eine Alternative stellen sogenannte Image- oder Pixel-Plots dar bei denen jeder Gridpunkt eine der Größe des Ergebniswerts entsprechende Farbcodierung zugewiesen bekommt (siehe Abb. 14).

### 2.3.3 Energie-Grids

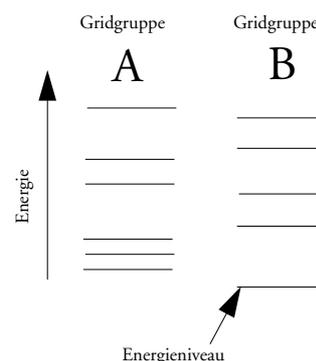
Ein weiteres wichtiges Anwendungsgebiet gridbasierter statistischer Analysen stellt die Generierung von Energieprofilen oder Energiemaps dar. Hierbei sind die Gridvariablen meist ein oder zwei Torsionswinkel. Als Gridgruppen-Ergebniswert wird dann entweder das Minimum (Minimum-Methode, Adiabatic Grid) oder der Mittelwert aller zu einer Gridgruppe gehörenden Energien berechnet. Die Verwendung des Energiemittelwertes bietet sich vor allem bei Moleküldynamik-Daten an, da hier große Fluktuationen in den Energien auftreten und das Energieminimum damit sehr anfällig gegen Ausreißer ist (siehe Abb. 15).



**Abb. 15:** Berechnung eines Energieprofils aus Moleküldynamik-Daten (300 K).

### 2.3.4 Überführung von Populations-Grids in Energie-Grids (Freie Gridgruppen-Energie)

Selbst wenn man für eine Gridgruppe die Minimumenergie reproduzierbar bestimmen kann, liegt ein entscheidender Nachteil der in Abschnitt 2.3.3 beschriebenen Minimum-Methode in der Tatsache, daß voneinander verschiedene Strukturen einer Gridgruppe, die nur eine geringfügig höhere Energie als das Gridgruppen-Minimum besitzen, bei der Analyse und damit bei der energetischen Betrachtung des Ensembles nicht berücksichtigt werden. Dies ist nur dann korrekt, wenn die Energieverteilung innerhalb der Gridgruppe für jede Gridgruppe dieselbe ist. Für den Fall, daß die Energien der verschiedenen Gridgruppen-Strukturen bei einer Gridgruppe sehr nahe beieinander und bei einer anderen sehr weit auseinander liegen (vgl. Abbildung rechts), sollte sich dieses in der Population der Gridgruppe bemerkbar machen. Sehr nahe beieinander liegende „Energieniveaus“ bedeutet, daß die Zahl der energetisch zugänglichen „Realisierungsmöglichkeiten“ und damit die Entropie (vgl. Gleichung 7) für eine Gridgruppe zunimmt, was zu einer Erniedrigung der Freien Energie für diesen Gridpunkt führen sollte.



Aus der Boltzmann-Gleichung 14 folgt, daß jeder Gridgruppe anhand ihrer relativen Population zu einer Referenz-Gridgruppe eine Energiedifferenz zugeordnet werden kann.<sup>1</sup>

$$\Delta E = E_i - E_{ref} = -kT \ln \frac{n_i}{n_{ref}} \quad \text{„Freie Gridgruppen-Energie“} \quad (14)$$

$E_i$ :	Energie der Gridgruppe i
$E_{ref}$ :	Energie der Referenz-Gridgruppe
$n_i$ :	Zahl der Strukturen, die zur Gridgruppe i gehören
$n_{ref}$ :	Zahl der Strukturen, die zur Referenz-Gridgruppe gehören
$k$ :	Boltzmann-Konstante
$T$ :	Temperatur [K]

Im Rahmen dieser Arbeit werden die Energien immer relativ zur Gridgruppe mit der größten Population (globales Minimum) berechnet [„Freie Gridgruppen-Energie“ (FGE)]. Voraussetzung der Methode ist, erstens sicherzustellen, daß sich innerhalb der Nichtgrid-Variablen ein thermisches Gleichgewicht eingestellt hat, d.h. daß sich die Gridgruppen selbst im thermischen Gleichgewicht befinden und zweitens, daß die Populationsverteilung zwischen den Gridgruppen  $n_i/n_{ref}$  einer Verteilung im thermischen Gleichgewicht entspricht.

#### Thermisches Gleichgewicht zwischen den Gridgruppen

Teilt man für ein System mit nur einem stetigen (nicht gequantelten) Freiheitsgrad (Variable) diesen in Gridgruppen ein, so wird sich im thermischen Gleichgewicht zwischen diesen Gridgruppen ein Populationsgleichgewicht einstellen, das durch den mittleren Energieunterschied der Gridgruppen bestimmt wird.<sup>2</sup> Es ist dabei unerheblich, ob der gesamte Wertebereich populiert ist, entscheidend ist lediglich, daß die zu analysierenden Gridgruppen (Zustände) miteinander im Gleichgewicht sind.

1. Die Gleichung 14 gilt nur für Systeme, die sich im thermischen Gleichgewicht befinden

2. Sofern der Energiemittelwert die Energieverteilung innerhalb der Gridgruppe realistisch repräsentiert.

Besitzt das System jedoch mehr als einen Freiheitsgrad, was praktisch bei allen Systemen der Fall ist, so ist es wichtig herauszufinden, ob zwischen der Gridvariablen und den anderen Variablen eine Abhängigkeit besteht.

Angenommen die konformative Vielfalt eines Systems läßt sich durch die Variablen A und B beschreiben (vgl. Abb. 16, links). Zur Berechnung der Häufigkeitsverteilung von A werden die stetigen Werte von A in 1D-Gridgruppen eingeteilt. Es ist offensichtlich, daß die Häufigkeitsverteilung für Variable A von der Variable B abhängig ist. Je nachdem welchen Bereich der Variablen B ( $X [120,240]$ ,  $Y[0,120]$ ,  $Z[-120,0]$ ) man betrachtet, erhält man unterschiedliche Häufigkeitsverteilungen für A (Abb. 16, rechte Seite).

### Thermisches Gleichgewicht innerhalb einer Gridgruppe

Berechnet man Häufigkeitsverteilungen der B-Variablen für einzelne A-Gridgruppen, so zeigt sich, daß die Populationsverteilung in B-Richtung für die einzelnen A-Gridgruppen (z. B. X und Y) unterschiedlich ist (Abb. 17, rechte Seite). Die Frage ist nun, ob sich für die A-Gridgruppen X und Y das

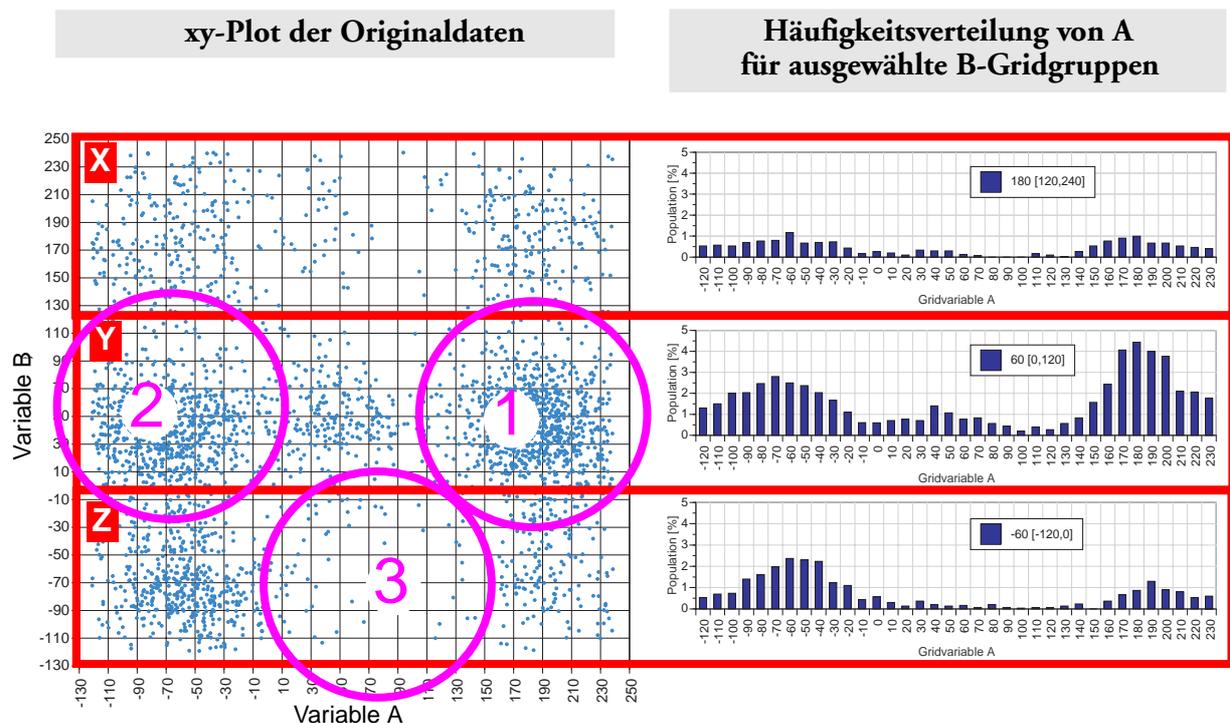
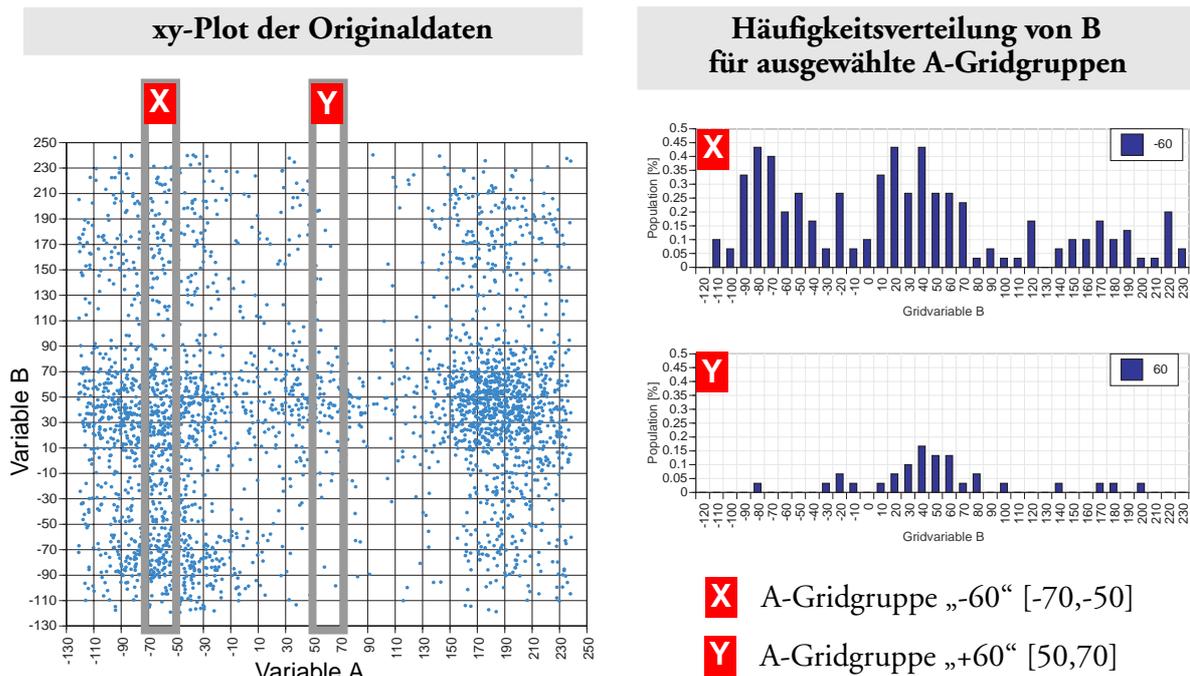


Abb. 16: Häufigkeitsverteilung von A in Abhängigkeit von B



**Abb. 17:** Häufigkeitsverteilung innerhalb der Gridgruppen von A

thermische Gleichgewicht der B-Gridgruppen eingestellt hat. Es ist offensichtlich, daß X eine deutlich höhere Gesamtpopulation aufweist als Y. In so einem Fall wäre es z.B. denkbar, daß sich für Y schon allein „aus Mangel an Daten“ keine zu X analoge Verteilung in B-Richtung eingestellt hat.

### Begrenzte thermische Gleichgewichte

Die Aussagekraft der Gesamthäufigkeitsverteilung von A ( $X+Y+Z$ ) (vgl. Abb. 16) hängt davon ab, inwieweit sich ein Populationsgleichgewicht für A *und* B (AB-Grid) eingestellt hat. Dabei sind folgende Szenarien denkbar:

1. Die Population aller AB-Gridgruppen entspricht einem thermischen Gleichgewicht („globales Grid-Gleichgewicht“). Damit entspricht die Populationsverteilung von A über den gesamten Wertebereich von A einem thermischen Gleichgewicht. Das aus der Populationsverteilung nach Gleichung 14 berechnete Energieprofil der FGE-Energie ist ohne Einschränkung gültig.
2. Die Population der AB-Gridgruppen ist nur für einen bestimmten Bereich von B und für alle zugehörigen A-Gridgruppen im thermischen Gleichgewicht (z.B. Bereich „Y“ in Abb. 16). Diesmal entspricht die Populationsverteilung von A über den gesamten Wertebereich von A einem thermischen Gleichgewicht mit der Randbedingung  $B[0, 120]$ . Das aus der Populationsverteilung nach Gleichung 14 berechnete Energieprofil der Freien Gridgruppen-Energie ist damit nur mit Einschränkung gültig (eingeschränktes Grid-Gleichgewicht).
3. Die Population der AB-Gridgruppen ist nur innerhalb bestimmter AB-Gridgruppen-Bereiche im thermischen Gleichgewicht (z.B. Bereiche „1“ und „2“ in Abb. 16; „lokales Grid-Gleichgewicht“). Aus diesen Daten läßt sich kein Energieprofil für A berechnen, das den gesamten Wertebereich von A einschließt. Es lassen sich jedoch für die einzelnen Bereiche lokale Energiekurven berechnen („lokale Minima“). Dabei sind die Energiekurven der Bereiche als unabhängig zu betrachten, solange nicht sichergestellt ist daß zwischen den Bereichen auch ein thermisches Gleichgewicht besteht.

## 2.4 Konformationsgleichgewichte

### 2.4.1 Definition eines Parameters zur Beschreibung der Einstellung von Konformationsgleichgewichten

Der Erfolg einer Moleküldynamik-Simulation hinsichtlich der statistischen Signifikanz des erzeugten Struktur-Ensembles hängt entscheidend davon ab, ob das Konformationsgleichgewicht für die zu untersuchenden Konformationsänderungen erreicht wurde. Es gilt nun einen empirischen Parameter EQ zu definieren, der die Einstellung eines Konformationsgleichgewichts anzeigt.

Ein Ensemble bestehend aus  $n$  Strukturen („Frames“) wird in  $k$  Konformationen eingeteilt, deren Konformationsgleichgewicht untersucht werden soll. Alle Strukturen, die sich keiner der Konformationen 1 bis  $k$  zuordnen lassen (z.B. Übergangszustände), werden in der „Gap-Gruppe“ zusammengefaßt (siehe Gleichung 15).

$$n = \sum_{i=1}^k n_i + n_{gap} \quad (15)$$

$$x_i = \frac{n_i}{n} \quad (16)$$

- $n$ : Zahl der Frames im Ensemble
- $k$ : Zahl der Konformationen deren Gleichgewichtseinstellung untersucht werden soll
- $n_i$ : Zahl der Frames die zur Konformation  $i$  gehört
- $n_{gap}$ : Zahl der Strukturen, die sich keiner der Konformationen 1 bis  $k$  zuordnen lassen
- $x_i$ : Molenbruch der Konformation  $i$

Als Kriterium für eine Gleichgewichtseinstellung zwischen zwei Konformationen (Zuständen, Grid-Gruppen) A und B gilt Gleichung 17, d.h. das Verhältnis der Molenbrüche muß unabhängig von der Zeit sein.

$$K_{AB}(t) = \frac{x_A(t)}{x_B(t)} = \textit{konstant} \quad (17)$$

- $K_{AB}$ : Gleichgewichtskonstante für das Konformationsgleichgewicht zwischen A und B
- $t$ : Zeit
- $x_A, x_B$ : Molenbruch der Konformation A bzw. B

Um die Einstellung des Konformationsgleichgewichts zu überprüfen, ist es also notwendig die zeitliche Konstanz aller Molenbrüche  $x_i$  zu prüfen. Da bei Moleküldynamik-Simulationen die Strukturen üblicherweise in regelmäßigen Zeitintervallen abgespeichert werden, ist der Frame-Index  $f$  mit der Simulationszeit  $t$  über Gleichung 18 verknüpft. Der Molenbruch der Konformation  $i$  zum „Zeit-

punkt“  $f$  ist gegeben durch Gleichung 19. Zur Bestimmung des Molenbruchs  $x_i(f)$  werden immer alle  $f$  Frames des Ensembles hinsichtlich der Zugehörigkeit zur Konformation  $i$  analysiert. Im Gegensatz zu „realen“ Konformationsgleichgewichten wird während der MD-Simulation die Zahl der zur Berechnung des Molenbruchs berücksichtigten Frames mit fortschreitender Simulation immer größer. Dies hat zur Folge, daß selbst wenn keine Konformationsübergänge erfolgen, sich alle Molenbrüche  $x_i$  ändern. Um die Molenbrüche  $x_i$  konstant zu halten, muß sich das Konformationsgleichgewicht laufend durch Konformationsübergänge neu einstellen.

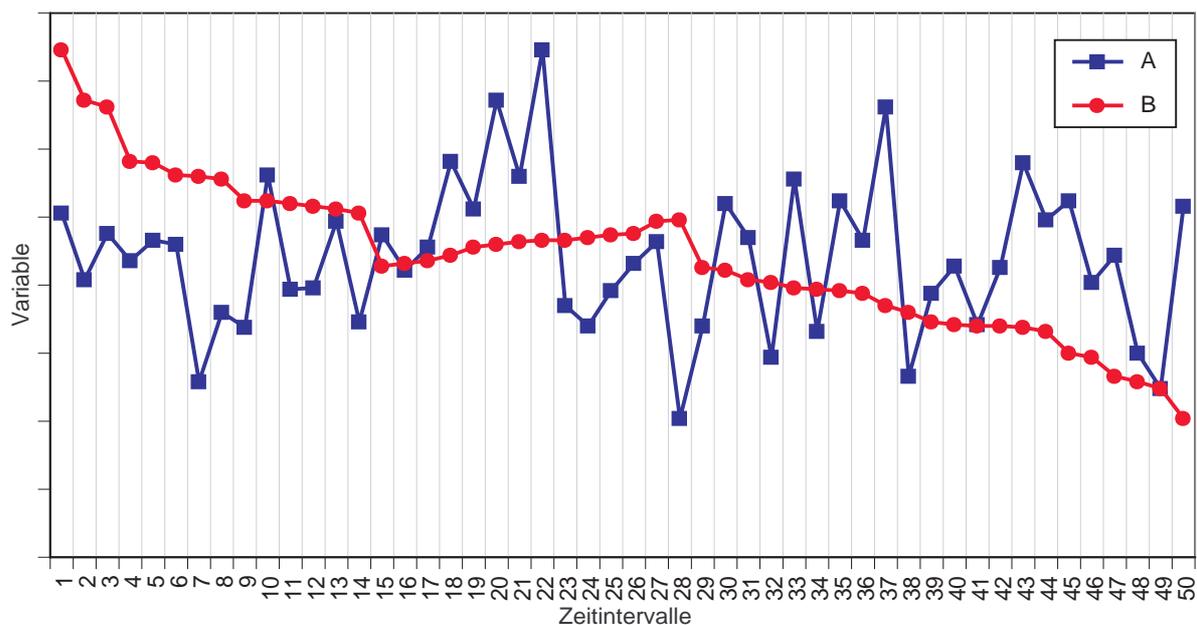
$$t = t_0 + f\Delta t_s \quad f = \frac{t-t_0}{\Delta t_s} = n(t) \quad (18)$$

$$x_i(f) = \frac{n_i(f)}{f} \quad (19)$$

- $f$ : Frame-Index
- $t$ : Simulationszeit
- $t_{Eq}$ : Equilibrierungszeit
- $\Delta t_s$ : Sampling Intervall
- $n(t)$ : Zahl der Frames im Ensemble zum Zeitpunkt  $t$
- $n_i$ : Zahl der Frames die zur Konformation  $i$  gehören
- $x_i$ : Molenbruch der Konformation  $i$

Zur Bewertung der zeitlichen Konstanz von  $x_i(f)$  sind daher drei Aspekte zu berücksichtigen: die Größe der kurzzeitigen Fluktuationen von  $x_i$ , eine evtl. vorhandene langsame Drift (vgl. Abb. 18) und die Anzahl der Konformationsübergänge im Beobachtungszeitraum  $[l_0+m, f]$ . Da mit fortschreitender Simulation die Gesamtanzahl der Strukturen zunimmt, wird der Einfluß einzelner Strukturen bzw. Übergänge immer geringer. Es ist daher zu erwarten, daß der Fluktuationsterm zunehmend an Bedeutung verliert und sich die Molenbrüche  $x_i$  nur noch langsam ändern (Drift). Eine Drift entsteht zum Beispiel auch wenn die Simulation „einfriert“ oder auch nur eine Konformation nicht mehr neu populiert wird (siehe Abb. 19). Der Molenbruch nimmt dann für diese Konformation mit  $x_i(f) = n_i/f$  ( $n_i = \text{konstant}$ ,  $f > n_i$ ) ab. Geschieht dies erst gegen Ende der Simulation, d.h.  $f$  und  $n_i$  sind sehr groß, so ist die resultierende Drift sehr klein. Da nun die Fluktuationen und die Drift sehr klein sind, würde in so einem Fall das Vorliegen eines Gleichgewichts „vorgetäuscht“. Aus diesem Grund ist zur Beurteilung der Gleichgewichtseinstellung die Zahl der Konformationsübergänge im Beobachtungszeitraum zu berücksichtigen.

Zur Berechnung der Fluktuationen werden die Molenbrüche  $x_i(f)$  innerhalb des Beobachtungszeitraumes  $[l_0+m, f]$  in regelmäßigen (Sub-)Intervallen der Breite  $m$  berechnet. Der „Nullpunkt“ des Beobachtungszeitraumes ist definiert durch  $l_0 = f - qm$ , wobei  $q$  die Anzahl der Subintervalle ist. Da die Änderung der Gleichgewichtskonstanten  $K$  von der relativen Änderung der Molenbrüche abhängt (vgl. Abb. 20), werden die Änderungen von  $x_i$  im Fluktuationsterm  $EQ^F$  relativ zum Molenbruch des jeweils vorhergehenden Subintervalls  $x_i(l-m)$  berechnet ( $\rightarrow EQ^F_i$ ). Alternativ wäre als „Referenz-

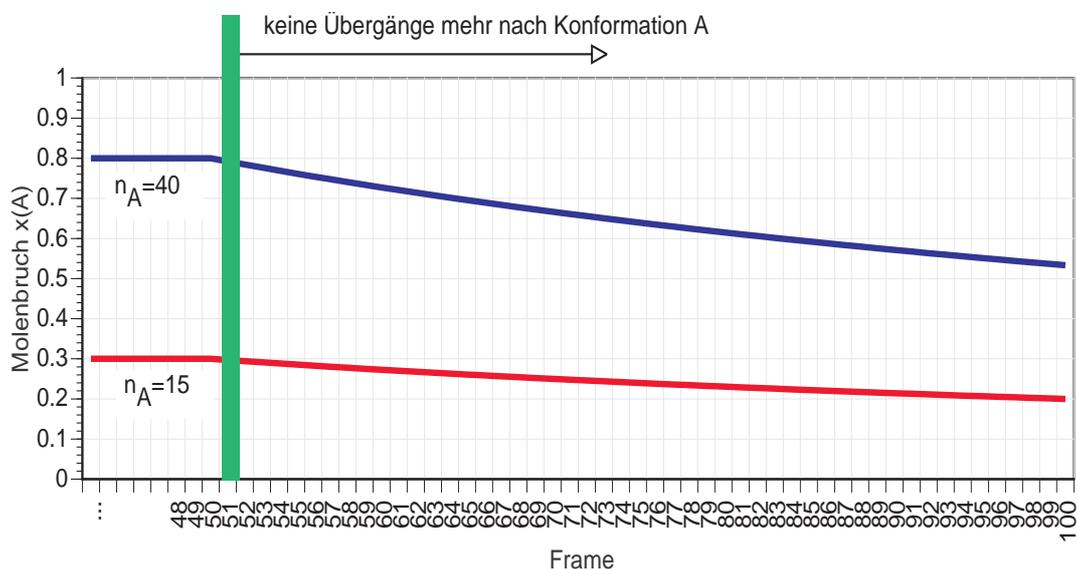


**Abb. 18:** Illustration der Zeitabhängigkeit von Variablen\*

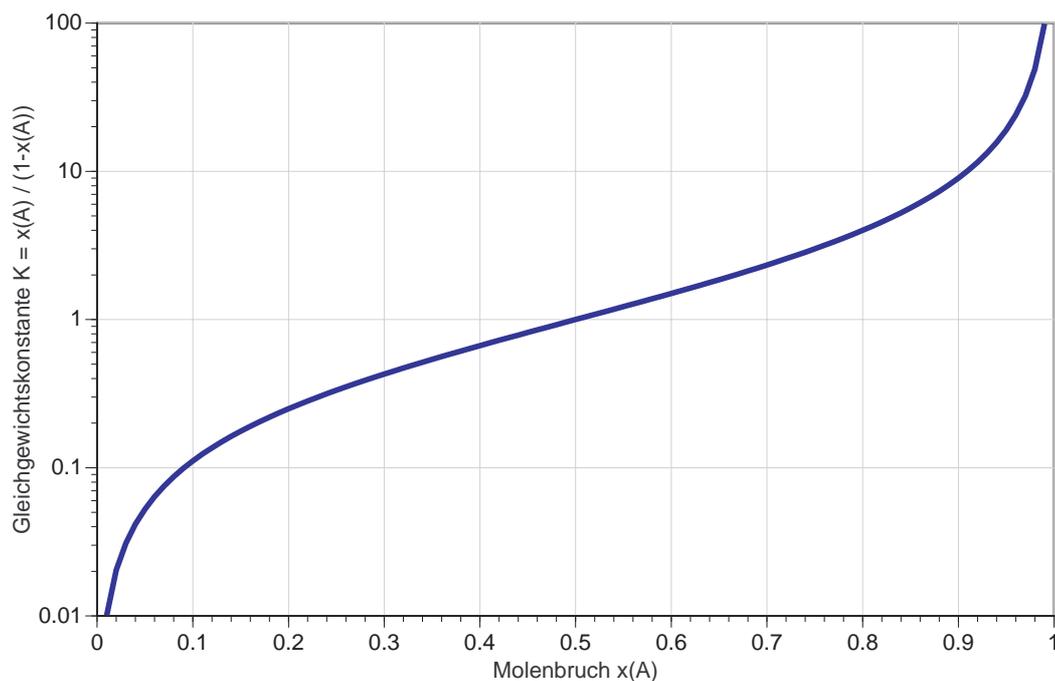
A: Beispiel für eine Fluktuation um einen Mittelwert (A, „lokale“ Zeitabhängigkeit)

B: Datenabfolge mit einer Tendenz (B, „globale“ Zeitabhängigkeit, „Drift“).

\* Die Werte der Variablen für A und B sind dieselben, sie unterscheiden sich nur in der zeitlichen Reihenfolge.



**Abb. 19:** Simulation der Änderung des Molenbruchs einer Konformation A, wenn diese nicht mehr von neuem populiert wird. (Dargestellt sind die Kurven für zwei verschiedene Ausgangsmolenbrüche von A).



**Abb. 20:** Änderung der Gleichgewichtskonstanten  $K_{AB}$  in Abhängigkeit vom Molenbruch  $x_A$  für ein Konformationsgleichgewicht zwischen einer Konformation A und B

molenbruch“ der Mittelwert  $\bar{x}_i[l_0 + m, f]$  aller  $x_i$  im Beobachtungszeitraum ( $\rightarrow EQ_i^{\bar{F}}(f)$ ) denkbar, dies hätte jedoch den Nachteil, daß die zeitliche Abfolge der Werte innerhalb des Beobachtungszeitraums verlorengeht.<sup>1</sup>

$$EQ_i^F(f) = \frac{1}{q} \sum_{l=l_0+m}^{f, m} \frac{|x_i(l) - x_i(l-m)|}{x_i(l-m)} \quad \begin{array}{l} l = (l_0 + m, l_0 + 2m, \dots, l_0 + qm) \\ l_0 = f - qm \end{array} \quad (20)$$

$$EQ_i^{\bar{F}}(f) = \frac{1}{q} \sum_{l=l_0+m}^{f, m} \frac{|x_i(l) - \bar{x}_i[l_0 + m, f]|}{\bar{x}_i[l_0 + m, f]} \quad (21)$$

- $m$ : Subintervallbreite (z.B. 100 Frames)
- $l_0$ : Frame-Index ab dem die Zeitabhängigkeit von  $x_i$  analysiert wird.
- $q$ : Anzahl der Subintervalle der Breite  $m$
- $\bar{x}_i[l_0 + m, f]$ : Mittelwert von  $x_i$  für alle Subintervalle im Bereich  $l_0+m$  bis  $f$

1. Bei den relativen Änderungen ist zu beachten, daß die Referenzmolenbrüche nicht Null sein dürfen (Division durch Null ist mathematisch nicht erlaubt). Dies wird in CAT dadurch abgefangen, daß entweder der aktuelle Molenbruch  $x_i(l)$  als Referenz gewählt wird oder die Berechnung von  $EQ_i(l)$  ausgesetzt wird. Die davon abhängigen Parameter werden den daraus resultierenden Verhältnissen angepaßt.

Die einfachste Möglichkeit einen Drift-Term  $EQ^D$  zu berechnen ist in Gleichung 22 gezeigt. Da hier die Differenz zwischen zwei expliziten Molenbrüchen  $x_i(f)$  und  $x_i(l_0+m)$  berechnet wird, ist  $EQ^D$  sehr empfindlich gegen Fluktuationen von  $x_i(f)$  und  $x_i(l_0+m)$ , daher ist es günstiger den Drift-Parameter auf Basis der Mittelwerte von  $x_i(f)$  über einen größeren Zeitbereich zu berechnen. So wird zur Berechnung des Drift-Terms  $EQ_{\bar{D}}$  (Gleichung 23) der Beobachtungszeitraum in zwei Hälften geteilt und für jede Hälfte der Mittelwert von  $x_i$  berechnet. Die relative Drift ist dann als doppelte Differenz der Mittelwerte der beiden Hälften dividiert durch den Mittelwert des gesamten Beobachtungszeitraums definiert. Multipliziert man anschließend mit 100 erhält man als „Einheit“ für den Drift-Parameter „Änderung in Prozent pro Beobachtungszeitraum“.

$$EQ_i^D(f) = \frac{|x_i(f) - x_i(l_0 + m)|}{x_i(f) + x_i(l_0 + m)} \quad (22)$$

$$EQ_i^{\bar{D}}(f) = \left( \frac{2 \cdot |\bar{x}_i[l_b, f] - \bar{x}_i[l_0 + m, l_b]|}{\bar{x}_i[l_0 + m, f]} \right) \quad l_b = l_0 + \frac{(q \cdot m)}{2} \quad (23)$$

$$EQ_i^{\bar{DL}}(f) = \left( \frac{|\bar{x}_i[l_0 + m, f] - \bar{x}_i[l_0 - qm, l_0]|}{\bar{x}_i[l_0 - qm, f]} \right) \quad \text{„Langzeit“-Drift} \quad (24)$$

Zur Prüfung, inwieweit es wirklich notwendig ist, die Änderungen der Molenbrüche  $x_i$  relativ zu einem Referenzmolenbruch zu berechnen, wurden absolute Fluktuationen  $EQ^{Fabs}$  in Form der Standardabweichung von  $\bar{x}_i[l_0 + m, f]$  berechnet (Gleichung 25). Der entsprechende Drift-Parameter ist in Gleichung 26 definiert. Beide wurden zum Gleichgewichtsparameter „EQ4“ kombiniert (Gleichung 30).

$$EQ_i^{Fabs}(f) = \left( \sqrt{\frac{1}{q-1} \sum_{l=l_0+m}^{f, m} (x_i(l) - \bar{x}_i[l_0 + m, f])^2} \right) \quad (25)$$

$$EQ_i^{\bar{D}abs}(f) = 2 \cdot (|\bar{x}_i[l_b, f] - \bar{x}_i[l_0 + m, l_b]|) \quad l_b = l_0 + \frac{(q \cdot m)}{2} \quad (26)$$

Ziel ist nun das Zeitverhalten der Molenbrüche aller Konformationen in einem einzigen empirischen Gleichgewichtsparameter  $EQ$  zusammenzufassen. Bei EQ1 (Gleichung 27), EQ3 (Gleichung 29) und EQ4 (Gleichung 30) wird diese Kombination durch einfache Mittelwertbildung erreicht. In EQ1 wird zusätzlich zur „normalen“ Drift im Beobachtungszeitraum noch eine „Langzeit“-Drift (Gleichung 24) berechnet. Eine weitere Optimierung wird durch Berücksichtigung der Anzahl der Konformationsübergänge im Beobachtungszeitraum erreicht. Ausgehend von der Bedingung, daß am Ende der Simulation aus jeder Konformation mindestens 100 Konformationsübergänge erfolgt sein

sollten, wird ein Bewertungsfaktor  $u_i^{trans}(f)$  definiert, der die Anzahl der Konformationsübergänge im Beobachtungszeitraum für jede Konformation auf die gesamte Simulationszeit extrapoliert. Ist für eine Konformation  $i$  die geforderte Anzahl von Konformationsübergängen noch nicht erreicht, ergibt sich  $u_i^{trans}(f) > 1$  und damit eine deutliche Erhöhung von EQ2.<sup>1</sup>

$$EQ1(f) = \frac{1}{3k} \sum_{i=1}^k (EQ_i^F(f) + EQ_i^{\bar{D}}(f) + EQ_i^{\bar{DL}}(f)) \quad (27)$$

$$EQ2(f) = \frac{1}{k} \sum_{i=1}^k EQ1(f) \cdot u_i^{trans}(f) \quad u_i^{trans}(f) = u_0 / (u_i(f) - u_i(l_0 + m)) \quad (28)$$

$$u_0 = 100 \cdot (f - l_0) / f$$

$u_i(f)$ : Anzahl der Übergänge aus Konformation  $i$  bis zum „Zeitpunkt“  $f$

$$EQ3(f) = \frac{1}{2} \sum_{i=1}^k ((EQ_i^F(f) + EQ_i^D(f))) \quad (29)$$

$$EQ4(f) = \frac{1}{2} \sum_{i=1}^k (EQ_i^{Fabs}(f) + EQ_i^{\bar{D}abs}(f)) \quad (30)$$

Zur sicheren Bewertung, ob und wie gut ein Konformationsgleichgewicht zum Zeitpunkt  $f$  eingestellt ist, wird schließlich nicht nur  $EQ(f)$  berücksichtigt, sondern aus allen  $EQ(l)$  im Beobachtungszeitraum der Mittelwert und die Drift berechnet (siehe Gleichung 31 und 32)

$$EQ^{mean}(f) = \frac{1}{q} \sum_{l=l_0+m}^{f,m} EQ(l) \quad (31)$$

$$EQ^{drift}(f) = 2 \cdot \overline{EQ}[l_b, f] - \overline{EQ}[l_0 + m, l_b] \quad l_b = l_0 + \frac{(q \cdot m)}{2} \quad (32)$$

Der Beobachtungszeitraum wurde bei allen Analysen in Abhängigkeit von  $f$  so variiert, daß er jeweils die letzten 25% der aktuellen Simulationszeit  $f$  umfaßt.

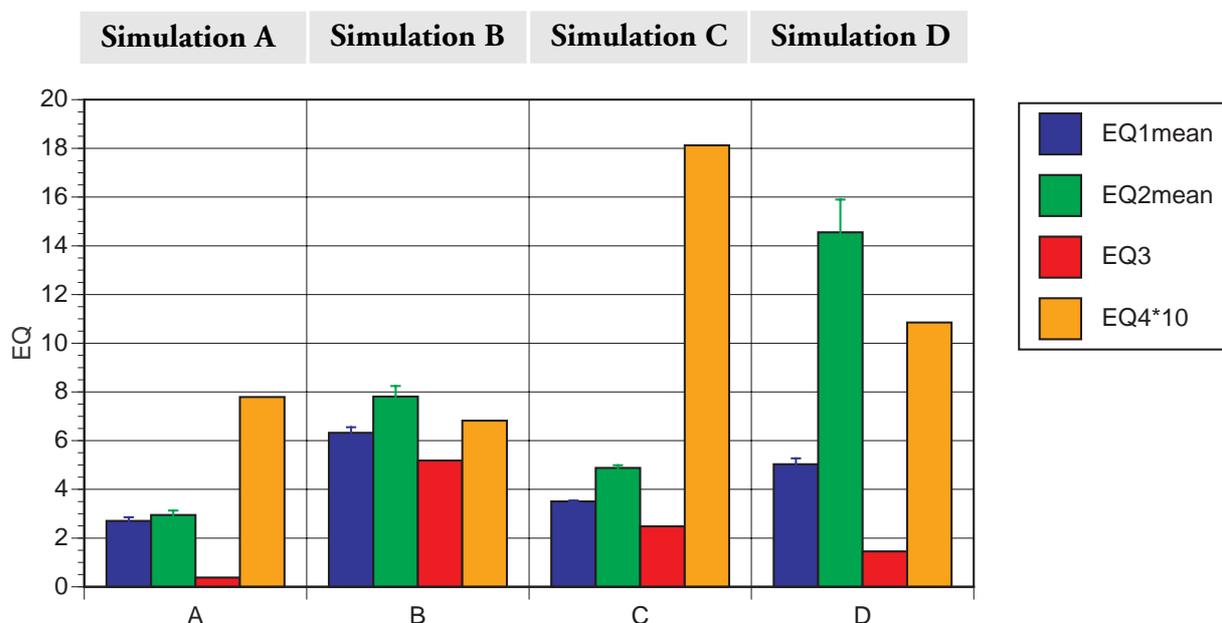
$$l_0 = f - qm \quad q = 0.25 \cdot \frac{f}{m} \quad q = \text{ganzzahlig} \quad (33)$$

1. Für den Fall, daß im Beobachtungszeitraum für eine Konformation  $i$  keine Konformationsübergänge stattfinden, wird  $u_i^{trans}(f) = 100$  gesetzt. Im Falle, daß überhaupt keine Konformationsübergänge stattfinden ist EQ2 undefiniert.

## 2.4.2 Bewertung der verschiedenen Gleichgewichtsparameter anhand der Einstellung eines Hydroxymethyl-Rotamerengleichgewichts

Die Daten der Langzeitsimulation (2 Mikrosekunden) von Maltose (100000 Frames, Sampling-Intervall = 20 ps) (vgl. Abschnitt 3.4.6) wurden in vier „unabhängige Simulationen“ aufgeteilt, so daß das Zeitverhalten der Rotamerenpopulation der Hydroxymethylgruppe bei vier Simulationen von jeweils 500 ns Simulationsdauer untersucht werden konnte (siehe Abb. X86 bis Abb. X89)<sup>1</sup>.

Die Berechnung der Gleichgewichtskonstanten zeigt deutlich, daß für Simulation B kein Konformationsgleichgewicht erreicht wurde, die Gleichgewichtskonstante  $K(gt/gg)$  fällt auch nach 400 ns Simulationszeit immer noch ab. Bei Simulation A (und auch evtl bei C) kann davon ausgegangen werden, daß sich das Konformationsgleichgewicht eingestellt hat. Bei Simulation D zeigen die Gleichgewichtskonstanten  $K(gt/gg)$  und  $K(gg/tg)$  am Ende der Simulation noch deutliche Schwankungen, was die Einstellung des Konformationsgleichgewichts in Frage stellt. Abb. 21 zeigt die verschiedenen berechneten Gleichgewichtsparameter EQ. Je kleiner EQ, desto besser sollte das Konformationsgleichgewicht eingestellt sein. Wegen der unterschiedlichen Berechnungsmethoden ist kein numerischer Vergleich zwischen verschiedenen EQs möglich. Es ist offensichtlich, daß EQ4 die Verhältnisse nicht richtig wiedergibt, da die Einstellung des Gleichgewichts bei Simulation A und B als praktisch „gleich gut“ bewertet werden. EQ3 ist sehr sensibel für Drifts, daher ergibt sich ein erhöhter Wert für EQ3 bei B. EQ1 und EQ2 unterscheiden sich in erster Linie bei Simulation D, hier scheinen kurzzeitig während der Simulation nicht genügend Übergänge vorhanden zu sein. Insgesamt beschreibt der Gleichgewichtsparameter EQ1 die Einstellung des Konformationsgleichgewichts im vorliegenden Fall am besten. Es ergibt sich für den Grad der Gleichgewichtseinstellung am Ende der jeweiligen Simulation die Reihenfolge  $A > C > D > B$ .



**Abb. 21:** Relative Bewertung der Einstellung des Gleichgewichts am Ende der Simulationen A bis D (MD Simulation von Maltose, Torsionswinkel L56O, CVFF,  $\epsilon = 4$ , 300 K, 500 ns)

1. Abbildungen mit der Kennzeichnung „X“ befinden sich in einem speziellen Abbildungsteil ab Seite 117.

## 3 Moleküldynamik-Simulationen als Methode zur effizienten Konformationsanalyse von Oligosacchariden

### 3.1 Einleitung und Zielsetzung

Die Konformationsanalyse von Oligosacchariden läßt sich in drei Bereiche einteilen:

1. Konformationssuche, d.h. das Aufspüren lokaler Minima der potentiellen Energie im n-dimensionalen Konformationsraum und deren statistische Analyse.
2. Analyse der Population des Konformationsraums im thermischen Gleichgewicht. Berechnung von Freien Energien aus der Rotamerenverteilung (Thermodynamik).
3. Analyse der Energiebarrieren für die Konformationsübergänge (Kinetik)

Ausgehend von dieser Einteilung sollen Simulations-Strategien und Analyse-Methoden an einfacheren Modellsubstanzen entwickelt werden, die es erlauben, basierend auf Moleküldynamik-Simulationen die konformative Vielfalt komplexer Oligosaccharide effizient zu analysieren. Weiterhin soll der Einfluß der Dielektrizitätskonstanten und Solvenseinflüsse untersucht werden.

### 3.2 Allgemeine Angaben zu den Simulationen

Für  $\beta$ -Maltose sind viele experimentelle Ergebnisse [179-181] wie auch berechnete Daten [128,133,134,152,178,182,183] in der Literatur veröffentlicht, daher eignet sich  $\beta$ -Maltose ausgezeichnet als Modell-Verbindung zur Entwicklung und Verfeinerung von Simulations- und Analyse-Methoden. Neben  $\beta$ -Maltose wurden auch für Sialyl-Lewis<sup>X</sup> (Neu5NAc $\alpha$ (2-3)Gal $\beta$ (1-4)[Fuc $\alpha$ (1-3)]GlcNAc), n-Butan, 1,1,3,3-Tetramethylcyclohexan, Glc $\alpha$ -O-iProp, Gal $\beta$ (1-3)Gal $\beta$  und Gal $\alpha$ (1-3)Gal $\beta$  MD-Simulationen durchgeführt.

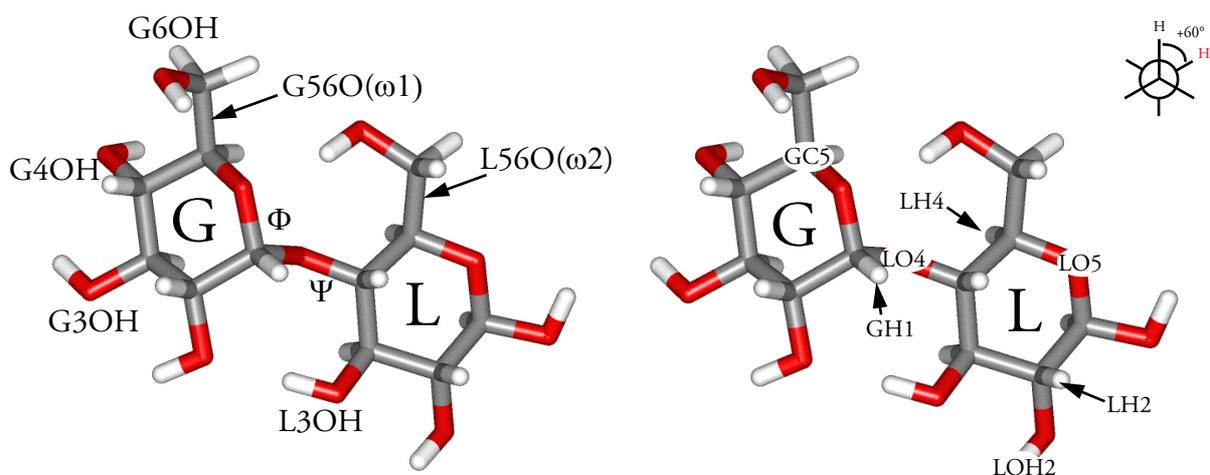
Zur Aufklärung des konformativen Verhaltens der aufgezählten Moleküle und zur Verifizierung diverser Methoden wurden mehrere MD-Simulationen mit unterschiedlichen Parametern durchgeführt. Die genauen Simulationszeiten und Temperaturen sind in den einzelnen Abschnitten angegeben. Der Integrationszeitschritt war grundsätzlich 1 fs. Als „Standard“-Kraftfeld wurde CVFF [174,184] gewählt, da mit diesem Kraftfeld für die lokalen  $\Phi/\Psi$ -Minima von Disacchariden gute Übereinstimmung mit dem Experiment gefunden wurde [185,186]. Desweiteren waren bereits eigene positive Erfahrungen bzgl. der Stabilität der MD-Simulationen mit diesem Kraftfeld vorhanden. Als Dielektrizitätskonstante für Simulationen ohne Solvens wurde meistens  $\epsilon = 4$  gewählt, da diese zum Modellieren von Kohlenhydraten ohne Solvens am besten geeignet ist [183].

Zur Überprüfung der Stabilität von Temperatur und Energie bei MD-Simulationen wurden grundsätzlich deren Mittelwerte für sequentiell aufeinanderfolgende Zeitintervalle berechnet und diese gegen die Zeit aufgetragen (Abb. X1)<sup>1</sup>. Ebenso wurde grundsätzlich eine Verteilung der relativen Energie und der Temperatur berechnet (Abb. X2).

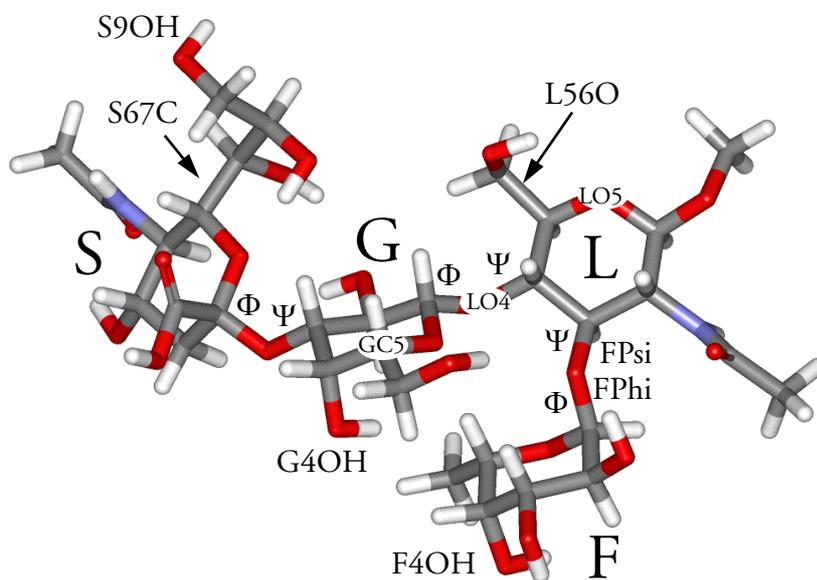
Alle Simulationen wurden mit dem Programm DISCOVER 2.9.7 (MSI/Biosym, San Diego) auf einer IBM SP2 (vgl. Abb. 10) durchgeführt und mit CAT2000 analysiert.

1. Abbildungen mit der Kennzeichnung „X“ befinden sich in einem speziellen Abbildungsteil ab Seite 117.

## Definition der Torsionswinkel und der Atomnamen der Modellverbindungen



**Abb. 22:** Definition der Torsionswinkel (links) und Atomnamen (rechts) von  $\beta$ -Maltose  
 $G56O = GO5-GC5-GC6-GO6$ ,  $G6OH = GC5-GC6-GO6-GOH6$ ,  
 $\Phi(\text{Phi})=GH1-GC1-LO4-LC4$ ,  $\Psi(\text{Psi})=GC1-LO4-LC4-LH4$   
 $G2OH =GH2- GC2-GO2-GOH2$ ,  $L1OH =LH1- LC1-LO1-LOH1$



**Abb. 23:** Definition der Einheiten, Torsionswinkel und Atomnamen von Sialyl-Lewis<sup>X</sup>  
 $S\text{Phi}=SC1-SC2-GO3-GC3$ ,  $S\text{Psi}=SC2-GO3-GC3-GH3$   
 $G\text{Phi}=GH1-GC1-LO4-LC4$ ,  $G\text{Psi}=GC1-LO4-LC4-LH4$   
 $F\text{Phi}=FH1-FC1-LO3-LC3$ ,  $F\text{Psi}=DC1-LO3-LC3-LH3$

### 3.3 MD-basierte Methoden zur Population des Konformationsraums von Oligosacchariden

#### 3.3.1 Hochtemperatur-MD-Simulationen (HTMD)

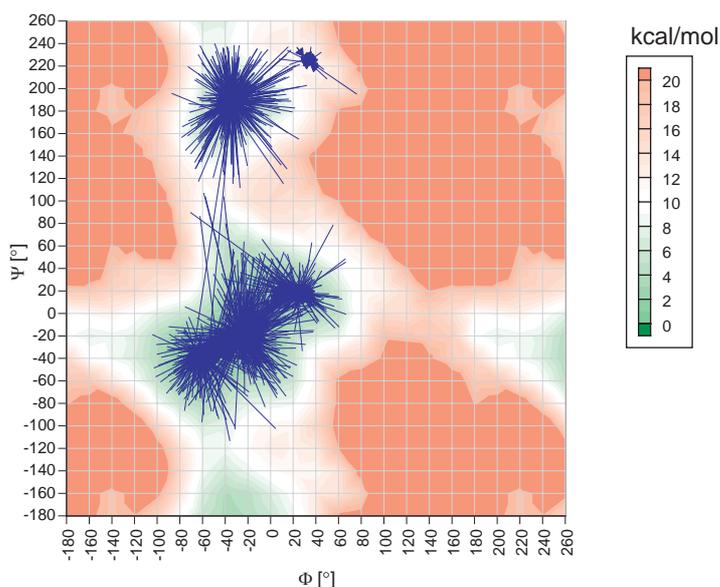
##### Überprüfung der Simulations-Methode am Modell-System $\beta$ -Maltose

Für  $\beta$ -Maltose wurde eine MD-Simulation mit einer Simulationsdauer von  $t_s = 10$  ns bei einer Simulationstemperatur von 1000 K (vgl. hierzu auch Abschnitt 3.4.7) durchgeführt (CVFF,  $\epsilon = 4$ , Dauer der Equilibrierungsphase  $t_0 = 20$  ps). Mit einem Sampling-Intervall von 1 ps wurden die Energien, Torsionswinkel und Atomabstände protokolliert und die Koordinaten in einem Strukturarchiv abgespeichert. Zur Vermeidung von Ringinversionen wurde an die Torsionswinkel „C1-C2-C3-C4“ und „C4-C5-O-C1“ mit der „Torsion Force“-Funktion von DISCOVER ein zusätzliches Potential angelegt (Kraftkonstante = 1000 kcal/rad<sup>2</sup>). Die Ringe waren damit in der <sup>4</sup>C<sub>1</sub>-Sesselkonformation praktisch eingefroren.

Die gesamte MD-Simulation von  $\beta$ -Maltose beanspruchte ca. 8h CPU-Zeit (IBM SP2, Single Processor). Als Resultat wurde ein Strukturarchiv (\*.arc File) mit 10000 nichtminimierten Strukturen (MD-Ensemble) und ein umfangreiches Simulationsprotokoll (\*.out File) erhalten, aus dem die Energien und Strukturparameter mit einem speziell dafür entwickelten Programm<sup>1</sup> extrahiert wurden.

Als Erweiterung der HTMD-Methode ist es sinnvoll, die während der MD-Simulation gespeicherten Strukturen anschließend ohne Constraints bzw. Restraints zu minimieren. Durch die Minimierung sammeln sich die  $\Phi/\Psi$ -Werte in den lokalen Minima der  $\Phi/\Psi$ -Energie-Karte („Conformational Clustering“ [128]; vgl. Abb. 24). Anhand der Cluster der minimierten Strukturen ist es einfacher die genaue Lage der  $\Phi/\Psi$ -Minima zu ermitteln, als dies anhand des MD-Ensembles möglich ist.

Die Häufigkeitsverteilung im  $\Phi/\Psi$ -Koordinatensystem („ $\Phi/\Psi$ -Raum“; Abb. X3) beweist, daß sich das System auf der  $\Phi/\Psi$ -Hyperfläche während der 1000 K-MD-Simulation ausschließlich entlang der Energietäler der Adiabatic-Map<sup>2</sup> bewegt. Auch die Häufigkeitsverteilung ausgewählter Torsions-



**Abb. 24:** „Conformational Clustering“ von HTMD-Strukturen ( $\beta$ -Maltose)  
Die Linien verbinden jeweils die  $\Phi/\Psi$ -Werte vor und nach der Minimierung.  
Hintergrund: Adiabatic Map berechnet durch einen  $\Phi/\Psi/\omega_1/\omega_2$ -Grid-Search

1. PAVDO („Programm zur Analyse von Discover Outputfiles“), Martin Frank, Heidelberg

winkel (Abb. X4) zeigt deutlich, daß sich das System für  $\Phi$ ,  $\Psi$  und  $\omega$  bei 1000K noch sehr stabil verhält<sup>1</sup>. Die drei Häufigkeitsmaxima für  $\omega$  bei  $-60^\circ$ ,  $+60^\circ$  und  $180^\circ$  treten deutlich hervor und die Population der Übergangszustände ist deutlich geringer.

Die Häufigkeitsverteilungen beliebiger einzelner Torsionswinkel- bzw. Torsionswinkelpaare<sup>2</sup> lassen sich über die Boltzmann-Beziehung in Energie-Profilen bzw. Maps („Freie Gridgruppen-Energie (FGE)“, siehe Abschnitt 2.3.4) überführen und man erhält so z.B. aus den MD-Daten  $\Phi/\Psi$ -Maps der Freien Energie<sup>3</sup>. Die FGE-Maps für  $\beta$ -Maltose sind selbst bei 1600 K noch in exzellenter Übereinstimmung mit den Adiabatic Maps (Abb. X5).

Aus der Häufigkeitsverteilung (Abb. X4) geht eindeutig hervor, daß HTMD-Simulationen nicht mehr gut geeignet sind, um die Rotamerenverteilung der OH-Gruppen zu beschreiben. Zwar stimmt das aus der Häufigkeitsverteilung bei 1000 K berechnete FGE-Profil (Abb. X6) noch qualitativ mit dem entsprechenden Profil aus einer 300 K MD-Simulation überein, der Verlauf des Profils ist jedoch bei einer Grid-Intervallbreite von  $10^\circ$  nicht mehr fließend.

### Anwendung ausgewählter Analyse-Methoden auf die Trajektorien-Daten

Ziel einer effizienten Konformationsuche wäre es, daß das zur Analyse der Trajektorien-Daten verwendete Programm anhand der Verteilung der  $\Phi/\Psi$ -Werte selbständig und automatisch die lokalen Minima findet und analysiert. Im Falle von  $\beta$ -Maltose ist es zwar noch relativ einfach, anhand der Trajektorien-Daten die syn- und die anti-Konformation voneinander zu separieren (vgl. Abb. X3), die Auftrennung der syn-Konformation in die aus der Literatur [133,134,152,178] bekannten „Sub-Minima“ ist programmtechnisch jedoch nicht trivial. Nahe beieinanderliegende Energieminima (lokale Populationsmaxima) lassen sich bei diesen hohen Temperaturen nicht mehr einfach separieren, da die zwischen ihnen liegenden energiearmen Übergangszustände ebenfalls sehr stark populiert sind. In CAT2000 wurden eine Reihe von Methoden implementiert, die es ermöglichen, MD-Daten automatisch zu clustern. Entscheidend für die erfolgreiche Anwendung der Cluster-Algorithmen ist die Wahl der (Start-) Parameter. Da deren Wert vom konkret vorliegenden Problem abhängt, erfordert der Einsatz von automatischen Cluster-Methoden eine gewisse Erfahrung. Als Beispiel wurde die „Variable-RMS-Clustering“-Methode (VRMS) (siehe Abschnitt 7.1) auf die  $\Phi/\Psi$ -Werte der 1000K-MD-Simulation angewendet (Abb. X7). Anhand der Populationsdichten findet der Algorithmus fünf Cluster, die programmintern sequentiell durchnummeriert werden. Die zu jedem Cluster gehörenden Strukturen werden in der Trajektorie mit der Clusternummer markiert.

Hat man die einzelnen  $\Phi/\Psi$ -Cluster (A bis D) (Abb. X7) definiert, können die zu den einzelnen Clustern gehörenden Eigenschaften (z.B. mittlere Energien) berechnet werden. Berechnet man z.B. die Häufigkeitsverteilungen der Hydroxymethylgruppen (Abb. X8) für jedes Cluster getrennt, erkennt man sofort aufgrund der gewählten Darstellung, daß für Cluster C' ( $-20^\circ/+20^\circ$ ) die Einstellung L56O = tg ( $180^\circ$ ) nicht stabil ist. Dies läßt sich damit erklären, daß sich bei dieser Anordnung die beiden O-Atome LO6 und GO5 zu nahe kommen. D.h. die Orientierung von L56O hat einen Einfluß auf  $\Phi/\Psi$ . Um zu überprüfen, ob dies auch für andere  $\Phi/\Psi$ -Bereiche gilt, wurden die Trajektorien-Daten basierend auf der Orientierung von L56O in drei Gruppen aufgeteilt (gg= $[-120^\circ, 0^\circ]$ , gt= $[0^\circ, 120^\circ]$ , tg= $[120^\circ, 240^\circ]$ ) anschließend wurde für jede Gruppe eine FGE-Map für  $\Phi/\Psi$  berech-

---

2. Die Adiabatic Map wurde durch einen Grid-Search erhalten, bei dem  $\Phi$  und  $\Psi$  in  $20^\circ$ -Schritten und die beiden  $\omega$ -Winkel in  $120^\circ$ -Schritten (gg, gt, tg) variiert wurden. Aus den so erhaltenen neun  $\Phi/\Psi$ -Maps wurde für jeden  $\Phi/\Psi$ -Gridpunkt die niedrigste Energie in der Adiabatic Map aufgetragen.

1. Für eine detaillierte Diskussion der „Temperatureffekte“ siehe Abschnitt 3.4.7

2. Im Prinzip lassen sich mit CAT2000 beliebig dimensionierte Populations-Grids in FGE-Grids überführen, es lassen sich jedoch nur maximal dreidimensionale Grids (z.B. 2 Torsionswinkel und Energie) graphisch darstellen.

3. Diese Aussage gilt nur allgemein, wenn das System das thermische Gleichgewicht erreicht hat (vgl. Abschnitt 2.3.4).

net (Abb. X9). Es zeigt sich, daß die Einstellung der Hydroxymethylgruppe L56O nicht nur den  $\Phi/\Psi$ -Bereich um  $(-20^\circ/+20^\circ)$ , sondern darüber hinaus auch die Lage des globalen Minimums beeinflusst.

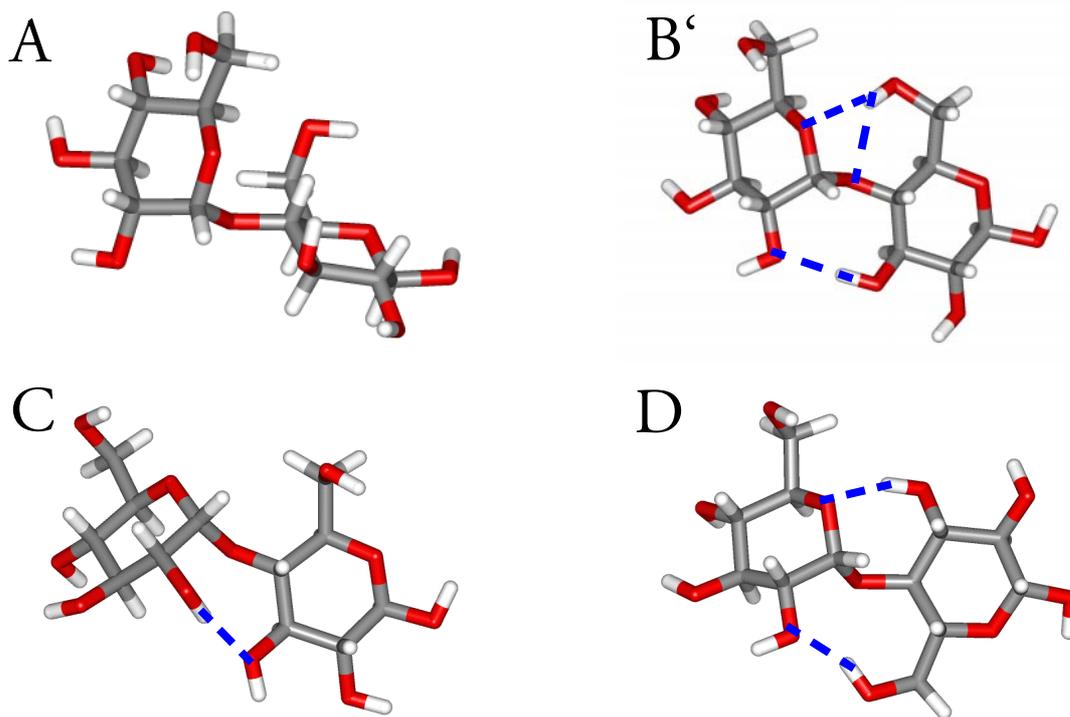
Eine Rotationsanalyse, bei der die Torsionswinkel hinsichtlich vollständiger Rotationen um  $360^\circ$  untersucht werden, ergab, daß das System bei 1000 K mit 5 Rotationen pro Nanosekunde um  $\Psi$  rotiert. Die Rotationsfrequenz um  $\Phi$  ist mit  $0.3 \text{ ns}^{-1}$  deutlich geringer (Abb. X10, rechts). Dies ist auch wegen der höheren Übergangsbarriere in  $\Phi$ -Richtung (Abb. X5) zu erwarten. Zur Visualisierung der Rotationen in Form einer „Rotationstrajektorie“ (Abb. X10, links) wurden die Torsionswinkelskala durch Addition oder Subtraktion von  $360^\circ$  so erweitert, daß in der zeitlichen Abfolge keine Änderungen der Torsionswinkel größer als  $180^\circ$  auftreten.

Zur Konformationsaufklärung mittels NMR-Methoden sind die interresidualen H-H-Abstände von besonderem Interesse<sup>1</sup>. Die Häufigkeitsverteilungen der Abstände im MD-Ensemble bieten einen schnellen Überblick welche Abstände für ein bestimmtes H-H-Paar überhaupt auftreten können (Abb. X11). Sind viele verschiedene Abstände zu analysieren, so sind die klassischen Häufigkeits-Plots in Form von Profilen unübersichtlich. Es hat sich daher bewährt, die Häufigkeits-Diagramme als „Pseudo-3D“-Histogramme darzustellen (Abb. X12), bei denen in x-Richtung die Torsionswinkelskala und die einzelnen Torsionswinkel in der y-Richtung angeordnet sind. Die Population (z-Richtung) wird durch unterschiedliche Farbgebung kodiert. Mit dieser Art der Darstellung lassen sich auch Verteilungen für verschiedene  $\Phi/\Psi$ -Cluster noch übersichtlich anordnen (Abb. X13). Alternativ können die mittleren Abstände oder die  $\langle r^6 \rangle$ -gemittelten Abstände pro Cluster (Abb. X14) berechnet werden um Informationen darüber zu erhalten, welche NOE-Kontakte für die einzelnen Cluster zu erwarten sind und welche davon die Population einer bestimmten Konformation eindeutig belegen. Für Maltose würde beispielsweise das Vorliegen der anti-Konformation eindeutig durch einen starken NOE für GH1-LH3 **und** GH1-LH5<sup>2</sup> bewiesen werden. Eine Bevorzugung von  $\Phi/\Psi$ -Winkeln im Bereich von  $-60^\circ/-40^\circ$  (Cluster A) würde sich im Vorhandensein eines mittelstarken NOE für GH1-LH3 **und** GH1-LH4 zeigen, während Cluster B und C einen starken NOE für GH1-LH4 hervorrufen würden.

### Anwendung ausgewählter Analyse-Methoden auf die minimierten Strukturen

Wie bereits anfangs erwähnt, bietet sich als Erweiterung der HTMD-Strategie die „Conformational Clustering“-Methode an. Die  $\Phi/\Psi$ -Werte sammeln sich in den lokalen Minima zu Clustern und man erhält damit genaue Informationen über mögliche Minimum-Konformationen (Abb. X15). Im Falle von Maltose erhält man unter den gewählten Bedingungen (CVFF,  $\epsilon=4$ , Conjugate Gradient Minimizer,  $0.001 \text{ kcal}/\text{\AA}$ ) drei deutlich voneinander getrennte lokale Minima D ( $-35^\circ/190^\circ$ , anti-Konformation), E ( $195^\circ/-20^\circ$ ), F ( $30^\circ/230^\circ$ ) und ein ausgedehntes  $\Phi/\Psi$ -Minimum (syn), das sich diagonal von  $(-70^\circ/-40^\circ)$  bis  $(30^\circ/20^\circ)$  erstreckt (Abb. X16). Durch sequentielle Anwendung verschiedener Clustering-Methoden („Hand Over Hand (HOH)“, „Fixed RMS“)<sup>3</sup> ließ sich der syn-Bereich in die Cluster A, B, B' und C aufteilen. Interessant hierbei ist die Frage, wodurch sich die Aufspaltung in B und B' erklären läßt. Eine Analyse der Torsionswinkel-Verteilung für die einzelnen Cluster (Abb. X17) zeigt, daß bei B' nur L56O=tg vorkommt und bei B entsprechend nur L56O=gg bzw. gt.

1. Weitergehende Möglichkeiten der Anwendung von MD-Simulationen zur Unterstützung der Strukturaufklärung mit NMR-Methoden werden im Abschnitt 3.3.3 detailliert dargestellt.
2. Zur Nomenklatur siehe Abb. 22
3. Beschreibung der Methoden siehe Abschnitt 7.1



**Abb. 25:** Lokale Minima von  $\beta$ -Maltose

Die Analyse der H-Brücken<sup>1</sup> (Abb. X18) zeigt für B' eine starke Interaktion<sup>2</sup> von LOH6...GO5 und LOH6...LO4. Charakteristische H-Brücke für Cluster B ist die interresiduale H-Brücke LOH3...GO2. Für Cluster C ist diese H-Brücke zwar auch möglich, die Bildung der entgegengerichteten H-Brücke GOH2...LO3 ist jedoch bevorzugt. Für Cluster A existieren keine interresidualen H-Brücken.

Die Ursache der Aufspaltung in Cluster B und B' ist also die Ausbildung einer zusätzlichen H-Brücke bei B' ausgehend von LOH6 (vgl. Abb. 25).

Die Analyse der energetischen Reihenfolge<sup>3</sup> der Minima ergibt: B (0 kcal/mol) < A (0.4 kcal/mol) < B' (0.5 kcal/mol) < C (0.8 kcal/mol) < D (2.8 kcal/mol) < E (10.4 kcal/mol)

Ein Vergleich der gefundenen  $\Phi/\Psi$ -Minima mit der Kristallstruktur von  $\beta$ -Maltose ( $\Phi/\Psi$  im Bereich von  $0^\circ \pm 10^\circ$ ) [179,187] zeigt, daß das globale Minimum B mit  $\Phi=-25\pm 4^\circ$  und  $\Psi=-21\pm 3^\circ$  nicht sehr weit davon entfernt ist. Weiterhin belegen die Kristallstrukturen das Vorhandensein einer H-Brücke zwischen GO2 und LO3.

Für eine detaillierte vergleichende Diskussion der, mit anderen Kraftfeldern, berechneten  $\Phi/\Psi$ -Minima für Maltose und experimentellen Ergebnissen aus NMR-Messungen bzw. der Messung der optischen Rotation [181] sei auf [152] verwiesen. Danach kann eine einzelne Struktur nicht in Übereinstimmung mit den experimentellen Befunden gebracht werden. Es wurde sogar das Vorhandensein der anti-Konformation vorgeschlagen [188]. Erst kürzlich wurde eine Kristallstruktur von komplex

1. Die H-Brücken wurden mit DISCOVER durch den Befehl „print hbonds“ ausgegeben. Eine H-Brücke liegt dann vor, wenn  $r(\text{OH}\cdots\text{O}) < 2.5\text{\AA}$  und  $\phi(\text{O-H}\cdots\text{O}) > 120^\circ$  ist (geometrisches Kriterium).
2. Die „Stärke“ einer H-Brücke ist durch ihre mittlere Population im Ensemble definiert.
3. jeweils die energieärmste Struktur

gebundener Maltose veröffentlicht (PDB-Code=1b8z) [180] in der vier Maltose-Moleküle mit leicht verschiedenen  $\Phi/\Psi/\omega_1/\omega_2$ -Orientierungen gebunden sind ( $-38^\circ/-18^\circ/\text{gg}/\text{gt}$ ,  $-11^\circ/9^\circ/\text{gg}/\text{gg}$ ,  $-59^\circ/-27^\circ/\text{gg}/\text{gg}$ ,  $-18^\circ/-9^\circ/\text{gt}/\text{gt}$ ).

Insgesamt läßt sich daher für  $\beta$ -Maltose von einem breiten Minimum ausgehen, welches sich im  $\Phi/\Psi$ -Bereich ( $-60^\circ/-40^\circ$ ) bis ( $-10^\circ/10^\circ$ ) erstreckt (Minimum A und B). Minimum C, das durch intramolekulare H-Brücken stabilisiert wird, dürfte in aprotischen Lösungsmitteln eine wichtige Rolle spielen. Dafür spricht auch der Befund, daß sich die optische Rotation ändert, wenn man von Wasser zu DMSO als Lösungsmittel übergeht [189]. Für die anti-Konformation ergibt sich aus den Rechnungen eine relative Energie von 3-4 kcal/mol. Diese sollte daher im freien Zustand nicht populiert sein. Die durch die Konformationssuche mit der HTMD-CC-Methode (Hochtemperatur-MD-Simulation Conformational Clustering) gefundenen  $\Phi/\Psi$ -Minima repräsentieren in bemerkenswerter Weise die Variabilität der experimentell gefundenen Strukturen.

Die HTMD-CC-Methode ist also geeignet, für ein Disaccharid alle wichtigen lokalen  $\Phi/\Psi$ -Minima innerhalb einer Simulationsdauer von 10 ns bei 1000 K zu populieren.

### Anwendung der Methode auf ein biologisch wichtiges Oligosaccharid: Sialyl-Lewis<sup>X</sup>

Als Beispiel für ein biologisch wichtiges Oligosaccharid wurde das Tetrasaccharid Sialyl-Lewis<sup>X</sup> [35,190,191] mit der HTMD-CC-Methode untersucht. Aus einer 1000K-MD-Simulation des neutralen Zuckers (CVFF,  $\epsilon = 80$ ,  $t_s = 10$  ns,  $\Delta t_s = 1$  ps) wurden die Daten von 10000 Frames erhalten. Durch die räumliche Nähe der beiden Zuckereinheiten Gal und Fuc ist anzunehmen, daß die Einstellung der glycosidischen Bindung Gal $\beta$ (1-4)GlcNAc nicht unabhängig von Fuc $\alpha$ (1-3)GlcNAc sein wird (Abb. 23). Ziel war es, alle wichtigen Konformationen zu ermitteln, die durch Kombination der verschiedenen lokalen  $\Phi/\Psi$ -Minima der einzelnen glycosidischen Bindungen entstehen. Dadurch, daß die Methode auf Trajektorien-Daten des Gesamtmoleküls beruht, sind, im Gegensatz zu systematischen Kombinationsmethoden (z.B. Disaccharid-Methode, vgl. Abschnitt 1.3), alle Effekte, die auf der Wechselwirkung nicht direkt miteinander verbundener Monosaccharide beruhen, bereits berücksichtigt.

Der erste Schritt in der Konformationsanalyse von Oligosacchariden ist die visuelle Bestimmung der lokalen  $\Phi/\Psi$ -Minima durch Berechnung und Darstellung einer FGE-Map für jede der glycosidischen Bindungen getrennt (Abb. X19, Abb. X20, Abb. X21). Dadurch erhält man einen Eindruck in welchem Umfang der Konformationsraum für die einzelnen glycosidischen Bindungen durchsucht wurde. Der nächste Schritt besteht im Auffinden von stabilen Konformationen für alle  $\Phi/\Psi$ -Paare kombiniert. Um stabile Konformationen und deren Eigenschaften (Mittelwerte, Verteilungen, usw.) berechnen zu können, müssen die Trajektorien-Daten zuerst in die Konformationen gruppiert werden, erst anschließend kann über die Gruppen-Codes, wie schon bei  $\beta$ -Maltose beschrieben, eine statistische Analyse der Gruppen erfolgen. Das automatische Aufspüren von Populationsmaxima (lokale Energieminima, Konformationen) im sechsdimensionalen  $\Phi/\Psi$ -Hyperraum ist zwar theoretisch möglich, die visuelle Überprüfung des Erfolgs des Clustering-Algorithmus ist jedoch nicht möglich. Um eine sichere Definition der stabilen Konformationen zu gewährleisten, werden daher im ersten Schritt die Daten der drei  $\Phi/\Psi$ -Paare unabhängig geclustert, z.B. durch die FRMS-Clustering-Methode<sup>1</sup> von CAT2000 (Abb. X24). Der Erfolg des Clustering- oder Gruppierungs-Algorithmus kann, wie schon erwähnt, durch Vergleich mit den FGE-Maps (Abb. X21) überprüft werden. Sind die lokalen Minima für jede glycosidische Bindung bestimmt, werden durch Kombination der einzelnen lokalen Gruppen die kombinierten PP-Gruppen erzeugt (Definition siehe Abb. X25). In der Praxis hat es sich bewährt, sich nicht auf automatische Clustering-Algorithmen zu verlassen, sondern die Daten mit der „Tem-

1. Beschreibung der Methoden siehe Abschnitt 7.1

plate-Grouping“-Methode (siehe Abschnitt 7.1) zu gruppieren. Ein konkretes Beispiel soll das Vorgehen noch einmal verdeutlichen: Für  $\text{Fuc}\alpha(1-3)\text{GlcNAc}$  werden mit Hilfe der Minima der FGE-Map (Abb. X21) drei Templates („virtuelle Konformationen“) mit  $\Phi/\Psi$ -Werten ( $50^\circ/30^\circ$ ), ( $-30^\circ/-25^\circ$ ) und ( $35^\circ/185^\circ$ ) definiert und ausgehend von diesen  $\Phi/\Psi$ -Zentren mit einer RMS-Toleranz von  $30^\circ$  drei Gruppen erzeugt, die einen frei definierbaren numerischen Gruppen-Code (Sub-PP-Code) zugeordnet bekommen. Alle Frames, die keiner der drei Template-Gruppen zugeordnet werden können, werden der Gruppe „Null“ („Gap“-Gruppe, undefinierte Gruppe) zugeordnet. Zum Vergleich: Der „Populationsgrid-basierte-Fixed-RMS-Clustering-Algorithmus (PFRMS)“<sup>1</sup> (siehe Abschnitt 7.1) liefert für die glycosidische Bindung  $\text{Fuc}\alpha(1-3)\text{GlcNAc}$  vier Cluster mit  $\Phi/\Psi$ -Zentren bei ( $60^\circ/30^\circ$ ), ( $-25^\circ/-25^\circ$ ), ( $10^\circ/175^\circ$ ) und ( $50^\circ/220^\circ$ ) (Abb. X24). Ein Vergleich mit der zugehörigen Adiabatic Map (Abb. X22)<sup>2</sup> läßt es jedoch als sinnvoll erscheinen, die beiden letzten Cluster zusammenzufassen. Analog dazu werden auch die Gruppen (GPP und SPP) für die beiden anderen glycosidischen Bindungen  $\text{Gal}\beta(1-4)\text{GlcNAc}$  und  $\text{Neu5Ac}(2-3)\text{Gal}$  erzeugt (vgl. Abb. X25). Jeder Frame besitzt nun einen SPP-, einen GPP- und einen FPP-Code. Durch Kombination der Sub-PP-Codes werden neue „kombinierte“ PP-Gruppen erzeugt, die in der Reihenfolge ihres Auftretens entlang der Trajektorie sequentiell durchnummeriert werden (Abb. X26). Alle PP-Gruppen, die durch Kombination mit einem Sub-PP-Code „Null“ (Gap-Gruppe) erhalten wurden, werden nachträglich zur PP-Gruppe „Null“ zusammengefaßt und damit praktisch eliminiert. Aus Abb. X26 ist ersichtlich, daß nach 2 ns Simulationszeit bei 1000 K die Zahl der neuen PP-Gruppen nur noch sehr langsam zunimmt. Von den theoretisch möglichen 120 Kombinationen (48 ohne Gap-Gruppen) wurden nach  $t_s = 10$  ns bei 1000 K zweidrittel zumindest einmal populiert. Die nichtpopulierten Kombinationen sind wahrscheinlich zu energiereich und daher physikalisch nicht relevant<sup>3</sup>. Die Populations-Analyse ergibt ein Maximum der Population für Gruppe 16 und 42 (Abb. X27). Eine Rückübersetzung des PP-Codes anhand der Tabellen in Abb. X25 ergibt, daß Konformationen mit  $\text{G}\Phi = 50^\circ$ ,  $\text{G}\Psi = 10^\circ$ ,  $\text{F}\Phi = 50^\circ$ ,  $\text{F}\Psi = 30^\circ$  am stabilsten sein sollten. Für  $\text{Neu5Ac}\alpha(2-3)\text{Gal}$  existieren zwei Einstellungen im Bereich  $\text{S}\Phi = -70^\circ$ ,  $\text{S}\Psi = 5^\circ$  und  $\text{S}\Phi = -160^\circ$ ,  $\text{S}\Psi = -20^\circ$ . Dies ist im Einklang mit den aus NMR-Messungen erhaltenen Ergebnissen [149,193]. Eine Analyse der Konformationsübergänge entlang der PP-Trajektorie zeigt, daß mehr als dreimal soviel Übergänge zwischen PP-Gruppe 16 und 42 stattfinden als zwischen irgendeinem anderen PP-Gruppen-Paar (Abb. X28). Bei diesem Übergang handelt es sich um einen SPP-Übergang zwischen  $\text{SPP} = 1$  ( $-70^\circ/5^\circ$ ) und  $\text{SPP} = 2$  ( $-160^\circ/-20^\circ$ ). Auch bei den in der Übergangshäufigkeit folgenden beiden PP-Übergangs-Paaren ( $41 \leftrightarrow 54$ ) und ( $35 \leftrightarrow 36$ ) handelt es sich um  $\text{SPP}(1 \leftrightarrow 2)$ -Übergänge. Bei den PP-Gruppen (35, 36, 41, 54), die zusammen eine Häufigkeit von 30% aufweisen, liegt für die glycosidische Bindung  $\text{Fuc}\alpha(1-3)\text{GlcNAc}$  eine Orientierung von  $\text{F}\Phi = -30^\circ$ ,  $\text{F}\Psi = -25^\circ$  vor. Ein Konformationsgleichgewicht wurde für die meisten PP-Gruppen innerhalb von 10 ns nicht erreicht, daher kann aus den Populationen der PP-Gruppen auch nicht auf deren Freie Energie geschlossen werden<sup>4</sup>.

1. Der PFRMS-Algorithmus findet die Zentren maximaler lokaler Population im n-dimensionalen Raum basierend auf einem n-dimensionalen Histogramm.
2. Für den Lewis<sup>X</sup>-Teil des Moleküls wurde für die beiden glycosidischen Bindungen eine Adiabatic Map durch einen zweistufigen Grid-Search berechnet: zur Berechnung der Adiabatic Map für  $\text{Gal}\beta(1-4)\text{GlcNAc}$ , wurden zuerst die Torsionswinkel  $\text{F}\Phi$  und  $\text{F}\Psi$  systematisch um  $20^\circ$  variiert (Rigid Map). Die so erzeugten 324 Strukturen wurden anschließend ohne Constraints minimiert („Conformational Clustering“ [128,192]). Für jedes der erhaltenen fünf  $\text{F}\Phi/\text{F}\Psi$ -Cluster wurde die Struktur mit der niedrigsten Energie aus dem Original-Archiv extrahiert und anschließend für jede der fünf extrahierten (Start-)Strukturen eine  $\text{G}\Phi/\text{G}\Psi$ -Relaxed-Map berechnet. Aus dem Energieminimum für jeden  $\text{G}\Phi/\text{G}\Psi$ -Gridpunkt erhält man schließlich eine Adiabatic Map für  $\text{G}\Phi/\text{G}\Psi$  (Abb. X23). Zur Berechnung der Adiabatic Map für  $\text{F}\Phi/\text{F}\Psi$  wurde analog vorgegangen (Abb. X22).
3. Die Möglichkeit, daß die fehlenden Kombinationen aus kinetischen Gründen (zu hohe Übergangsbarrieren) nicht auftreten, kann anhand der Daten nicht ausgeschlossen werden. Durch die Ergebnisse der 100 ns MD-Simulation bei 1000 K (siehe folgende Ausführungen) ist dies jedoch unwahrscheinlich.

Die Minimierung<sup>1</sup> der 10000 gespeicherten Strukturen (Conformational Clustering) ergab die in Abb. X29 bis Abb. X31 dargestellten  $\Phi/\Psi$ -Werte für die Minimumstrukturen. Für Neu5Ac $\alpha$ (2-3)Gal finden sich im Bereich der FGE-Energie  $<2$  kcal/mol (Abb. X29) drei bis vier lokale Minima A ( $-70^\circ/0^\circ$ ), B ( $-150^\circ/-20^\circ$ ), C ( $-100^\circ/-40^\circ$ ) und A' ( $-120^\circ/30^\circ$ ). In energiereicheren Regionen sind weitere lokale Minima (D bis G) zu finden. Alle diese Minima sind in der Literatur bereits beschrieben, für Minima A, B, C und D liegen auch experimentelle Hinweise vor [149,151,193,194].

Für Gal $\beta$ (1-4)GlcNAc (Abb. X30) befinden sich lokale (Haupt-)Minima im Bereich A ( $60^\circ/20^\circ$ ), B ( $-40^\circ/-20^\circ$ ) C ( $50^\circ/180^\circ$ ) und D ( $180^\circ/20^\circ$ ). Minimum C und D zeigen in  $\Phi$ -Richtung eine ausgeprägte Flexibilität. Die (Haupt-)Minima für Fuc $\alpha$ (1-3)GlcNAc (Abb. X31) befinden sich im Bereich A [ $(65^\circ/40^\circ)$ ,  $(40^\circ/20^\circ)$ ,  $(20^\circ/30^\circ)$ ], B ( $-25^\circ/-25^\circ$ ), C ( $20^\circ/-25^\circ$ ) und D ( $30^\circ/180^\circ$ ).

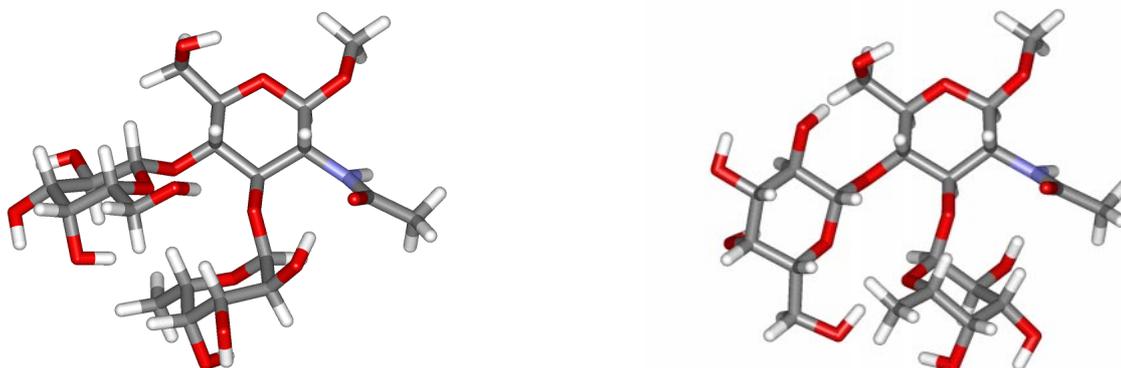
Die Kristallstruktur von Lewis<sup>X</sup> zeigt für Gal $\beta$ (1-4)GlcNAc und Fuc $\alpha$ (1-3)GlcNAc  $\Phi/\Psi$ -Werte von ( $45^\circ/14^\circ/45^\circ/19^\circ$ )[195] bzw. ( $42^\circ/19^\circ/46^\circ/18^\circ$ )[196] und liegt damit wie auch die NMR-Struktur ( $65 \pm 10^\circ/0 \pm 15^\circ/75 \pm 10^\circ/25 \pm 10^\circ$ ) [59] für beide glycosidischen Bindungen im Bereich A.

Um schließlich ein Konformationsgleichgewicht (vgl. Abschnitt 2.4) für die PP-Gruppen zu erreichen wurde eine 1000K-MD-Simulation des neutralen Zuckers mit einer Simulationsdauer von 100 ns (CVFF,  $\epsilon=80$ ,  $\Delta t_s=1$  ps) durchgeführt. Der Vergleich der Adiabatic Maps (Abb. X22, Abb. X23) mit den aus den MD-Daten berechneten FGE-Maps (Abb. X20, Abb. X21) ergibt eine gute Übereinstimmung was die Lage der globalen und lokalen Minima betrifft. Die energetische Bewertung der Nebenminima unterscheidet sich teilweise jedoch deutlich. So besitzt das Minimum B der FPhi/FPsi-Map (vgl. Abb. X31) in der Adiabatic Map eine relative Energie von 2 kcal/mol in der FGE-Map ist der Energieunterschied jedoch deutlich geringer (0.3 kcal/mol). Das Minimum C ist in beiden Maps nicht sichtbar.

Der Lewis<sup>X</sup>-Teil von Sialyl-Lewis<sup>X</sup> gilt allgemein als unflexibel [59,194]. Eine Analyse der LX-Gruppen (vgl. Abb. X32) zeigt, daß nach 25 ns bei 1000K alle theoretisch möglichen 12 Gruppen mindestens einmal populiert wurden (Abb. X33). Weiterhin wurde innerhalb von  $t_s = 100$  ns für alle LX-Gruppen außer 12 das Konformationsgleichgewicht erreicht (Abb. X34, Abb. X37). Dies ist nicht weiter verwunderlich, denn Gruppe 12 besitzt eine relative Energie von  $>15$  kcal/mol und hat somit keine physikalische Relevanz. Unter diesem Aspekt sind bereits nach weniger als 5 ns alle relevanten LX-Gruppen populiert, so daß eine 1000K-MD-Simulation der Länge 10 ns ausreichend ist, um für den Lewis<sup>X</sup>-Teil alle relevanten Konformationen im  $\Phi/\Psi$ -Hyperraum zu populieren. Die LX-Gruppe mit der höchsten Population (LX-Code = 3) (vgl. Abb. 26) entspricht einer Konformation wie sie auch experimentell gefunden wurde [59,195,196]. Die gegenüber den anderen Zuständen deutlich erhöhte Lebensdauer von 46 ps (bei 1000 K) (Abb. X36) weist auf hohe Übergangsbarrieren hin und unterstützt damit den Befund, daß bei 300K-MD-Simulationen keine  $\Phi/\Psi$ -Konformationsübergänge im Lewis<sup>X</sup>-Teil gefunden wurden [149]. Die meisten Übergänge finden zwischen LX-Gruppe 1 und 3 statt (Abb. X35). Es handelt sich hierbei um einen FPP(1 $\leftrightarrow$ 2)-Übergang bei dem Fuc $\alpha$ (1-3)GlcNAc zwischen Minimum A und B wechselt. Der Energieunterschied zwischen LX-Gruppe 1 und 3 beträgt 1.9 kcal/mol. Dieser Wert ergibt sich sowohl aus den Differenzen der Mittelwerte der potentiellen Systemenergie als auch aus den anhand der Populationsunterschiede berechneten Freien Energie (Abb. X32). Der Energieunterschied von 1.9 kcal/mol entspricht auch in etwa dem Wert wie er sich aus der Adiabatic Map von FPhi/FPsi ergibt (Abb. X22). Ein Grund warum Minimum B (FPP-

4. Aus einem Vergleich mit einer 1000 K MD-Simulation der Länge  $t_s = 100$  ns ergab sich, daß die aus den Daten der 10 ns-MD-Simulation berechneten „FGE-Maps“ jedoch schon eine gute Näherung an die Gleichgewichts-FGE-Map darstellen.

1. Durchschnittlich sind 1500 Minimierungsschritte bis zum Konvergenzkriterium von 0.001 kcal/Å notwendig ( $\epsilon=80$ )



**Abb. 26:** Globales Minimum LX=3 (links) und Nebenminimum LX=9 (rechts) von Sialyl-Lewis<sup>X</sup> (vgl. Abb. X32)

Code = 2) in der FGE-Map von Fuc $\alpha$ (1-3)GlcNAc deutlich energieärmer ist als in der Adiabatic Map, ist entropischer Natur: Für FPP-Code = 2 gibt es zwei Realisierungsmöglichkeiten mit Energien im Bereich bis 2 kcal/mol (LX-Gruppen 1 und 9) (vgl. Abb. 26); die nächste Realisierungsmöglichkeit für FPP-Code = 1 (Minimum A) liegt bereits 4 kcal/mol über dem globalen Minimum (vgl. Abb. X32) und trägt daher zur Population von Minimum A praktisch nichts bei. Insgesamt ist das FPP-Populationsverhältnis (B/A) bei 1000 K gleich 0.75. Daraus errechnet sich ein Energieunterschied von 0.6 kcal/mol. Dieser Energieunterschied liegt schon sehr nahe am Energieunterschied wie er sich aus der FGE-Map ergibt.

Insgesamt gesehen sind Hochtemperatur-MD-Simulationen<sup>1</sup> eine effektive Methode zur Konformationsanalyse von Oligosacchariden, vorausgesetzt man hat eine entsprechend leistungsfähige Analyse-Software, um die gewünschten Informationen aus den Trajektoriendaten zu gewinnen. Der wesentliche Vorteil der Methode beruht darauf, daß das Gesamtmolekül durch eine einzige MD-Simulation analysiert werden kann. Sämtliche sterischen Wechselwirkungen im Molekül sind berücksichtigt. Weiterhin nimmt die CPU-Zeit durch eine zunehmende Größe des Oligosaccharids nicht exponentiell zu wie bei Grid-Search-Methoden<sup>2</sup>.

- 
1. Die C-C- und C-O-Bindungen von Sialyl-Lewis<sup>X</sup> zeigen bei 1200K eine Standardabweichung von 0.06Å, die glycosidischen Bindungswinkel C-O-C fluktuieren durchschnittlich im Bereich von  $\pm 6^\circ$
  2. Bei Verwendung eines Cutoffs ( $\approx 10\text{\AA}$ ) für die nichtbindenden Wechselwirkungen ergibt sich eine annähernd lineare Zunahme der CPU-Zeit mit der Größe des Oligosaccharids.

### 3.3.2 Kombination von Grid-Search- und MD-Methoden zur Konformationsanalyse von Oligosacchariden (GCMD-Methode)

Nicht immer sind Hochtemperatur-MD-Simulationen geeignet, um bestimmte Effekte angemessen zu beschreiben. So wurde bereits gezeigt, daß die OH-Gruppen bei hohen Temperaturen keine fließenden Rotationsprofile mehr zeigen. Der Nachteil von MD-Simulationen bei 300 K ist, wie bereits erwähnt, daß extrem lange Simulationszeiten notwendig sind um Konformationsübergänge im  $\Phi/\Psi$ -Raum zu beobachten. MD-Simulationen bei 300K sind jedoch sehr gut geeignet um innerhalb eines  $\Phi/\Psi$ -Minimums die OH-Gruppen zu equilibrieren (vgl. Abschnitt 3.4.2). Um einen n-dimensionalen Konformationsraum vollständig systematisch abzusuchen, bietet sich die Grid-Search-Methode an. Unglücklicherweise ist die Lage der  $\Phi/\Psi$ -Minima je nach konkretem Verknüpfungstyp für die Zuckereinheiten sehr unterschiedlich, so daß man den Konformationsraum in sehr kleinen Schritten absuchen müßte, um alle Minima genau zu lokalisieren. Andererseits ist die Zahl der Minima pro glycosidische Bindung meistens deutlich geringer als zehn. Um alle wichtigen Minima zu finden, sollte es daher genügen, für jede glycosidische Bindung die beiden Torsionswinkel in  $120^\circ$ -Schritten zu rotieren und die so erzeugten neun Strukturen anschließend zu minimieren. Damit hat man zwei Probleme bereits beseitigt: erstens hat man den Konformationsraum weiträumig abgesucht und zweitens die genaue Lage der  $\Phi/\Psi$ -Minima durch die Minimierung ermittelt (GSCC-Methode). Die Eigenschaften der lokalen Minima (z.B. H-Brücken, Rotationsprofile der OH-Gruppen) werden im Anschluß durch eine MD-Simulation bei 300 K ermittelt. Die gesamte Sequenz GS-CC-MD wird als GCMD-Methode bezeichnet.

Die Methode wurde für Sialyl-Lewis<sup>X</sup> zur Analyse der lokalen Minima des Lewis<sup>X</sup>-Teils angewandt. Dazu wurden durch systematische Variation von GPhi/GPsi/FPhi/FPsi insgesamt 81 Strukturen erzeugt und minimiert. Aus den so erzeugten Strukturen wurden zuerst die mehrfach vorkommenden entfernt<sup>1</sup>, die verbleibenden 76 Strukturen wurden anschließend als Startstrukturen für 300 K- MD-Simulationen mit jeweils  $t_s = 100$  ps verwendet (Abb. X38). Insgesamt ergab sich damit eine Simulationsperiode der Länge 7.6 ns. Die aus den  $\Phi/\Psi$ -Populationsdaten berechneten FGE-Maps sind nicht mehr global gültig, da zwischen weit entfernten Minima (wahrscheinlich) kein Konformationsgleichgewicht erreicht wurde. Die FGE-Map gilt daher nur für die unmittelbare Umgebung des Minimums (lokale Gridgleichgewichte, vgl. Abschnitt 2.3). Die so erhaltenen, nur lokal gültigen, FGE-Maps<sup>2</sup> (Abb. X40, Abb. X41) beschreiben viel besser die Lage der lokalen Minima, die durch die HTMDCC-Methode erhalten wurden, als die Hochtemperatur FGE-Maps (Abb. X20, Abb. X21). Insbesondere ist nun Minimum C für Fuc $\alpha$ (1-3)GlcNAc in der Karte zu sehen.

Die Methode eignet sich also um sehr schnell alle lokalen Minima für nicht allzu große Oligosaccharide zu erzeugen (pro glycosidische Bindung wächst die Zahl der Startstrukturen maximal um den Faktor neun). Die lokalen Eigenschaften der Minima werden dann durch MD-Simulation analysiert. Da die MD-Simulation für jede Startstruktur mindestens 100 ps lang sein sollte, läßt sich die Methode praktisch bis zum Tetrasaccharid ( $t_s \approx 70$  ns) routinemäßig anwenden. Die auf diese Weise erzeugten Ensembles lassen sich, wie bei den HTMD-Simulationen bereits beschrieben, in  $\Phi/\Psi$ -Gruppen einteilen und die Eigenschaften der einzelnen Gruppen statistisch hinsichtlich relevanter Größen wie z.B. H-H-Abstände analysieren und vergleichen (siehe nächster Abschnitt).

1. Zur Identitätsprüfung werden ausgewählte Torsionswinkel (im vorliegenden Fall alle  $\Phi/\Psi$ -Winkel) gerastert (z.B.  $10^\circ$ ) und die Kombinationsgruppen der einzelnen so erzeugten Gridgruppen berechnet. Für jede Kombinationsgruppe wird die Struktur mit der geringsten Energie aus den Strukturarchiv extrahiert.
2. Aus den Daten der gesamten Simulation wird nur eine Map berechnet

### 3.3.3 Analyse von MD-Ensembles zur Unterstützung der Strukturaufklärung mittels NMR-Methoden

Die Analyse von interglycosidischen Transfer-NOEs ist eine der wichtigsten Methoden zur Strukturaufklärung von Oligosacchariden in Lösung [59,149,194,197]. Die relativen NOEs lassen sich unter Anwendung der ISPA-Näherung [198] in Abstandsbeschränkungen umwandeln, welche als Randbedingungen für die Generierung eines Strukturvorschlags mittels Computersimulationen verwendet werden können. Eine Methode, um herauszufinden, welche  $\Phi/\Psi$ -Kombinationen im Einklang mit den NOE-Messungen sind, ist die Generierung sogenannter Distance-Maps [97,192-194,199]. Dabei werden die  $\Phi/\Psi$ -Torsionswinkel systematisch variiert und die zugehörigen H-H-Abstände gemessen. Die Contourlinien der experimentell ermittelten unteren und oberen Abstandsgrenzen werden in  $\Phi/\Psi$ -Diagramme eingetragen.  $\Phi/\Psi$ -Kombinationen, die mehrere experimentelle Randbedingungen erfüllen, liegen im Schnittbereich der entsprechenden Contourlinien. Eine alternative Methode besteht in der Generierung einer großen Anzahl von Strukturen ( $10^6$ ) durch Metropolis-Monte-Carlo (MMC)-Simulation mit einem hohen Temperaturparameter (2000K) und anschließender Selektion derjenigen Konformationen, die alle Abstandsbeschränkungen erfüllen [197]. Ein analoges Verfahren kann auch auf Basis von Hochtemperatur-Simulationen realisiert werden: Aus dem Ensemble einer 1000K-MD-Simulation von Sialyl-Lewis<sup>X</sup> mit  $t_s = 10$  ns wurden diejenigen Strukturen extrahiert, die einer Liste von NOE-Randbedingungen genügen. Für den Lewis<sup>X</sup>-Teil wurden folgende NMR-Constraints zur Extraktion verwendet (vgl. Abb. X42): GH1-LH4=[2.2, 2.8], GH1-LH6R=[2.0, 3.0], FH1-LH3=[2.2, 2.8], FH5-GH2=[2.2, 2.8], FH6-GH2=[2.3, 3.3] [194]. Alle extrahierten Strukturen liegen für Gal $\beta$ (1-4)GlcNAc und Fuc $\alpha$ (1-3)GlcNAc im  $\Phi/\Psi$ -Bereich um ( $45^\circ/20^\circ$ ) (Abb. X46). Die gemessenen NOEs können also durch eine Konformation erklärt werden, die dem globalen Minimum (LX-Code = 3, vgl. Abb. X32) entspricht. Für Neu5Ac $\alpha$ (2-3)Gal wurden anhand von ROESY-Spektren ROEs für SH8-GH3 und für SH3<sub>ax</sub>-GH3 gefunden [149]. Dieser Befund kann nicht durch eine einzelne Konformation erklärt werden, da nicht beide Abstände gleichzeitig kleiner als 3.3 Å sein können (vgl. Abb. X47). Verwendet man einmal die SH8-GH3-Randbedingung [1.8, 3.3] zur Extraktion der Strukturen aus dem HTMD-Ensemble, so erhält man ausschließlich Strukturen, die in den  $\Phi/\Psi$ -Bereich um ( $-160^\circ/-20^\circ$ ) fallen. Bei Anwendung des SH3<sub>ax</sub>-GH3-Kriteriums [1.8, 3.3] erhält man ausschließlich Strukturen mit einem  $\Phi/\Psi$ -Bereich um ( $-80^\circ/-10^\circ$ ).

Aus den Daten der HTMD-Simulationen lassen sich noch weitere hilfreiche Informationen gewinnen: Aus den Häufigkeitsverteilungen (Abb. X43) kann man, wie bereits bei Maltose beschrieben, abschätzen, für welche H-H-Abstände überhaupt NOEs auftreten können. Gruppiert man die Daten anhand der Minima der FGE-Maps (vgl. Abb. X32) so lassen sich Voraussagen über potentielle NOEs von lokalen Minima machen. Für die in Abschnitt 3.3.1 definierten lokalen Minima für Lewis<sup>X</sup> (LX-Code) läßt sich so die Voraussage machen, daß das Vorliegen einer Struktur, bei der sich beide glycosidischen Bindungen im Minimum B (vgl. Abb. X30) befinden, sich in einem deutlichen NOE für Gal(H1)-Fuc(H1) zeigen müßte (Abb. X44). Zwar geht aus den oben aufgeführten NOE-Constraints eindeutig hervor, daß das globale Minimum (LX-Code=3) in Lösung vorliegt (GH2-FH5 ist selektiv für das globale Minimum), das Vorhandensein des Minimums (LX-Code=3), bei dem sich Fuc $\alpha$ (1-3)GlcNAc im Minimum B (vgl. Abb. X31) befindet, kann anhand der in Abb. X44 aufgeführten H-H-Abstände jedoch nicht ausgeschlossen werden. Keiner der gemessenen H-H-Abstände läßt selektiv auf das Vorhandensein des Minimums (LX-Code=3; Minimum B für Fuc $\alpha$ (1-3)GlcNAc) schließen. Um dessen Existenz auszuschließen ist ein diskriminierender NOE jedoch notwendig, denn nur wenn ein für das Minimum (LX-Code=3) selektiver NOE im Spektrum fehlt, kann

die Existenz dieses Minimums mit großer Wahrscheinlichkeit ausgeschlossen werden. Hinweise, daß eine Konformation, bei der  $\text{Fuc}\alpha(1-3)\text{GlcNAc}$  eine Orientierung wie in Minimum B einnimmt, in Lösung vorliegen könnte, finden sich in [200].

Alternativ zur Generierung des MD-Ensembles durch eine HTMD-Simulation kann auch die in Abschnitt 3.3.2 vorgestellte GCMD-Methode verwendet werden. Der Vorteil eines 300K-Ensembles liegt in der stabileren Rotamerenverteilung der OH-Gruppen. Dies ist vor allem dann wichtig, wenn Transfer-NOE-Messungen in DMSO oder unterkühltem Wasser ( $-15^\circ$ ) durchgeführt wurden. Unter diesen Bedingungen können NOEs zu OH-Protonen beobachtet werden. Deren Wert hängt jedoch von der Orientierung der OH-Gruppe ab. Da sich die OH-Gruppen innerhalb eines lokalen  $\Phi/\Psi/\omega$ -Minimums bei 300 K equilibrieren lassen (siehe Abschnitt 3.4.2) sind die 300K-GCMD-Ensembles sehr gut geeignet um z.B. die Interpretation von NOESY-Spektren aus NMR-Messungen in DMSO zu erleichtern.

Die Berechnung der H-Brücken-Matrix (Abb. X44) aus dem MD-Ensemble liefert weitere wertvolle Informationen über intraresiduale und interresiduale H-Brücken in übersichtlicher Form. Die für die Strukturaufklärung wichtigen interresidualen H-Brücken zeigen sich in der gewählten Form der Darstellung analog den „Kreuzpeaks“ bei 2D-NMR-Spektren außerhalb der Matrixdiagonalen.

## 3.4 Analyse der Rotamerengleichgewichte

### 3.4.1 Einleitung

Während es im vorangegangenen Abschnitt hauptsächlich darum ging, lokale Minima im  $\Phi/\Psi$ -Raum aufzuspüren und ihre lokalen Eigenschaften zu untersuchen, ist das Hauptziel dieses Abschnitts zu untersuchen wie lange es dauert, bis das Konformationsgleichgewicht zwischen den Konformationen („Gridgruppen“) eingestellt ist. Dabei wird hierarchisch zwischen OH-, Hydroxymethyl- und  $\Phi/\Psi$ -Rotameren unterschieden, da diese in unterschiedlichen Zeitskalen equilibrieren. Unter der Voraussetzung, daß das Rotamerengleichgewicht eingestellt ist, lassen sich aus den relativen Populationen die Energieunterschiede mit Hilfe der Boltzmann-Formel (Gleichung 14) berechnen. Die in Abschnitt 2.3 und Abschnitt 2.4 gemachten Erläuterungen bilden die Grundlagen zum Verständnis dieses Abschnitts. Als Modellsystem wurde  $\beta$ -Maltose verwendet.

### 3.4.2 Rotationsprofile, Rotamerenpopulation und Gleichgewichtseinstellung der OH-Gruppen<sup>1</sup>

Zur Bestimmung der Simulationszeit ( $t_s$ ), die notwendig ist, um das Rotamerengleichgewicht ( $-60^\circ$ ,  $+60^\circ$ ,  $180^\circ$ ) einzelner OH-Gruppen bei 300 K einzustellen, wurde eine MD-Simulation mit  $t_s = 3$  ns durchgeführt (CVFF,  $\epsilon=4$ ,  $\Delta t_s = 50$  fs). Ausgewählte Trajektorien von Torsionswinkeln sind in Abb. X48 bis Abb. X50 dargestellt.

Die Rotameren der OH-Gruppen wurden wie folgt definiert: „ $-60^\circ$ “ =  $[-90^\circ, -30^\circ]$ , „ $+60^\circ$ “ =  $[30^\circ, 90^\circ]$ , „ $180^\circ$ “ =  $[150^\circ, 210^\circ]$  (vgl. auch Abb. X53).

Die berechnete Rotamerenpopulation und die Gleichgewichtsparameter EQ für die einzelnen OH-Gruppen sind in Abb. X51 dargestellt<sup>2</sup>. Aus dem Wert von EQ (vgl. Abschnitt 2.4.1) erhält man einen ersten Hinweis, daß für die primären OH-Gruppen (G6OH und L6OH) auch nach 3 ns das Konformationsgleichgewicht evtl. nicht eingestellt ist. Auffällig unterschiedlich ist die Rotamerenpo-

1. Definition der Torsionswinkel für die OH-Gruppen:  $\text{C}_{n-1}-\text{C}_n-\text{O}-\text{H}$  (Ausnahme L1OH:  $\text{O}5-\text{C}1-\text{O}-\text{H}$ )

2. Die Prozentzahlen beziehen sich auf  $n_{\text{eff}} = n - n_{\text{gap}}$

pulation der benachbarten Torsionswinkel G2OH und G3OH. Während bei G3OH praktisch Gleichverteilung zwischen den OH-Rotameren vorliegt, ist bei G2OH der Rotameren-Zustand  $-60^\circ$  zu mehr als 50% populiert. Dieser muß daher gegenüber den beiden anderen Rotameren eine deutliche Stabilisierung erfahren. Der maximale Energieunterschied zwischen zwei OH-Rotameren liegt bei 0.85 kcal/mol (Abb. X52).

Die Gleichgewichtsanalyse des Rotamerengleichgewichts für G3OH (siehe Abb. X55 bis Abb. X57) zeigt, daß sich das Gleichgewicht nach 500 ps eingestellt hat ( $EQ1 < 5$ ). Nach 500 ps sind bereits aus allen drei Rotameren mehr als 200 Übergänge erfolgt (Abb. X58), welches sich in einem kleinen Wert für EQ2 zeigt. Auch der weitere Verlauf von EQ1 zeigt, daß das Rotamerengleichgewicht für G3OH sehr stabil ist (Abb. X56). Nach einer Simulationszeit von  $t_s = 1.2$  ns sinkt der Wert für EQ1 sogar unter 2 und bleibt dort für den Rest der Simulationszeit.

Das aus der Häufigkeitsverteilung (Abb. X53) für G3OH berechnete Profil der Freien Gridgruppen-Energie (Abb. X54) (vgl. Abschnitt 2.3.4) zeigt zwei Minima nahezu identischer Energie bei  $-50^\circ$  und  $+175^\circ$ . Das dritte Minimum liegt bei  $+60^\circ$  und besitzt eine um 0.22 kcal/mol höhere Energie. Die Rotationsbarrieren liegen bei 0.85 kcal/mol bzw. 1.37 kcal/mol. Die Absenkung der beiden globalen Minima und der Energiebarrieren bei  $0^\circ$  und  $120^\circ$  kann mit einer Coulomb-Anziehung zwischen dem OH3-Proton und dem O2-Atom bzw. O4-Atom erklärt werden. Um sicher zu gehen, daß das Energieprofil auch wirklich auf einem globalen oder eingeschränkten Gridgruppengleichgewicht beruht, müßte für alle 36 G3OH-Gridgruppen auf denen es basiert, die Einstellung des Populations-Gleichgewichts geprüft werden. Ausgehend von dem Ergebnis der Analyse des Rotameren-Gleichgewichts und der Tatsache, daß die Gridgruppe mit der geringsten Population ( $P(-120^\circ) = 0.6\%$ ) immerhin durch 300 Frames repräsentiert wird, kann angenommen werden, daß das Populations-Gleichgewicht über den gesamten Wertebereich von G3OH eingestellt ist. Die Detailanalyse der Rotamerenübergänge (Abb. X60) bestätigt den Verlauf des Energieprofils. Es erfolgen ungefähr dreimal soviel Übergänge über die Energiebarrieren bei  $0^\circ$  und  $120^\circ$  wie über die Barriere bei  $-120^\circ$ . Ca. 10% der Rotameren-Lebensdauern sind gleich dem Sampling-Intervall von 50 fs (Abb. X59) (siehe hierzu Abschnitt 3.5).

Das Profil der Freien Gridgruppen-Energie (Abb. X54) für den Torsionswinkel G2OH zeigt die Stabilisierung des Zustands  $-60^\circ$  um 0.3 kcal/mol. Ursache hierfür ist, neben der Wechselwirkung mit dem glycosidischen O-Atom, die zusätzlich mögliche interresiduale Wechselwirkung mit dem O3-Atom der reduzierenden Glucose-Gruppierung. Diese interresiduale Wechselwirkung ist wahrscheinlich auch der Grund für die Schwankungen in der Rotamerenverteilung von G2OH (Abb. X61) und L3OH (Abb. X62). Es ist anzunehmen, daß die Wechselwirkungen zwischen G2OH und L3OH von  $\Phi/\Psi$  abhängen (vgl. Abschnitt 3.3.1).

Die Abhängigkeit der Rotamerenverteilung von der Simulationszeit für die primären OH-Gruppen ist in Abb. X63 und Abb. X64 gezeigt. Im Bereich bis  $t_s = 500$  ps findet man den üblichen Verlauf der Einstellung des Rotameren-Gleichgewichts. Im Bereich zwischen  $t_s = 500$  ps und  $t_s = 800$  ps wird das Rotamerengleichgewicht für G6OH und L6OH jedoch offensichtlich gestört. Ein Blick auf die Trajektorie von  $\omega_1$  (G56O) und  $\omega_2$  (L56O) (Abb. X49) offenbart die Ursache der Störung: Nach ungefähr 700 ps Simulationszeit erfolgt ein Übergang für G56O von tg nach gg. Der Torsionswinkel L56O zeigt im Bereich um 600 ps einen Übergang von gt nach tg, welcher nach weiteren 400 ps wieder rückgängig gemacht wird. Insgesamt finden für G56O innerhalb von 3 ns sechs Rotamerenübergänge statt, für L56O sind es deren zwölf. Die Rotamerengleichgewichte der primären OH-Gruppen sind also von der Einstellung des Hydroxymethylgleichgewichts abhängig, welches sich aber viel langsamer einstellt. Aussagekräftige Rotationsprofile bzw. Rotamerenverteilungen können für die primären OH-Gruppen nur erhalten werden, indem man entweder die Simulationszeit solange verlängert, bis sich das Hydroxymethylgleichgewicht eingestellt hat, oder man verhindert durch Anwendung von

zusätzlichen Potentialen (Restraints), daß ein Rotamerenübergang der Hydroxymethylgruppen erfolgt. Im letzteren Fall erhält man für die primäre OH-Gruppe ein Rotationsprofil mit der Randbedingung einer „eingefrorenen“ Hydroxymethyl-Orientierung („eingeschränktes Gridgruppengleichgewicht“). Ein Vergleich der G6OH-Energieprofile mit und ohne eingestelltem G56O-Rotamerengleichgewicht ist in Abb. X82 gezeigt. Wegen der geringen Population des gt-Zustands für G56O bei der 3ns-MD-Simulation ist die Einstellung  $-60^\circ$  für G6OH, die für den gt-Zustand die stabilste ist, im Ensemble unterrepräsentiert. Dies führt im vorliegenden Fall zu einer starken Erhöhung der relativen Energie dieses Zustandes, so daß der G6OH( $+60^\circ$ )-Zustand zum energieärmsten wird.

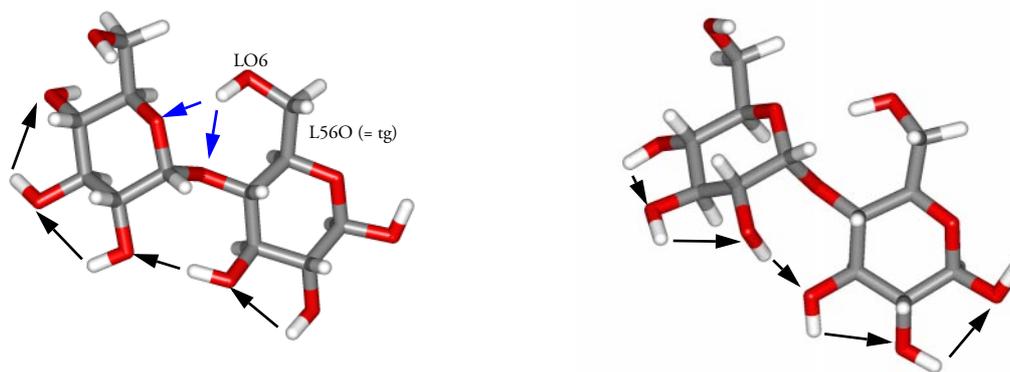
Anhand der wenigen Übergänge, die man für die Hydroxymethylgruppen bei 300 K innerhalb von  $t_s = 3$  ns beobachtet, läßt sich nur eine sehr grobe Abschätzung für die Dauer der Einstellung des Hydroxymethylgleichgewichts machen: Unter der Annahme, daß für jedes der drei Rotameren 100 Übergänge erfolgen sollen und die mittlere Lebensdauer für jedes Rotamer bei 500 ps liegt, ergibt sich für die Zeit, die zur Einstellung des Rotamerengleichgewichts bei 300K benötigt wird, mindestens 150 ns. Die Überprüfung dieser Abschätzung erfolgte durch eine 300K-MD-Simulation mit  $t_s = 2$   $\mu$ s. Eine detaillierte Analyse der Gleichgewichtseinstellung (siehe Abschnitt 3.4.4) ergab, daß mindestens 400 ns Simulationszeit notwendig sind, bis sich das Rotamerengleichgewicht der Hydroxymethylgruppen bei 300 K eingestellt hat.

### 3.4.3 „Nachbargruppeneffekte“

In diesem Abschnitt soll kurz darauf eingegangen werden, wie sich die Orientierung einer OH-Gruppe auf die Orientierung der benachbarten OH-Gruppen auswirkt.

Abb. X69 zeigt Korrelations-Contourplots der Freien Gridgruppen-Energie für verschiedene Kombinationen von Torsionswinkeln zweier benachbarter (und auch einiger weiter entfernter) OH-Gruppen. Es zeigt sich eindeutig, daß Information (Stellung einer bestimmten OH-Gruppe) nicht nur zwischen unmittelbar benachbarten OH-Gruppen desselben Ringes, sondern auch über die Ringgrenzen hinweg zum benachbarten Ring „transportiert“ wird. So bewirkt ein Torsionswinkel  $G2OH = 180^\circ$  mit hoher Wahrscheinlichkeit nicht nur eine Orientierung  $G3OH = 180^\circ$  und  $G4OH = 180^\circ$ , sondern auch  $L3OH = 180^\circ$  und  $L2OH = 180^\circ$ . Die OH-Gruppen sind damit gleichsinnig im Uhrzeigersinn orientiert (Abb. 27, links). Eine entsprechende Orientierung gegen den Uhrzeigersinn (Abb. 27, rechts) ist auch möglich, diese besitzt ebenfalls eine hohe Wahrscheinlichkeit, da eine derartige Anordnung die Zahl der intramolekularen H-Brücken maximiert und daher einen stark negativen Coulomb-Term aufweist. Die Orientierung der OH-Gruppen in einer Richtung sollte mit abnehmender Dielektrizitätskonstante immer stärker dominieren [133,178,201]. Die dirigierende Wirkung der G2OH-Gruppe läßt sich auch durch einfache Häufigkeitsplots nachweisen (Abb. X70 und Abb. X71). Dazu werden die Frames anhand des Torsionswinkels G2OH in drei Gridgruppen aufgeteilt:  $-60^\circ$  [ $-120^\circ, 0^\circ$ ],  $+60^\circ$  [ $0^\circ, 120^\circ$ ],  $180^\circ$  [ $120^\circ, 240^\circ$ ]. Für jede Gridgruppe wird ein Populationsprofil der benachbarten Torsionswinkel G3OH und L3OH berechnet. Es zeigt sich sehr deutlich, daß die Orientierung von G2OH Einfluß auf die Orientierung der Nachbargruppen hat.

Kombiniert man jeweils die drei Rotameren-Gridgruppen der Torsionswinkel G2OH, G3OH, G4OH, L2OH und L3OH zu neuen 5D-Kombinations-Gridgruppen (GG-Code), so ergeben sich theoretisch 243 verschiedene Kombinationen. Die Analyse der Daten einer 300K-MD-Simulation von  $\beta$ -Maltose (CVFF,  $\epsilon=4$ ,  $t_s = 1$  ns,  $\Delta t_s = 20$  fs) ergab, daß 230 Kombinationen innerhalb der Simulationsperiode realisiert werden (Abb. X72). Es hat sich herausgestellt, daß die Anordnung im Uhrzeigersinn (GG-Code = 17 und GG-Code = 50) und die Anordnung entgegen dem Uhrzeigersinn (GG-



**Abb. 27:**  $\beta$ -Maltose: Orientierung der OH-Gruppen im Uhrzeigersinn (links) und entgegen den Uhrzeigersinn (rechts). Mögliche H-Brücken von LO6 (Donor) wenn der Torsionswinkel L56O die tg-Stellung einnimmt.

Code = 71) am häufigsten populiert sind. Beide zusammen ergeben eine Population von 14%, d.h. 86% der Strukturen zeigen keine Anordnung bei der alle fünf OH-Gruppen in einer Richtung orientiert sind.

### 3.4.4 Rotationsprofile, Rotamerenpopulation und Gleichgewichtseinstellung der Hydroxymethylgruppen<sup>1</sup>

Basis der Analyse bildet eine Langzeit-MD-Simulation von  $\beta$ -Maltose bei 300 K (CVFF,  $\epsilon = 4$ ,  $t_s = 2 \mu\text{s}$ ,  $\Delta t_s = 20 \text{ ps}$ ). Die Trajektorien der  $\omega$ -Torsionswinkel (G56O, L56O) sind in Abb. X74 dargestellt. Die Rotameren der Hydroxymethylgruppen wurden wie folgt definiert: „gg“=[-115,-5], „gt“=[5,115] und „tg“=[125,235] (vgl. Abb. X75). Trägt man die relative Population der Konformationen ( $100 \cdot x_i$ ) wie üblich in Abhängigkeit von der Simulationszeit auf (Abb. X77), so sieht man deutlich die starken Änderungen innerhalb der ersten 250 ns (entspricht 12500 Frames). Im Bereich zwischen 250 ns und 500 ns bleibt die Population aller Rotamere praktisch konstant. Anschließend zeigt sich innerhalb der nächsten ca. 600 ns für gg und gt eine leichte Instabilität des Gleichgewichts in Form einer reversiblen Drift, woraus für diesen Zeitbereich ein Fehler von ca. 25% für diejenigen Gleichgewichtskonstanten ergibt, an denen gg beteiligt ist (siehe Abb. X78).

Abb. X79 zeigt die Änderung der Gleichgewichtsparameter EQ in Abhängigkeit von der Simulationszeit. Sowohl EQ1 als auch EQ2 zeigen besonders deutlich das Erreichen des Gleichgewichtszustandes bei 300-400 ns durch einen starken Abfall von  $EQ^{mean}$  auf einen Wert kleiner 5. Ab einer Simulationszeit größer 600 ns ist auch die Konformationsübergangsbedingung so gut erfüllt, daß  $EQ2^{mean} < EQ1^{mean}$  wird. EQ3 zeigt die erwartete „Unruhe“ im Drift-Term selbst noch bei langen Simulationszeiten. EQ4 nimmt zwar mit zunehmende Simulationszeit ebenfalls ab, das Erreichen des Gleichgewichtszustandes ist jedoch nicht so deutlich sichtbar wie bei den nach relativen Werten berechneten EQ-Parametern. Dies wird auch bei Betrachtung der einzelnen Terme (Abb. X80) deutlich.

Als Rotamerenpopulation der Hydroxymethylgruppen nach Simulationszeit von  $t_s = 2 \mu\text{s}$  unter den beschriebenen Simulationsbedingungen erhält man: G56O [gg(19%), gt(59%), tg(22%)]; L56O [gg(30%), gt(53%), tg(17%)]

1. Definition der Torsionswinkel für die Hydroxymethylgruppen: O5-C5-C6-O6

Das für G56O berechnete Profil der Freien Gridgruppen-Energie (Abb. X76) ist in ausgezeichneter Übereinstimmung mit dem aus Grid-Search-Daten berechneten Profil (vgl. Abschnitt 4.3.5). Lediglich die Übergangszustände sind nicht ausreichend oft populiert. Mit einer relativen Häufigkeit  $< 0.01\%$  (Intervallbreite =  $10^\circ$ ) werden diese selbst bei  $10^5$  gespeicherten Frames durch weniger als 10 Frames populiert. Hier zeigt sich deutlich, daß MD-Simulationen bei 300K nicht geeignet sind um, mit den derzeitigen technischen Möglichkeiten, Zustände mit einer relativen Energie  $> 4$  kcal/mol in statistisch ausreichender Anzahl zu populieren.

### 3.4.5 Rotationsprofile, Rotamerenpopulation und Gleichgewichtseinstellung der glycosidischen Torsionswinkel

Basis der Analyse bildet wiederum die Langzeit-MD-Simulation von  $\beta$ -Maltose bei 300 K (CVFF,  $\epsilon = 4$ ,  $t_s = 2 \mu\text{s}$ ,  $\Delta t_s = 20$  ps). Die Trajektorien der  $\Phi$ - und  $\Psi$ -Winkel sind in Abb. X73 dargestellt. Es ist offensichtlich, daß die Rotationen um  $\Phi$  und  $\Psi$  nicht mehr einfach durch ein dreizähliges Potential beschrieben werden kann. Ein dreizähliges Potential bildet zwar sicherlich noch die Basis des Energieprofils ( $sp^3$ ), der entgültige Verlauf des Profils in Abhängigkeit vom Torsionswinkel wird jedoch durch übergeordnete Faktoren wie Destabilisierung bestimmter Torsionswinkelwerte durch räumliche Überlappung von Atomen oder Absenkung der Energie durch Ausbildung von interresidualen H-Brücken bestimmt.

Aus dem Energieprofil der Freien Gridgruppen-Energie (Abb. X81) kann man erkennen, daß bei 300 K nicht der gesamte Torsionswinkelbereich von  $\Phi$  und  $\Psi$  populiert ist. Das Energieprofil für den Torsionswinkel besitzt ein Minimum bei  $\Phi = -20^\circ$  und zwei Wendepunkte bei  $\Phi = -60^\circ$  und  $\Phi = +20^\circ$ . Für  $\Psi$  erhält man zwei Minima bei  $\Psi = -30^\circ$  und  $\Psi = +190^\circ$  und einen Wendepunkt bei  $\Psi = +10^\circ$ . Die Rotameren für  $\Psi$  wurden daher wie folgt definiert: „syn“ =  $[-90, +60]$ , „anti“ =  $[+150, +230]$ .

Innerhalb der Simulationsperiode von  $t_s = 2 \mu\text{s}$  finden bei 300 K genau vier Übergänge von der syn- in die anti-Konformation statt. Die Lebensdauern für die anti-Konformationen sind: 1660 ps, 340 ps, 560 ps, 4960 ps. Der erste Übergang erfolgt nach 111 ns Simulationszeit, die folgenden nach weiteren 89 ns, 525 ns und 661 ns Simulationszeit. Wegen der großen mittleren Lebensdauer der syn-Konformation kann man davon ausgehen, daß es derzeit technisch nicht möglich ist, das syn/anti-Gleichgewicht mit einer MD-Simulation bei 300K einzustellen. Das Energieprofil (Abb. X81) ist daher wahrscheinlich für  $\Psi$  nur im Bereich der syn-Konformation korrekt. Die Gleichgewichtsanalyse (Abb. X84, Abb. X85) zeigt abrupte Änderungen der Gleichgewichtskonstanten beim Übergang  $\text{syn} \leftrightarrow \text{anti}$ . Der Gleichgewichtsparameter EQ1 bleibt deutlich größer als 5 und zeigt daher den „Nichtgleichgewichtszustand“ richtig an, EQ2 ist außerhalb des erlaubten Wertebereichs  $[0, 20]$ . EQ4 ist dagegen vollkommen ungeeignet, um Aussagen über die syn/anti-Gleichgewichtseinstellung zu machen, da hier Änderungen unter 1% auftreten. Diese kleinen Änderungen ziehen jedoch deutliche Fluktuationen der Gleichgewichtskonstanten nach sich, da die anti-Konformation nur zu ca. 0.5% populiert ist.

Berechnet man die korrelierte Häufigkeitsverteilung für  $\Phi$  und  $\Psi$  (Abb. X83) kann man, wie bereits in Abschnitt 3.3.1 gefunden, erkennen, daß innerhalb der syn-Konformation eine Abhängigkeit zwischen  $\Phi$  und  $\Psi$  besteht. Das Häufigkeitsmaximum befindet sich bei  $\Phi = -25^\circ / \Psi = -25^\circ$ . Die Population nimmt nicht symmetrisch ausgehend von diesem Maximum ab, es gibt zwei „Ausläufer“ erhöhter Population mit Zentren bei  $\Phi = -60^\circ / \Psi = -40^\circ$  und  $\Phi = 10^\circ / \Psi = 10^\circ$ .

Durch MD-Simulationen bei 300K kann also für Maltose selbst mit Simulationszeiten von  $2 \mu\text{s}$  kein Konformationsgleichgewicht im  $\Phi/\Psi$ -Raum eingestellt werden, es kann lediglich das lokale Gleichgewicht innerhalb der syn-Konformation analysiert werden.

### 3.4.6 Reproduzierbarkeit von Rotamerengleichgewichten

Um zu überprüfen, welche Simulationszeit notwendig ist, bis sich ein Rotamerenverhältnis der Hydroxymethylgruppen innerhalb bestimmter Fehlergrenzen reproduzierbar eingestellt hat, wurden aus den 100000 Datensätzen der 2  $\mu$ s-MD-Simulation von Maltose einzelne MD-Trajektorien unterschiedlich langer Simulationszeit erzeugt. Im einzelnen wurden aus den Daten 4 Trajektorien à 500 ns, 5 Trajektorien à 400 ns, 8 Trajektorien à 250 ns, 10 Trajektorien à 200 ns, 20 Trajektorien à 100 ns, 40 Trajektorien à 50 ns und 100 Trajektorien à 20 ns erhalten. Für alle 187 Einzeltrajektorien wurde die Population der Rotameren der Hydroxymethylgruppe am reduzierenden Ende (L56O) berechnet. Die Abhängigkeit des Gleichgewichtsparameters EQ1 von der Simulationszeit zeigt Abb. X90.

Für Simulationsdauern von 20 ns ergibt sich für die Population der Rotameren ein eindeutiges Bild (siehe Abb. X91, Abb. X92, Abb. X93): die Gleichgewichtskonstanten  $K(\text{gg}/\text{gt})$  und  $K(\text{gg}/\text{tg})$  variieren über einen Bereich von drei Größenordnungen und sind daher durch Einzelsimulationen der Länge 20 ns definitiv nicht sicher reproduzierbar. Wenn man bedenkt, daß Simulationsdauern von 20 ns mehr als doppelt so lang sind, als die meisten, die bis jetzt in der Literatur beschriebenen MD-Trajektorien von Kohlenhydraten [149,152], so drängt sich die Frage auf, inwieweit die publizierten Populationsverteilungen, die meist „in Übereinstimmung mit dem Experiment“ sind, überhaupt einen theoretischen Wert besitzen. Es bleibt unklar, wie die fast durchweg gute Übereinstimmung mit dem Experiment zustandekommt. Es sind folgende Möglichkeiten denkbar: Erstens, es wird von einer experimentell gefundenen Start-Struktur ausgegangen und die Konformation ändert sich während der MD-Simulation nicht grundlegend, zweitens, das Ergebnis ist wirklich zufällig „richtig“ und drittens, es werden tendenziell nur Simulationsergebnisse publiziert, wenn diese in Übereinstimmung mit dem Experiment sind.

Für das CVFF-Kraftfeld ergibt sich bei 300 K folgendes Ergebnis: Die Reihenfolge der Rotamerenpopulation ( $\text{tg} < \text{gg} < \text{gt}$ ) kann erst für Simulationszeiten größer als 200 ns einigermaßen sicher reproduziert werden.

Aus den vier 300K-MD-Simulationen mit  $t_s = 500$  ns (siehe auch Abb. X86 bis Abb. X88) erhält man für die Energiedifferenzen der L56O-Rotameren:  $\Delta E(\text{gt}-\text{tg}) = 0.70 \pm 0.03$  kcal/mol,  $\Delta E(\text{gg}-\text{gt}) = 0.35 \pm 0.15$  kcal/mol,  $\Delta E(\text{tg}-\text{gg}) = 0.35 \pm 0.15$  kcal/mol. Energiedifferenzen mit Beteiligung der gg-Konformation lassen sich mit einer deutlich schlechteren Genauigkeit bestimmen als für das gt/tg-Gleichgewicht (vgl. Abb. X94).

Insgesamt kann davon ausgegangen werden, daß mit Erreichen des Rotamerengleichgewichtszustands ( $\text{EQ1} < 5$ ) die Energiedifferenzen ungefähr mit einer Genauigkeit von 0.1 kcal/mol bestimmt werden können. Eine entsprechende Analyse für die OH-Gruppen bestätigt dies: mit Erreichen des Gleichgewichtszustands bei 500 ps ist die Standardabweichung für die Energiedifferenz der Rotameren auch dort 0.1 kcal/mol (vgl. Abb. X95).

### 3.4.7 Einfluß der Simulationstemperatur auf die Einstellung der Rotamerengleichgewichte

Wie in den vorangegangenen Abschnitten gezeigt, sind MD-Simulationen bei Raumtemperatur nicht geeignet, um das Konformationsgleichgewicht für die  $\Phi$ ,  $\Psi$  und  $\omega$ -Winkel mit vertretbarem technischen Aufwand zu erreichen. Problematisch sind hierbei die bei diesen Winkeln vorhandenen Rotationsbarrieren, die größer als 3 kcal/mol sind. Es gibt zwei Möglichkeiten um dieses Problem zu umgehen: Erstens durch Anlegen eines negativen Zusatzpotentials die Energiebarrieren abzusenken (Umbrella Sampling) [157,202] oder man erhöht die Simulationstemperatur und gibt dem System dadurch mehr kinetische Energie, die es ihm erlaubt, die Barrieren schneller zu überwinden.

Zur Analyse, welche Erhöhung der Simulationstemperatur notwendig ist, um das Konformationsgleichgewicht für die  $\Phi$ ,  $\Psi$  und  $\omega$ -Winkel innerhalb einer für Routine-MD-Simulationen akzeptablen Simulationszeit ( $< 20$  ns) zu erreichen, wurden folgende MD-Simulation von  $\beta$ -Maltose (CVFF,  $\epsilon = 4$ ) durchgeführt:

400 K: Simulationszeit=300 ns, Sampling-Intervall=3 ps

500 K: Simulationszeit=100 ns, Sampling-Intervall=1 ps

600 K: Simulationszeit=30 ns, Sampling-Intervall=0.5 ps

Entscheidendes Kriterium für den Einsatz von „Hochtemperatur“-MD-Simulationen zur Beschreibung von Konformationsgleichgewichten ist, daß sich das System weiterhin „physikalisch im Sinne des zugrundeliegenden Modells“ verhält, auch wenn im biologischen Sinne „unphysikalische“ Simulationsbedingungen zum Einsatz kommen. Als Test für die Konsistenz des Modells wurde der Verlauf der Energie-Profile der Freien Gridgruppen-Energie gewählt. Diese sollten in erster Näherung temperaturunabhängig sein.<sup>1</sup>

### Hydroxymethylgruppen

Die aus den Häufigkeitsverteilungen (Abb. X98) der  $\omega$ -Winkel berechneten Energieprofile (Abb. X99, Abb. X100) zeigen eine ausgezeichnete Temperaturunabhängigkeit. Minimale Abweichungen (0.2 kcal/mol) findet man im gg-Zustand, etwas größer (0.4 kcal/mol) sind die Abweichungen in den Übergangszuständen. Es dürfte sich hier wahrscheinlich wegen der hohen Übergangsbarrieren um ein Sampling-Problem<sup>2</sup> und nicht um ein methodisches Problem handeln.

Die Gleichgewichtsanalyse der Hydroxymethylgruppen (Abb. X96, Abb. X97) ergab folgende Simulationszeiten, die zur Einstellung des Rotamerengleichgewichts notwendig sind: 400 K : 40-100 ns; 500K: 10-50 ns; 600 K: 5-10 ns (zum Vergleich 300K: 400 - 500 ns).

### OH-Gruppen

Daß sich bei diesen hohen Temperaturen und langen Simulationszeiten das Rotamerengleichgewicht der OH-Gruppen eingestellt hat, steht außer Zweifel (vgl. Abschnitt 3.4.2). Entscheidender für die OH-Gruppen ist die Frage nach der „Stabilität“ der Energie-Profile und ob H-Brücken-Effekte auftreten. Die aus den Simulationen erhaltenen FGE-Profile (Abb. X101 bis Abb. X107) zeigen abhängig von der betrachteten OH-Gruppe eine Temperaturunabhängigkeit nicht über den gesamten Wertebereich des Torsionswinkels. Eine auffällige Stabilisierung bei niedriger Temperatur tritt beim Torsionswinkel L3OH im Bereich um  $-60^\circ$  (Abb. X101) auf ( $-0.3$  kcal/mol); etwas weniger ausgeprägt ist die Stabilisierung bei  $180^\circ$ . Bei der benachbarten OH-Gruppe (L2OH) tritt ein analoger Effekt bei  $+60^\circ$  auf. Sowohl L1OH wie auch G2OH zeigen keinen Temperatur-Effekt, die bevorzugte Orientierung dieser Gruppen paßt jedoch zu einer Anordnung der OH-Gruppen gegen den Uhrzeigersinn, wie sie in Abb. 27 rechts dargestellt ist. Eine Erklärung für diesen Befund wäre, daß sich bei höheren Temperaturen die Anordnung aller sekundären OH-Gruppen gegen den Uhrzeigersinn wegen der erhöhten kinetische Energie der Atome nur sehr selten einstellt und daher die besondere Stabilisierung, die diese Anordnung darstellt, nicht mehr zum Tragen kommt.

1. Eine Temperaturabhängigkeit ist zu erwarten, wenn sich bei einer Nichtgridvariablen das eingeschränkte Konformationsgleichgewicht (z.B. durch Population neuer Zustände) ändert oder wenn ein temperaturabhängiger physikalischer Prozess (z.B. H-Brückenbindung) bestimmend für das Profil ist.
2. Zu wenige Strukturen populieren die Übergangszustände: von 100000 gespeicherten Frames populieren bei 400 K nach Boltzmann nur sieben Frames eine Gridgruppe der mittleren relativen Energie von 5 kcal/mol (Gridintervall =  $10^\circ$ )

## Glycosidische Torsionswinkel

Sowohl für  $\Phi$  als auch für  $\Psi$  erhält man FGE-Profile (Abb. X108, Abb. X109), die über weite Bereiche unabhängig von der Temperatur sind. Interessanterweise unterscheidet sich die Energie der anti-Konformation aus dem 300 K-Profil von  $\Psi$  nicht wesentlich von dem Energiewert, den man aus den Hochtemperatur-FGE-Profilen erhält, obwohl bei der 300K-MD-Simulation definitiv kein syn/anti-Konformationsgleichgewicht für  $\Psi$  erreicht wurde (vgl. Abschnitt 3.4.5). Da die Energie der anti-Konformation im 300K-Profil niedriger ist, kann angenommen werden, daß das System Schwierigkeiten hatte bei 300 K die anti-Konformation wieder zu verlassen und so eine hohe Population der anti-Konformation erzeugt wurde, die rein zufällig annähernd der Gleichgewichtspopulation entspricht. Problematisch bzgl. der FGE-Profile von  $\Phi$  ist lediglich, daß selbst bei 600 K immer noch nicht der gesamte Torsionswinkelbereich für  $\Phi$  populierte ist. Wegen des steilen Anstiegs der Energie auf über 8 kcal/mol ist jedoch nicht anzunehmen, daß sich im Bereich  $[90^\circ, 240^\circ]$  noch ein lokales Minimum niedriger Energie verbirgt. Die Berechnung der Freien Gridgruppen-Energie in Abhängigkeit von  $\Phi$  und  $\Psi$  ( $\Phi/\Psi$ -FGE-Maps) (Abb. X110) zeigt ebenfalls die Unabhängigkeit der Freien Gridgruppen-Energie von der Temperatur selbst noch bis 1800 K. Damit läßt sich schlußfolgern, daß Hochtemperatur-Simulationen (HTMD) für  $\Phi$ ,  $\Psi$  und  $\omega$  unter den gewählten Simulationsbedingungen (harmonisches Potential für die Bindungslängen) konsistente Resultate liefern und daher zur Beschreibung des Konformationsraums von Kohlenhydraten geeignet sind. Weiterhin kann durch HTMD-Simulationen bei 1000 K innerhalb weniger Nanosekunden Simulationszeit das Konformationsgleichgewicht erreicht werden.

## 3.5 Analyse von Rotamerenübergängen

### 3.5.1 Einleitung

Intramolekulare Bewegungsprozesse können mit dem Modell des Austauschs zwischen Bewegungszuständen (z.B. Konformeren) im thermischen Gleichgewicht beschrieben werden [203]. Für intramolekulare Bewegungsprozesse beschreibt die Geschwindigkeitskonstante  $k_{ij}$  den Wechsel vom Bewegungszustand  $i$  zum Bewegungszustand  $j$  als Prozeß erster Ordnung (vgl. Gleichung 34). Es wird angenommen, daß die Übergänge schnell im Vergleich zur mittleren Lebensdauer eines Einzelzustands sind.

$$\frac{dn_i}{dt} = \frac{dn_{ij}}{dt} + \frac{dn_{ik}}{dt} + \frac{dn_{il}}{dt} + \dots = (k_{ij} + k_{ik} + k_{il} + \dots) \cdot n_i = k^{eff} \cdot n_i \quad (34)$$

- $dn_{ij}/dt$ : Zahl der Übergänge von Zustand  $i$  nach  $j$  pro Zeitintervall
- $k_{ij}$ : Geschwindigkeitskonstante für den Übergang von Zustand  $i$  nach  $j$
- $n_i$ : Zahl der Strukturen die zum Zustand  $i$  gehören

Geschwindigkeitskonstanten aus Trajektorien-Daten werden berechnet, indem man die Zahl der Übergänge ( $u$ ) pro Sampling-Intervall ( $\Delta t_s$ ) aus einem Zustand  $i$  durch die Zahl der Frames, die zum Zustand  $i$  gehören ( $x_i \cdot t_s / \Delta t$ ), dividiert (Gleichung 35).

$$\frac{u_{ij}}{\Delta t_s} = k_{ij} \cdot x_i \frac{t_s}{\Delta t_s} \quad k_{ij} = \frac{u_{ij}}{x_i t_s} \quad (35)$$

$u_{ij}$ : Zahl der Übergänge von Zustand  $i$  nach  $j$   
 $t_s$ : Simulationszeit  
 $\Delta t_s$ : Sampling-Intervall  
 $x_i$ : Molenbruch des Zustands  $i$  im Konformationsgleichgewicht

Falls nur eine Übergangsmöglichkeit besteht, gilt  $k = 1/\tau_i$  und die Übergangsbarriere für diesen Übergang kann direkt über die Arrhenius Gleichung (Gleichung 36) aus der mittleren Lebensdauer des Ausgangszustands berechnet werden, sofern der präexponentielle Faktor  $A$  bekannt ist.

$$k = A e^{-E^\ddagger/RT} \quad \text{Arrhenius Gleichung} \quad (36)$$

$k$ : Geschwindigkeitskonstante  
 $A$ : präexponentieller Faktor  
 $E^\ddagger$ : Übergangsenergie  
 $R$ : Gaskonstante  
 $T$ : Temperatur

Für Konformationsübergänge von Kohlenhydraten bestehen in der Regel jedoch mehrere Übergangsmöglichkeiten. Selbst im einfachsten Fall eines Rotamerengleichgewichts einer sekundären OH-Gruppe gibt es zwei Richtungen in denen ein Rotamerenübergang erfolgen kann.

Für die Beschreibung von Konformationsübergängen bei Kohlenhydraten sind zwei Aspekte von besonderem praktischen Interesse:

1. Die mittlere Lebensdauer (kinetische Stabilität) einer bestimmten Konformation (vgl. Abschnitt 3.5.2, Abschnitt 3.6.2)
2. Der bevorzugte „Reaktionsweg“ für Konformationsübergänge zwischen zwei Konformationen (vgl. Abschnitt 3.6.4)

Die mittlere Lebensdauer einer Konformation wird durch die sie umgebenden Energiebarrieren auf der Potentialhyperfläche bestimmt. Da bei komplexen Systemen wie Oligosacchariden normalerweise weder die Anzahl der einzelnen Übergangsmöglichkeiten noch deren Übergangsbarrieren bekannt sind, können Konformationsübergänge auf der Potentialhyperfläche in erster Näherung nur mit Hilfe einer effektiven Geschwindigkeitskonstanten  $k_i^{eff}$  beschrieben werden. Für Konformationsübergänge aus einer Konformation  $i$  heraus soll gelten  $k_i^{eff} = 1/\tau_i$ .

Wie die einzelnen Übergangsmöglichkeiten zu  $k_i^{eff}$  beitragen, wird durch die Energien der einzelnen Übergangszustände und deren präexponentiellem Faktor  $A$  bestimmt. Es ist anzunehmen, daß die Übergangsmöglichkeiten mit den niedrigsten Übergangsenergien geschwindigkeitsbestimmend sind. Die effektive Geschwindigkeitskonstante  $k_i^{eff}$  ist über die Arrhenius-Gleichung mit einer effektiven Übergangsenergie  $E^{eff}$  und einem effektiven präexponentiellen Faktor  $A^{eff}$  verknüpft (Gleichung 37).

$$\ln k_i^{eff} = \ln A_i^{eff} - \frac{E_i^{eff}}{RT} \quad (37)$$

Berechnet man die mittlere Lebensdauer eines Zustand  $i$  aus MD-Simulationen bei verschiedenen Simulationstemperaturen, so kann man durch Auftragung von  $\ln(\tau^{-1})$  gegen  $T^{-1}$  aus der Steigung der Geraden  $E_i^{eff}$  und aus dem Achsenabschnitt bei  $T^{-1} = 0$  den präexponentiellen Faktor  $A_i^{eff}$  bestimmen.

Um ein effizientes Modellierung zu ermöglichen, sollte ein Simulationsprotokoll zum Einsatz kommen, das alle benötigten Daten durch eine einzige Simulation liefert. Die Temperaturänderung wird daher entweder durch schrittweise Abkühlung des Systems („Simulated Annealing“) [204] oder durch Aufheizen („Simulated Heating“) <sup>1</sup> erreicht.

Kennt man die einzelnen Konformationen (z.B. Rotamere,  $\Phi/\Psi$ -Minima), so kann man für definierte Übergänge zwischen zwei Konformationen A und B die Geschwindigkeitskonstante aus der Zahl der beobachteten Übergänge (A→B) geteilt durch die Simulationszeit, die das System im Ausgangszustand verbracht hat, berechnen. Aus der Abhängigkeit  $k(T)$  läßt sich über eine Arrhenius-Analyse die Übergangsenergie und der präexponentielle Faktor  $A$  für diesen Übergang bestimmen (Abschnitt 3.6). Da keinerlei Einschränkungen bzgl. der Dimensionalität vorliegen, sollte sich die Methode auch auf Übergänge im n-dimensionalen Hyperraum anwenden lassen.

Um die Methode der Bestimmung der Übergangsenergien durch Arrhenius-Analyse von Trajektorien-Daten zu entwickeln und ihre Anwendbarkeit zu überprüfen, wurde sie zuerst auf einfache Modellsysteme (n-Butan, Cyclohexanderivate) angewandt und die Resultate mit experimentellen Werten verglichen (Abschnitt 3.6.1).

Eine Weiterentwicklung der Methode besteht in der Analyse des Übergangsmechanismus (Übergangswege) für Konformationsübergänge im Konformationshyperraum (Abschnitt 3.6.4). Dabei werden zuerst mögliche Übergangspfade im Konformationshyperraum durch die „Template-Grouping“-Methode markiert und anschließend aus den Daten die Geschwindigkeitskonstanten für die einzelnen Wege bestimmt.

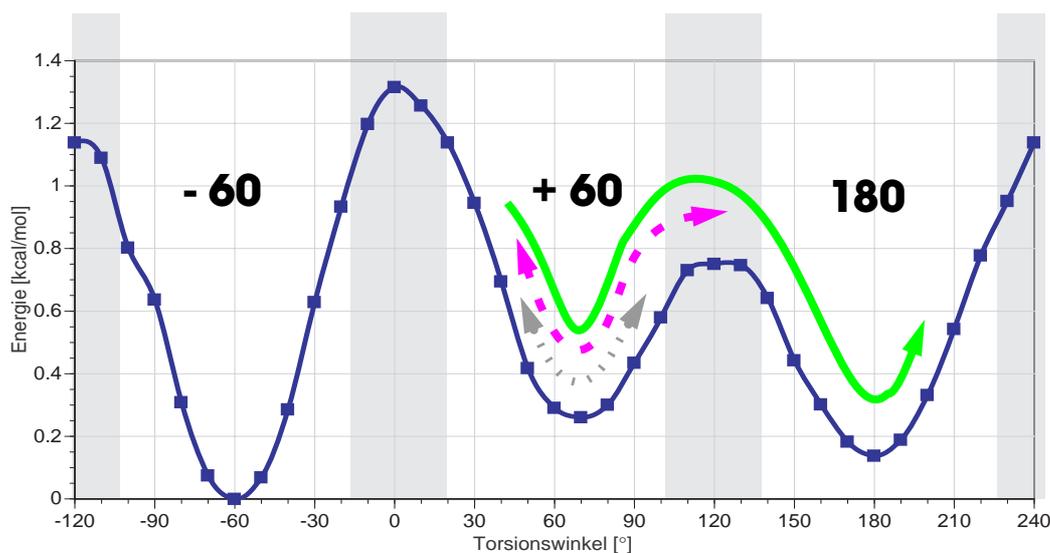
### 3.5.2 Bestimmung der mittleren Lebensdauern von Rotameren

#### Festlegung des Sampling-Intervalls

Erster Schritt der Bestimmung der Lebensdauern von Konformationen ist die Festlegung des optimalen Sampling-Intervalls ( $\Delta t_s$ ). Um die mittlere Lebensdauer eines Zustands ausreichend genau bestimmen zu können, ist es notwendig, daß das Sampling-Intervall kürzer ist als die minimale Lebensdauer des Zustands. Im Prinzip könnte man die Daten nach jedem Intergrationsschritt (z. B. jede Femtosekunde) abspeichern, auf diese Weise würde man jedoch sehr schnell an technische Grenzen stoßen, denn bereits nach einer Nanosekunde hätte man eine Million Datenpunkte pro Variable zu analysieren.

Zur Festlegung des optimalen Sampling-Intervalls für Rotationen der OH-Gruppen wurde zuerst die maximale Geschwindigkeit der Rotationen um die C–O-Bindung bestimmt. Hierzu wurde eine 300K-MD-Simulation von  $\beta$ -Maltose durchgeführt und die Torsionswinkel in einem Intervall von

1. Es hat sich gezeigt, daß bei der Verwendung von DISCOVER bei einer Temperaturerhöhung vor jeder Sampling-Phase eine Temperatur-Equilibrierungsphase durchlaufen werden muß, eine Temperaturniedrigung jedoch ohne eine solche Equilibrierungsphase durchgeführt werden kann.



**Abb. 28:** Rotationsprofil einer OH-Gruppe, schematische Darstellung der Schwingungen innerhalb eines Energieminimums (.....), der Pseudo-Übergänge (- - -) und eines „richtigen“ Rotameren-Übergangs (—).

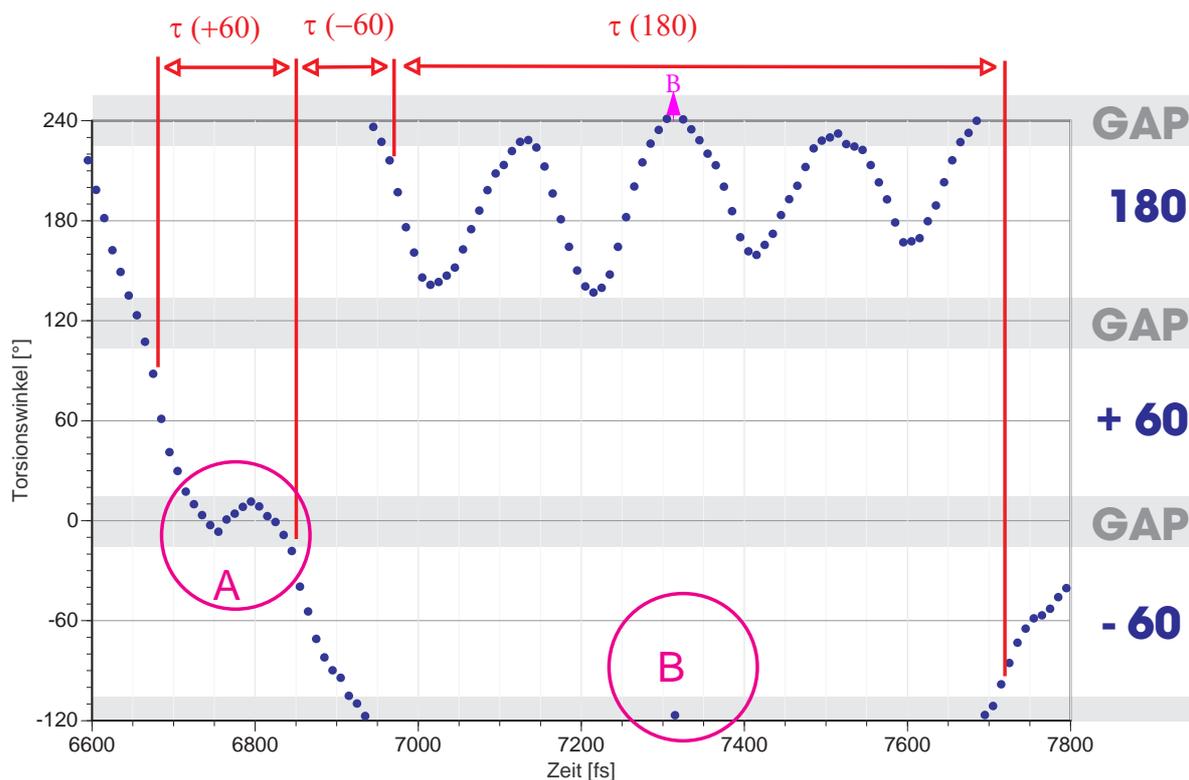
10 fs gespeichert ( $t_s = 10$  ps). Aus der Trajektorie des G3OH-Torsionswinkels läßt sich direkt eine maximale Winkelgeschwindigkeit von 2.4 Grad/fs bestimmen (siehe Abb. X111). Es ist sinnvoll anzunehmen, daß für jede OH-Gruppe drei Rotameren ( $-60^\circ$ ,  $+60^\circ$ ,  $180^\circ$ ) existieren<sup>1</sup>. Teilt man die Winkelskala in drei „Rotameren-Sektoren“, ergibt sich eine minimale Aufenthaltsdauer von 50 fs pro Rotamer für den Fall, daß die OH-Gruppe mit maximaler Winkelgeschwindigkeit rotiert. Ein Beispiel für ein solches Verhalten einer OH-Gruppe ist bei der G3OH-Trajektorie im Zeitintervall [2, 2.5] ps zu finden (siehe Abb. X111). Damit ergibt sich für Rotameren-Übergänge von OH-Gruppen bei 300 K eine obere Grenze des Sampling-Intervalls von ca. 50 fs.

### Unterdrückung von Pseudo-Übergängen

Befindet sich das System z.B. im Rotamerenzustand  $+60^\circ$  (lokales Minimum der Energie), so schwingt es mit einer bestimmten Frequenz um einen Torsionswinkel-Mittelwert (siehe Abb. 28). Hat man als Grenzen zwischen den Rotameren die Werte  $0^\circ$ ,  $120^\circ$  und  $240^\circ$  definiert, so kann es vorkommen, daß das System z.B. die Grenze  $120^\circ$  minimal überschreitet und kurz danach wieder in den Ausgangszustand zurückkehrt (Pseudo-Übergang).<sup>2</sup> Dadurch würde man für den Rotamerenzustand  $180^\circ$  eine extrem kurze Lebensdauer erhalten, die nicht die wirklichen Verhältnisse widerspiegelt (siehe auch Abb. 29).

Eine Lösung des Problems ist die Verwendung von Pufferzonen (Gaps), die sozusagen als „variable Grenzen“ wirken. Eine Struktur wird solange dem Ausgangszustand zugeordnet bis das Gap vollständig durchquert ist (siehe Abb. 29). Die Breite des Gaps hängt vom System und den Simulationsbedingungen ab. Da man im Voraus nie genau weiß, wo genau sich die Energieminima und -maxima

1. Es gibt auch Rotationsprofile mit nur zwei Minima. Dies ist z.B. für G2OH ( $\epsilon=1$ ) der Fall; hier ist der Zustand  $-60^\circ$  so stark durch H-Brücken-Bindung stabilisiert, daß die Übergangsbarriere zwischen  $-60^\circ$  und  $+60^\circ$  verschwindet.
2. Dies ließe sich zum Beispiel dadurch zu erklären, daß sich im Moment unmittelbar nach Überschreiten des Übergangszustands die anderen Freiheitsgrade des Systems so verändert haben, daß eine Rückkehr in den Ausgangszustand günstiger ist als den Weg ins benachbarte lokale Energieminimum fortzusetzen.



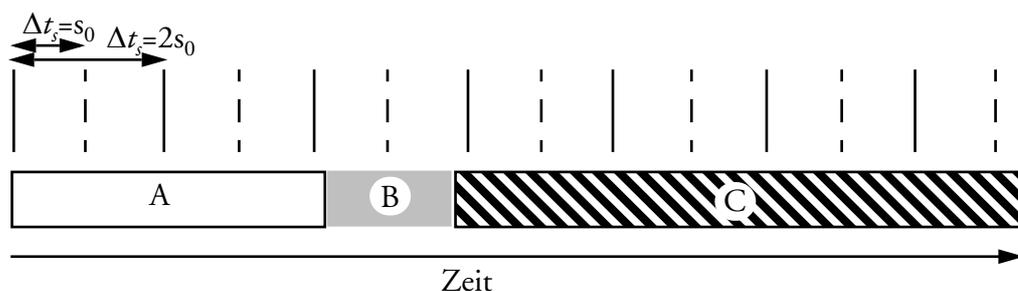
**Abb. 29:** Bestimmung der Lebensdauern von Rotameren, Definition von Gaps zur Unterdrückung von Pseudo-Übergängen (siehe markierte Bereiche „A“ und „B“)

eines Rotationsprofils befinden, dienen die Winkel  $0^\circ$ ,  $120^\circ$  und  $240^\circ$  immer als Mittelpunkt der Gaps. Die Breite der Gaps muß so gewählt werden, daß alle Energiemaxima (Übergangszustände) durch die einzelnen Gaps erfaßt werden, die Energieminima aber nicht in einem Gap zu liegen kommen. Als weiteres Kriterium soll gelten, daß die Zeit, die das System in einem Gap verbringt, viel kleiner als die mittlere Lebensdauer der Rotamerenzustände sein sollte. Dieses Kriterium gewinnt besonders bei erhöhten Simulationstemperaturen an Bedeutung, da hier die Übergangszustände stärker populiert werden. Die Gap-Breite wird als Prozentwert bzgl. der Intervallbreite angegeben.

### Abhängigkeit der mittleren Lebensdauer vom Sampling-Intervall

Um zu überprüfen inwieweit die Ergebnisse für die mittleren Lebensdauern durch ein zu großes Sampling-Intervall verfälscht werden, wurde eine MD-Simulation von  $\beta$ -Maltose (CVFF, 300 K,  $\epsilon=4$ ,  $t_s = 1$  ns,  $\Delta t_s = 20$  fs) durchgeführt. Basierend auf dem so erzeugten Ensemble von 50000 Strukturen wurden MD-Trajektorien mit größerem Sampling-Intervall ( $x \cdot \Delta t_s$ ) dadurch simuliert, daß nur jede  $x$ -te Struktur in die Analyse einbezogen wurde.

Die Zahl der Rotamerenübergänge nimmt mit zunehmendem Sampling-Intervall in guter Näherung mit  $y=a/(b+x)$  ( $R^2 \approx 0.9$ ) ab (siehe Abb. X114). Die Rotameren-Population ist praktisch unabhängig vom Sampling-Intervall (vgl. Abb. X112).



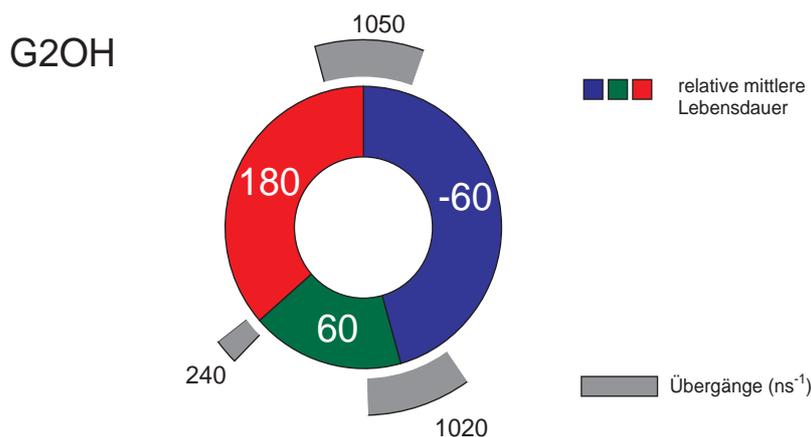
**Abb. 30:** Schematische Darstellung des „Überspringens von Übergängen“ bei der Erhöhung des Sampling-Intervalls

Trägt man die mittlere Lebensdauer eines Rotamerenzustands gegen das Sampling-Intervall auf, so zeigt sich ein annähernd linearer Anstieg bis das Sampling-Intervall ( $\Delta t_s$ ) ungefähr gleich der mittleren Lebensdauer  $t_0$  (siehe Gleichung 38) ist. Wird das Sampling-Intervall größer als  $t_0$ , so besteht immer noch ein annähernd linearer Zusammenhang, die Steigung der Geraden ist aber etwas flacher (vgl. Abb. X113).

$$\tau_0 = \lim_{\Delta t_s \rightarrow 0} \tau(\Delta t_s) \quad (38)$$

Die Ursache für die Erhöhung der mittleren Lebensdauer mit zunehmendem Sampling-Intervall dürfte im „Überspringen von Übergängen“ zu suchen sein (vgl. Abb. 30). Geht ein System von A nach B und anschließend nach C in der in Abb. 30 schematisch dargestellten Weise über, so kann es vorkommen, daß keine Struktur gespeichert wird, die dem Zustand B entspricht. Voraussetzung hierfür ist, daß die Lebensdauer von B kleiner als das Sampling-Intervall  $\Delta t_s$  ist. Als Resultat würde man eine Erhöhung der Lebensdauer von A um  $\Delta t_s$  erhalten. Die Erhöhung der mittleren Lebensdauer von A müßte daher um so größer sein, je kleiner die Lebensdauer des Nachbarzustandes B und je höher die Übergangsfrequenz  $A \leftrightarrow B$  ist.

Diese Annahme läßt sich anhand der Simulationsdaten bestätigen: Beispielsweise findet man für den Torsionswinkel G2OH als Reihenfolge für die Abhängigkeit der mittleren Lebensdauer von Sampling-Intervall:  $60^\circ < 180^\circ < -60^\circ$  (siehe Abb. X115). Der Zustand  $-60^\circ$  [ $\tau_0(-60^\circ) = 424$  fs] besitzt den Nachbarzustand  $(+60^\circ)$  mit der kürzesten mittleren Lebensdauer [ $\tau_0(+60^\circ) = 178$  fs]. Weiterhin ist die Übergangsfrequenz für den Übergang  $(-60^\circ \leftrightarrow +60^\circ)$  mit  $1020 \text{ ns}^{-1}$  sehr hoch (vgl. Abb. 31). Damit ist die Wahrscheinlichkeit groß, daß mit zunehmendem Sampling-Intervall ein Übergang von  $-60^\circ$  nach  $+60^\circ$  nicht registriert wird. Wegen der viermal niedrigeren Übergangsfrequenz für den Übergang  $(+60^\circ \leftrightarrow 180^\circ)$  wird ein großer Anteil der Übergänge zurück in den Ausgangszustand  $-60^\circ$  erfolgen. Dadurch erhöht sich die Lebensdauer des vorhergehenden Zustands  $-60^\circ$  nicht nur um  $\Delta t_s$ , sondern um die Lebensdauer des Zustands  $+60^\circ$  plus der des aktuellen Zustands  $-60^\circ$ . Es ist, als hätte der Übergang nach  $+60^\circ$  nie stattgefunden. Daher besteht für den Zustand  $-60^\circ$  die größte Abhängigkeit vom Sampling-Intervall. Der Zustand  $+60^\circ$  hingegen besitzt nur längerlebige Nachbarzustände; zusätzlich besteht für Übergänge zum Nachbarzustand mit der kürzeren Lebensdauer [ $\tau_0(180^\circ) = 365$  fs] zusätzlich noch eine geringe Übergangswahrscheinlichkeit, so daß hauptsächlich Übergänge in den längerlebigen Zustand  $-60^\circ$  erfolgen. Die Wahrscheinlichkeit, daß das System vom Zustand  $+60^\circ$  in einen Zustand  $-60^\circ$  mit einer Lebensdauer kleiner als das Sampling-Intervall übergeht, ist also äußerst gering. Zusätzlich ist der Übergang  $(-60^\circ \leftrightarrow 180^\circ)$  genauso wahrscheinlich wie die Rückkehr

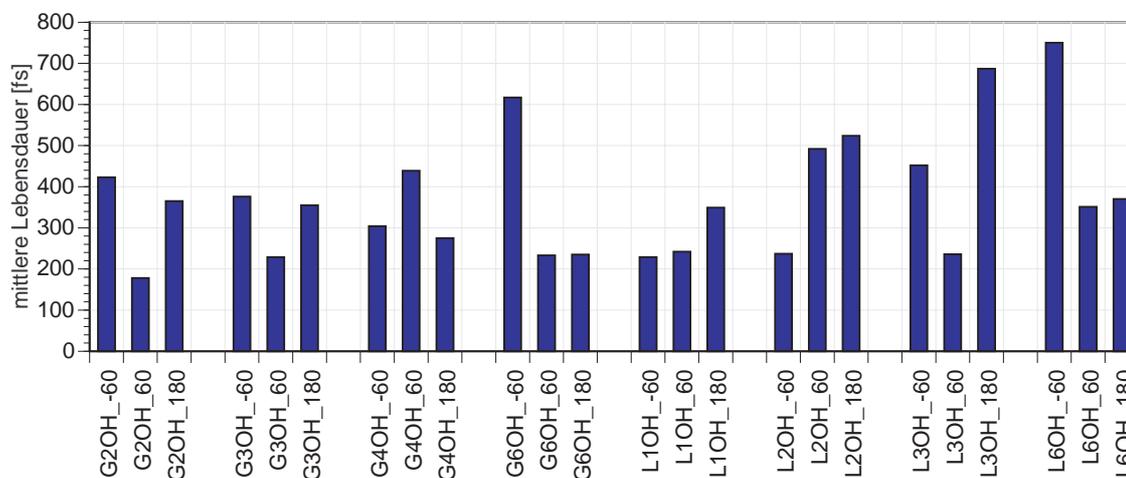


**Abb. 31:** Schematische Darstellung der relativen mittleren Lebensdauern und der Übergangsfrequenz für G2OH

in den Ausgangszustand +60°. Eine Verlängerung der mittleren Lebensdauer durch Rückkehr in den Ausgangszustand ist daher viel weniger wahrscheinlich als bei -60°. Dies schlägt sich in der deutlich geringeren Erhöhung der mittleren Lebensdauer mit zunehmendem Sampling-Intervall nieder.

### Die mittlere Lebensdauer von OH-Rotameren bei 300 K

Aus der weiter oben beschriebenen MD-Simulation von  $\beta$ -Maltose (CVFF,  $\epsilon = 4$ ,  $t_s = 1$  ns,  $\Delta t_s = 20$  fs) wurden die mittleren Lebensdauern  $\tau_0$  durch linearen Fit ermittelt (Abb. 32). Die meisten mittleren Lebensdauern von OH-Rotameren bei 300K bewegen sich im Bereich zwischen 200 fs und 500 fs.

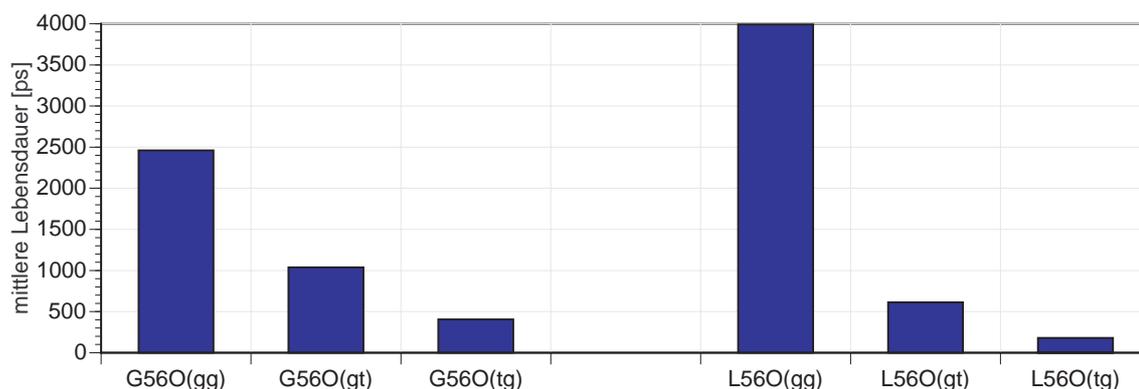


**Abb. 32:** mittlere Lebensdauern von OH-Rotamerenzuständen.

Auffällig lange mittlere Lebensdauern zeigen G6OH(-60°), L6OH(-60°) und L3OH(180°). Bei L3OH(180°) ist der Grund wahrscheinlich in der Ausbildung einer langlebigen interresidualen H-Brücke zu GO2 zu suchen, bei den anderen beiden bietet sich als Erklärung eine intraresiduale H-Brücke zum Ring-O-Atom an, da beide Hydroxymethylgruppen während der Simulation hauptsächlich in der gt-Orientierung vorliegen.

### Die mittlere Lebensdauer von Hydroxymethyl-Rotameren bei 300 K

Die Langzeit-MD-Simulation von  $\beta$ -Maltose (CVFF,  $\varepsilon = 4$ ,  $t_s = 2 \mu\text{s}$ ,  $\Delta t_s = 20 \text{ ps}$ ) bildet die Grundlage zur Bestimmung der mittleren Lebensdauern  $\tau_0$  (Abb. X116). Die mittleren Lebensdauern von Hydroxymethyl-Rotameren bei 300 K unterscheiden sich um mehr als eine Größenordnung (Abb. 33). Die kürzesten Lebensdauern besitzen die tg-Zustände. Auffällig ist die Halbierung der mittleren Lebensdauer bei L56O [ $\tau_0(\text{tg}) = 181 \text{ ps}$ ] gegenüber G56O [ $\tau_0(\text{tg}) = 406 \text{ ps}$ ]. Die Häufung von O-Atomen (GO5, LO4, LO6) scheint die tg-Stellung von L56O kinetisch zu destabilisieren (vgl. Hassel-Ottar-Effekt [134,205]). Die lange mittlere Lebensdauer von 2.5 ns bzw. 4 ns für den gg-Zustand deutet auf hohe Übergangsbarrieren hin, die diesen Zustand umgeben. Das Profil der Freien Gridgruppen-Energie (Abb. X76) bestätigt dies: es muß eine Energiebarriere von ca. 6 kcal/mol überwunden werden um den gg-Zustand zu verlassen.



**Abb. 33:** Mittlere Lebensdauern von Hydroxymethyl-Rotamerenzuständen bei 300K (CVFF,  $\varepsilon=4$ ).

### Reproduzierbarkeit von mittleren Lebensdauern

Die in Abschnitt 3.4.6 beschriebene Aufteilung der Langzeit-MD-Simulation von  $\beta$ -Maltose in Trajektorien unterschiedlich langer Simulationszeit bildet die Datenbasis zur Analyse der Streuung der mittleren Lebensdauer in Abhängigkeit von der Simulationszeit (Abb. X117). Trägt man die Standardabweichung der mittleren Lebensdauer gegen die Simulationszeit auf, so zeigt sich, daß die gg-Konformation selbst nach 500 ns noch eine große Schwankung in den mittleren Lebensdauern aufweist (gg:  $4135 \pm 1036 \text{ ps}$ ; gt:  $692 \pm 85 \text{ ps}$ ; tg:  $227 \pm 19 \text{ ps}$ ) (siehe Abb. X118).

Eine entsprechende Analyse für die OH-Gruppen ist in Abb. X119 und Abb. X120 gezeigt. Nach 500 ps Simulationszeit schwanken die mittleren Lebensdauern noch um ca. 100 fs (10-25%). Beispiel G3OH:  $-60^\circ$ :  $570 \pm 94 \text{ fs}$ ;  $+60^\circ$ :  $340 \pm 39 \text{ fs}$ ;  $180^\circ$ :  $585 \pm 130 \text{ fs}$

### 3.5.3 Die mittlere Lebensdauer von $\Phi/\Psi$ -Konformationen bei 300 K

Aus der Langzeit-MD-Simulation von  $\beta$ -Maltose (Abb. X73) geht klar hervor, daß Übergänge zwischen syn- und anti-Konformation zwar auch bei 300 K stattfinden, diese wegen der großen Lebensdauer der syn-Konformation (vier Übergänge in  $2 \mu\text{s}$ ) extrem selten sind. Um eine statistische Signifikanz der mittleren Lebensdauer zu erreichen, sollte die Lebensdauer des syn-Zustand mindestens zwanzigmal bestimmt werden. Dies würde jedoch eine Simulationsdauer von  $10 \mu\text{s}$  erfordern. Eine statistisch verlässliche Bestimmung der mittleren Lebensdauer der syn-Konformation ist daher durch eine MD-Simulation bei 300 K zur Zeit technisch nicht realisierbar. Aus den vier (anti  $\rightarrow$  syn)-Übergängen läßt sich die mittlere Lebensdauer der anti-Konformation zu ca. 2 ns abschätzen. Diese

Größenordnung der mittleren Lebensdauer ist noch kurz genug, damit sie mehrfach bestimmt werden kann. Um dies zu realisieren, wurden zwanzig 3ns-MD-Simulationen und zwanzig 5ns-MD-Simulationen<sup>1</sup> bei 300 K ( $\epsilon=4$ ) in der anti-Konformation gestartet. Um einen eventuell vorhandenen Solvenseinfluß zu untersuchen, wurde eine anti-Konformation solvatisiert (Periodic Boundary Conditions (PBC), Wasserbox  $20\text{\AA}\times 20\text{\AA}\times 20\text{\AA}$ , 242 Wassermoleküle,  $\epsilon=1$ ) und die Wassermoleküle durch eine 100 ps-MD-Simulation ( $\beta$ -Maltose: Schweratome fixiert) equilibriert und anschließend minimiert (ohne Fixierung). Ausgehend von der so erzeugten Startstruktur wurden zehn 250 ps-MD-Simulationen und zehn 500 ps-MD-Simulationen bei  $400\text{ K}^2$  durchgeführt. Während der Equilibrierungsphase (20 ps) wurde bei allen Simulationen an den  $\Phi$ - und  $\Psi$ -Winkel ein zusätzliches Potential (torsion force) von  $1000\text{ kcal/rad}^2$  angelegt, um zu verhindern, daß ein (anti  $\rightarrow$  syn)-Übergang schon während der Equilibrierungsphase<sup>3</sup> erfolgt. Dieses Potential wurde während der ersten 5 ps der Sampling Phase linear auf Null reduziert.

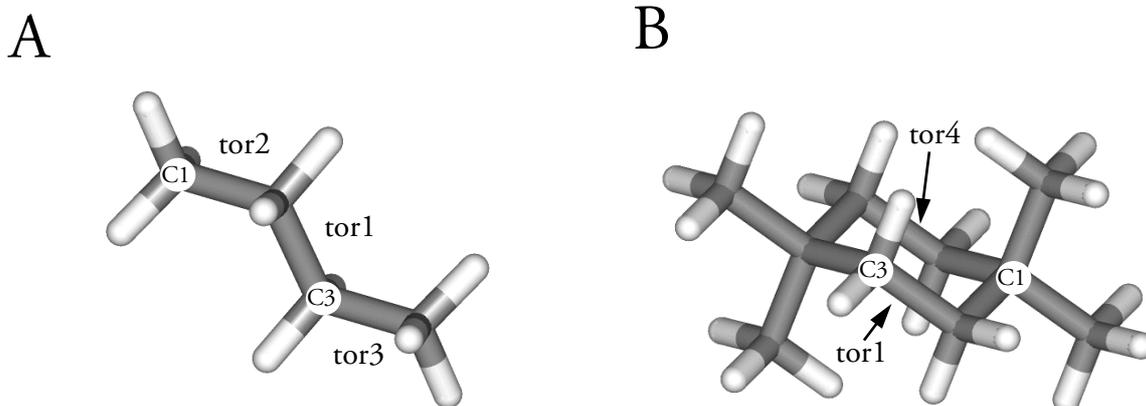
Als mittlere Lebensdauer für den anti-Zustand bei 300K ( $\epsilon=4$ ) wurde 900 ps erhalten, wobei die einzelnen Lebensdauern über fast zwei Größenordnungen variieren (Abb. X121). Bei 400 K (PBC) erhält man eine mittlere Lebensdauer von 90 ps. Mit Hilfe der Arrhenius-Gleichung (37) errechnet sich daraus in beiden Fällen eine Übergangsbarriere von  $3.7\text{ kcal/mol}$  ( $A = 2\cdot 10^{12}$ , vgl. Abschnitt 3.6.3). Das explizit vorhandene Wasser hat also wahrscheinlich keinen Einfluß auf die anti $\rightarrow$ syn-Übergangsbarriere. Weiterhin ist die erhaltene Übergangsbarriere aus den „Multiple-Start“-MD-Simulationen in ausgezeichneter Übereinstimmung mit den Barrieren die durch Arrhenius-Analyse von Hochtemperatur-Simulationen bestimmt wurden (Abschnitt 3.6.3).

- 
1. Bei den derzeit zur Verfügung stehenden Simulationsprogrammen wird bei methodischen Untersuchungen, wie im vorliegenden Fall, ein Großteil der Simulationszeit dadurch „verschwendet“, daß nach dem erfolgten Übergang anti $\rightarrow$ syn die Simulation fortgesetzt wird. Es gibt keine Möglichkeit die Simulation konditionell im Bezug auf strukturelle Eigenschaften (Torsionswinkel, Abstände) zu steuern. Auf der „Wunschliste“ für zukünftige Erweiterungen der Simulationsprogramme wie DISCOVER, TINKER, AMBER, usw. steht damit eine Funktion wie: if (torsion > wert) then goto ...
  2. Bei 300K wäre eine Gesamtsimulationszeit von ca. 40 ns notwendig um 20 anti-syn-Übergänge zu erhalten. Das ist aufgrund der Größe des Systems (5472 Atome) technisch zu aufwendig, so daß die Simulations-Methode („Multiple-Start“-MD-Simulation) zuerst bei 400 K erprobt wurde.
  3. Der in DISCOVER implementierte Temperatur-Equilibrierungsalgorithmus arbeitet nicht sehr „schonend“. Die kinetische Energie der einzelnen Atome wird teilweise so drastisch variiert, daß während der Equilibrierungsphase selbst Konformationsübergänge bei niedrigen Temperaturen erfolgen können, die sonst praktisch nie beobachtet werden. So ist es mehrmals vorgekommen, daß während der 300K-Equilibrierungsphase von Maltose innerhalb von 20 ps ein Übergang von syn nach anti erfolgt ist.

### 3.6 Arrhenius Analyse zur Bestimmung von Übergangsenergien

#### 3.6.1 Bestimmung von Übergangsbarrieren ausgewählter Modellsysteme und Vergleich mit experimentellen Werten

Eine Überprüfung, inwieweit die durch Arrhenius-Analyse aus Trajektorien-Daten berechneten Übergangsbarrieren in Übereinstimmung mit experimentellen Werten sind, wurde anhand zweier Modellsysteme (Abb. 34), von denen experimentell bestimmte Übergangsbarrieren vorliegen, durchgeführt.



**Abb. 34:** Modellverbindungen n-Butan (A) und 1,1,4,4-Tetramethylcyclohexan (B)

A: tor1=C1-C2-C3-C4, tor2=H1-C1-C2-C3, tor3=C2-C3-C4-H4

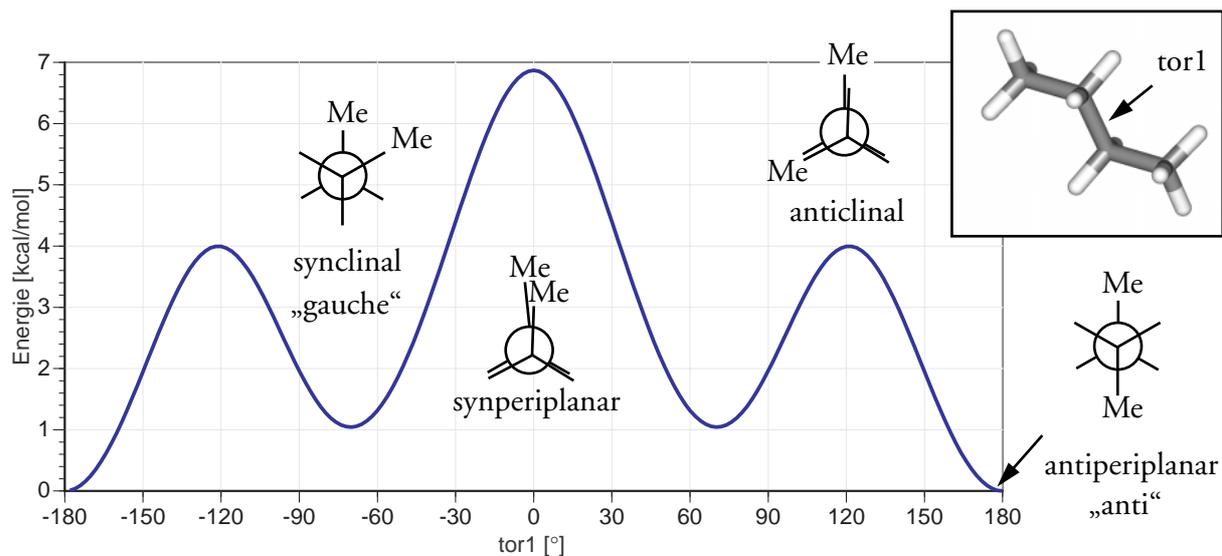
B: tor1=C1-C2-C3-C4, tor4=C4-C5-C6-C1

#### Rotationsbarrieren von n-Butan

Der einfachste Kohlenwasserstoff, der energetisch unterschiedliche Konformationen einnehmen kann, ist n-Butan. Experimentell wurde gefunden, daß die anti-Konformation gegenüber der gauche-Konformation um 0.97 kcal/mol energieärmer ist [106]. Für den Übergang anti  $\rightarrow$  gauche findet man in der Literatur [206,207] einen Wert von 3.4 kcal/mol. Die Angaben für die synperiplanare Anordnung der Methylgruppen variieren im Bereich 4-6 kcal/mol (4.6 kcal/mol [203], S. 107; 6.2 kcal/mol [208], S.107)

Eine systematische Rotation um den zentralen Torsionswinkel „tor1“ in 2°-Schritten mit anschließender Minimierung (CVFF,  $\epsilon = 10$ , Conjugate Gradient, Abbruchkriterium = 0.001 kcal/Å, Torsion Force (tor1) = 1000 kcal/rad<sup>2</sup>) ergibt ein Rotationsprofil (Abb. 35), welches bereits eine gute Annäherung an die experimentellen Werte zeigt:  $\Delta E(\text{gauche-anti}) = 1.04$  kcal/mol,  $\Delta E(\text{anticlinal-anti}) = 4.0$  kcal/mol,  $\Delta E(\text{synperiplanar-anti}) = 6.9$  kcal/mol

Um die Energiebarrieren basierend auf MD-Daten zu bestimmen, wurde ein Simulated-Annealing-Protokoll (CVFF,  $\epsilon = 10$ ) durchlaufen, bei dem die Simulationstemperatur im Bereich von 1800 K bis 500 K schrittweise erniedrigt wurde. Die Simulationszeit pro Temperatur betrug 2 ns, das Sampling-Intervall war 100 fs. Die Temperatur-Equilibrierungszeit betrug wie üblich 20 ps. Insgesamt beruhen die durchgeführten Arrhenius-Analysen für tor1, tor2 und tor3 damit auf den Daten von 200000 Frames. Zur Beschreibung der chemischen Bindungen wurde bei allen MD-Simulationen ein harmonisches Potential verwendet, so daß die Bindungslängen symmetrisch um einen Mittelwert variieren (Abb. X122). Bei 1400 K beträgt die Standardabweichung der Bindungslängen 0.07 Å.



**Abb. 35:** Rotationsprofil für „tor1“ (C1-C2-C3-C4) aus Grid-Search (n-Butan, CVFF,  $\epsilon = 10$ , Intervall =  $2^\circ$ , Minimierung mit Torsion Force (1000 kcal/rad<sup>2</sup>))

Die Gleichgewichtsanalyse für die Rotameren ( $-60^\circ$  [-120°, 0°],  $+60^\circ$  [0°, 120°],  $180^\circ$  [120°, 240°]) (Abb. X123, Abb. X124) zeigt, daß sich für 500 K und 600 K (evtl. auch bei 700 K) das Konformationsgleichgewicht innerhalb von 2 ns nicht eingestellt hat. Für die Arrhenius-Analyse (Abb. X125) wurde daher nur der Temperaturbereich zwischen 700K und 1800K berücksichtigt. Die Berechnung der FGE-Profile für tor1 im Temperaturbereich zwischen 500K und 1800K zeigt deutlich, daß nur bei Einstellung des Konformationsgleichgewichts die FGE-Profile näherungsweise temperaturunabhängig sind (Abb. X126). Das FGE-Profil für 500K und 600K weicht deutlich von den restlichen Profilen ab. Die FGE-Profile für Simulationstemperaturen größer als 700K zeigen, wie auch die Ergebnisse der Arrhenius-Analyse (Abb. X127), eine gute Übereinstimmung mit den experimentellen Werten. Ein Zusammenfassung der Ergebnisse zeigt Tabelle 1.

**Tabelle 1: Eckdaten der Rotationsprofile von n-Butan [kcal/mol]**

	Grid-Search	FGE-Profil	Arrhenius-Analyse	Experiment
$\Delta E(\text{gauche-anti})$	1.04	1.1	1.3 (4.0-2.7) ( $A=4.1 \cdot 10^{12}$ ; $R^2=0.99$ )	0.97
$\Delta E(\text{anticlinal-anti})$	4.0	3.5	4.0 ( $A=4.3 \cdot 10^{12}$ ; $R^2=0.99$ )	3.4
$\Delta E(\text{synperiplanar-anti})$	6.9	6.8	7.3 (1.3 + 6.0) ( $A=5.5 \cdot 10^{12}$ ; $R^2=0.98$ )	4.6/6.2

## Ringinversionsbarrieren von Cyclohexan-Derivaten

Stellvertretend für einen komplexen Übergang mit einer hohen Übergangsbarrriere wurden die Ringinversionsbarrieren für 1,1,3,3-Tetramethylcyclohexan und 1,1,4,4-Tetramethylcyclohexan durch Arrhenius-Analysen aus Trajektorien-Daten bestimmt. Die experimentell bestimmte Barriere liegt für 1,1,3,3-Tetramethylcyclohexan bei 9.6 kcal/mol und für 1,1,4,4-Tetramethylcyclohexan bei 11.7 kcal/mol [209].

Die Definition der Konformationen A und A' erfolgt über die Torsionswinkel  $\text{tor1}(\text{C1-C2-C3-C4})$  und  $\text{tor4}(\text{C4-C5-C6-C1})$ . Alle Strukturen, die im  $\text{tor1-tor4}$ -Koordinatensystem einen RMS-Wert kleiner als  $30^\circ$  bzgl. des Referenzpunktes (-50/50) besitzen, werden willkürlich der Konformation A zugeordnet (Template Grouping). Analog dazu ist die Konformation A' im  $\text{tor1-tor4}$ -Koordinatensystem durch den Referenzpunkt (50/-50) und einem RMS-Radius von  $30^\circ$  definiert (vgl. Abb. X128).

Das Simulated-Annealing-Protokoll (CVFF,  $\epsilon = 4$ ) erstreckte sich über einen Temperaturbereich von 1800 K bis 500 K. Die Simulationszeit pro Temperatur betrug 3 ns, das Sampling-Intervall war 500 fs. Die Temperatur-Equilibrierungszeit betrug wie üblich 20ps. Für die Arrhenius-Analyse (Abb. X129) wurde nur der Temperaturbereich von 1000 K bis 1800 K verwendet.

Aus der Arrhenius-Analyse für 1,1,4,4-Tetramethylcyclohexan ergab sich für die Übergangsbarrriere ein Mittelwert von 11.95 kcal/mol ( $A = 2.5 \cdot 10^{13}$ ,  $R^2 = 0.97$ ). Die Arrhenius-Analyse für 1,1,3,3-Tetramethylcyclohexan ergab eine berechnete Barriere von 9.2 kcal/mol ( $A = 1.4 \cdot 10^{13}$ ,  $R^2 = 0.98$ ) und ist damit ebenfalls sehr nah am experimentellen Wert. Die Abschätzung der Inversionsbarriere für 1,1,4,4-Tetramethylcyclohexan aus der  $\text{tor1-tor4}$ -FGE-Map (Abb. X128) ergibt eine Barriere von ca. 11-12 kcal/mol.

Insgesamt hat sich gezeigt, daß Arrhenius-Analysen basierend auf Hochtemperatur-MD-Simulationen sehr gut geeignet sind, um Ringinversionsbarrieren zu bestimmen.

### 3.6.2 Bestimmung der effektiven Übergangsbarrieren von Kohlenhydraten aus der mittleren Lebensdauer einer Konformation

#### Bestimmung der effektiven Übergangsbarrieren für OH-Rotamere

Um die Übergangsbarrieren für die OH-Rotameren zu berechnen, wurde eine MD-Simulation (CVFF,  $\epsilon=4$ ) durchgeführt bei der die Temperatur von 300 K auf 700 K in 100 K-Schritten erhöht wurde. Die Simulationszeit (Sampling-Phase) für jede Temperatur betrug 1 ns, Energie- und Torsionsdaten wurden im Intervall von 50 fs gespeichert (Abb. X130, Abb. X131). Eine Temperatur-Equilibrierungsphase von 20 ps war jeder Sampling-Phase vorgeschaltet. Die Analyse der Systemenergie zeigt, daß die, durch die Temperaturerhöhung zugeführte, kinetische Energie in erster Linie zur Erhöhung der Valenzenergie des Systems führt (Abb. X132, Abb. X133).

Wie zu erwarten hängt die mittlere Lebensdauer der OH-Rotameren von der Temperatur ab (vgl. Abb. X134). Es hat sich gezeigt, daß die mittlere Lebensdauer weiterhin vom Sampling-Intervall (vgl. Abb. X113) und von der Breite des Gaps abhängt (Abb. X135). Zudem ist der Einfluß der Gap-Breite und des Sampling-Intervalls temperaturabhängig (Abb. X136). Alle diese Effekte lassen sich auf das Auftreten von Pseudo-Übergängen und auf das Überspringen von Zuständen (siehe Abschnitt 3.5.2) zurückführen. Beide Einflüsse sind um so größer, je höher die Simulationstemperatur ist. Es handelt sich also hier um Sampling-Effekte. Eine prinzipielle Lösung des Problems der Verfälschung der mittleren Lebensdauer durch Sampling-Effekte wäre eine extreme Verkürzung des Sampling-Intervalls ( $\Delta t_s$ ) auf  $0.01\tau$  (die Lebensdauern schwanken um bis zu zwei Größenordnungen). Hierbei würde sich

jedoch ein enormes technisches Problem ergeben: Zum Erreichen des Konformationsgleichgewichts<sup>1</sup> ist eine Simulationszeit größer als  $100 \cdot k \cdot \tau$  notwendig ( $k$  = Zahl der Konformationen,  $\tau$  = mittlere Lebensdauer der Konformationen), vorausgesetzt alle Konformationen im Gleichgewicht haben denselben Molenbruch und besitzen diesselbe mittlere Lebensdauer und jede Konformation soll hundertmal von Neuem populiert<sup>2</sup> werden (vgl. Gleichung 39). Für ein OH-Rotameren-Gleichgewicht (drei Konformationen) sind daher für  $\Delta t_s = 0.01\tau$  die Daten von 30000 Frames pro Simulations-Temperatur zu analysieren.

$$t_{s100i} \geq 100 \left( \frac{\tau_i}{x_i} \right) \quad (39)$$

$t_{s100i}$ : Simulationszeit, die benötigt wird um 100 Übergänge aus Konformation  $i$  zu beobachten  
 $\tau_i$ : mittlere Lebensdauer der Konformation  $i$   
 $x_i$ : Molenbruch der Konformation  $i$  im Gleichgewicht

Aus dieser Abschätzung wird deutlich, daß mit  $\Delta t_s = 0.01\tau$  selbst bei den einfachsten Konformationsgleichgewichten schon eine riesige Datenmenge zu verarbeiten wären. Um mit dem aktuellen Stand der Technik eine effiziente Arrhenius-Analyse durchführen zu können, ist es daher unumgänglich die Datenmenge zu reduzieren. Da zum Erreichen des Gleichgewichtszustands eine bestimmte „Mindestsimulationszeit“ notwendig ist, kann die Simulationszeit nur bedingt verkürzt werden. Es ist daher notwendig mit einem Sampling-Intervall von größer als  $0.01\tau$  zu arbeiten und die Sampling-Effekte durch Wahl einer geeigneten Gap-Breite „in den Griff zu bekommen“. Ein Anhaltspunkt zur Wahl einer geeigneten Gap-Breite bietet der „Anteil der undefinierten Lebensdauern“, also der Lebensdauern die kleiner oder gleich dem Sampling-Intervall sind (vgl. Abb. 137, unten) und daher nicht bestimmt werden können.

Aus der Verteilung der Lebensdauern (Abb. X138, Abb. X139) wird deutlich, daß für OH-Gruppen durch Erhöhung der Gap-Breite auf 50% (= 60°) der Anteil der undefinierten Lebensdauern für 300 K auf unter 10% reduziert werden kann.

Für die OH-Rotamere aller OH-Gruppen wurde für verschiedene Gap-Breiten  $\ln(\tau^{-1})$  gegen  $1/T$  aufgetragen (Beispiele siehe Abb. X140, Abb. X141) und durch linearen Fit  $E^{eff}$  und  $A^{eff}$  bestimmt (Abb. X142). Es zeigt sich, daß die erhaltenen Übergangsenergien und präexponentiellen Faktoren mit zunehmender Gap-Breite immer kleiner werden. Vergleicht man die Übergangsbarrieren mit den aus den Populationsanalysen erhaltenen FGE-Profilen, so läßt sich feststellen, daß diese zwar qualitativ gut in Einklang zu bringen sind, die effektiven Übergangsbarrieren aus der Arrhenius-Analyse aber um ca. 0.2 kcal/mol höher als die Übergangsbarrieren der FGE-Profile sind. In Anbetracht der technischen Schwierigkeiten, die bei der Bestimmung der mittleren Lebensdauern vorhanden sind, kann ein Energieunterschied von 0.2 kcal/mol als „Übereinstimmung“ bewertet werden.

- 
1. Allgemein wird die Simulationszeit bis zum Erreichen des Gleichgewichtszustandes durch die Konformation mit den größten Verhältnis ( $\tau_i/x_i$ ) bestimmt.
  2. Damit die mittlere Lebensdauer eine statistische Signifikanz aufweist, wird angenommen, daß mindestens 100 Beobachtungen notwendig sind, um eine Normalverteilung der Lebensdauern zu erhalten.

## Bestimmung der effektiven Übergangsbarrieren für Hydroxymethylgruppen

Zur Bestimmung der Übergangsbarrieren der Hydroxymethyl-Rotameren wurden ebenfalls mehrere Serien von MD-Simulationen (CVFF,  $\epsilon = 4$ ), bei denen verschiedene Temperaturbereiche mit unterschiedlich langer Simulationszeit durchlaufen wurden, durchgeführt:

A: Simulated Annealing (1800K - 900K/-300K) --/2.5ns/50fs<sup>1</sup>

B: Simulated Annealing (1400K - 600K/-200K) --/5ns/100fs

C: Simulated Annealing (600K - 400K/-100K) --/10ns/100fs

D: Simulated Heating (300K - 700K/100K) 20ps/1ns/50fs

E: Simulated Heating (300K - 700K/100K) 20ps/2ns/100fs

F: Simulated Heating (900K - 1800K/300K) 20ps/2.5ns/100fs

G: Einzelsimulationen 300K, 20ps/2 $\mu$ s/20ps; 400K, 20ps/300ns/3ps; 500K, 20ps/100ns/1ps; 600K, 20ps/30ns/0.5ps

Die mittlere Lebensdauer der Hydroxymethylgruppen nimmt ebenfalls, wie erwartet, mit der Temperatur ab (Abb. X143). Trägt man für die verschiedenen Simulationen  $\ln(\tau^{-1})$  gegen  $1/T$  auf, so zeigt sich auch für die Hydroxymethylgruppen eine lineare Abhängigkeit (Abb. X144). Insgesamt kann man feststellen, daß bei niedrigen Temperaturen stärkere Abweichungen von der Geraden auftreten können (Abb. X145, Abb. X146). Interessanterweise liegt für den tg-Zustand von G56O (Abb. X144) bei 300 K der Datenpunkt der 1ns-MD-Simulation näher an der Geraden als der Datenpunkt der 2ns-MD-Simulation. Wie aus Abschnitt 3.4.7 hervorgeht ist bei 300 K auch nach 2 ns noch kein Hydroxymethylgleichgewicht eingestellt, so daß die  $\ln(\tau^{-1})$ -Werte für Simulationstemperaturen kleiner 600K selbst für 10ns-MD-Simulationen noch sehr zufällig sind.

Vergleicht man die aus der Arrhenius-Analyse gewonnenen effektiven Übergangsbarrieren mit den FGE-Profilen (Abb. X99, Abb. X100), so läßt sich, wie schon für die OH-Gruppen gefunden, eine gute qualitative Übereinstimmung finden (Abb. X147). Insgesamt sind wieder die aus den mittleren Lebensdauern erhaltenen effektiven Übergangsbarrieren etwas höher (ca. 0.5 kcal/mol) als man anhand der FGE-Profile erwarten würde. Weiterhin läßt sich feststellen, daß die Barrieren, die man erhält, wenn man nur Temperaturen >700K in die Analyse einbezieht, tendenziell etwas niedriger sind als die Barrieren, die sich z.B. aus den „Langzeitsimulationen“ (Simulationen „G“) bei niedriger Temperatur ergeben.

### 3.6.3 Bestimmung der Übergangsbarrieren definierter Konformationsübergänge

#### Übergangsbarrieren für Hydroxymethyl-Rotationen

Die Geschwindigkeitskonstanten für alle möglichen Übergänge zwischen den Hydroxymethyl-Zuständen gg, gt und tg wurden mittels Gleichung 35 aus den Trajektorien-Daten der MD-Simulationen A bis G (siehe voriger Abschnitt) berechnet. Die  $\ln(k)$ -Werte der einzelnen Übergänge wurden gegen  $1/T$  aufgetragen (Abb. X148 bis Abb. X149) und die Übergangsbarrieren aus der Steigung der Geraden bestimmt (Abb. X150, Abb. X151).

1. Reihenfolge der angegebenen Simulationsparameter: (Starttemperatur-Endtemperatur/Temperaturintervall) Equilibrierungszeit/Simulationszeit pro Temperatur/Sampling-Intervall

Vergleicht man die aus der Arrhenius-Analyse gewonnenen Übergangsbarrieren mit den FGE-Profilen (Abb. X99, Abb. X100), so läßt sich, wie schon für die effektiven Übergangsbarrieren gefunden, eine gute qualitative Übereinstimmung feststellen. Insgesamt sind die aus den Arrhenius-Analysen erhaltenen Übergangsbarrieren wieder etwas höher (ca. 0.5 kcal/mol) als man anhand der FGE-Profile erwarten würde. Bezieht man jedoch nur die Daten bei Temperaturen > 700 K in die Analyse ein, so erhält man Übergangsbarrieren die in ausgezeichneter Übereinstimmung mit den FGE-Profilen sind.

### Übergangsbarrieren für Rotationen um die glycosidischen Torsionswinkel

Zur Arrhenius-Analyse des syn-anti-Übergangs wurde eine MD-Simulation von  $\beta$ -Maltose (CVFF,  $\epsilon = 4$ ,  $\Delta t_S = 100$  fs,  $t_0 = 20$  ps) mit variabler Simulationstemperatur (Simulated-Heating Protokoll; 900 K - 1800 K / 300 K / 2.5 ns) durchgeführt (Abb. X153). Die  $\Psi$ -FGE-Profile (Abb. X152), sowie die  $\Phi/\Psi$ -FGE-Maps (Abb. X110) zeigen selbst bei Temperaturen bis zu 1800 K noch einen konsistenten Verlauf. Die  $\Phi/\Psi$ -FGE-Map ist selbst bei diesen extrem hohen Temperaturen in guter Übereinstimmung mit der  $\Phi/\Psi$ -FGE-Map, die man aus Grid-Search Daten (vgl. Abschnitt 4.2.2) erhält (Abb. X5). Für den syn $\rightarrow$ anti-Übergang ergibt sich aus der Arrhenius-Analyse eine Barriere von 7.2 kcal/mol ( $A = 1.3 \cdot 10^{12}$ ,  $R^2 = 0.98$ ). Zusammen mit der Barriere für den anti $\rightarrow$ syn-Übergang von 3.8 kcal/mol ( $A = 2.6 \cdot 10^{12}$ ,  $R^2 = 0.90$ ) errechnet sich die Energie-Differenz  $\Delta E(\text{anti-syn})$  zu 3.4 kcal/mol. Dies ist in guter Übereinstimmung mit den  $\Psi$ -FGE-Profilen und den  $\Phi/\Psi$ -FGE-Maps.

#### 3.6.4 Bestimmung bevorzugter Übergangswege

##### Direkte Übergänge im $\Phi/\Psi$ -Raum

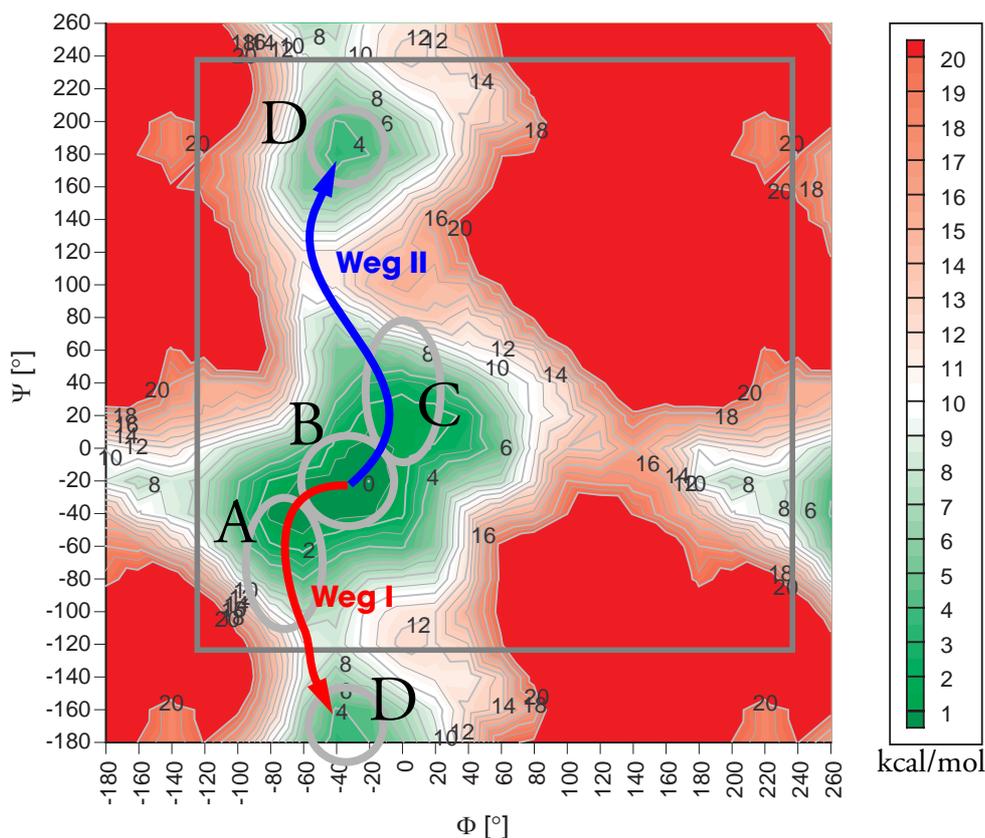
Wie aus Abb. 36 ersichtlich ist, existieren für  $\beta$ -Maltose zwei Wege auf der  $\Phi/\Psi$ -Karte um vom globalen Minimum (syn) zum etwa 3-4 kcal/mol energiereicheren Nebenminimum (anti) zu gelangen.

Der (Hin-)Weg I (B $\rightarrow$ A $\rightarrow$ D) entspricht einer Rotation um  $\Psi$  im Uhrzeigersinn und sollte nach den Energiewerten der Adiabatic Map um 3-4 kcal/mol energie günstiger sein als der Alternativ-Weg II (B $\rightarrow$ C $\rightarrow$ D).

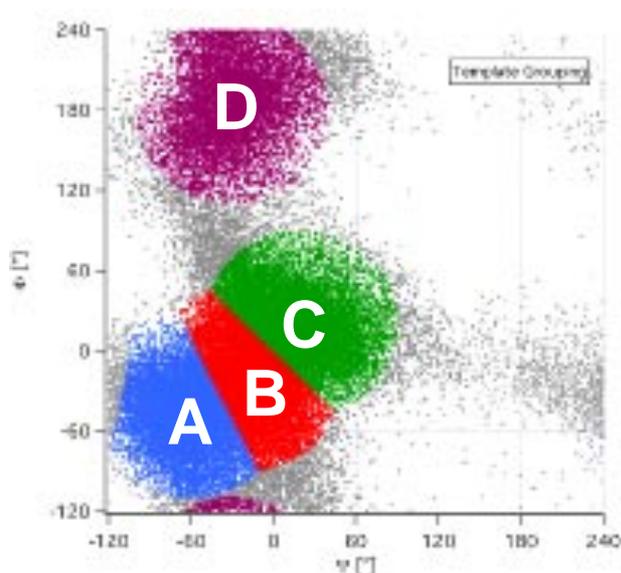
Da das System in der syn-Konformation um das Zentrum B ( $-20^\circ/-20^\circ$ ) in Richtung (A $\leftrightarrow$ B $\leftrightarrow$ C) schwingt, kann davon ausgegangen werden, daß beim Übergang in die anti-Konformation immer die Region B passiert wird und dann entweder über die Region A oder C der Übergang vollendet wird. Um nun anhand der MD-Daten zu analysieren, welchem Weg das System während einer MD-Simulation folgt, ist es notwendig die alternativen Übergangswege im Frame-Ensemble zu markieren. Dazu werden alle Frames, die zu einer „Region X“ im  $\Phi/\Psi$ -Koordinatensystem gehören sollen, mit der „Template-Grouping“-Methode (siehe Abschnitt 7.1) markiert.

Um den Energieunterschied der Übergangsbarriere für die beiden syn $\rightarrow$ anti-Übergangswege für Maltose aus Trajektorien-Daten zu berechnen, wurde wieder eine Serie von MD-Simulationen (CVFF,  $\epsilon = 4$ ,  $t_S = 2.5$  ns,  $\Delta t_S = 50$  fs,  $t_0 = 20$  ps) mit unterschiedlicher Simulationstemperatur (Simulated-Heating Protokoll; 900 K - 1800 K / 300 K / 2.5 ns) durchgeführt und die Anzahl der Übergänge für die beiden Wege bestimmt. Aus dem Verhältnis läßt sich aus Gleichung 40 der Energieunterschied für die Übergangsbarrieren berechnen. Alternativ kann aus der Temperaturabhängigkeit von  $\ln(k_I/k_{II})$  durch eine Arrhenius(analoge)-Analyse ebenfalls  $\Delta E$  bestimmt werden.

Innerhalb der Simulationszeit von 2.5 ns erfolgen Übergänge entlang des Weges II erst ab einer Simulationstemperatur von 600 K (Abb. X154).



**Abb. 36:** Mögliche Übergangswegen für einen syn→anti-Übergang. Adiabatic  $\Phi/\Psi$ -Map ( $\beta$ -Maltose, CVFF,  $\epsilon = 4$ ,  $\Phi/\Psi/\omega_1/\omega_2$ -Grid-Search).



Template Definition für  $\Phi/\Psi$  ( $\alpha$ -Maltose)

Gruppe	$\Phi$	$\Psi$	RMS-Toleranz
A	$-60^\circ$	$-40^\circ$	70
B	$-20^\circ$	$-20^\circ$	70
C	$20^\circ$	$20^\circ$	70
D	$-30^\circ$	$180^\circ$	70

**Abb. 37:** Anwendung der „Template-Grouping“-Methode zur Analyse der Übergangswegen für einen syn-anti-Übergang von  $\beta$ -Maltose (CVFF,  $\epsilon=4$ , 1800K-Trajektorien-Daten,  $t_s = 2.5$  ns).

$$\Delta E = -R \cdot T \cdot \ln \frac{k_I}{k_{II}} = -R \cdot T \cdot \ln \frac{u_I}{u_{II}} \quad (40)$$

$k_I$ : Geschwindigkeitskonstante für Weg I  
 $u_I$ : Zahl der Übergänge Weg I

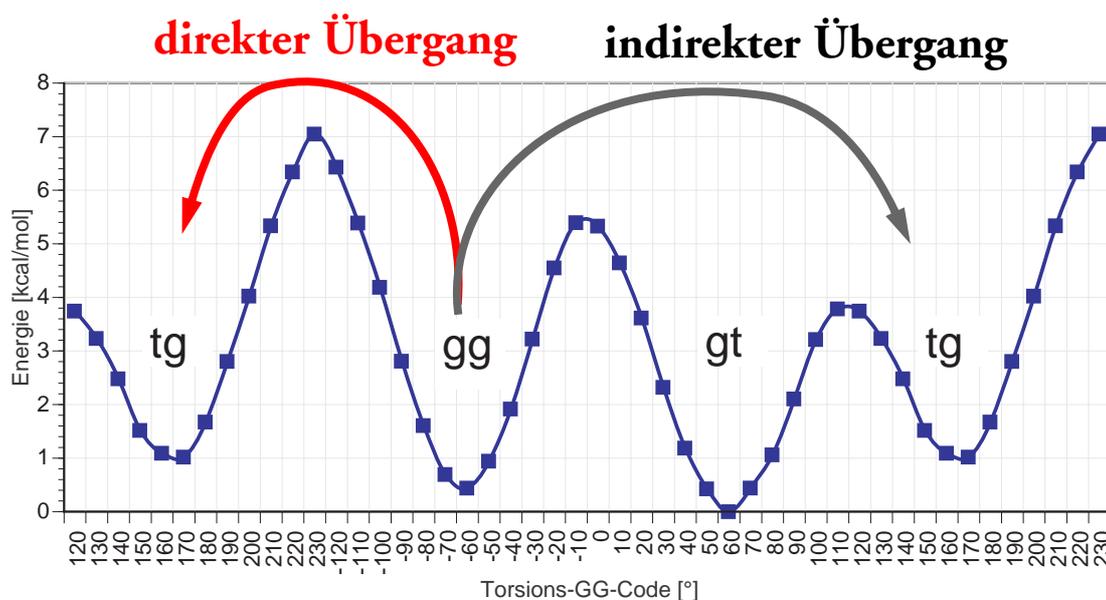
Berechnet man für die verschiedenen Simulationen nach Gleichung 40 den Energieunterschied der Übergangsbarriere für die beiden Wege, wobei Hin- und Rückweg in die Analyse mit einbezogen wurden, so ergibt sich ein Mittelwert von  $\Delta E = 3.1 \pm 0.6$  kcal/mol. Aus der Temperaturabhängigkeit  $\ln(u_I/u_{II})$  ergibt sich  $\Delta E = 3.8$  kcal/mol ( $R^2 = 0.85$ ).

Insgesamt stimmen die aus kinetischen Daten ermittelten Energiedifferenzen der Übergangsbarrieren für syn-anti-Übergangswegen von Maltose sehr gut mit den anhand der Adiabatic Map geschätzten Energien überein.

### Indirekte Übergänge

Existieren für ein System mehrere mögliche Wege um von einem Zustand A zu einem Zustand B zu gelangen, so wird das System hauptsächlich demjenigen folgen, der die kleineren Übergangsbarrieren besitzt. Damit ist es theoretisch denkbar, daß der direkte Übergang zwischen den Zuständen A und B eine höhere Energiebarriere und damit eine niedrigere Geschwindigkeitskonstante aufweist, als der „indirekte“ Weg über einen oder mehrere Zwischenzustände. Ein solcher Fall liegt für die  $gg \leftrightarrow tg$ -Übergänge der Hydroxymethylgruppen von Maltose vor (Abb. 38).

Basierend auf einer MD-Simulation von  $\beta$ -Maltose (CVFF,  $\epsilon = 4$ ,  $\Delta t_S = 100$  fs,  $t_0 = 20$  ps) mit variabler Simulations-Temperatur (Simulated-Heating-Protokoll; 300 K - 700 K / 100 K / 2 ns; 900 K - 1800 K / 300 K / 2.5 ns) wurden eine Detailanalyse des  $gg \leftrightarrow tg$ -Übergangs (Abb. X155, Abb. X156) und des  $gt \leftrightarrow tg$ -Übergangs durchgeführt (Abb. X157, Abb. X158). Es zeigt sich, daß erst ab einer



**Abb. 38:** Übergangswegen für einen  $gg \rightarrow tg$ -Übergang. FGE-Profil für den Torsionswinkel L56O (Maltose, CVFF,  $\epsilon=4$ , 1500 K,  $t_s = 2.5$  ns).

Simulationstemperatur größer als 1500 K der direkte  $gg \leftrightarrow tg$ -Übergang gegenüber dem indirekten Übergang über den  $gt$ -Zustand bevorzugt ist (Abb. X155). Für den  $gt \leftrightarrow tg$ -Übergang hat selbst bei 1800 K der indirekte Übergang über den  $gg$ -Zustand wegen der hohen  $gg \leftrightarrow tg$ -Barriere nur eine untergeordnete Bedeutung (25-30%) (Abb. X157). Die Dauer des indirekten Übergangs ist für  $gt \leftrightarrow tg$  auch deutlich länger als für  $gg \leftrightarrow tg$ . Grund hierfür ist die lange Lebensdauer des  $gg$ -Zustands im Vergleich zum  $gt$ -Zustand. Die Ergebnisse der Arrhenius-Analyse (Abb. X156, Abb. X158) sind in Tabelle 2 zusammengestellt.

**Tabelle 2: Übergangsbarrieren für Hydroxymethyl-Rotamerenübergänge (Maltose, L56O)**

Übergang	$E^\ddagger$ [kcal/mol]	$A$ [ $s^{-1}$ ]	$R^2$
$gg \rightarrow tg$	7.04	$3.56E+12$	0.999
$gg \rightarrow gt \rightarrow tg$	4.89	$1.45E+12$	0.993
$tg \rightarrow gg$	6.82	$3.56E+12$	0.993
$tg \rightarrow gt \rightarrow gg$	4.88	$1.60E+12$	0.991
$gt \rightarrow tg$	4.26	$1.77E+12$	0.993
$gt \rightarrow gg \rightarrow tg$	7.32	$1.95E+12$	0.995
$tg \rightarrow gt$	3.07	$2.16E+12$	0.976
$tg \rightarrow gg \rightarrow gg$	6.29	$1.60E+12$	0.995

### 3.7 Einfluß der Dielektrizitätskonstanten und Solvenseffekte

#### 3.7.1 Radial-Histogramme zur Analyse der Solvensstruktur

##### Radial-Histogramme kleiner solvatisierter Moleküle

Die Verteilung von Solvens-Atomen um ein bestimmtes Zentralatom kann durch Radialverteilungsfunktionen beschrieben werden [129,178,210]. Radial-Histogramme beschreiben die Wahrscheinlichkeit ein Atom im Abstand  $r$  von einem Zentralatom zu finden. Sie werden in CAT2000 berechnet, indem die mittlere Anzahl der Atome eines vorgegebenen Typs ermittelt werden, die sich in einer Kugelschale mit Radius  $r$  und Dicke  $\Delta r$  befinden. Anschließend wird durch das Volumen der Schale ( $4\pi r^2 \Delta r$ ) dividiert.

Als Referenzbeispiele wurden jeweils ein Wasser-, ein Methan- und ein Methanolmolekül mit Wasser solvatisiert (PBC,  $20\text{\AA} \times 20\text{\AA} \times 20\text{\AA}$ , 265 Wassermoleküle). Mit allen drei Startstrukturen wurde eine MD-Simulation mit fixiertem Zentral-O-Atom durchgeführt (CVFF, 300 K,  $t_0 = 50$  ps,  $\epsilon = 1$ ,  $t_s = 300$  ps,  $\Delta t_s = 1$  ps). Aus den jeweils 300 gespeicherten Strukturen wurden die Radial-Histogramme für verschiedene Atompaare berechnet (Abb. X159, Abb. X160)

Das zentrale Wassermolekül zeigt das typische Verhalten eines guten Wasserstoffbrückenbindungspartners: Sowohl für  $O/H_{\text{solvens}}$  als auch für  $H/O_{\text{solvens}}$  findet man im Radial-Histogramm bei  $1.7 \text{\AA}$  einen Peak, gefolgt von einem deutlichen Minimum bei  $2.3 \text{\AA}$ . Im Bereich  $< 3 \text{\AA}$  findet man für O/

O<sub>solvens</sub> vier Nachbarn (Abb. X161, links). Aus dem Verlauf des Radial-Histogramms läßt sich also aus der Sicht des zentralen Wassermoleküls auf eine Nah-Ordnung des Solvenswasser im Bereich zwischen 1.5 Å und 4 Å schließen. Bei Methanol ist das Akzeptorverhalten des O-Atoms deutlich weniger stark ausgeprägt als beim Wasser. Die Donor-Eigenschaften von Methanol und Wasser sind jedoch vergleichbar. Die Radial-Population bleibt ab einem Radius  $r > 5$  Å konstant („Bulk“-Wasser) und spiegelt das H:O-Verhältnis im Solvens (2:1) wieder.

Aus den Radial-Histogrammen für die Schweratome (Abb. X160) ist ersichtlich, daß Methanol eine Zwischenstellung zwischen Wasser und Methan einnimmt. Das C-Atom der Methylgruppe „sieht“ praktisch dieselbe Solvensumgebung wie das C-Atom des Methans. Die Solvens-O-Atome sind vom C-Atom des Methan(ol)s um ca. 0.5 Å weiter entfernt als vom O-Atom des Wassers bzw. Methanols. Im Bereich  $< 3.2$  Å findet man für Methan durchschnittlich nur ein Solvens-O-Atom (Abb. X161, rechts), ein Indiz für den hydrophoben Charakter von Methan.

Die Radial-Histogramme eignen sich also sehr gut um Nahordnungen des Solvens in unmittelbarer Umgebung ausgewählter Atome zu untersuchen.

### Interaktion mit dem Solvens in Abhängigkeit von der Orientierung der Hydroxymethylgruppen

Experimentell [211,212] wurde gefunden, daß die Hydroxymethyl-Gruppen von Glucose-Einheiten die tg-Orientierung praktisch nicht populieren. Führt man Kraftfeldrechnungen im Vakuum ( $\epsilon=1$ ) durch, so erhält man jedoch aufgrund der starken OH6-O4- bzw. OH4-O6-Wechselwirkung (H-Brücke) praktisch ausschließlich „tg“. Durch Erhöhung der Dielektrizitätskonstanten läßt sich der Anteil zwar auf ca. 10-20% tg (CVFF) reduzieren, es liegt aber damit immer noch ein nicht zu vernachlässigender Anteil an tg vor, was nicht im Einklang mit den experimentellen Ergebnissen ist.

Um zu Überprüfen, ob der gg-Zustand auf Kosten des tg-Zustands durch spezifische Interaktion mit Solvensmolekülen stabilisiert wird [201], wurden drei MD-Simulation von Glc $\alpha$ -O-iProp mit verschieden orientierter Hydroxymethylgruppe durchgeführt (CVFF, 300K, PBC (20x20x20),  $\epsilon = 1$ ,  $t_0 = 50$  ps,  $t_s = 300$  ps,  $\Delta t_s = 1$  ps). Die Orientierung der Hydroxymethylgruppe wurde durch eine zusätzliche Kraftkonstante von 10 kcal/rad<sup>2</sup> („torsion force“) festgehalten. Die Analyse der Solvensumgebung für die Atome O4, OH4, O5, O6 und OH6 durch Radial-Histogramme (Abb. X162) zeigt, daß die Wahrscheinlichkeit ein Solvens-O-Atom im Abstand 1.7 Å von OH4 und OH6 zu finden, im gt-Zustand größer sind als in den beiden anderen Zuständen. Man kann daraus schließen, daß OH4 und OH6 im gg-Zustand und noch etwas mehr im tg-Zustand dazu neigen intramolekulare H-Brücken auszubilden. Eine Bestätigung dieser Schlußfolgerung erhält man durch Bestimmung der mittleren Anzahl der intra- und intermolekularen H-Brücken (Abb. X162, unten). Die mittlere Anzahl der intramolekularen H-Brücken nimmt minimal in der Reihenfolge gt<gg<tg zu. Insgesamt findet man im Mittel zehn H-Brücken für Glc $\alpha$ -O-iProp, wobei im gt-Zustand das intra/inter-Verhältnis ca. 1/9 ist, im tg-Zustand dagegen 2/8 und der gg-Zustand liegt dazwischen. Aus den Radial-Histogrammen für O4/H<sub>solvens</sub> ergibt sich eine deutliche Erniedrigung der Akzeptoreigenschaft des O4-Atoms gegenüber den Solvens-H-Atomen für den tg-Zustand. Dies ist leicht nachzuvollziehen, da im tg-Zustand auch im Solvens die intramolekulare H-Brücke zwischen O4 und O6 zu einem bedeutenden Anteil ausgebildet sein sollte. Die Detailanalyse der H-Brücken<sup>1</sup> für den tg-Zustand (Abb. X163) zeigt, daß während 85% der Simulationszeit zwischen O4 und O6 eine intramolekulare H-Brücke ausgebildet ist, wobei die Richtung O6-OH6→O4 deutlich bevorzugt (3:1) ist. Im Bezug auf die Interaktion mit Wasser erhält man aus der H-Brücken-Analyse, daß O4 wesentlich weniger die Rolle des H-Akzeptors einnimmt als O6. O4 wirkt gegenüber dem Wasser häufiger als H-Donor.

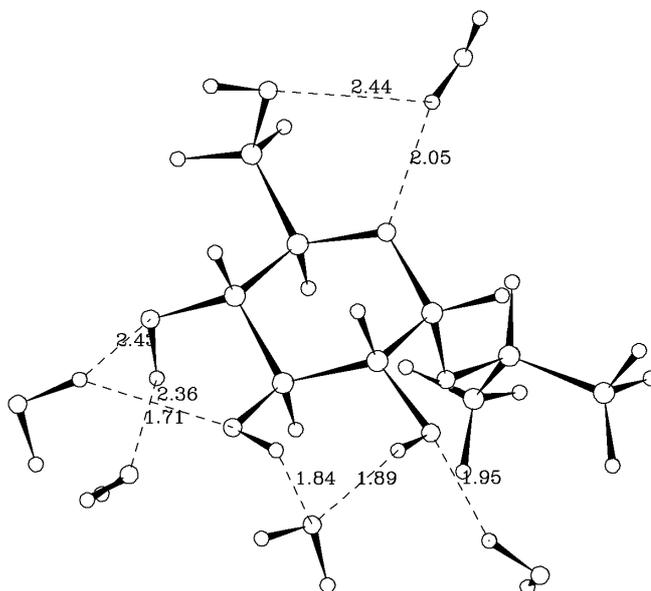
1. Die H-Brücken wurden mit Hilfe der DISCOVER-Funktion „print hbonds“ ausgegeben und anschließend mit CAT2000 analysiert.

Die Analyse der mittleren Energien für die drei Orientierungen gg, gt und tg anhand der MD-Trajektorien ergibt die energetische Reihenfolge  $gt < tg < gg$ . Auffällig ist, daß der tg-Zustand eine um ca. 1.5 kcal/mol niedrigere intramolekulare Energie aufweist als die anderen Zustände. Die Ursache hierfür dürfte in der Ausbildung der intramolekularen H-Brücke zwischen O6 und O4 liegen. Sowohl gg als auch gt weisen dafür eine günstigere intermolekulare Energie auf als tg.

**Tabelle 3: Mittlere Energien der Hydroxymethylrotameren von Glc $\alpha$ -O-iProp in Wasser**

Orientierung	Systemenergie [kcal/mol]	E(pot) [kcal/mol]	E(val) [kcal/mol]	E(intra) [kcal/mol]	E(inter) [kcal/mol]
tg	-1456.72	-4.411	24.448	-4.128	-24.731
gg	-1452.304	-4.145	23.912	-2.682	-25.376
gt	-1456.884	-4.575	24.49	-2.573	-26.491

Die Analyse der Strukturen durch 2D-Radial-Histogramme, bei denen die Solvensumgebung von zwei „Zentralatomen“ analysiert wird, ergab, daß gleichzeitige Interaktionen von einem Wassermolekül mit zwei Donor- bzw. Akzeptoratomen (Abb. 39) des Zuckers relativ häufig vorkommen (bis zu 50% der Simulationszeit für einzelne Atompaare) und daher in einer Diskussion über stabilisierende Effekte durch das Solvens berücksichtigt werden müssen.



**Abb. 39:** Verschiedene Möglichkeiten zur Interaktion zwischen Kohlenhydrat und Wasser.

### 3.7.2 Minimierungen von solvatisierten Molekülen

In diesem Abschnitt soll überprüft werden, inwieweit relative Energien einzelner minimierter Solvens-Strukturen Aussagekraft hinsichtlich der relativen Stabilität des solvatisierten Systems besitzen. Dies ist vor allem im Hinblick auf Protein-Kohlenhydrat-Komplexe wichtig, da hier oft relative Wechselwirkungsenergien zur Bewertung der Bindungsaffinität verwendet werden.

Im ersten Schritt wurde das „einfachste“ denkbare solvatisierte System untersucht: ein Wassermolekül in einer  $20\text{\AA} \times 20\text{\AA} \times 20\text{\AA}$ -Wasserbox. Dieses System wurde 50 ps bei 300 K equilibriert und anschließend minimiert (Konvergenz:  $0.006 \text{ kcal/\AA}$ , 1541 Schritte). Um diesen Prozeß möglichst oft zu wiederholen, wurde das System durch eine 10 ps-Equilibrierungsphase bei 200 K aus dem Minimum herausgehoben („gestört“) und durch eine aufwendige Minimierungssequenz (Steepest Descent / 100 / 0.1, Conjugate Gradient / 5000 / 0.01, Steepest Descent / 5000 / 0.1, Conjugate Gradient / 20000 / 0.01, BFGS (Newton Raphson) / 20000 / 0.001)<sup>1</sup> neu minimiert. Dieser Zyklus wurde zwanzigmal durchlaufen („MD-Minimierung“). Die einzelnen Energien der minimierten Strukturen sind in Abb. X164 gezeigt. Insgesamt nimmt die Systemenergie mit zunehmender Anzahl der durchlaufenen MD-Minimierungszyklen ab, jedoch ist weder die Systemenergie noch die Energie des zentralen  $\text{H}_2\text{O}$ -Moleküls mit akzeptabler Verlässlichkeit ( $\Delta E < 1 \text{ kcal/mol}$ ) reproduzierbar. Die Molekülenergie schwankt über einen Bereich von 5 kcal/mol, die Systemenergie gar über einen Bereich von 30 kcal/mol (Abb. X167). Konkret heißt das, daß schon die Energie eines solvatisierten Moleküls, bei dem nicht einmal Konformationsänderungen auftreten können, um mehrere kcal/mol schwankt, also in einem Bereich, der nach der Boltzmann-Formel schon über Existenz oder nicht Nichtexistenz eines Systemzustands entscheidet. Dies macht die relative Bewertung von minimierten Solvensstrukturen allein aufgrund ihrer Energie sehr fragwürdig. Die Aufschlüsselung der Molekülenergie in die einzelnen Terme (Abb. X166) zeigt, daß die Schwankungen in erster Linie bei der intermolekularen Wechselwirkungsenergie auftreten.

Um zu überprüfen, wie sich die Energierterme eines Zuckermoleküls bei der Minimierung in einer Wasserbox verhalten, wurde als Modellverbindung  $\text{Glc}\alpha\text{-O-iProp}$  in den drei Hydroxymethylorientierungen gg, gt und tg mit der bereits beschriebenen Methode (PBC, 250 Wassermoleküle) minimiert. Tendenziell nimmt auch hier die Systemenergie mit zunehmender Anzahl der durchlaufenen MD-Minimierungs-Zyklen ab, teilweise sogar unter Zunahme der Molekülenergie von  $\text{Glc}\alpha\text{-O-iProp}$  (Abb. X165). Die Coulomb-Stabilisierung des Gesamtsystems liegt bei ca.  $-1500 \text{ kcal/mol}$ . Angesichts dieser Zahl ist es leicht nachzuvollziehen, daß das System die Erhöhung der Energie eines Teilchens um ein paar kcal/mol „in Kauf nimmt“, wenn sich das Gesamtsystem dadurch stabilisieren kann. Die Energie des solvatisierten Moleküls wird also in erster Linie durch die Wechselwirkung mit dem Solvens bestimmt (Abb. X166) und diese ist wiederum davon abhängig wie sich das Gesamtsystem stabilisiert. Im Hinblick auf Protein/Kohlenhydrat-Komplexe ist es daher durchaus denkbar, daß es im Einzelfall bei einer Minimierung vorkommen kann, daß das Protein oder der Zucker durch das Solvens in eine sehr energiereiche Konformation gezwungen wird, die in keinsten Weise repräsentativ für das Molekül ist, aber vielleicht zufällig eine hohe Wechselwirkungsenergie zwischen Zucker und Protein verursacht. Mit dem Auswahlkriterium „maximale Wechselwirkungsenergie“ würde man so sicher nicht den richtigen Vorschlag für die Konformation im gebundenen Zustand machen. Als Schlußfolgerung aus den Rechnungen ergibt sich damit: Die Molekülenergie einzelner Strukturen schwankt im Bereich von einigen kcal/mol absolut zufällig. Ein Vergleich von verschiedenen Systemen oder Konformationen basierend auf wenigen minimierten Einzelstrukturen ist daher nicht repräsentativ. Daher kann im vorliegenden Fall auch keine sichere Aussage darüber gemacht werden, welche der drei Hydroxymethyl-Orientierungen von  $\text{Glc}\alpha\text{-O-iProp}$  am stabilsten ist. Die mittlere

1. Die Reihenfolge der Angaben ist: Minimierungsmethode / maximale Anzahl der Schritte / Konvergenzkriterium. Es hat sich gezeigt, daß sich durch Wechsel der Minimierungsmethoden bessere Minimierungsergebnisse erzielen lassen.

Energiedifferenz der verschiedenen Hydroxymethyl-Orientierungen ist viel kleiner als die Energieschwankungen innerhalb eines Hydroxymethyl-Zustands (Abb. X166). Energieunterschiede zwischen verschiedenen Konformationen eines solvatisierten Moleküls von wenigen kcal/mol können erst durch die Analyse einer großen Anzahl von minimierten Solvens-Strukturen statistisch verlässlich bestimmt werden. Eine gute Minimierung von Solvens-Strukturen ist jedoch technisch extrem aufwendig und daher nur in speziellen Fällen angebracht. Die Methode der MD-Minimierung hat sich in diesem Zusammenhang auch bei anderen Systemen bewährt. Für die Analyse der Wechselwirkungsenergie von Protein/Kohlenhydrat-Komplexen ergibt sich aus den in diesem Abschnitt erhaltenen Resultaten, daß es günstiger ist die Analyse der Wechselwirkungsenergie anhand einer möglichst großen Anzahl von MD-Strukturen durchzuführen. Dadurch entfällt die aufwendige Minimierung der Strukturen und der Solvens-Einfluß ist weniger anfällig für Ausreißer.

### 3.7.3 Einfluß der Dielektrizitätskonstanten und des Solvens auf die Freie Gridgruppen-Energie

Ein Vergleich der FGE-Profile der Hydroxymethylgruppen von  $\beta$ -Maltose für unterschiedliche Dielektrizitätskonstanten zeigt für  $\epsilon=1$  bei beiden  $\omega$ -Winkeln (G56O, L56O) eine Stabilisierung der tg-Konformation gegenüber der gg-Konformation um ca. -2 kcal/mol (Abb. X168, Abb. X169). Bei  $\epsilon=4$  und  $\epsilon=80$  besitzt die tg-Konformation ungefähr dieselbe Energie wie gg (G56O), bzw. ist um 1.0-1.4 kcal/mol gegenüber dieser destabilisiert (L56O). Ursache dieser Stabilisierung der tg-Konformation bei  $\epsilon=1$  ist die Ausbildung der H-Brücke GO4 $\cdot\cdot$ H $\cdot\cdot$ GO6 (G56O) bzw. eine Doppelinteraktion GO5 (LO4)  $\cdot\cdot$ H-LO6 (L56O) (siehe Abb. 27, links). Ein Vergleich mit den FGE-Profilen einer 12 ns-MD-Simulation, die in einer Wasserbox der Größe 20Å x 20Å x 20Å (PBC) bei 400 K und  $\epsilon = 1$  durchgeführt wurde, zeigt deutlich, daß die Stabilisierung der tg-Konformation aufgrund der Ausbildung intramolekularer H-Brücken durch Anwesenheit von expliziten Wassermolekülen signifikant reduziert wird, da für die beiden OH6-Protonen und das GOH4-Proton nun alternative H-Akzeptoren zur Verfügung stehen. Derselbe Effekt kann jedoch auch durch Erhöhung der Dielektrizitätskonstanten erreicht werden. Daß die Dielektrizitätskonstante jedoch nicht beliebig erhöht werden kann, zeigt sich an der relativen Energie des tg-Zustands. Die beste Übereinstimmung mit der Simulation in Wasser erhält man mit einer Dielektrizitätskonstanten im Bereich zwischen zwei und vier.

Analoge Ergebnisse erhält man aus der Analyse der Abhängigkeit des FGE-Profiles der OH-Gruppen von der Dielektrizitätskonstanten (Abb. X170). Hier reproduziert eine Dielektrizitätskonstante zwischen drei und vier ziemlich genau die Ergebnisse wie man sie aus einer Simulation in Wasser erhält.

Es gibt jedoch auch Effekte, die sich nicht durch Wahl einer geeigneten Dielektrizitätskonstante simulieren lassen. Für die OH-Gruppe L3OH ist ohne Solvens ( $\epsilon=4$ ) die Orientierung  $-60^\circ$  und  $180^\circ$  wegen der Möglichkeit zur Ausbildung intramolekularer H-Brücken gegenüber der Orientierung  $+60^\circ$  um ca. 0.3 kcal/mol stabilisiert<sup>1</sup> (Abb. X171). In Wasser ist hingegen die Orientierung  $+60^\circ$  (LOH3 in trans-Stellung zu LH3) genauso stabil wie  $-60^\circ$ . Hier ist die Orientierung  $180^\circ$  um 0.3 kcal/mol destabilisiert. Die Stabilisierung von  $+60^\circ$  läßt sich mit der Ausbildung von H-Brücken zum Wasser erklären. Bei einer Simulationstemperatur von 1000 K wird das FGE-Profil von L3OH instabil, außerdem ist der Solvenseffekt für  $+60^\circ$  nicht mehr vorhanden.

Die Analyse der FGE-Maps von  $\Phi/\Psi$  (Abb. X172) zeigt, daß auch die Lage des  $\Phi/\Psi$ -Minimums von Maltose von der Dielektrizitätskonstanten abhängt. Für  $\epsilon = 1$  ist das Minimum in Richtung positiver Werte für  $\Phi/\Psi$  verschoben (Minimum B und C). Vor allem im Minimum C ( $+20^\circ/+20^\circ$ ) ist die Anzahl der intramolekularen H-Brücken besonders hoch (Abb. X174).

---

1. Für  $\epsilon = 1$  ist die Stabilisierung von  $-60^\circ$  gleich 3.5 kcal/mol

Die FGE-Map von  $\Phi/\Psi$ , die aus einer 12 ns-Simulation in einer Wasserbox (PBC, 20x20x20) erhalten wurde (Abb. X172), zeigt eine Stabilisierung des Minimums C (+20°/+10°). Die 1 kcal/mol - Contourline, die bei einer Simulation ohne Solvens ausschließlich den Bereich von Minimum A und B einschließt, umschließt bei der MD-Simulation mit Solvens zusätzlich auch noch den Bereich von Minimum C. Eine Temperaturstudie (Abb. X173) zeigt deutlich, daß dieser „Solvenseffekt“ bei hohen Temperaturen (> 600 K) nicht mehr vorhanden ist. Eine Erklärungsmöglichkeit für den stabilisierenden Effekt des Wassers für  $\Phi/\Psi$ -Konformation C ist folgende: Durch die Orientierung der L3OH-Gruppe nach  $\pm 60^\circ$  und weniger nach  $180^\circ$  sind die Voraussetzung zur Ausbildung der interresidualen H-Brücke GOH2-LO3 günstiger als ohne Solvens. Wie in Abschnitt 3.3.1 gezeigt, stabilisiert diese H-Brücke Minimum C.

Aus den HTMD-FGE-Maps für  $\Phi/\Psi$  wird jedoch deutlich, daß die wesentliche Gestalt der FGE-Map in PBC auch bei Temperaturen von 1200 K noch erhalten bleibt.

Insgesamt kann jedoch aus den Ergebnissen geschlossen werden, daß es nicht immer unbedingt notwendig ist, die MD-Simulation eines nicht allzu großen Kohlenhydrats mit explizit vorhandenem Lösungsmittel durchzuführen. Sofern keine spezifischen Wechselwirkungen mit einzelnen Solvensmolekülen auftreten, lassen sich die elektrostatischen Bedingungen im Solvens auch durch die Wahl einer geeigneten Dielektrizitätskonstanten annähernd simulieren. Es hat sich gezeigt, daß eine Dielektrizitätskonstante  $\epsilon = 3-4$  die Verhältnisse in Wasser am besten beschreibt. Anders liegen die Verhältnisse für große langgestreckte Moleküle, da diese in Abwesenheit von Solvensmolekülen aufgrund der van-der-Waals-Anziehung „zusammenknäueln“. In diesem Fall sind die Solvensmoleküle als „Abstandshalter“ notwendig.

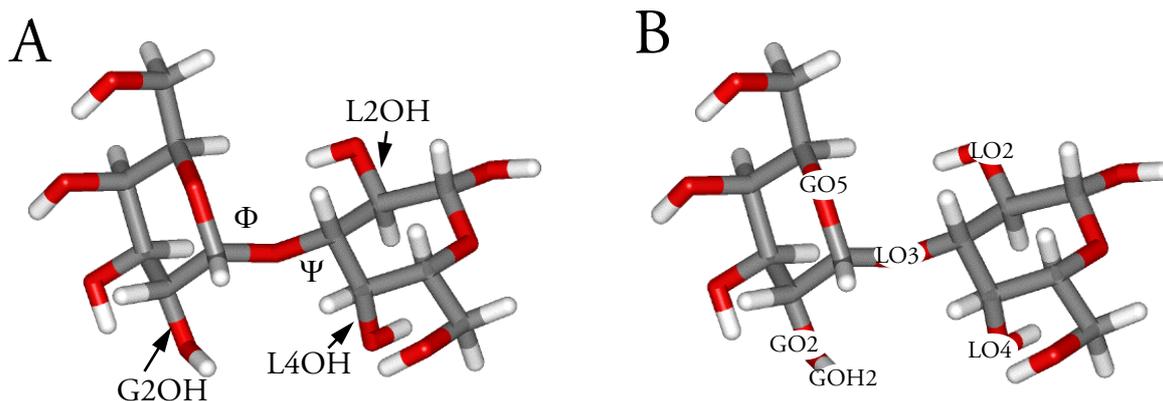
### 3.7.4 Analyse spezifischer Solvens-Effekte

Eine meßbare Stabilisierung bestimmter Zuckerkonformationen durch „Komplexierung“ einzelner Solvensmoleküle (vgl. Abb. 39) ist theoretisch sehr schwer vorherzusagen. Ein Grund dafür ist, daß die Solvensmoleküle, die eine Stabilisierung bewirken könnten, während einer MD-Simulation laufend austauschen und so eine große Anzahl von Molekülen beobachtet werden muß. Außerdem gibt es bei größeren Kohlenhydraten unzählige Möglichkeiten wo ein Solvens-Effekt theoretisch auftreten kann. Zum Schluß bleibt noch die Frage wodurch sich ein eventuell vorhandener Solvens-Effekt überhaupt bemerkbar macht, d.h. es muß ein Kriterium gefunden werden, das es erlaubt einen Solvens-Effekt zu detektieren.

Im Folgenden ist ein Beispiel beschrieben, bei dem ein spezifischer Solvens-Effekt gefunden wurde, der auch durch experimentelle Daten belegt werden kann:

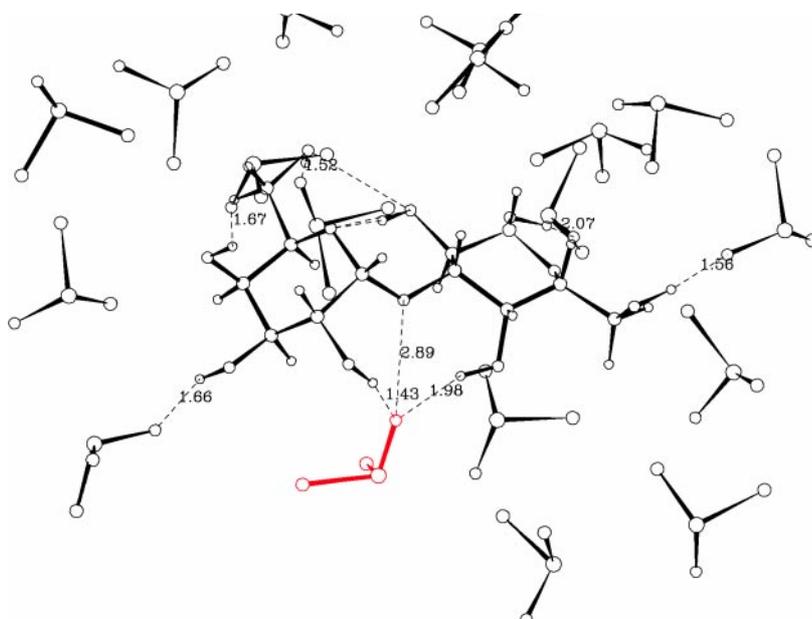
MD-Simulationen von Gal $\beta$ 1-3Gal $\beta$  und Gal $\alpha$ 1-3Gal $\beta$  wurden sowohl in der „Gasphase“ ( $\epsilon=4$ ) als auch in DMSO (PBC, 25Åx25Åx25Å)[167] durchgeführt (CVFF,  $t_s = 1$  ns, 300 K). Die Populations-Analyse der interresidualen H-H-Abstände ausgewählter OH-Gruppen (Abb. X175, Abb. X176) zeigt einen deutlichen „Solvens-Effekt“: Die Häufigkeitsverteilung des Abstands der Protonen GOH2-LOH4 besitzt für Gal $\alpha$ 1-3Gal $\beta$  in DMSO ein ausgeprägtes Maximum bei 2.6 Å. Für Gal $\beta$ 1-3Gal $\beta$  in DMSO erhält man bei diesem Abstand auch ein deutliches Maximum für den Abstand der Protonen GOH2-LOH2. Beide Maxima sind bei der Simulation in der Gasphase nicht vorhanden. Die Analyse der FGE-Profile ergibt für Gal $\alpha$ 1-3Gal $\beta$  in DMSO eine deutliche Stabilisierung (> 1kcal/mol) für L4OH = -60° und G2OH = -60° (Abb. X177). Die Dreizentren-Interaktions-Analyse<sup>1</sup> von CAT2000 (Abb. X178) ergibt, daß die Ursache für dieses Verhalten in einer chelat-arti-

1. Bei der Dreizentren-Interaktions-Analyse wird eine Interaktions-Trajektorie nach gleichzeitigen Wechselwirkungen dreier „Subsets“ abgesucht. „Subsets“ können einzelne Atome, funktionelle Gruppen oder ganze Moleküle sein.



**Abb. 40:** Definition der Torsionswinkel und Atome von Gal $\alpha$ 1-3Gal $\beta$  (A) und Gal $\beta$ 1-3Gal $\beta$  (B)  
 $\Phi$ (Phi)=GH1-GC1-LO3-LC3,  $\Psi$ (Psi)=GC1-LO3-LC3-LH3  
 G2OH =GC1- GC2-GO2-GOH2, L4OH =LC3- LC4-LO41-LOH4, usw.

gen Komplexbildung des Zuckers mit einem einzelnen DMSO-Molekül liegt (Abb. 41). Dabei bildet das O-Atom des DMSO-Moleküls gleichzeitig eine H-Brücke mit GOH2 und LOH4 aus und stabilisiert dadurch nicht nur die Orientierung der Torsionswinkel G2OH und L4OH bei  $-60^\circ$ , sondern fixiert damit den Abstand der beiden OH-Atome bei  $2.6 \text{ \AA}$  was sich experimentell durch einen starken NOE der beiden Atome zeigen sollte. Weiterhin führt die Komplexierung dazu, daß im  $\Phi/\Psi$ -Bereich  $10^\circ/0^\circ$  (Abb. X179) eine Energiebarriere entsteht und so das, in der Gasphase schon leicht angedeutete, Minimum  $+30^\circ/+10^\circ$  neben dem globalen Minimum bei  $-40^\circ/-20^\circ$  deutlich hervortritt. Experimentell konnten die Rechnungen durch einen starken NOE für G2OH-L4OH bei Gal $\alpha$ 1-3Gal $\beta$  bestätigt werden [213].



**Abb. 41:** Stabilisierung einer  $\Phi/\Psi$ -Konformation durch spezifische Interaktion mit einem Solvensmolekül (Gal $\alpha$ 1-3Gal $\beta$  in DMSO)

### 3.8 Zusammenfassung

Wie in den vorangegangenen Abschnitten gezeigt wurde, verhält sich die Rotamerenverteilung der glycosidischen Torsionswinkel  $\Phi$  und  $\Psi$  und der Hydroxymethylgruppen selbst bei Simulationstemperaturen weit über 1000 K konsistent mit dem zugrundeliegenden Modell. Auf Basis der FGE-Maps bzw. der FGE-Profile konnte gezeigt werden, daß bei Hochtemperatur-Simulationen von Oligosacchariden in der Gasphase ( $\epsilon=4$ ) selbst bei Temperaturen von 1800 K keine Artefakte hinsichtlich der Population des  $\Phi, \Psi, \omega$ -Torsionsraums auftreten. Voraussetzung für die Simulation bei extrem hohen Temperaturen ist lediglich die Verwendung eines harmonischen Potentials für die Bindungslängen um extreme Bindungslängen zu vermeiden. Weiterhin ist es angebracht, die bei diesen Temperaturen auftretende Ringinversion durch Anlegen eines Zusatzpotentials („torsion force“) zu unterbinden. HTMD-Simulationen stellen eine schnelle und effiziente Methode dar, um den Konformationsraum nach stabilen Konformationen zu durchsuchen. Die hierfür benötigte Simulationszeit hängt von den Energiebarrieren für die Rotationsübergänge ab. Sie liegt für  $\beta$ -Maltose und Sialyl-Lewis<sup>X</sup> bei 1000 K in der Größenordnung von 10-20 ns, eine Konformationsanalyse benötigt daher eine CPU Zeit von ca. 10-20 Stunden (IBM SP2, Single Processor). Mit der GCMD-Methode wurde eine Möglichkeit zur effizienten Konformationsuche bei 300 K vorgestellt. Diese Methode ist geeignet, um schnell alle  $\Phi/\Psi$ -Minima eines mittelgroßen Oligosaccharids zu populieren und für die lokalen  $\Phi/\Psi$ -Minima ein Rotamerengleichgewicht der OH-Gruppen einzustellen. Daher ist es möglich durch GCMD die lokalen Eigenschaften von  $\Phi/\Psi$ -Minima bei Raumtemperatur zu berechnen. Sowohl HTMD- als auch GCMD-Methode besitzen das Potential wichtige Informationen zur Strukturaufklärung von Oligosacchariden mittels NMR-Methoden beizutragen (vgl. Abschnitt 3.3.3).

Weiterhin wurde mit EQ ein Parameter entwickelt, der es erlaubt, die Einstellung des Konformerengleichgewichts unabhängig von der Anzahl der miteinander im Gleichgewicht stehenden Konformationen anzuzeigen. Aus den Trajektorien Daten eines Gleichgewichtsensembles lassen sich mit CAT2000 vollautomatisch alle Rotationsprofile der Freien Energie, FGE-Maps, deskriptive Statistiken, mittleren Lebensdauern, Populationswahrscheinlichkeiten usw. sowohl für die gesamte Simulationszeit, als auch für einzelne sequentielle Perioden oder pro Gruppe (Konformation) berechnen. Es wurden Cluster- und Gruppierungsmethoden entwickelt, die umfangreiche mechanistische Studien basierend auf Trajektorien-Daten ermöglichen. Die Durchführung von MD-Simulationen bei verschiedenen Simulationstemperaturen ermöglicht es aus den mittleren Lebensdauern der Konformationen deren effektive Übergangsbarrrieren durch Arrhenius-Analysen zu bestimmen. Hierzu wurde die technische Umsetzung entwickelt und in CAT2000 implementiert. Weiterhin wurden Methoden zum Aufspüren von spezifischen Solvenseffekten entwickelt. Schlußendlich ist es durch Anwendung einer der vorgestellten Sampling-Methoden (HTMD-CC oder GCMD) und der anschließenden Analyse der Daten mit CAT2000 möglich, eine Konformationsanalyse eines (theoretisch) beliebig komplexen Oligosaccharids innerhalb weniger Tage durchzuführen. Dieses war bis jetzt, je nach Größe des Oligosaccharids, technisch unmöglich oder die Durchführung hat mehrere Wochen Arbeits- und Rechenzeit in Anspruch genommen. So daß insgesamt eine deutliche Effizienzsteigerung in der Konformationsanalyse von Oligosacchariden realisiert werden konnte.

## 4 Berechnung von Rotationsprofilen der Freien Energie basierend auf Grid-Search-Methoden

### 4.1 Einleitung und Problemstellung

Die experimentell durch NMR-Methoden gefundene Rotamerenverteilung [211,212,214,215] der Konformere, die durch Rotation der CH<sub>2</sub>OH-Gruppe („Hydroxymethylgruppe“) um die C5-C6 Einfachbindung erhalten wird, stellt nach wie vor eines der großen ungelösten Probleme in der kraftfeldbasierten Konformationsanalyse von Kohlenhydraten dar.

Für Glucose wurde experimentell, durch Auswertung der vicinalen Kopplungskonstanten (vgl. Abschnitt 7.2) zwischen H5 und H6R bzw. H5S (vgl. Abb. 7), kein Hinweis dafür gefunden, daß eine Konformation, bei der die Hydroxymethylgruppe die tg-Stellung einnimmt, in Lösung in nennenswerter Konzentration vorliegt („anti-tg-Effekt“). Theoretisch wird die Instabilität einer tg-Konformation für Glucose mit einer ungünstigen syn-axialen bzw. peri-Wechselwirkung durch die parallele Anordnung der C6-O6- und C4-O4-Bindungen erklärt [216]. Ein nach dieser Argumentation ebenfalls zu erwartender „anti-gg-Effekt“ für Galactose existiert nicht.

Kraftfeldrechnungen in Vakuum ( $\epsilon=1$ ) weisen für Glucose, wegen der Möglichkeit zur Ausbildung einer starken intramolekularen H-Brücke zwischen OH6 und O4, immer die tg-Konformation als energieärmste Konformation aus [201,217,218]. Die meisten Autoren machen für die fehlende Übereinstimmung mit dem Experiment „Solvenseffekte“ verantwortlich [157,201,217,219-221] und manche davon sind der Meinung durch MD Simulationen in Wasser von wenigen hundert Picosekunden Hinweise auf eine Instabilität der tg-Konformation gefunden zu haben [201].

Diverse Kraftfeld-Entwickler versuchen das Problem durch eine spezielle Parametrisierung zu lösen. Beispiele hierfür sind GROMOS [157,222,223] und AMBER/HOMANS [147].

Das Problem ist aber nach wie vor nicht zufriedenstellend gelöst. Falls wirklich das Lösungsmittel einen destabilisierenden Einfluß auf die tg-Konformation haben sollte, dann wäre eine spezielle Kraftfeldparametrisierung, welche die tg-Konformation destabilisiert, ziemlich unsinnig, denn dies würde bedeuten, daß man bei einer Simulation von Glucose in einer Wasserbox mit einem modifizierten Kraftfeld den „anti-tg-Effekt“ praktisch doppelt berücksichtigt.

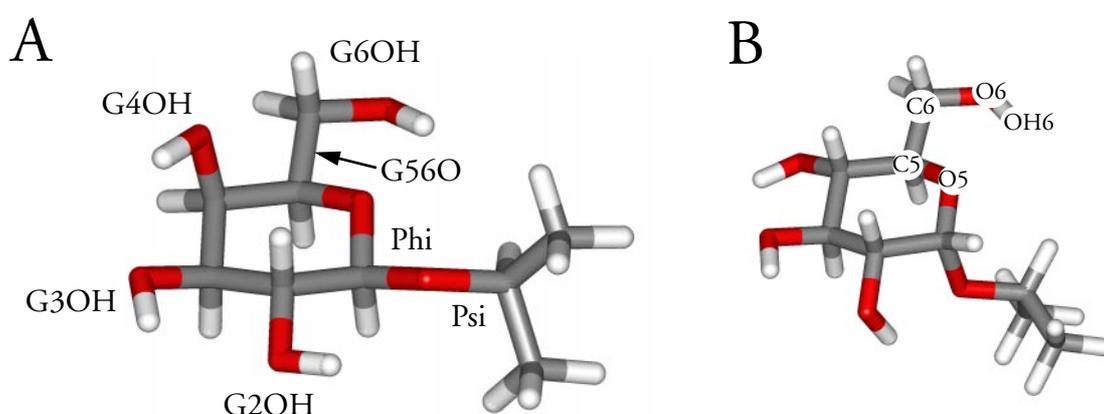
Spezielle Kraftfeldparametrisierungen sind nur dann sinnvoll, wenn intramolekulare Eigenschaften durch ein Kraftfeld nicht richtig beschrieben werden. Falls es sich beim „anti-tg-Effekt“ um einen Solvenseffekt handelt, dann muß man die Simulationsbedingungen ändern und nicht das Kraftfeld.

Es gilt nun mit Hilfe der in DISCOVER implementierten Kraftfelder (CVFF [173,174], CFF91 [224], AMBER/Homans [225,226]) für verschiedene Dielektrizitätskonstanten das Rotamerengewicht der Hydroxymethylgruppen für Glucose und für Galactose mit Grid-Search-Methoden zu berechnen. Die Ergebnisse sollen anschließend mit Ergebnissen aus Ab-Initio-Rechnungen verglichen werden um eine evtl. vorliegende unzureichende Parametrisierung der Kraftfelder aufzuspüren.

## 4.2 Methoden

### 4.2.1 Beschreibung der Methode

Durch die Möglichkeit des OH6-Protons intramolekulare Wasserstoffbrücken mit O4 und O5 auszubilden, ist anzunehmen, daß das Rotationsprofil von der Stellung der OH-Gruppe am C6-Atom abhängt. Ziel ist es nun durch Anwendung der Grid-Search-Methode ein Energieprofil für die Rotation um die C5–C6-Bindung herauszuarbeiten, das diese Effekte berücksichtigt. Der Einfluß einer potentiell möglichen intramolekularen OH4 → O6 Wasserstoffbrücke muß bei einer vollständigen Analyse ebenfalls in Betracht gezogen werden. Da damit aber ein vierdimensionales Problem (drei Torsionswinkel plus Energie) vorliegen würde, wird der Anschaulichkeit wegen der Einfluß der 4OH-Gruppe zunächst als konstant betrachtet. In einem zweiten Schritt soll der Einfluß der 4OH-Gruppe auf das Rotationsprofil analysiert werden. Alle Rechnungen wurden an den Modellverbindungen Glc $\alpha$ -O-iProp und Gal $\beta$ -O-iProp durchgeführt (siehe Abb. 42).



**Abb. 42:** Definition der Torsionswinkel von Gal $\beta$ -O-iProp (A) und Glc $\alpha$ -O-iProp (B)  
 G56O = O5-C5-C6-O6, G6OH = C5-C6-O6-OH6, Phi=H1-C1-O-C, Psi=C1-O-C-H  
 G2OH =H2- C2-O2-OH2, G3OH =H3- C3-O3-OH3, G4OH =H4- C4-O4-OH4

Ausgehend von der Startstruktur wurden systematisch die Torsionswinkel G56O und G6OH in 10°-Schritten rotiert und so alle möglichen Kombinationen von G56O/G6OH-Torsionswinkeln erzeugt. Die G56O-G6OH-Diagramme („Maps“) wurden direkt durch Analyse der 1296 Strukturen - ohne vorherige Minimierung („Rigid Grid“) - hinsichtlich Energie und Struktur (Abstände) generiert. Die Möglichkeit einer OH4 → O6 Wasserstoffbrücke wurde durch die Verwendung einer Startstruktur mit G4OH(Gal) = 87° bzw. G4OH(Glc) = -77° bewußt ausgeschlossen (siehe Abb. 42).

Um den Einfluß des verwendeten Kraftfelds und der Dielektrizitätskonstanten zu untersuchen, wurden die Rechnungen für CVFF, CFF91 und AMBER/Homans (jeweils mit den Dielektrizitätskonstanten  $\epsilon=1, 4$  und 80) durchgeführt, so daß insgesamt für jede Modellverbindung neun Sätze à 1296 „Rigid Grid“-Strukturen erzeugt und analysiert wurden. (CPU Zeit: 3 min. / Grid Search)

Um den Einfluß der Stellung der 4OH-Gruppe auf die Rotamerenverteilung zu untersuchen, wurde exemplarisch für Gal $\beta$ -O-iProp ein vierdimensionaler Grid-Search durchgeführt. Dabei wurden die Torsionswinkel G56O, G6OH, G4OH und G3OH systematisch um 20° rotiert (siehe hierzu Abschnitt 4.3.5).

Alle angegebenen Energiewerte in den Karten sind relative Energiewerte. Die Systemenergie wurde jeweils relativ zum globalen Energieminimum berechnet, dies ermöglicht eine sehr gute Vergleichbarkeit der Energiewerte innerhalb einer Energie-Karte. Energievergleiche zwischen verschiedenen Grid-Ensembles (Maps) sind auf dieser Basis jedoch nicht möglich, da das globale Minimum nicht notwendigerweise identisch ist. Die Valenz-, van-der-Waals- und Coulomb-Energie wurde daher relativ zum Frame 1 ( $G56O = 180^\circ$  /  $G6OH = 180^\circ$ ) berechnet. Dadurch sind Energievergleiche zwischen den einzelnen Kraftfeldern möglich, da alle Energien relativ zu derselben Konformation berechnet wurden. Die Winkelskala des Torsionswinkels  $G56O$  wurde in den Bereich  $[+90^\circ, 450^\circ]$  verschoben, damit die Minima der Coulomb-Energie in der Mitte der Karten zu liegen kommen. Technischer Hintergrund für dieses Vorgehen ist, daß Contour-Plots mit Extrema in der unmittelbaren Nähe der Skalengrenzen unschöne Randeffekte aufweisen.

#### 4.2.2 Berechnung von Rotationsprofilen der Freien Energie aus multidimensionalen Grids

Energiekarten (dreidimensional) sind für eine umfangreichere quantitative vergleichende Diskussion unpraktisch, da schon eine Überlagerung oder Differenzbildung von drei Karten nicht mehr möglich ist. Da in der vorliegenden Problemstellung sowieso nur die  $G56O$ -Dimension interessiert, stellt sich die Frage, wie man am sinnvollsten die Verhältnisse im  $G56O$ - $G6OH$ -Energie-Raum in die  $G56O$ -Energie-Ebene abbildet. Die Projektion des  $G56O$ - $G6OH$ -Energie Raumes in die  $G56O$ -Energie-Ebene bewirkt, daß die Dimensionalität des, der Analyse zugrundeliegenden, Grids reduziert wird. Die Variable  $G56O$  bleibt weiterhin eine Gridvariable und  $G6OH$  wird zu einer „Nichtgrid-Variablen“ (vgl. Abschnitt 2.3). Die einfachsten Methoden der Reduktion der Dimensionalität eines Grids sind Minimumsuche (Minimum-Methode, Adiabatic Grid) und Mittelwertbildung (Mittelwert-Methode). Voraussetzung für die Mittelwert-Methode ist die Annahme, daß die Energieverteilung in den anderen (Nichtgrid-)Dimensionen für jede Gridgruppe dieselbe ist, was jedoch oft nicht der Fall ist. Die Minimum-Methode beruht auf der Tatsache, daß nach Boltzmann diejenige Struktur die größte Wahrscheinlichkeit besitzt, die die niedrigste Energie aufweist. In der praktischen Anwendung wird daher bei der Minimum-Methode aus dem Pool der erzeugten Grid-Strukturen für jede Gridgruppe diejenige Struktur mit der geringsten Energie herausgesucht. Sofern sichergestellt ist, daß das Energieminimum für jede  $G56O$ -Gridgruppe sicher, d.h. reproduzierbar, bestimmt werden kann, was bei Grid-Search-Methoden, im Gegensatz zu Moleküldynamik-Methoden, normalerweise der Fall ist, ist die Minimum-Methode der Mittelwert-Methode vorzuziehen. Der Nachteil der Minimum-Methode besteht in der Tatsache, daß voneinander verschiedene Strukturen einer Gridgruppe, die nur eine geringfügig höhere Energie als das Gridgruppen-Minimum besitzen, bei der Projektion und damit bei der energetischen Betrachtung des Ensembles nicht berücksichtigt werden. Wie im folgenden noch im Detail gezeigt wird (vgl. Abb. 43, rechts), ist die Energieverteilung in der  $G6OH$ -Dimension nicht für jede  $G56O$ -Gridgruppe gleich. Dies hat zur Folge, daß für eine  $G56O$ -Gridgruppe die Energieniveaus sehr nahe beieinanderliegen können (wie bei  $G56O = tg$ ) und für einen anderen sehr weit auseinander (wie bei  $G56O = gg$  und  $G56O = gt$ ). Sehr nahe beieinander liegende „Energieniveaus“ bedeutet jedoch, daß die Zahl der energetisch zugänglichen „Realisierungsmöglichkeiten“ für eine Gridgruppe, und damit die Entropie für eine Gridgruppe zunimmt, was schließlich zu einer Erniedrigung der Freien Energie für diese Gridgruppe führen sollte.

$$\frac{N_i}{N_j} = e^{-\frac{E_i - E_j}{kT}} \quad \text{Boltzmannsches Verteilungsgesetz} \quad (41)$$

Jeder Struktur läßt sich mit Hilfe der Boltzmann-Formel (41) anhand ihrer Energiedifferenz zum globalen Minimum eine relative Wahrscheinlichkeit bzw. relative Häufigkeit bei einer bestimmten Temperatur  $T$  zuordnen. Diese relativen Häufigkeiten lassen sich verwenden um die relative Population einer Gridgruppe niedrigerer Dimensionalität bei  $T$  zu berechnen, welcher wieder über Formel (14) (vgl. Abschnitt 2.3.4) eine relative Energie zugeordnet werden kann.

Konkret werden durch einen Grid-Search alle Kombinationsmöglichkeiten bzw. „Realisierungsmöglichkeiten“ für die Gridvariablen erzeugt. Gewichtet man diese anschließend über ihre Systemenergie und die Temperatur mit Hilfe der Boltzmann-Verteilungsfunktion, kann für eine beliebige Gridvariable eine Häufigkeitsverteilung berechnet werden, welche wiederum über die Boltzmann-Formel und die Temperatur (zurück) in ein Energieprofil („Boltzmann-gewichtetes Energieprofil“, „Freie Gridgruppen-Energie“) umwandelbar ist (Abb. 44).

Mit der Methode des boltzmann-gewichteten Energieprofils lassen sich Grids beliebiger Dimensionalität auf eindimensionale Grids reduzieren. In der Freien Gridgruppen-Energie sind die Energieinformationen aller ursprünglichen Gridvariablen enthalten. Der Vorteil des rigid-grid-basierten Ansatzes zur Erzeugung eines Energieprofils ist, daß alle erzeugten Strukturen voneinander verschieden sind und sich außer den Gridvariablen an der Struktur des Moleküls nichts ändert.

## 4.3 Ergebnisse aus dem Grid-Search

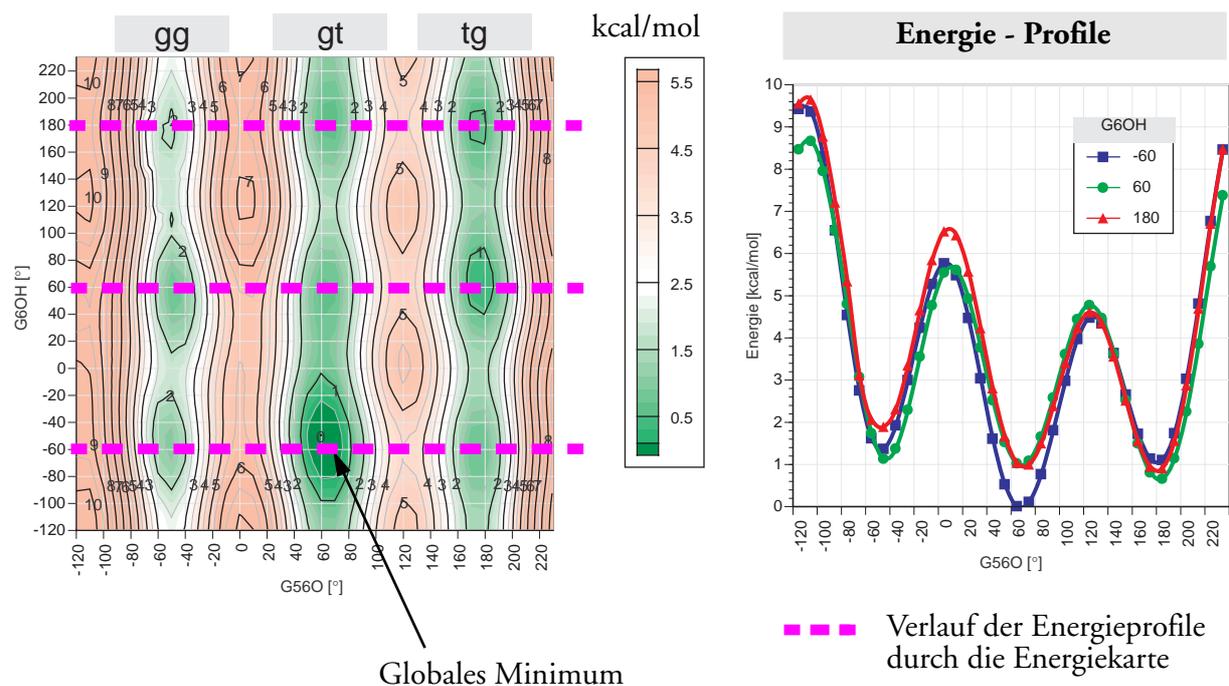
### 4.3.1 Zweidimensionale Energiekarten

Da die primäre Zielsetzung dieses Abschnitts die Rotamerenverteilung von G56O und nicht die Abhängigkeit der Energie von G56O/G6OH ist, soll an dieser Stelle nur kurz beispielhaft auf die generierten Energiekarten von Gal $\beta$ -O-iProp eingegangen werden um ein paar wesentliche Punkte zu veranschaulichen:

Aus den Energiekarten (vgl. Abb. 43; Details siehe Abb. X180 bis Abb. X185) ist deutlich ersichtlich, daß das Energieprofil für den Torsionswinkel G56O wie vermutet vom Torsionswinkel G6OH abhängt. Das globale Minimum liegt für alle getesteten Kraftfelder ( $\epsilon=4$  bzw.  $4^*r$  für AMBER/Homans) bei G56O = 60° (gt) und G6OH = -60°. Schnitte durch die Energiekarte entlang der G56O-Richtung bei G6OH = -60°, +60°, 180° (siehe Abb. 43, rechts) zeigen deutlich eine Stabilisierung des globalen Minimums gegenüber den anderen (lokalen) Minima um  $\geq 0.8$  kcal/mol. Wesentlicher Punkt ist jedoch, daß für G56O = 180° (tg) die drei Profile eine ähnliche Energie aufweisen, während für G56O = -60° (gg) das G6OH [180]-Profil eine deutlich höhere (0.5 kcal/mol) Energie aufweist, als die beiden anderen (G6OH [-60] und G6OH [60]). Dies dürfte einen deutlichen Einfluß auf das Rotationsprofil von G56O haben (siehe Abschnitt 4.3.2).

Trägt man die einzelnen Energie-Terme (Valenz-, Van-der-Waals und Coulomb-Energie) als Funktion von G56O und G6OH auf (vgl. Abb. X180 bis Abb. X185), so fallen sofort die zwei Minima der Coulomb-Energie auf (A: G56O = 260° (-100°) / G6OH = 10°, B: G56O = 360° (0°) / G6OH = -10°). Diese lassen sich eindeutig der Wechselwirkung des Protons OH6 mit O4 (Minimum A) bzw. mit O5 (Minimum B) zugeordnen, da sich für G56O = 260° nur OH6 $\cdot$ O4 und für G56O = 0° nur OH6 $\cdot$ O5 ausreichend nahe kommen ( $< 2.5\text{\AA}$ ) (vgl. Abb. X188). Der Einfluß des Kraftfelds auf die Energiekarten wurde durch Differenz-Maps analysiert (siehe Abb. X186, Abb. X187).

Im Vergleich zum CVFF-Kraftfeld zeigt CFF91 (bei  $\epsilon = 4$ ), deutlich höhere Übergangsbarrieren zwischen gg und gt (Interaktion O6  $\leftrightarrow$  O5), wohingegen die Übergangsbarriere zwischen gg und tg (Interaktion O6  $\leftrightarrow$  O4/C4) beim CFF91-Kraftfeld niedriger ist (Abb. X181). Ein Torsionswinkel



**Abb. 43:** Galβ-O-iProp: Abhängigkeit der Systemenergie von den Torsionswinkeln G560 und G6OH. (Gridintervall = 10°; Kraftfeld = CVFF;  $\epsilon=4$ .)

G6OH  $\approx 0^\circ$  wird vom CFF91 nahezu über den gesamten G560-Torsionsbereich als energiereicher bewertet als beim CVFF-Kraftfeld. Eine Aufschlüsselung der Differenz-Maps in die einzelnen Energierterme zeigt als Ursache eine ungünstigere Valenzenergie (siehe Abb. X187), wohingegen als Ursache für die niedrigere gg  $\leftrightarrow$  tg - Barriere die Van-der-Waals-Energie zu finden ist. Bemerkenswert ist noch die Region erhöhter Energie bei G560 = 270° / G6OH = 0°. Ursache hierfür ist eine van-der-Waals-Repulsion zwischen OH6 und O4, die sich beim CFF91 schon leicht abzeichnet ( $\approx 2$  kcal/mol) und beim AMBER/Homans-Kraftfeld zum dominierenden Faktor wird. Beim AMBER-Kraftfeld schießt die Energie in diesem Bereich um über 100 kcal/mol in die Höhe, wodurch die Region G560 [220°, 300°] / G6OH [-60°, 60°] auf der Energiekarte förmlich zur „verbotenen Zone“ erklärt wird. Dadurch wird z.B. verhindert, daß bei einer Erniedrigung der Dielektrizitätskonstanten auf  $\epsilon = 1$  das globale Minimum, wie bei den beiden anderen Kraftfeldern, nach G560 = -60° (gg) und G6OH = -40° wechselt. In dieser Region befindet sich (auch beim AMBER-Kraftfeld) das weiter oben schon erwähnte stark negative Coulomb-Potential, das durch eine OH6  $\cdot\cdot$  O4 Anziehung (H-Brücke) zustande kommt.

Für die G560-G6OH-Energiekarten von Gl $\alpha$ -O-iProp (siehe Abb. X196 bis Abb. X200) gilt im Prinzip dasselbe wie für Galβ-O-iProp. Auffälligste Unterschiede sind die Verschiebung des Minimums A der Coulomb-Energie von G560 = 260° / G6OH = 10° in den Bereich um G560 = 200° / G6OH = 0° und die Halbierung der van-der-Waals-Abstoßung im Bereich G560=210° / G6OH = 0° (für CFF91 und AMBER/Homans). Ersteres ist die Folge der Tatsache, daß die H-Brücke OH6  $\rightarrow$  O4 nunmehr in der tg-Stellung ausgebildet wird, letzteres kann mit den höheren Atomabständen (siehe Tabelle 4) erklärt werden.

**Tabelle 4: Atomabstände (vgl. Abb. X188)**

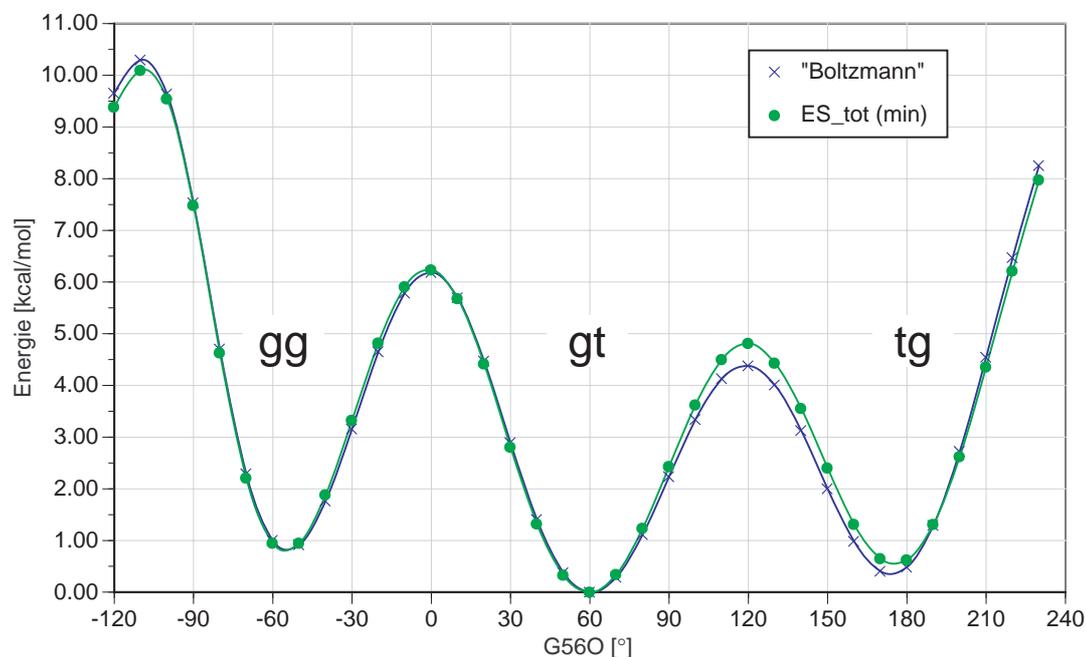
System	$(O4...O6)_{\min}/$ $(O5...O6)$ [Å]	$(O4...OH6)_{\min}$ [Å]	$(O5...O6)_{\min}/$ $(O4...O6)$ [Å]	$(O5...OH6)_{\min}$ [Å]
Gal	2.40 / 3.26	1.56	2.48 / 3.65	1.88
Glc	2.58 / 3.61	1.66	2.48 / 4.35	1.88

### 4.3.2 FGE-Profile

Mit der Methode des boltzmann-gewichteten Energieprofils lassen sich Grids beliebiger Dimensionalität auf eindimensionale Grids reduzieren (Abschnitt 4.2.2). Als Ergebnis findet man für Gal $\beta$ -O-iProp eine Erniedrigung der Energie für tg um 0.2 kcal/mol (0.9 kJ/mol) gegenüber dem Energieprofil, das nach der Minimum-Methode bestimmt wurde (Abb. 44).

Für Glc $\alpha$ -O-iProp ergibt sich eine Erhöhung der Energie für tg gegenüber gg und gt um 0.2 kcal/mol (0.9 kJ/mol) (Abb. X201)

Die Aufspaltung der Energieprofile ist abhängig von der Dielektrizitätskonstanten (Abb. X190).



**Abb. 44:** Gal $\beta$ -O-iProp: Rotationsprofil des Torsionswinkels G56O gewonnen aus „Rigid Grid“-Daten. CVFF,  $\epsilon = 4$

„Boltzmann“ = Freie Grid-Energie des G56O-Profiles (siehe Text) bei 300K

ES\_tot(min) = Energie des G56O-Profiles nach der Minimum-Methode

### 4.3.3 Abhängigkeit der Rotamerenverteilung vom Kraftfeld und der Dielektrizitätskonstanten

Auf Basis der Rotationsprofile der Freien Gridgruppen-Energie (FGE-Profile) läßt sich der Einfluß des Kraftfelds (Abb. X192) und der Dielektrizitätskonstanten (Abb. X191) quantitativ diskutieren (vgl. Tabelle 5, Tabelle 6). Dies wird im Folgenden für Gal $\beta$ -O-iProp im Detail durchgeführt, für Glc $\alpha$ -O-iProp werden entsprechende Effekte beobachtet. Eine vergleichende Diskussion findet sich in Abschnitt 4.3.4.

Für Gal $\beta$ -O-iProp unterscheiden sich die Kraftfelder in erster Linie in der energetischen Bewertung der tg-Konformation. Während AMBER/Homans die tg-Konformation als ungünstigste Konformation ausgibt, ist diese für das CFF91-Kraftfeld das globale Minimum (vgl. auch Häufigkeitsverteilung Abb. X194). Experimentell [211,212] findet man für D-Galactose-Derivate je nach Lösungsmittel und Stellung der Galactose-Einheit im Gesamtmolekül unterschiedliche Populationsverteilungen (gg(15-25%), gt(30-60%), tg (20-50%)). Für Methyl- $\alpha$ -D-galactopyranosid wurde in D<sub>2</sub>O ein Rotamerenverhältnis von (gg(14%), gt(47%), tg(39%)) bestimmt [212]. Zusammenfassend ergibt sich aus den NMR-Messungen die Tendenz  $E(\text{gg}) > E(\text{tg})$ ,  $E(\text{gt})$ , wobei das gg/gt-Verhältnis von Fall zu Fall vollkommen unterschiedlich ausfallen kann.

**Tabelle 5: Relative Energien für Gal $\beta$ -O-iProp ( $\epsilon=4$ )(vgl. Abb. X192)**

Zustand	CVFF [kcal/mol]	CFF91 [kcal/mol]	AMBER [kcal/mol]
gg	0.8	1.1	0.8
gt	0	0	0
tg	0.4	-0.1	1
ÜZ(gg $\leftrightarrow$ gt)	6.2	8.4	2.4
ÜZ(gt $\leftrightarrow$ tg)	4.3	4.3	2.9
ÜZ(tg $\leftrightarrow$ gg)	10.3	8.6	10.5

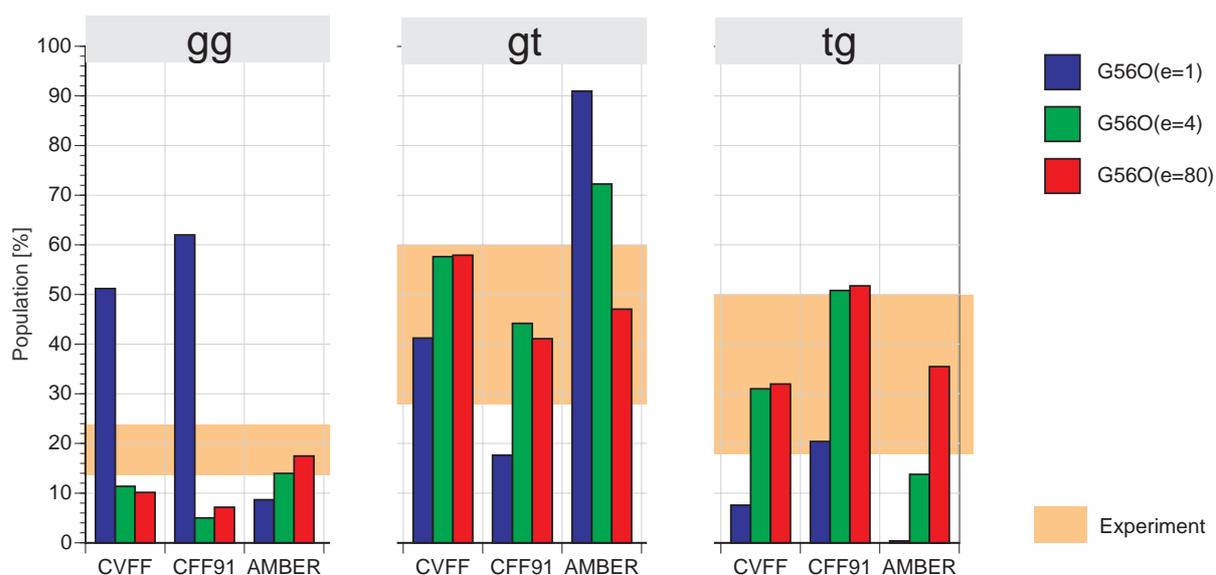
**Tabelle 6: Relative Energien für Glc $\alpha$ -O-iProp ( $\epsilon=4$ ) (vgl. Abb. X193)**

Zustand	CVFF [kcal/mol]	CFF91 [kcal/mol]	AMBER [kcal/mol]
gg	0.7	-0.1	0.3
gt	0	0	0
tg	0.5	1.1	1.2
ÜZ(gg $\leftrightarrow$ gt)	6.5	8.6	2.2
ÜZ(gt $\leftrightarrow$ tg)	4.3	4.3	3.1
ÜZ(tg $\leftrightarrow$ gg)	8.9	6.7	6.9

Wie schon bei den Differenz-Maps (Abschnitt 4.3.1) gefunden, zeigt das CFF91 (bei  $\epsilon = 4$ ) im Vergleich zum CVFF-Kraftfeld, eine deutlich (2 kcal/mol) höhere Übergangsbarriere zwischen gg und gt (Interaktion O6  $\leftrightarrow$  O5), wohingegen die Übergangsbarriere zwischen gg und tg (Interaktion O6  $\leftrightarrow$

O4/C4) um 1.7 kcal/mol niedriger ist. AMBER/Homans zeigt sowohl für den (gt  $\leftrightarrow$  gg)- als auch für den (gt  $\leftrightarrow$  tg)-Übergangszustand eine auffällig niedrigere Energie als die beiden anderen Kraftfelder. Mit 2.4 bzw. 2.9 kcal/mol sind diese sogar kleiner als die experimentell bestimmte von Ethanol (Tabelle 14), was als äußerst unwahrscheinlich zu bewerten ist, da im Ethanol eine ekliptische Stellung von H-C-C-O und beim Zucker eine ekliptische Stellung von O-C-C-O vorhanden ist, welche eine höhere Energie besitzen sollte. Vor allem die Unabhängigkeit der (gt  $\rightarrow$  gg)-Übergangsbarriere von der Dielektrizitätskonstanten (Abb. X191) ist gedanklich schwer nachzuvollziehen. Dieses Verhalten ist angesichts der Tatsache, daß sich hier zwei elektonegative Sauerstoffatome (O6, O5) „aneinander vorbeischieben“ sehr verwunderlich. Zwar ist eine Stabilisierung durch das Proton OH6 anzunehmen, die gerade die Coulomb-Abstoßung zwischen den O-Atomen kompensiert, doch sollte dieser Effekt auch bei der Übergangsbarriere (gt  $\leftrightarrow$  tg) auftreten können (Wechselwirkung mit O4), hier besteht jedoch eine deutliche Abhängigkeit der Höhe der Barriere von der Dielektrizitätskonstanten.

Die berechneten Rotamerenverteilungen für verschiedene Kraftfelder und Dielektrizitätskonstanten sind für Gal $\beta$ -O-iProp in Abb. 45 zusammengefaßt. Durch die Möglichkeit eine intramolekulare Wasserstoffbrücke zwischen O6 und O4 bzw. O5 auszubilden, kommt es zu einem extremen Ansteigen der Population des gg-Zustands bei CVFF und CFF91 für  $\epsilon=1$ . Dies entspricht zwar nicht dem experimentell gefundenen Wert (wobei zu beachten ist, daß die NMR-Messungen bei höheren Dielektrizitätskonstanten durchgeführt wurden), ist aber vom zugrunde liegenden Modell ein logisches Ergebnis. Das AMBER/Homans-Kraftfeld reproduziert für  $\epsilon=80^*r$  sehr gut die Rotamerenverteilung, die experimentell gefunden wurde, für  $\epsilon=1^*r$  versagt es jedoch komplett, da hier das gg-Rotamer deutlich unterrepräsentiert ist. Hier zeigt sich, daß eine zusätzliche Parametrisierung des AMBER/Homans-Kraftfelds vorliegt, die einzig und allein deshalb vorgenommen wurde (vgl. auch Abschnitt 4.3.1) um das Rotamerengleichgewicht, daß bei niedrigen Dielektrizitätskonstanten durch eine OH6 $\rightarrow$ O4 Wasserstoffbrückenbindung (und damit gg) dominiert würde, in Richtung gt und tg und damit experimenteller Resultate zu verschieben. Experimentell gibt es Hinweise, daß die Population der gg-Konformation beim Übergang zu Lösungsmitteln mit niedrigeren Dielektrizitätskonstanten zunimmt (gg-Population: H<sub>2</sub>O( $\epsilon=80$ ): 20%, DMSO( $\epsilon=47$ ): 29%, Dimethylformamid ( $\epsilon=37$ ):



**Abb. 45:** Rotamerenverteilung für Gal $\beta$ -O-iProp berechnet aus der boltzmann-gewichteten relativen Gridgruppen-Population (300K) für ein G56O-G6OH-Grid. gg=G56O[-120, 0], gt=G56O[0, 120], tg=G56O[120,240]

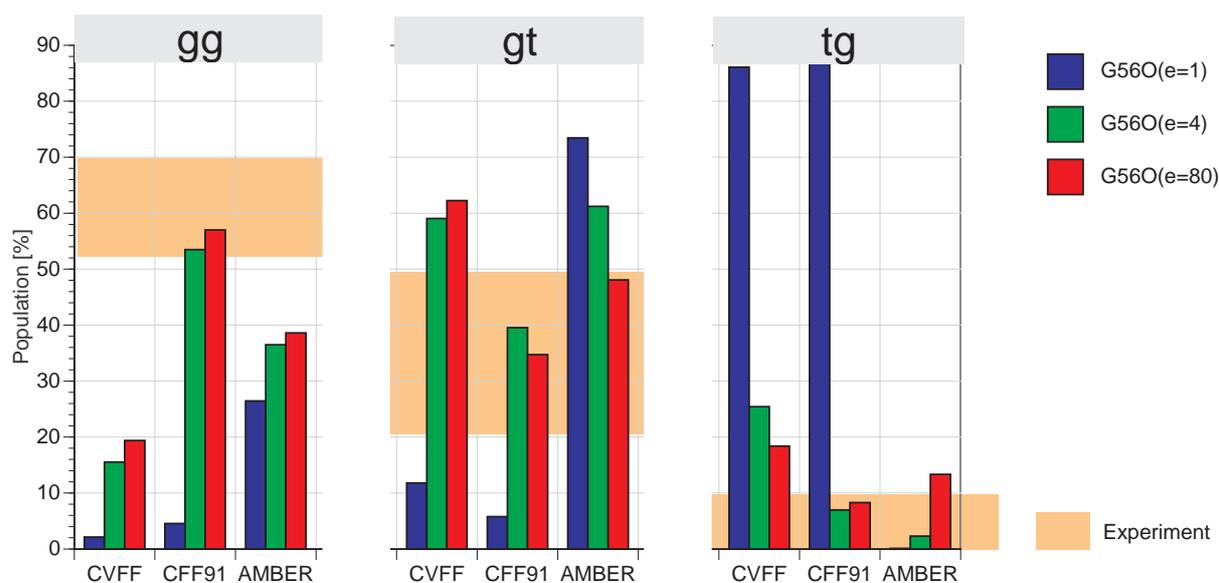
42 %) [211]. Es soll jedoch nicht verschwiegen werden, daß der Trend nicht immer so ausgeprägt ist, wie im angeführten experimentellen Beispiel. Einen Trend zunehmender gg-Population mit abnehmender Dielektrizitätskonstanten zeigt lediglich das CVFF-Kraftfeld.

#### 4.3.4 Vergleich und Bewertung der Ergebnisse für Glc $\alpha$ -O-iProp und Gal $\beta$ -O-iProp

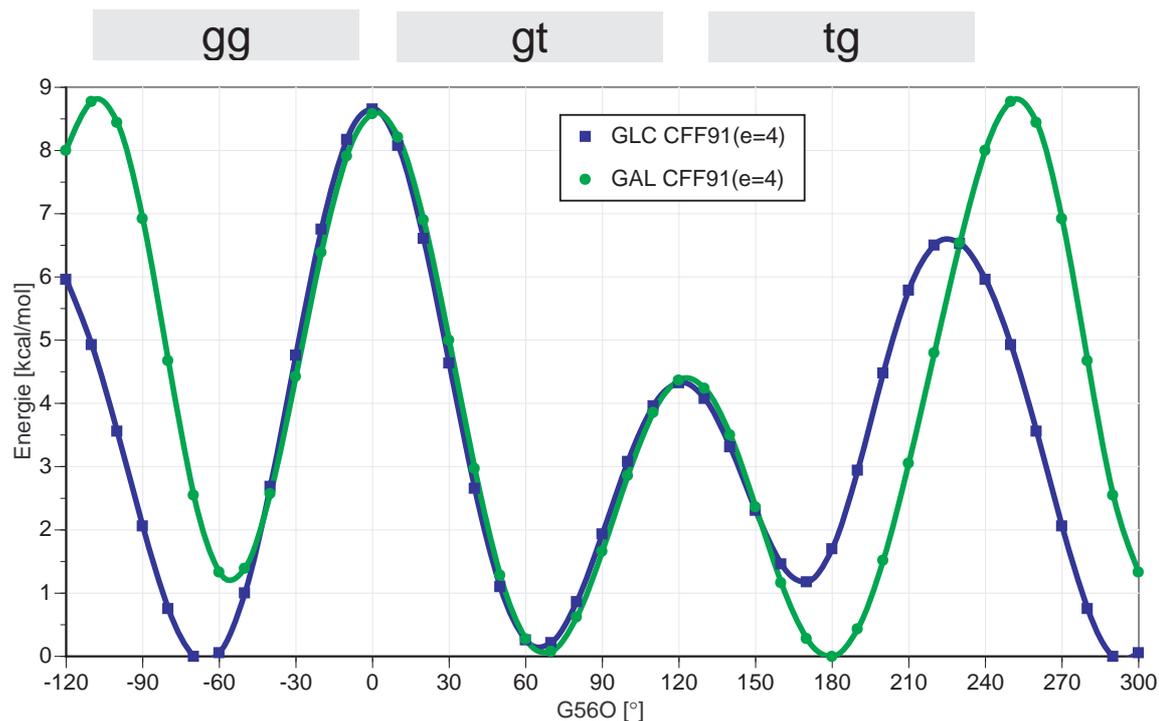
Der Vergleich mit den experimentell gefundenen Ergebnissen (vgl. Abb. 45 und Abb. 46) ergibt eindeutig, daß Rechnungen mit einer Dielektrizitätskonstante  $\epsilon = 1$  für alle Kraftfelder die schlechteste Übereinstimmung mit dem Experiment ergeben, was auch zu erwarten war. Eine Dielektrizitätskonstante  $\epsilon = 80$  ergibt zwar meistens keine wesentlich anderen Ergebnisse als  $\epsilon = 4$ , ein Blick auf den Coulomb-Term der G56O-G6OH-Karten zeigt jedoch, daß dieser für  $\epsilon = 80$  fast überall auf Null reduziert ist. Damit werden die intramolekularen elektrostatischen Wechselwirkungen praktisch ausgeblendet, was sicherlich nicht sehr realistisch ist. Die experimentell gefundene Rotamerenverteilung für Glucose wird am besten vom CFF91-Kraftfeld ( $\epsilon = 4$ ) reproduziert (vgl. auch Abb. 47). Für Galactose gibt es keinen eindeutigen „Sieger“.

Da CFF91 die experimentellen Ergebnisse bei Verwendung einer Dielektrizitätskonstanten  $\epsilon > 1$  auch ohne spezielle Parametrisierung schon relativ gut reproduziert, ist eine Parametrisierung des Kraftfelds durch Einführung eines gauche-Potentials, wie es erst kürzlich auch bei GROMOS [157] durchgeführt wurde, nicht notwendig. Die Auswirkungen einer solchen „Spezialparametrisierung“ kann man bei AMBER/Homans sehen: dieses liefert zwar gute Übereinstimmung mit dem Experiment für  $\epsilon = 80^*r$ , ist aber, wegen der zu niedrigen Übergangbarrieren (siehe Tabelle 5 und Tabelle 6) und dem extremen Anstieg der van-der-Waals-Energie in dem Bereich, in dem eine intramolekulare H-Brücke zwischen OH6 und O4 zu erwarten wäre (Abb. 48), kritisch zu betrachten. CFF91 liefert für die G56O-Rotation diejenigen Energieprofile, die insgesamt am nächsten am Experiment und auch theoretisch am nachvollziehbarsten sind (Abb. 47).

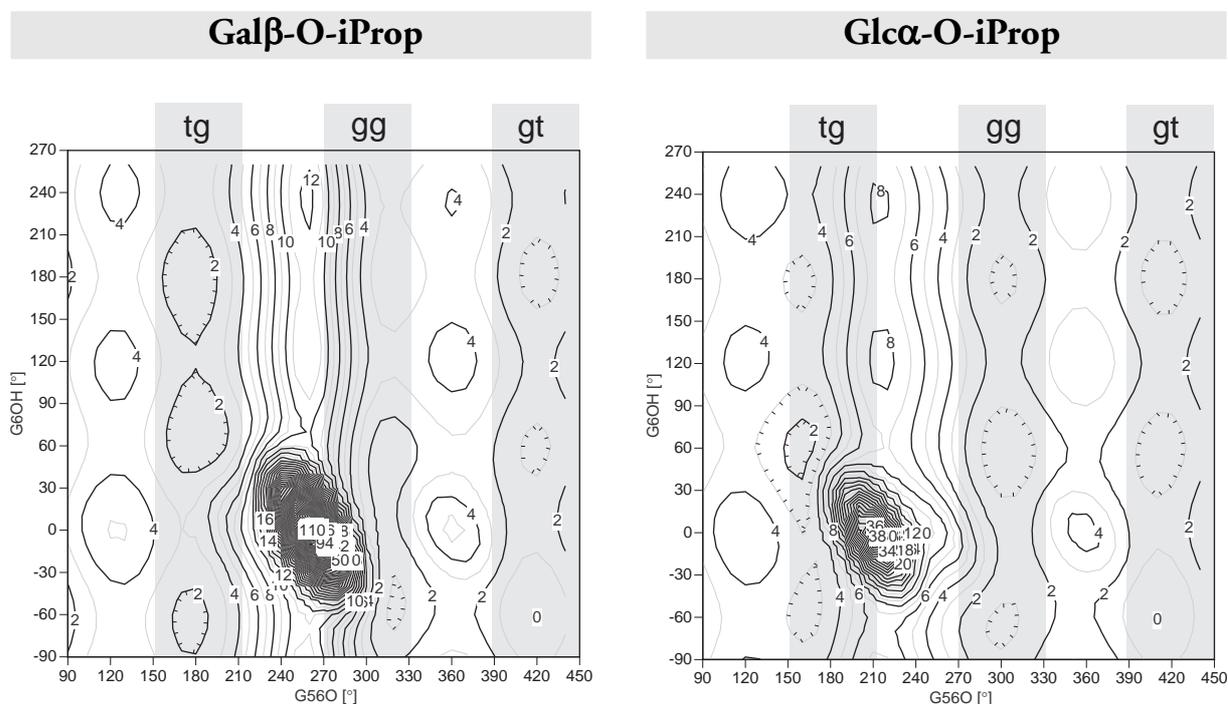
Theoretisch nicht nachzuvollziehen ist jedoch beim CFF91-Kraftfeld die starke Population des G6OH-Torsionswinkels im Bereich von  $180^\circ$  (siehe Abb. X194, Abb. X195). Diese Stellung sollte die instabilste sein, da hier keine Stabilisierung durch intramolekulare H-Brücken erfolgen kann. Diese



**Abb. 46:** Rotamerenverteilung für Glc $\alpha$ -O-iProp berechnet aus der boltzmann-gewichteten relativen Gridpopulation (300K) für ein G56O-G6OH-Grid. gg=G56O[-120, 0], gt=G56O[0, 120], tg=G56O[120,240]



**Abb. 47:** Energieprofile von G56O berechnet aus Grid-Search Daten. „Freie Gridgruppen-Energie“, Datenbasis: G56O/G6OH Grid (Gridintervallbreite = 10°,  $\varepsilon = 4$ , CFF91) GAL = Gal $\beta$ -O-iProp, GLC = Glc $\alpha$ -O-iProp



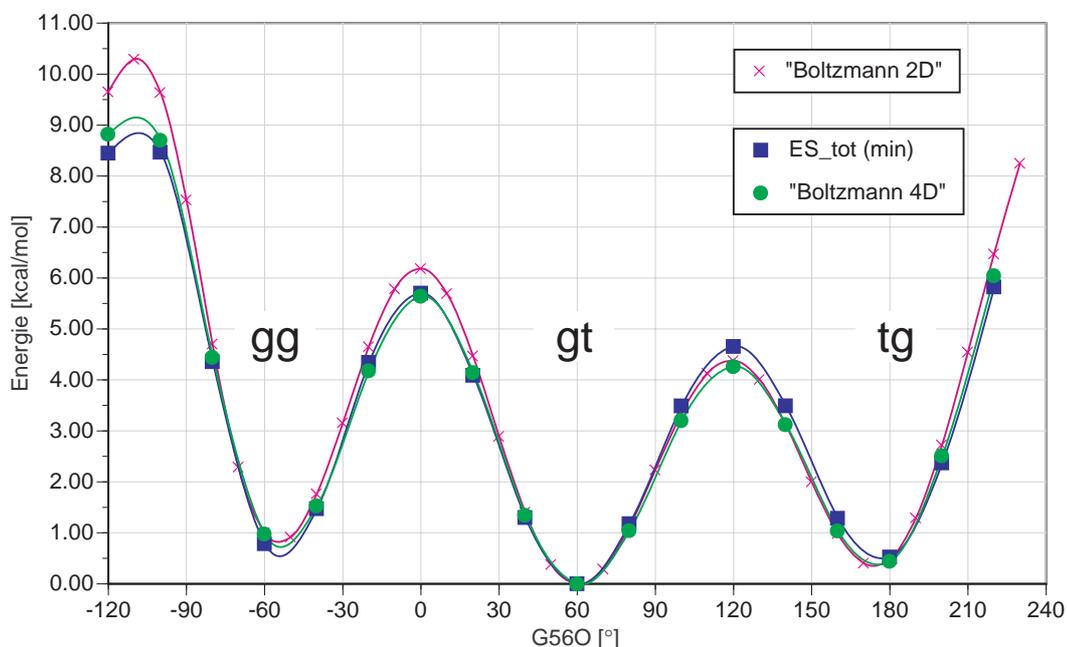
**Abb. 48:** Abhängigkeit der Systemenergie von den Torsionswinkeln G56O und G6OH. Gridintervall = 10°; Kraftfeld = AMBER/Homans;  $\varepsilon=4^*r$ .

logische Annahme wird am besten durch das CVFF-Kraftfeld bestätigt. Insgesamt, d.h. hinsichtlich Rotamerenverteilung, Übergangszustände und G6OH-Profil, liefert das CVFF-Kraftfeld zwar einigermaßen konsistente Ergebnisse, versagt aber hinsichtlich der experimentellen Ergebnisse im Bezug auf den Unterschied in der Rotamerenverteilung der Hydroxymethylgruppe zwischen Galactose und Glucose vollständig - diese ist nämlich, wenn man dem CVFF-Kraftfeld Glauben schenken soll, praktisch diesselbe.

#### 4.3.5 Einfluß der Orientierung der 4OH-Gruppe auf die Rotamerenverteilung

Während beim G56O-G6OH-Grid der Einfluß der Stellung von 4OH, 3OH als konstante Größe betrachtet wurde, soll nun geklärt werden, inwieweit die Rotation von 4OH und 3OH sich auf die Gestalt der G56O-G6OH-Karten und G56O-Profile auswirkt. Bei einem systematischen Grid-Search für drei Torsionswinkel mit einer Gridintervallbreite von  $10^\circ$  würden 46656 Strukturen erzeugt werden. Um aber an die Grenzen des momentan praktisch machbaren vorzustoßen, wurde noch eine weitere Dimension (3OH) in den Grid-Search einbezogen, was bei einer Gridintervallbreite von  $20^\circ$  104976 Strukturen ergibt. Die reine CPU Zeit (IBM SP2, Single Processor) um diese Strukturen zu erzeugen lag bei 5 Stunden, die gesamte Rechnung dauerte jedoch mehr als einen Tag, da viele Festplattenzugriffe erfolgten. Insgesamt wurde von DISCOVER eine Strukturarchiv der Größe 320 MB und eine Protokolldatei von ca. 3 GB erzeugt.

Der vierdimensionale Grid-Search (G56O, G6OH, G4OH, G3OH) wurde für Gal $\beta$ -O-iProp bei  $\epsilon = 4$  durchgeführt, Kraftfeld war CVFF. Um aus den Daten eine G56O-G6OH-Energiekarte bzw. G56O-Energieprofile zu erzeugen, muß wie bei den G56O-Profilen bereits beschrieben, eine Projektion des G56O-G6OH-G4OH-G3OH-Grid-„Hyperraumes“ in die G56O-G6OH-Ebene bzw. G56O-Dimension durchgeführt werden. Hierzu wurde die Freie Gridgruppen-Energie für die G56O-G6OH-Gridgruppen bzw. G56O-Gridgruppen berechnet (vgl. Abschnitt 4.2.2). Die resultie-

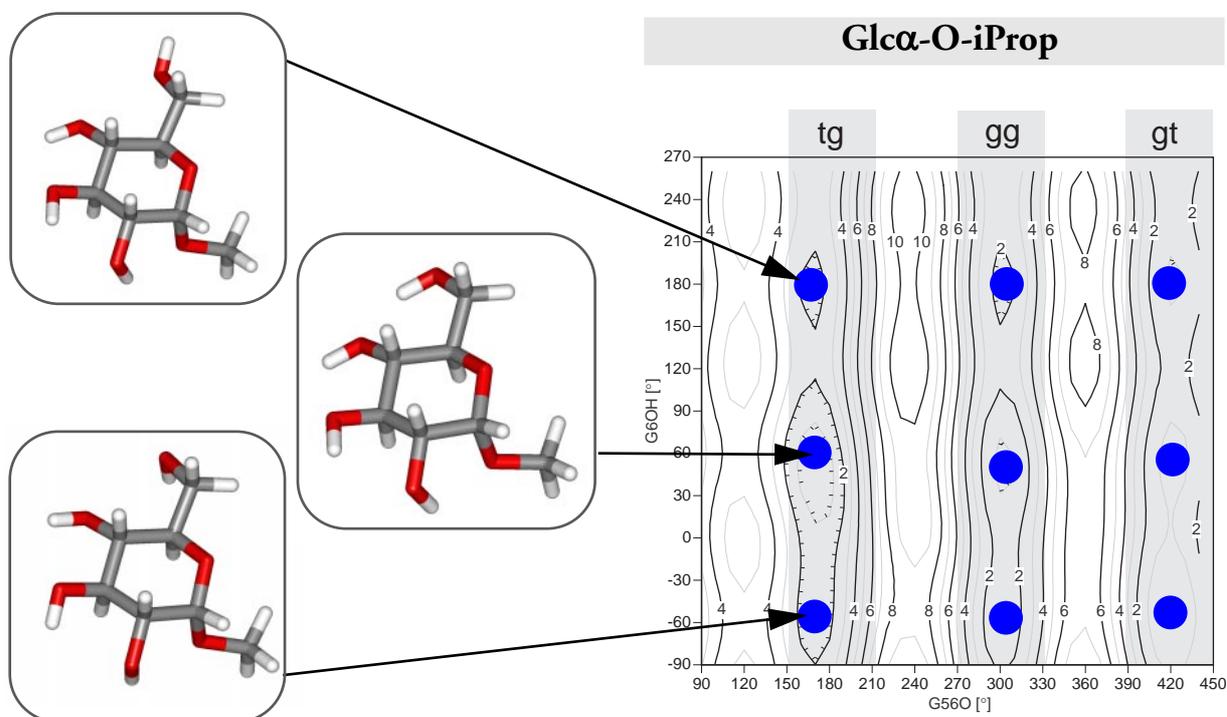


**Abb. 49:** Gal $\beta$ -O-iProp: Energieprofil von G56O berechnet aus Grid Search Daten.  
 Boltzmann: „Freie Grid-Energie“, ES-tot(min): Minimumenergie-Methode („Adiabatic Grid“)  
 2D: G56O-G6OH Grid (Gridintervallbreite =  $10^\circ$ , Kraftfeld = CVFF,  $\epsilon = 4$ )  
 4D: G56O-G6OH-G4OH-G3OH Grid (Gridintervallbreite =  $20^\circ$ , Kraftfeld = CVFF,  $\epsilon = 4$ )

rende Energiekarte ist in Abb. X203 dargestellt. Die Differenzkarte zum 2D-Grid-Search (Abb. X203, rechts) zeigt deutlich den Einfluß der 4OH-Gruppe: im Bereich des Übergangszustands  $gg \leftrightarrow tg$  erfolgt eine deutliche Stabilisierung für den Bereich G6OH [100°, 280°]. Dies ist nur durch eine  $OH4 \rightarrow O6$  H-Brücke zu erklären, da in diesem Bereich das Proton OH6 selbst keine Möglichkeit zu Ausbildung einer H-Brücke hat. Die Analyse der Orientierung von G4OH in Abhängigkeit von G56O und G6OH zeigt eindeutig: in dem Bereich, in dem eine Stabilisierung erfolgt, ist G4OH im Bereich +60° [0°, -120°], d.h. in Richtung O6 orientiert (vgl. Abb. X204, links). Daß die Stabilisierung elektrostatischer Natur ist, zeigt sich auch im Coulomb-Term der G56O-G6OH-Karte (vgl. Abb. X204, rechts). Der Bereich negativen Potentials ist gegenüber dem 2D-Search genau in die Region hin erweitert in der eine Stabilisierung beobachtet wird. Das Rotationsprofil von G56O aus dem 4D-Grid-Search ist in Abb. 49 gezeigt. Aus dem Vergleich des Profils für den 2D-Grid-Search mit denen des 4D-Grid-Search ergibt sich die Schlußfolgerung, daß sich der Einfluß der 4OH-Gruppe im wesentlichen in der Stabilisierung der Übergangszustände  $gg \leftrightarrow tg$  und  $gg \leftrightarrow gt$  zeigt, eine Änderung der Rotamerenverteilung ist nicht festzustellen (vgl. Abb. X205).

#### 4.4 Vergleich mit Ab-Initio-Rechnungen<sup>1</sup>

Zur Überprüfung der Güte der Kraftfelder im Bezug auf die Energie der Hydroxymethyl-Rotameren auf dem höchsten zur Verfügung stehenden technischen Level wurden für Gal $\beta$ -O-iProp und Gl $\alpha$ -O-iProp jeweils 9 lokale Minima (CVFF) (vgl. Abb. 50) mit Gaussian94 durch Ab-Initio-Methoden minimiert (Restricted Hartree Fock (RHF) Basissatz 6-31G<sup>\*\*</sup>; DFT Basissatz 6-31G<sup>\*\*</sup>).



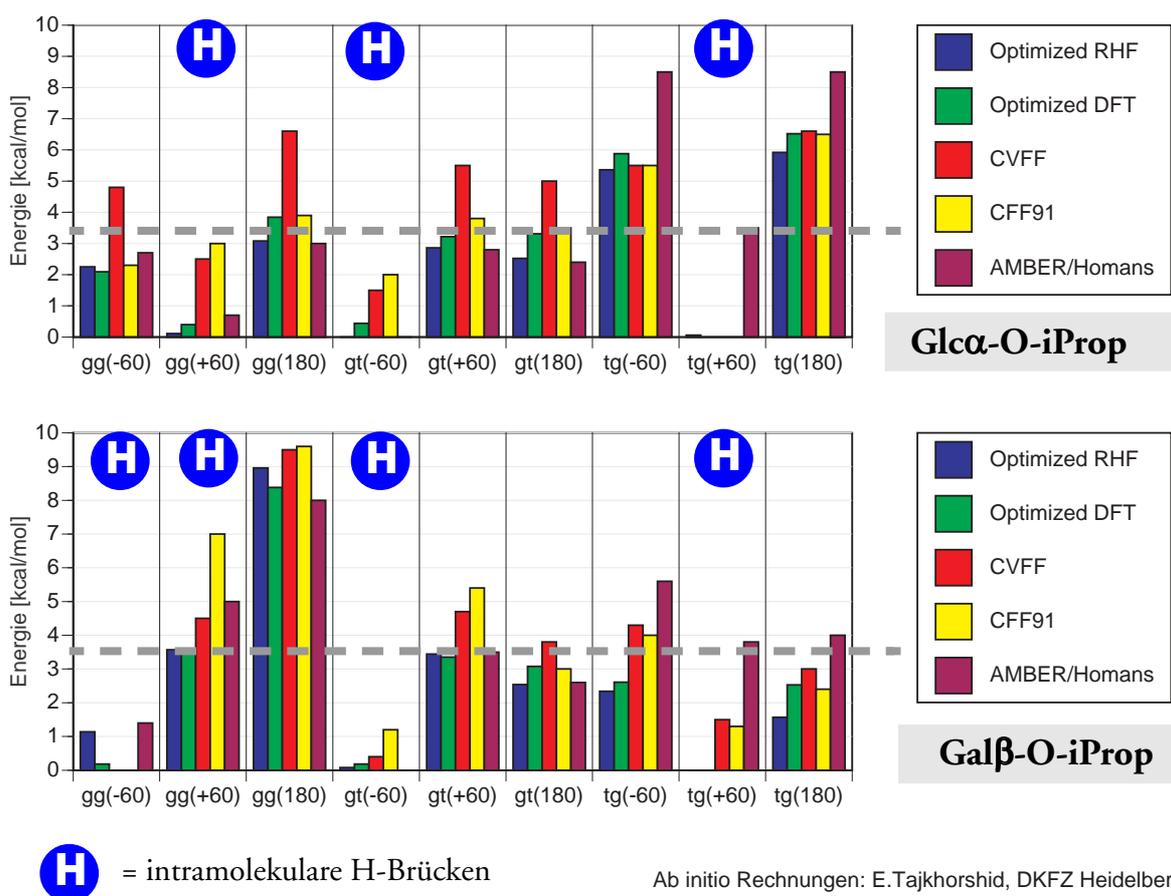
**Abb. 50:** Neun ausgewählte Startstrukturen von Gl $\alpha$ -O-iProp für die Minimierung mit Ab-Initio-Methoden, Contourlinien: G56O-G6OH-Grid-Search (Gridintervall = 10°; Kraftfeld = CVFF;  $\epsilon=4$ ).

Der Vergleich der Energien der einzelnen G56O/G6OH-Rotameren unter Einbeziehung der entsprechenden Ab-Initio-Energien zeigt deutlich, daß für Gl $\alpha$ -O-iProp der Zustand tg(+60) (Ab-Initio-Minimum liegt bei G56O=168°/G6OH=34° für alle Methoden außer AMBER/Homans das globale Minimum ist. Insofern kann davon ausgegangen werden, daß diese Anordnung im Vakuum definitiv die stabilste Konformation ist. Damit bleibt die Frage, warum mit NMR-Messungen für Glucose keine tg-Konformation gefunden wurde, nach wie vor offen. Vergleicht man die Ab-Initio-Energien der drei tg-Konformere, so stellt man fest, daß tg(-60) und tg(180) im Vergleich zu tg(+60) sehr energiereich sind (5.5 bis 6.5 kcal/mol). Im Gegensatz dazu sind die Unterschiede bei den gg- und gt-Konformeren nicht so stark ausgeprägt. Geht man davon aus, daß im Solvens durch Interaktion mit dem Wasser diejenigen Konformationen, die keine intramolekularen H-Brücken ausbilden und daher im Vakuum eine hohe Energie besitzen, stabilisiert werden (z.B. um 3.5 kcal, vgl. gestrichelte Linie), so wäre es möglich, daß für gg und gt alle drei Orientierungen nach Boltzmann populiert sind, während für tg weiterhin nur eine populiert ist, da die beiden anderen tg-Konformationen immer noch zu energiereich sind. Es kommt daher zusätzlich zur Energiestabilisierung von gg und gt noch eine entropische Stabilisierung hinzu.

1. Die Ab-Initio-Rechnungen wurden von Emadeddin Tajkhorshid (DKFZ Heidelberg) durchgeführt.

In analoger Weise läßt sich auch für Gal $\beta$ -O-iProp argumentieren. Im Unterschied zu Glc $\alpha$ -O-iProp gibt es aber nur einen Zustand der eine extrem hohe Energie besitzt ( $\rightarrow$  gg(180°)). Daher sollte es im Solvens zwei energiearme Realisierungsmöglichkeiten für gg geben. Experimentell findet man für Galactose unter Auswertung der Cambridge Structural Data Base 7% gg und anhand der NMR-Daten 18% gg [157]. Obwohl also im gg-Zustand auch eine syn-axiale Anordnung von O4 und O6 vorhanden ist und damit theoretisch dieselben destabilisierenden Kräfte wirken sollten wie im tg-Zustand von Glucose, ist der gg-Zustand bei Galactose experimentell nachweisbar und der tg-Zustand für Glucose nicht. Nach den Resultaten aus den Ab-Initio-Rechnungen ist anzunehmen, daß es sich bei Gal(gg) gegenüber Glc(tg) um eine entropische Stabilisierung handelt.

Vergleicht man anhand von Abb. 51 wie gut die einzelnen Kraftfelder im Einklang mit den Ab-Initio-Resultaten sind, so stellt man fest, daß AMBER ohne Homans, d.h. ohne die spezielle tg-Parametrisierung sehr gut mit den Ab-Initio-Werten übereinstimmt. CFF91 ergibt ebenfalls noch eine gute Übereinstimmung. Mit CVFF erhält man für die gg und gt Zustände von Glucose eine zu hohe Energie. Mit CVFF erhält man für die gg und gt Zustände von Glucose eine zu hohe Energie.



**Abb. 51:** Abhängigkeit der Systemenergie von den Torsionswinkeln G56O und G6OH. Energievergleich: Kraftfeldmethoden (CVFF, CFF91, AMBER/Homans,  $\epsilon=1$ ) versus Ab Initio Methoden (Gaussian 94): RHF (6-31G\*\*), DFT (6-31G\*\*)

## 5 Untersuchungen zur Bindungsaffinität von Protein-Kohlenhydrat-Komplexen

### 5.1 Einleitung

NMR-Messungen in aprotischen Lösungsmitteln wie DMSO gewinnen wegen der zusätzlichen Information, die man durch NOEs mit OH-Protonen erhält, zunehmend an Bedeutung in der Strukturanalyse von Oligosacchariden [65-67,213]. Nachdem sowohl die Löslichkeit von Proteinen in DMSO [227] als auch der Erhalt von wesentlichen Teilen der Sekundär- und Tertiärstruktur dokumentiert ist [68,69,228,229], stellt sich die Frage inwieweit der Übergang zu einem aprotischen Lösungsmittel wie DMSO Auswirkungen auf die Bindungsaffinität von Protein-Kohlenhydrat-Komplexen besitzt [213]. Eine vorhandene, meßbare Bindung ist Voraussetzung für die Strukturanalyse von Protein-Kohlenhydrat-Komplexen in DMSO. Die Analyse des Lösungsmiteleinflusses auf die Komplexbildung ist jedoch nicht nur unter NMR-methodischen Aspekten von Interesse, sondern verspricht auch einem wichtigen Beitrag im Zusammenhang zur nach wie vor kontroversen Diskussion um dem Einfluß des Lösungsmittels auf die Bindungsaffinität [81,84-86].

Als Modellsystem zur Untersuchung des Lösungsmiteleinflusses und sonstiger, die Bindungsaffinität beeinflussender Faktoren bietet sich Hevein [230,231] an, da dieses Protein das kleinste und damit überschaubarste bekannte Lektin (43 Aminosäuren) ist. Weiterhin liegt für Hevein-Chitobiose- und Hevein-Chitotriose-Komplexe umfangreiches experimentelles und theoretisches Datenmaterial vor [70,75,76,79,125,155,168], so daß sich dieses System, wie kein anderes, für umfangreichere theoretische Untersuchungen basierend auf MD-Simulationen<sup>1</sup> eignet.

### 5.2 Modellierung von Hevein-Chitotriose-Komplexen

#### 5.2.1 Problemstellung und Zielsetzung

Die dreidimensionale Struktur des Hevein-Chitobiose-Komplexes ist aus NMR-Messungen bekannt [75] (vgl. Abb. 53). Beim Übergang von GlcNAc, Chitobiose zu Chitotriose als Ligand für Hevein erhöht sich die Bindungskonstante jeweils um eine Größenordnung [75,79]. Es wird angenommen, daß die für die Bindungsaffinität von Chitobiose wesentlichen Interaktionen zu den Aminosäuren TRP21, TRP23, TYR30 und SER19 ausgebildet werden. Weiterhin ist bekannt, daß eine hydrophobe Wechselwirkung zwischen Zuckerringen und aromatischen Aminosäuren durch eine schichtartige Anordnung besonders zur Stabilisierung von Protein-Kohlenhydrat-Komplexen beitragen [54]. So läßt sich die Zunahme der Bindungskonstante beim Übergang von GlcNAc zu Chitobiose folgerichtig damit erklären, daß hier eine zusätzliche lipophile Interaktion zwischen der weniger polaren  $\alpha$ -Seite des reduzierenden Endes von Chitobiose und dem aromatischen Ringssystem von TRP21 vorhanden ist. Beim Übergang zur Chitotriose kann diese Argumentationslinie jedoch nicht fortgesetzt werden, da außer HIS35 keine weitere aromatische Aminosäure im Protein vorhanden ist und die schichtartige Anordnung des dritten Zuckerrings mit HIS35 sterisch (wahrscheinlich) nicht möglich ist. Da die Struktur des Hevein-Chitotriose-Komplexes nicht bekannt ist, soll diese in einer Art wissenschaftsbasiertem Ansatz aus dem Hevein-Chitobiose-Komplex durch Optimierung der Wechselwirkungsenergie modelliert werden, um so Hinweise darüber zu erhalten, durch welche zusätzlichen Interaktionen die Erhöhung der Bindungsaffinität beim Übergang zu Chitotriose verursacht werden

1. Ein wichtiger praktischer Aspekt ist in diesem Zusammenhang auch die Tatsache, daß Hevein aufgrund seiner geringen Größe vollständig in eine Solvensbox der Größe 40Åx40Åx40Å paßt.

könnte. Dazu ist es notwendig die Interaktionsanalyse zuerst am bekannten Hevein-Chitobiose-Komplex zu testen. Desweiteren soll untersucht werden inwieweit durch eine Interaktionsanalyse Aussagen über den Einfluß von Lösungsmittelleffekten gemacht werden können. Konkret soll dazu eine Wechselwirkungsanalyse basierend auf dem modellierten Hevein-Chitotriose-Komplex in Wasser und DMSO durchgeführt werden.

### 5.2.2 Methoden

Detaillierte Interaktionsanalysen [28,124,126] bei denen die mittlere Wechselwirkungsenergie von einzelnen Residue-Paaren für eine größere Anzahl von Strukturen berechnet wird, sind in der Literatur extrem selten. Der Grund hierfür dürfte darin liegen, daß es keines der bekannten Modelling Softwarepakete ermöglicht, eine Interaktionsanalyse für ein Ensemble von Strukturen komfortabel und übersichtlich durchzuführen. Zwar bietet das Programm DISCOVER Vers. 2.95 die Möglichkeit, mit verschiedenen Befehlen (print residue-residue interaction, enclose, interaction) eine detaillierte Interaktionsanalyse durchzuführen, die Anwendung dieser Funktionen führt jedoch zur Ausgabe einer wahren Datenflut im Simulationsprotokoll, die nur für kleine Einzelstrukturen manuell analysierbar ist. Wie schon in Abschnitt 3.7.2 dargelegt, ist die Berechnung der mittleren Wechselwirkungsenergie basierend auf wenigen minimierten Solvensstrukturen mit einem großen Unsicherheitsfaktor behaftet. Da andererseits die Minimierung einer ausreichend großen Anzahl von Solvensstrukturen technisch extrem aufwendig ist, ist eine Interaktionsanalyse basierend auf minimierten Solvensstrukturen momentan nicht sinnvoll zu realisieren. Vielversprechender erscheint eine Interaktionsanalyse basierend auf einer möglichst großen Anzahl von MD-Strukturen, da sich hier die starken Solvenseffekte herausmitteln sollten. Zur technischen Umsetzung dieser Methode wurde zuerst ein Programm<sup>1</sup> geschrieben, das die Unzulänglichkeit von DISCOVER dahingehend behebt, indem es aus dem Simulationsprotokoll<sup>2</sup> u.a. die wesentlichen Interaktionsenergien extrahiert und in einem verwertbaren Format abspeichert.

Als Ausgabeformat für Wechselwirkungen (Energien, H-Brücken, usw.) wurde dafür das „CAT 2D-Category-Format“ entwickelt, welches mit einem Header beginnt, der Informationen über die gewählten Simulationsparameter enthält und anschließend die Interaktionen in der Form „Residue1/Residue2/Frame/Werte/...“ auflistet. Zur Durchführung einer Interaktionsanalyse, basierend auf diesen Interaktions-Listen, wurden spezielle Funktionen in CAT2000 integriert, die es erlauben, für eine (theoretisch) unbegrenzte Anzahl von Strukturen z.B. eine Residue-Residue-Interaktionsmatrix der mittleren Wechselwirkungsenergie für ausgewählte Residues zu berechnen. Die technischen Grenzen liegen momentan lediglich in der Limitierung des zur Verfügung stehenden Festplattenspeichers.

Für die folgenden Analysen wurde jeweils, soweit nicht anders vermerkt, eine 200 ps MD-Simulation mit vorangestellter 50 ps Equilibrierungsphase (CVFF, PBC,  $\epsilon=1$ , Double Cutoff=11/13 Å) durchgeführt und jede Picosekunde eine Struktur gespeichert. Die Interaktionsanalyse wurde nur für die letzten hundert Strukturen mit den DISCOVER-Befehlen „print molecule-molecule interaction“ und „print residue-residue interaction“ durchgeführt. Da die Sekundärstruktur von Hevein bei der Minimierung (und damit auch bei der MD-Simulation) nicht stabil ist (siehe Abschnitt 5.2.3) wurden alle MD-Simulationen mit im Raum fixierten  $C_{\alpha}$ -Atomen durchgeführt. Die Schweratome des Liganden wurden während der Temperatur-Equilibrierungsphase ebenfalls fixiert, da DISCOVER das System sehr abrupt aufheizt und es dadurch oft zur Dissoziation des Komplexes kommt. Während der Sampling-Phase war der Ligand jedoch frei beweglich.

---

1. PAVDO (Programm zur Analyse von Discover Outputfiles)

2. Die Größe des Output-Files der Interaktionsanalyse von 100 Strukturen beträgt ca. 500 MByte

### 5.2.3 Überprüfung der Stabilität der Proteinstruktur bei MD-Simulationen

Wichtige Voraussetzung für eine erfolgreiche Durchführung einer MD-Simulation von Protein-Kohlenhydrat-Komplexen ist die Verwendung geeigneter Simulationsbedingungen. Im Zusammenhang mit MD-Simulationen von Proteinen stellt sich dabei grundsätzlich immer die Frage, inwieweit es notwendig ist, die Beweglichkeit des Proteins durch Fixierung einzelner Atome im Raum einzuschränken.

Um zu überprüfen, ob die Sekundärstruktur von Hevein während einer MD-Simulation stabil bleiben würde, wurde zunächst eine dreistufige Minimierung der Ausgangsstruktur<sup>1</sup> [75] durchgeführt. In der ersten Stufe wurden alle Schweratome im Raum fixiert (Optimierung der H-Atome), im zweiten Schritt waren nur noch die  $C_{\alpha}$ -Atome fixiert und schließlich wurde die Minimierung ohne Fixierung fortgesetzt. Das Konvergenzkriterium war jeweils 0.01 kcal/Å. Falls sich die Struktur bereits bei einer Minimierung instabil zeigt, ist davon auszugehen, daß sie während einer MD-Simulation erst recht ihre Tertiärstruktur verlieren würde. Im Zusammenhang mit dieser Fragestellung sollen in diesem Abschnitt mehrere Analysemethoden vorgestellt werden, die es prinzipiell erlauben, die Stabilität von Proteinstrukturen während einer MD-Simulation oder Minimierung zu verfolgen. Alle diese Funktionen wurden in CAT2000 implementiert.

Eine Methode, die es erlaubt die Stabilität der Proteinstruktur während einer MD-Simulation oder Minimierung zu untersuchen ist die Analyse der Abstandsmatrix der  $C_{\alpha}$ -Atome der Aminosäuren<sup>2</sup>: Aminosäuren, die miteinander in Wechselwirkung treten und so evtl. Sekundärstrukturelemente aufbauen, zeichnen sich durch einen Abstand der  $C_{\alpha}$ -Atome aus, der sich in einem Bereich zwischen 4 Å und 7 Å bewegt. Markiert man in einem Matrix-Diagramm, bei dem die einzelnen Aminosäuren in x- und y-Richtung aufgetragen sind, nur diejenigen Aminosäurepaare deren  $C_{\alpha}$ -Abstände sich in diesem Abstandsbereich befinden, z.B. durch Graufächen („Intensität“), so lassen sich parallel verlaufende Aminosäureketten (z.B.  $\beta$ -Sheets) durch einen senkrecht zur Matrixdiagonalen verlaufenden schmale Intensitätsverlauf erkennen (Abb. X206).  $\alpha$ -Helices erkennt man dagegen an einen Intensitätsverlauf unmittelbar parallel neben der Matrixdiagonalen. Wesentlich für die Beschreibung von Tertiärstrukturelementen sind die Interaktionen von sequentiell weit voneinander entfernten Aminosäurepaaren. Diese zeigen sich im Diagramm als isolierte graue Flächen („Kreuzpeaks“). Ordnet man jedem Aminosäurepaar deren  $C_{\alpha}$ -Abstände in den bestimmten Toleranzbereich (4-7 Å) fallen die Intensität=1 zu, so kann man die Intensitäten der Aminosäurepaare während einer MD-Simulation aufsummieren und so die Stabilität eines Abstands relativ zur Simulationszeit in Prozent angeben.

Will man dagegen nur einen Ausgangs- und einen Endzustand vergleichen, wie im Falle einer einfachen Minimierung, läßt sich die Stabilität der  $C_{\alpha}$ -Abstände auch durch einfache Differenzbildung in einer Matrix darstellen (Abb. X207, oben). Anhand der Änderung der  $C_{\alpha}$ -Abstände erhält man einen sehr schnellen zweidimensionalen Überblick über die relative Änderung der Position einer Aminosäure. So wird z.B. deutlich, daß sich ASN15 während der Minimierung von den Aminosäuren 1 bis 11 entfernt und in Richtung THR27 wandert. Anders verhält es sich mit HIS35. Diese Aminosäure wandert in Richtung der Aminosäuren 1 bis 11.

Alternativ kann man die Stabilität des Proteins auch durch Berechnung des Rigid-Body-RMS untersuchen. In CAT2000 wurde hierfür zusätzlich zur Berechnung des RMS bzgl. einer festen Referenzstruktur (RRMS)<sup>3</sup> eine spezielle Variante implementiert, die in einer Abfolge von Strukturen immer

1. freundlicherweise zur Verfügung gestellt von Jesús Jiménez-Barbereo, Madrid
2. CAT2000 berechnet die Abstandsmatrix direkt aus dem Koordinatenarchiv
3. Diese Funktion ist z.B. auch dafür geeignet um die Stabilität der räumlichen Orientierung eines Liganden in der Bindungstasche zu untersuchen. Dazu ist jedoch notwendig Teile des Proteins im Raum zu fixieren.

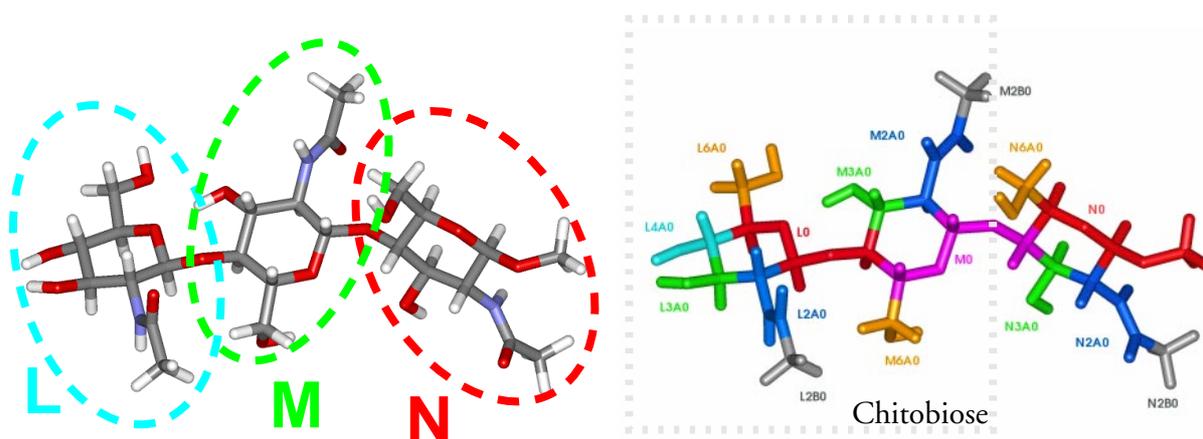
die jeweils vorhergehende als Referenz-Struktur verwendet (PRMS). Diese CAT-Funktion wurde im vorliegenden Fall verwendet um die Strukturänderungen in der oben angegebenen Minimierungssequenz zu untersuchen (Abb. X207, unten).

Sowohl die RMS- als auch die  $C_{\alpha}$ -Abstands-Methode zeigen, daß die Struktur von Hevein bei einer Minimierung ohne Fixierung über weite Bereiche nicht stabil ist. Interessanterweise ist der Aminosäurebereich der die Bindungstasche für N-Acetylglucosamine bildet (Aminosäuren 17 bis 32) weitgehend stabil. Um jedoch sicherzugehen, daß sich das Hevein während der MD-Simulation nicht entfaltet, ist es sinnvoll alle Simulationen und Minimierungen mit fixierten  $C_{\alpha}$ -Atomen durchzuführen.

#### 5.2.4 Interaktions-Analyse des Hevein/Chitobiose-Komplexes

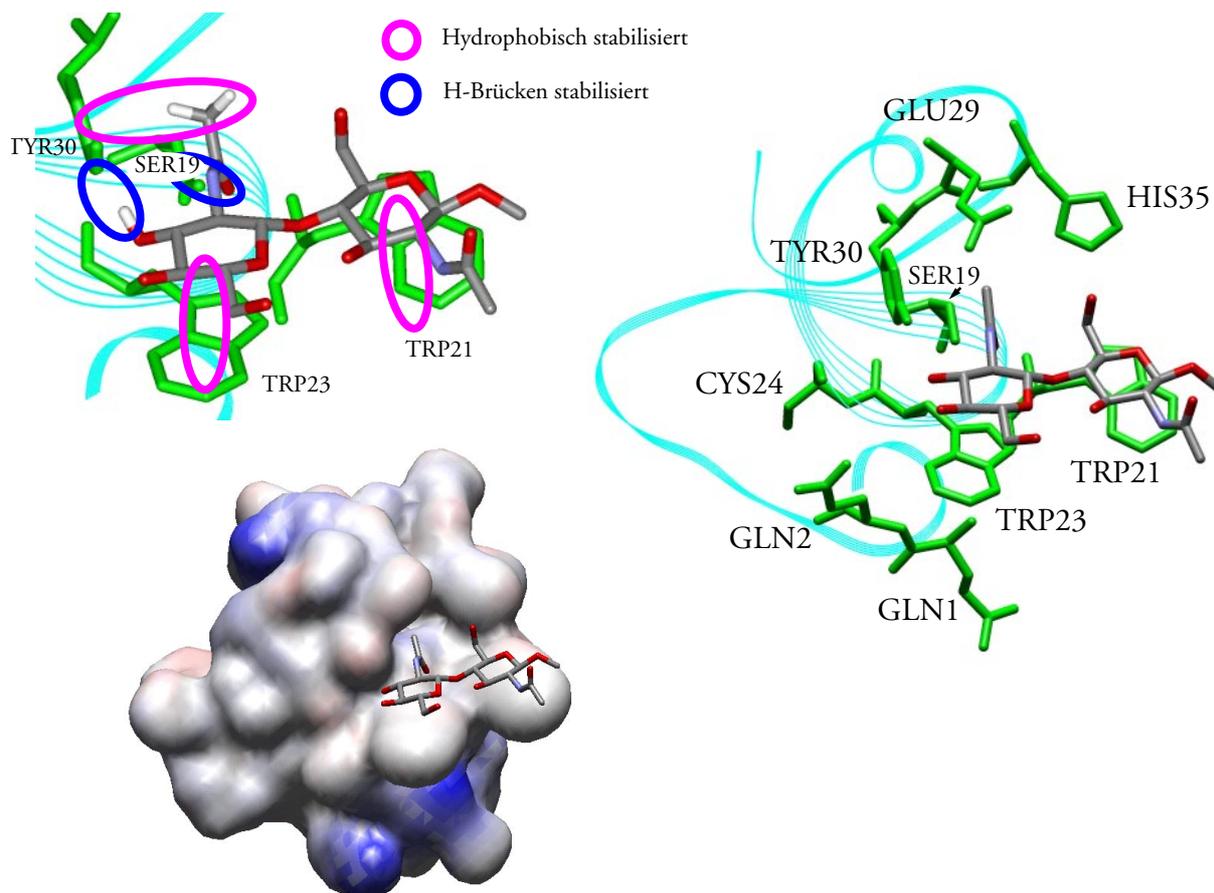
Die NMR-Struktur des Hevein/Chitobiose-Komplexes [75][155] (vgl. Abb. 53) wurde als Startstruktur zur Interaktionsanalyse des Hevein/Chitobiose-Komplexes verwendet. Aus den NMR-Messungen [75] sind mehrere intermolekulare NOEs bekannt: die Methylgruppe der Acetamidogruppe am nichtreduzierenden Zucker zeigt Kontakte zum He von TYR30 und zu H2 von TRP21, weiterhin existieren für das H4-Proton derselben Zuckereinheit Kontakte zu H2 und H4 von TRP23. Zusätzlich wird von einer H-Brücke zwischen OH von SER19 und dem Carbonyl-O-Atom der nichtreduzierenden Zuckereinheit ausgegangen (vgl. Abb. 53).

Zur Generierung einer Startstruktur für die MD-Simulation wurde die NMR-Struktur<sup>1</sup> mit 1420 Wassermolekülen solvatisiert (PBC,  $32\text{\AA} \times 50\text{\AA} \times 32\text{\AA}$ , Cutoff =  $11\text{\AA}$ ) und anschließend optimiert. Dazu wurde zuerst die Wasserbox für 30 ps equilibriert (mit fixierten Schweratomen von Hevein und Chitobiose), anschließend wurden zehn Zyklen eines MD-Minimierungsprotokolls (vgl. Abschnitt 3.7.2) durchlaufen (10 ps, 200K,  $C_{\alpha}$ -Atome des Heveins fixiert). Aus den minimierten Strukturen wurde diejenige mit der geringsten Energie (vgl. Abb. 53) als Startstruktur für die zur Berechnung der mittleren Interaktionsenergie durchgeführte MD-Simulation verwendet.



**Abb. 52:** Definition und systematische Nomenklatur der Einheiten (links) und der Residues (rechts) für Chitobiose und Chitotriose

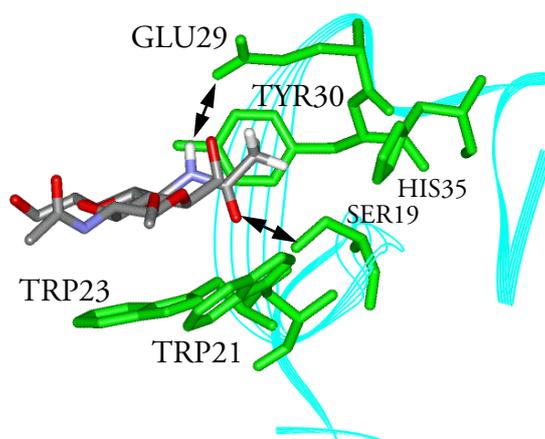
1. Die Aminosäuren wurden entsprechend ihren Zustand bei pH=7 protoniert bzw. deprotoniert.



**Abb. 53:** Startstruktur zur Berechnung der Wechselwirkungsenergie für den Hevein/Chitobiose - Komplex (MD-minimierte NMR-Struktur [75])

Die Analyse der Residue-Residue-Wechselwirkungsenergie (Abb. X208) und H-Brücken-Matrix (Abb. X209)<sup>1</sup> für den Hevein/Chitobiose-Komplex zeigt, daß die stärksten Wechselwirkungen zu TRP21 und TRP23 vorhanden sind (ca. -12 kcal/mol). Deutliche Wechselwirkungen findet man auch noch für TYR30, GLU29, SER19 und GLN1. Während die Wechselwirkung mit SER19 und GLU29 überwiegend auf Coulomb-Wechselwirkung zurückzuführen ist, beruht die Interaktion mit TRP21 und TRP23 hauptsächlich auf Van-der-Waals-Wechselwirkung. Die Wechselwirkung mit TYR30 beruht einerseits auf einer Van-der-Waals-Interaktion mit der Acetamido-Gruppe (Residues L2A0 und L2B0) und andererseits auf einer Coulomb-Interaktion mit L3A0. Der Ligand-Residue mit der höchsten Wechselwirkungsenergie ist L2A0 (C2H2-NH-CO) verursacht durch die starke Wechselwirkung mit SER19, TYR30, TRP23 und GLU29 (Abb. 54). Im Vergleich zur analogen Gruppierung M2A0 zeigt L2A0 nur eine geringe Wechselwirkung mit dem Solvens. Ein Hinweis darauf, daß die Gruppe im Komplex wenig Solvenszugänglichkeit besitzt. Die Analyse der H-Brücken-Matrix<sup>2</sup> zeigt deutlich eine starke<sup>3</sup> intermolekulare H-Brücke zwischen SER19(Donor) und dem Carbonyl-O-Atom von L2A0. Weitere H-Brücken werden zwischen den Residue-Donor-Akzeptor-Paa-

1. Definition und Nomenklatur der Residues siehe Abb. 52, die Bezeichnung der Atome in den Abbildungen wird automatisch erzeugt und ist nicht immer systematisch, daher wird eine eigene Atomnomenklatur im Text verwendet.
2. Die Matrix wird automatisch basierend auf einer einzigen Residueliste berechnet und ist daher symmetrisch. Dies bedeutet aber auch, daß alle Atome, die nicht sowohl als Donor als auch als Akzeptor wirken können, unberechtigtweise in einer der beiden Listen („Donor-Atome“ bzw. „Akzeptor-Atome“) aufgeführt sind.
3. Stärke eine H-Brücke = (Häufigkeit im Ensemble) x (Skalierungsfaktor, der Abstand und Winkel der H-Brücke berücksichtigt)



**Abb. 54:** Lage der N-Acetyl-Gruppe im modellierten Hevein/Chitobiose -Komplex

ren TYR30/L3A0, L3A0/TYR30 und TRP23/M3A0 beobachtet. Eine starke intramolekulare H-Brücke wird zwischen MOH3 und dem Ring-O-Atom LO5 des nicht-reduzierenden Zuckers ausgebildet. Dies zeigt sich auch in der Interaktion von M3A0 mit dem Solvens: Der Residue besitzt nur ein Drittel der Wechselwirkungsenergie mit dem Solvens wie der vergleichbare Residue L3A0. Insgesamt berechnet sich die mittlere Wechselwirkungsenergie für Hevein und Chitobiose zu  $-52$  kcal/mol (Abb. X215). Dieser Wert kann jedoch nicht direkt mit der experimentell bestimmten Bindungsenthalpie von  $-7.5$  kcal/mol verglichen werden, denn zur Bindungsenthalpie tragen noch weitere Effekte (z.B. Solvatisierungseffekte) bei (vgl. Abschnitt 1.2.2). Insgesamt erscheinen die Ergebnisse logisch nachvollziehbar und sind qualitativ in Übereinstimmung mit den experimentellen Daten [75]. Die Analyse der Protein-Ligand-Wechselwirkung basierend auf Residue-Residue-Interaktionsmatrizen hat sich somit als eine effiziente Methode erwiesen um schnell die wesentlichen Interaktionen eines Protein-Kohlenhydrat-Komplexes herauszufinden. Weiterhin besitzt der Ansatz für das Design hochaktiver Liganden ein enormes Potential, da die Auswirkung der Substitution einer funktionellen Gruppe auf die Wechselwirkungsenergie schnell und detailliert untersucht werden kann. Das technische Problem der zu analysierenden riesigen Datenmenge ist mit CAT2000 gelöst.

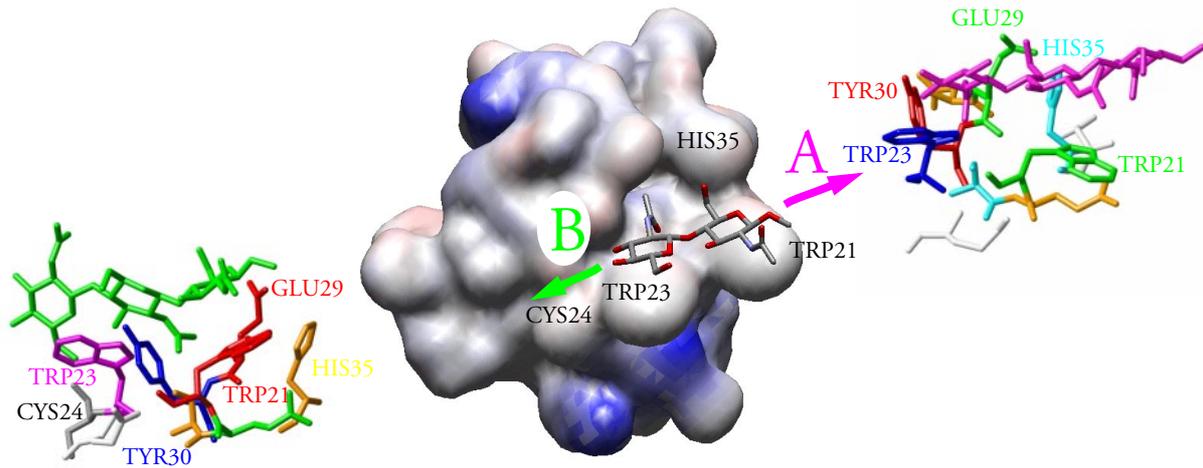
### 5.2.5 Modellierung des Hevein-Chitotriose-Komplexes

#### Generierung der Startstrukturen

Prinzipiell existieren zwei Möglichkeiten um unter Beibehaltung der prinzipiellen Interaktion von Chitobiose zum Chitotriose-Komplex zu gelangen: Eine Verlängerung des Moleküls am reduzierenden Ende (A) und eine „Verlängerung“ in Richtung CYS24 (B) (siehe Abb. 55).

Ausgehend von der NMR-Struktur des Hevein/Chitobiose-Komplexes wurde daher die Chitotriose durch Überlagerung mit den entsprechenden Atomen der Chitobiose in die Bindungstasche positioniert (Orientierung A und B) und die Chitobiose anschließend aus den Strukturen A und B entfernt. Die beiden Komplexe wurden solvatisiert (PBC (40x40x40); 1820 Wassermoleküle) und anschließend mit der bereits beim Hevein-Chitobiose-Komplex beschriebenen Methode optimiert. Anschließend wurde jeweils eine MD-Simulation bei 300K durchgeführt (CVFF,  $t_s=200$  ps). CPU Zeit pro Simulation: 250 h (IBM SP2, Single Processor).

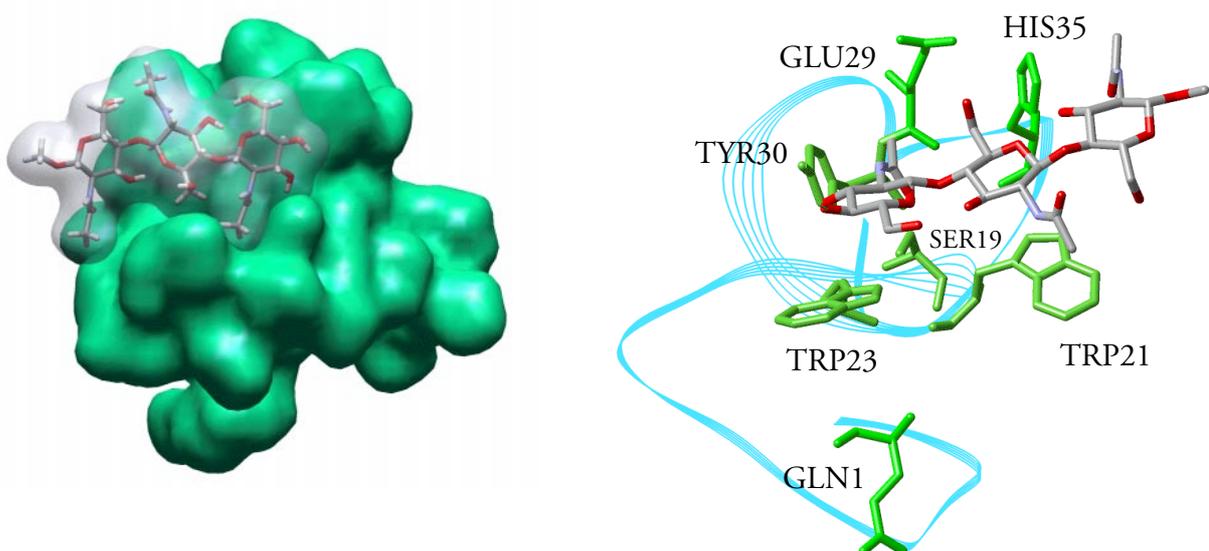
Um zu überprüfen, welche der beiden Orientierungen die günstigste Wechselwirkung mit dem Protein zeigt, wurde eine Wechselwirkungsanalyse, wie bereits bei Chitobiose beschrieben, durchgeführt.



**Abb. 55:** Möglichkeiten zur „Verlängerung“ der Chitobiose zur Chitotriose

### Interaktions-Analyse für Orientierung A

Aus den Residue-Residue-Wechselwirkungsenergien (Abb. X210) und der H-Brücken-Matrix (Abb. X212) erkennt man deutlich, daß für Orientierung A eine sehr starke Wechselwirkung mit HIS35 existiert, die weitgehend elektrostatischer Natur ist. Eine ebenfalls sehr starke Coulomb-Anziehung besteht zur Aminosäure GLU29. Im Gegensatz zum Chitobiose-Komplex ist die Interaktion zu TRP23 deutlich reduziert, was bedeutet, daß der Ligand mit der „L-Einheit“ (Abb. 52) nicht mehr genau über TRP23 liegt. Betrachtet man die optimierte Ausgangsstruktur (Abb. 56), so zeigt sich, daß die Chitotriose etwas in Richtung HIS35 gewandert ist und die TYR30-Seitenkette (wegen der starken Wechselwirkung mit der N-Acetylgruppe) mitgezogen hat. Die elektrostatische Wechselwirkung mit dem flexiblen N-terminalen Ende der Aminosäurekette (GLN1) trägt wahrscheinlich in der Realität wenig zur Bindungsaffinität bei, da anzunehmen ist, daß die Lage relativ zum Liganden aufgrund der hohen Beweglichkeit von GLN1 sehr stark variiert. Insgesamt findet man für die Orientierung A

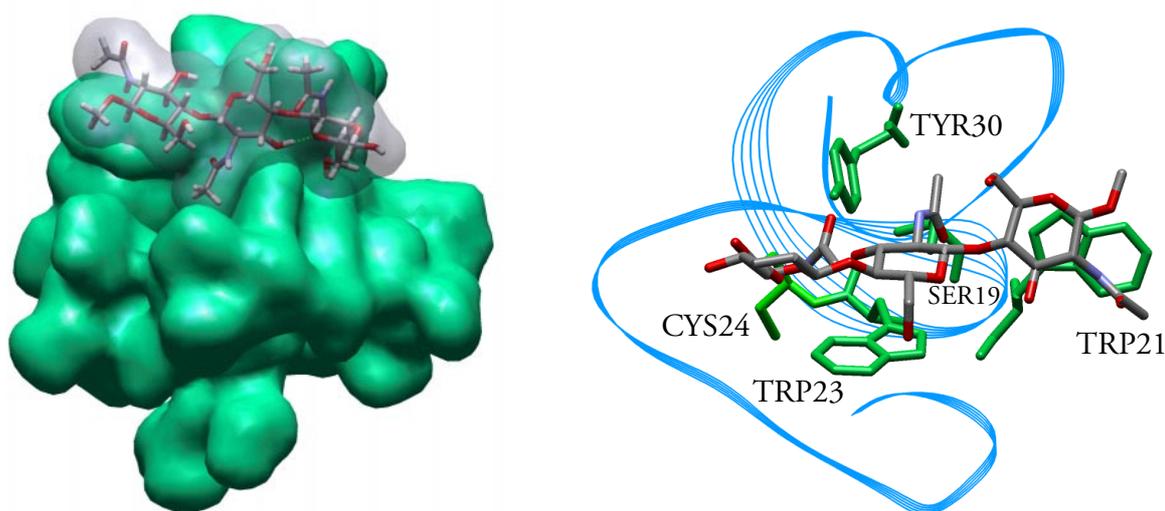


**Abb. 56:** Strukturvorschlag A für den Hevein/Chitotriose -Komplex

dieselben Interaktion wie bei Chitobiose plus die zusätzliche starke Wechselwirkung mit HIS35. Die Analyse der H-Brücken zeigt die Ausbildung einer zusätzlichen H-Brücke zwischen dem NH-Proton von HIS35 und MO6 der zentralen Zuckereinheit.

### Interaktions-Analyse für Orientierung B

Bei der Betrachtung der Residue-Residue-Wechselwirkungsmatrix (Abb. X211) ist zu beachten, daß die zentrale Zuckereinheit (M-Einheit) nun die Lage der L-Einheit in der Bindungstasche in Orientierung A einnimmt. Auffällig für Orientierung B ist eine sehr starke Wechselwirkung von GLU29 mit verschiedenen Ligandenresidues, die weitgehend auf Coulomb-Anziehung beruht. Eine ebenfalls sehr starke Coulomb-Anziehung besteht außer zu SER19 zur Aminosäure CYS24. Die Interaktion zu TRP23 und TRP21 ist in derselben Größenordnung wie beim Chitobiose-Komplex. Die Interaktion mit TYR30 ist im Vergleich zur Orientierung A etwas erhöht (-4 kcal/mol). Auffällig ist weiterhin die insgesamt geringere Interaktion mit dem Solvens. Zusammenfassend findet man für die Orientierung B alle wichtigen Interaktion wie bei Chitobiose plus eine zusätzliche starke Wechselwirkung mit CYS24. Die Analyse der H-Brücken (Abb. X212, unten) beweist, daß die Coulomb-Interaktion mit CYS24 auf der Ausbildung einer H-Brücke von LOH6 zum Carbonyl-O-Atom beruht.



**Abb. 57:** Strukturvorschlag B für den Hevein/Chitotriose -Komplex

### Vergleichende Diskussion

Die Berechnung der Molekül-Molekül-Wechselwirkungsenergie (Abb. X215) ergibt für die Orientierung B eine um ca. -6 kcal/mol günstigere Wechselwirkungsenergie als für die Orientierung A. Entscheidender für eine Voraussage der höheren Bindungsaffinität ist jedoch, daß -4 kcal/mol der Komplexstabilisierung allein durch die Erhöhung der Van-der-Waals-Wechselwirkung zustande kommt. Wie experimentell gezeigt werden konnte, trägt diese Art der Wechselwirkung hauptsächlich zur Stabilisierung von Protein-Kohlenhydrat-Komplexen in realen Systemen bei [54]. Die bei Orientierung B gefundene starke Wechselwirkung mit GLU29 dürfte, wie auch die starke Interaktion mit HIS35 für Orientierung A, in realen Systemen wenig zur Erhöhung der Bindungsenthalpie beitragen,

da es sich bei beiden Aminosäuren um gut solvenszugängliche geladene Aminosäuren handelt. Es ist daher anzunehmen, daß die Desolvatisierung von GLU(-) bzw. HIS(+) mindestens denselben Energiebetrag benötigt, wie bei der Coulomb-Interaktion mit dem Liganden wieder frei wird. Den entscheidenden, für eine Komplexbindung günstigen, Beitrag liefert die van-der-Waals-Interaktion zwischen den Zuckerringen und TRP21 bzw. TRP23 sowie die Interaktion zwischen der Methylgruppe der N-Acetyl-Gruppierung mit dem aromatischen Ring von TYR30. Günstigerweise entsteht dabei eine Anordnung des Zuckers in der Bindungstasche, die das Carbonyl-O-Atom einer N-Acetylgruppe in die Nähe des OH-Protons von SER19 positioniert und dieses gleichzeitig vom umgebenden Solvens abschirmt (vgl. Abb. 54). Dadurch kann sich eine starke H-Brücke ausbilden. Wegen der insgesamt höheren van-der-Waals Stabilisierung ist anzunehmen, daß die Orientierung B in Wasser die stabilere Anordnung sein sollte, was jedoch nicht ausschließt, daß Anordnung A auch eingenommen wird.

### 5.2.6 Interaktionsanalyse des Hevein-Chitotriose-Komplexes in DMSO

Um den Einfluß eines aprotischen Lösungsmittels auf die Protein-Kohlenhydrat-Wechselwirkungsenergie zu untersuchen, wurde eine MD-Interaktions-Analyse für den Hevein-Chitotriose-Komplex in Orientierung B nach der bereits beschriebenen Methode durchgeführt.

Während der 300ps-MD-Simulation (CVFF, 300K) in DMSO<sup>1</sup> (PBC, 45Åx45Åx45Å)[167] bleibt das Chitotriose-Molekül in der Bindungstasche gebunden. Lediglich kurzzeitig löst es sich mit einem Ende vom Protein ab. Die Residue-Residue-Wechselwirkungsanalyse (Abb. X213) und die Analyse der H-Brücken (Abb. X214) ergibt, daß im Vergleich zur MD-Simulation in Wasser vor allem die Coulomb-Interaktionen reduziert sind. Berechnet man die gesamte Wechselwirkungsenergie für die beiden Moleküle (Abb. X215), so findet man eine Reduktion der Coulomb-Energie um 20 kcal/mol gegenüber einer Reduzierung der van-der-Waals-Interaktion um ca. 2 kcal/mol beim Übergang von Wasser nach DMSO.

#### Qualitative Vorhersage der Änderung der Bindungsaffinität des Hevein-Chitotriose-Komplexes beim Wechsel zu DMSO als Lösungsmittel auf der Basis von Wechselwirkungsenergien

Anhand der Molekül-Molekül-Wechselwirkungsenergien (Abb. X215) würde man qualitativ erwarten, daß sich die Bindungskonstante deutlich erhöht, wenn man von Chitobiose zu Chitotriose übergeht, diese aber wieder fast auf den Wert von Chitobiose absinkt, wenn man zu DMSO als Lösungsmittel übergeht. Ob dies zutrifft wurde im experimentellen Teil dieser Arbeit durch NMR-Titrations-Experimente (Abschnitt 5.3) untersucht. Vergleicht man die Änderung der Interaktion mit den wichtigen Aminosäuren TRP21, TRP23, TYR30 und SER19 (Abb. X216), so zeigt sich, daß auch in DMSO die wichtigen Interaktionen noch vorhanden sind, wenn auch teilweise mit einer unterschiedlichen Gewichtung.

---

1. Das verwendete DMSO-Modell beruht auf einem „United Atom“-Ansatz bei dem die Methylgruppen durch ein Pseudo-Atom repräsentiert werden.

## 5.3 Experimentelle Bestimmung der Bindungsaffinität von Hevein-Chitotriose-Komplexen in Wasser/DMSO-Mischungen<sup>1</sup> mittels NMR-Methoden

### 5.3.1 Problemstellung und Zielsetzung

Wie in Abschnitt 5.1 bereits dargestellt, hat die Verwendung von aprotischen Lösungsmitteln wie DMSO für NMR-Messungen den Vorteil, daß man auch Informationen über die chemische Umgebung von OH-Protonen erhält. Dies ist von großem Wert für die Strukturaufklärung von Oligosacchariden. Zur Überprüfung, ob die Methode auch zur Analyse von Protein-Kohlenhydrat-Komplexen geeignet ist, war erstens (qualitativ) zu überprüfen, ob sich die Hevein-Struktur durch Beimischung von DMSO verändert und zweitens welchen Einfluß das DMSO auf die Bindungsaffinität besitzt.

Der Einfluß des DMSO auf die Hevein-Struktur sollte qualitativ durch Vergleich der <sup>1</sup>H-NMR-Spektren von Hevein mit zunehmender DMSO-Konzentration in einem ersten Experiment überprüft werden (Abschnitt 5.3.4). Im zweiten Schritt sollte dann die Bindungskonstante in einer DMSO-d<sub>6</sub>/Wasser-Mischung von 4:1 bei 305 K in Analogie zur in [75] beschriebenen Methode bestimmt werden (Abschnitt 5.3.5) und anschließend, sofern eine ausreichend große Bindungskonstante für den Komplex gefunden werden sollte, auch die Bestimmung der Bindungsenthalpie und -entropie durch eine van't Hoff-Analyse durchgeführt werden. Schließlich war geplant durch die Bestimmung der Bindungskonstanten in einer 1:1 DMSO-d<sub>6</sub>/Wasser-Mischung eine Abschätzung der Bindungsaffinität in 100% DMSO vorzunehmen.

### 5.3.2 Beschreibung der Methode

Die wichtigste nichtkalorimetrische Methode zur Bestimmung von Bindungskonstanten sind NMR-Titrations [54,75,80]. Voraussetzung für die Anwendung der Methode ist ein schneller Austausch des Liganden zwischen gebundenem und freiem Zustand relativ zur NMR-Zeitskala. Dies ist normalerweise für Kohlenhydrat-Protein Komplexe erfüllt. Aus einer NMR-Messung erhält man Informationen über die chemische Umgebung eines bestimmten magnetischen Kerns (z.B. <sup>1</sup>H, <sup>13</sup>C, <sup>15</sup>N, <sup>31</sup>P, <sup>19</sup>F). Da sich die chemische Umgebung der direkt an der Komplexbindung beteiligten Atome von der im unkomplexierten Zustand unterscheidet, ist die gemessene chemische Verschiebung (Gleichung 44) abhängig von der Menge an komplexgebundenem Ligand (Gleichung 42) und diese wiederum hängt über die Gleichgewichtskonstante der Komplexbindung von der Ligandenkonzentration in der Lösung ab (Gleichung 43). Beobachtet man bei Zugabe kleiner Portionen von Zuckerlösung zu einer Lektin-Lösung eine Änderung der chemischen Verschiebung bestimmter Signale, so ist dies ein deutlicher Hinweis darauf, daß eine Komplexbildung vorhanden ist.



$$K = \frac{[P \cdot L]}{[P] \cdot [L]} \quad (43)$$

$$\delta_{\text{beobachtet}} = \delta_{\text{frei}} + (\delta_{\text{gebunden}} - \delta_{\text{frei}}) \cdot \frac{[P \cdot L]}{[P] + [P \cdot L]} \quad (44)$$

1. Die experimentellen Arbeiten wurden am Instituto Quimica Organica, CISC, Madrid. Spanien mit freundlicher Unterstützung durch Jesús Jiménez-Barbero und Francisco Javier Cañada durchgeführt.

$$\frac{[P] \cdot [L]}{[P] + [P \cdot L]} = \frac{1}{[P]} \left[ \frac{\left( [P] + [L] + \frac{1}{K} \right)}{2} - \sqrt{\frac{\left( [P] + [L] + \frac{1}{K} \right)^2}{4} - ([P] \cdot [L])} \right] \quad (45)$$

Erhöht man die Konzentration des Liganden systematisch und quantitativ (Titrationsen), so können aus der Änderung der chemischen Verschiebung in Abhängigkeit von der Ligandenkonzentration Bindungskonstanten bestimmt werden [80]. Dabei wird die Konzentration des Proteins üblicherweise konstant gehalten.

Praktisch wird dazu die Proteinlösung auf zwei NMR-Röhrchen verteilt. In einer der beiden Lösungen wird eine möglichst große Menge an Ligand gelöst. Sofern man die Änderung der chemischen Verschiebung von Proteinsignalen verfolgen möchte, wird zuerst ein  $^1\text{H}$ -NMR-Spektrum der reinen Protein-Probe aufgenommen ( $\delta_{\text{free}}$ ) und anschließend definierte kleine Mengen der konzentrierten Ligand/Protein-Lösung zugegeben, wobei nach jeder Zugabe ein  $^1\text{H}$ -NMR-Spektrum aufgenommen wird. Durch Auftragen von  $\delta_{\text{beobachtet}}$  gegen die Ligandenkonzentration erhält eine Titrationskurve aus deren Verlauf sich die Bindungskonstante  $K_a$  und die chemische Verschiebung im gebundenen Zustand  $\delta_{\text{bound}}$  durch nicht-linearen Fit aus Gleichung 44 und 45<sup>1</sup> ermitteln lassen. Sind die Proteinsignale, die eine Änderung der chemischen Verschiebung zeigen, zugeordnet, kann läßt sich durch diese Methode auch das aktive Zentrum des Proteins bestimmen.

Führt man diese NMR-Titrationsen bei verschiedenen Temperaturen durch, so läßt sich durch Auftragung von  $R \log K_a$  gegen  $1/T$  (van't Hoff Analyse) die Bindungsenthalpie und die Bindungsentropie bestimmen.[232]. Diese Methode ist nicht sehr genau, da bei der van't Hoff Analyse angenommen wird, daß  $\Delta C_p$  temperaturunabhängig ist und dies bei Protein-Zucker-Komplexen normalerweise nicht der Fall ist [79]. Man erhält mit der van't Hoff Analyse jedoch zumindest eine semiquantitative Abschätzung, ob die Komplexbildung enthalpisch oder entropisch begünstigt ist.

Alle  $^1\text{H}$ -NMR-Spektren wurden auf einem 500 MHz Varian Unity Spektrometer aufgenommen. Die Zuordnung der zur Bestimmung der Bindungskonstanten verwendeten Proteinsignale wurde nach [75] vorgenommen.

---

1. Gleichung 45 wird durch Umformen und Einsetzen von Gleichung 43 erhalten.

### 5.3.3 Vorbereitung und Reinigung der Hevein-Probe

Eine aus *Hevea brasiliensis* Latex isolierte Hevein-Probe [230] war bereits vorhanden, so daß direkt mit der Vorbereitung der NMR-Probe begonnen werden konnte.

Die Konzentration der Originalprobe (Puffer: 0.05 M Na-Phosphat, 0.1 M NaCl, 15% D<sub>2</sub>O, pH=5.9) wurde durch Messung der UV-Absorption bei 280 nm bestimmt. Es wurden 5 µl der Originalprobe mit 300 µl H<sub>2</sub>O verdünnt (1:61) und ein UV-Spektrum aufgenommen. Die Absorption bei 280 nm betrug 0.262. Daraus errechnet sich die Heveinkonzentration der Original-Probe zu 1.26 mM ( $\epsilon(280\text{nm})=12660 \text{ M}^{-1} \text{ cm}^{-1}$ ).

Zur Reinigung des Heveins und zur Rückgewinnung von Hevein aus Hevein/Zuckergemischen von Titrations-Experimenten wurde Reverse-Phase-HPLC eingesetzt.

Lösungsmittelgradient H<sub>2</sub>O -> CH<sub>3</sub>CN/CF<sub>3</sub>COOH(pH2); Flowrate = 2.2 ml/min, Heliumfluß = 100 ml/min; Druck 800-1000 Psi; Detektion: UV 280 nm, Retentionszeit des Hevein = 35-36 min.

Das Lösungsmittel wurde durch Gefriertrocknung (Lyophilisierung) von Hevein (M=4.73 kDa) entfernt.

### 5.3.4 Änderung der chemischen Verschiebung der Protonensignale von Hevein durch Zugabe von DMSO

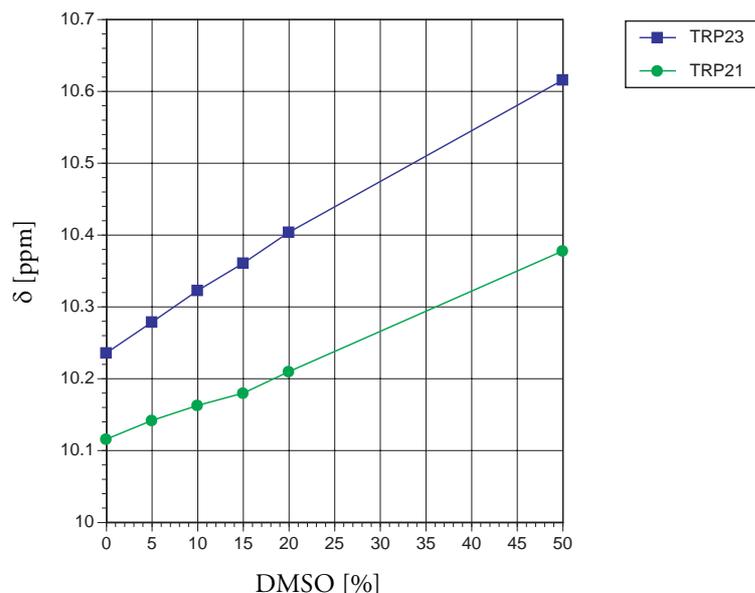
#### Beschreibung des NMR-Experiments

Aus der Original-Probe wurden 175 µl entnommen und mit 275 µl Puffer (0.05 M Na-Phosphat, 0.1 M NaCl, 15% D<sub>2</sub>O, pH=5.9) verdünnt (1:2.57). Die Ausgangskonzentration an Hevein in der Meßprobe betrug damit 0.49 mM.

Zur Meßprobe wurden nacheinander 24 µl (= 5%), 26 µl (= 10%), 30 µl (= 15%) und 32.5 µl (= 20%) DMSO-d<sub>6</sub> zugefügt und jeweils ein 1D-<sup>1</sup>H-NMR-Spektrum bei 32°C aufgenommen. Die Konzentration von Hevein nach der letzten DMSO-Zugabe betrug 0.39 mM (Volumen der Probe = 562.5 µl).

#### Ergebnis

Die Spektren (Abb. X217) sind selbst bei 50% DMSO noch gut aufgelöst, so daß technisch gesehen nichts gegen die Bestimmung der Bindungskonstanten in DMSO/Wasser-Mischungen spricht. Die Signale der für den Bindungsprozess wichtigen Protonen (NH-Protonen von TRP21, TRP23 und GLN20;  $\gamma$ -Proton von GLN20) zeigen eine Tieffeldverschiebung mit zunehmender DMSO Konzentration (Abb. 58).



**Abb. 58:** Chemische Verschiebung von TRP23(NH1) und TRP21(NH1) in Abhängigkeit von der DMSO-Konzentration

### 5.3.5 Hevein/Chitotriose: Bestimmung der Bindungskonstanten in 20% DMSO

#### Vorbereitung der Probe

Aus der 20%-DMSO-Hevein-Probe (562.5  $\mu$ l; 0.39 mM Hevein) (vgl. Abschnitt 5.3.4) wurden 100  $\mu$ l entnommen und darin 1.6 mg Chitotriose (Sigma Chemical Co.) aufgelöst (Konzentration der Chitotriose-Lösung = 25.52 mM; 1mM = 0.627 mg/ml).

#### Beschreibung des NMR-Experiments

Zur Ausgangsprobe wurden nacheinander 5  $\mu$ l, 6  $\mu$ l, 10  $\mu$ l, 20  $\mu$ l, 25  $\mu$ l und 34  $\mu$ l der Chitotriose-Lösung zugefügt und jeweils ein 1D- $^1$ H-NMR-Spektrum bei 305 K aufgenommen. Die Titration wurde mit konstanter Hevein-Konzentration durchgeführt.

#### Ergebnisse

Die Spektren (Abb. X218) zeigen mit zunehmender Chitotriose-Konzentration eine deutliche Linienverbreiterung der Signale der NH1-Protonen von TRP21 ( $\delta=10.21$  ppm) und TRP23 ( $\delta=10.40$  ppm). Weiterhin zeigt sich eine Tieffeldverschiebung des NH1-Protons von TRP21 ( $\delta=10.21$  ppm) und der NH-Protonen von GLN20 ( $\delta=9.77$  ppm), GLY25 ( $\delta=9.19$  ppm) und GLN6<sup>1</sup> ( $\delta=9.05$  ppm), desweiteren eine Hochfeldverschiebung des  $\gamma$ -Protons von GLN20 ( $\delta=0.68$  ppm). Durch multivariaten Fit nach Gleichung 44 und 45 wurde aus der Abhängigkeit der chemischen Verschiebung von der Chitotriose-Konzentration (Abb. X219) die Bindungskonstante  $K_a$  bestimmt (Tabelle 7).

1. Die Zuordnung der Signale wurde basierend auf den in [75] publizierten chemischen Verschiebungen vorgenommen.

Eine sichere Zuordnung der Signale anhand der 1D- $^1$ H-Spektren ist nicht möglich, da die publizierten Werte in Wasser gemessen wurden und daher nur eine grobe Näherung für die chemische Verschiebung in DMSO darstellen. Da im vorliegenden Zusammenhang nur die Bestimmung der Bindungskonstanten interessiert, ist die Zuordnung der Peaks zu einzelnen Protonen nur von untergeordneter Bedeutung. Die Bezeichnungen sind daher in erster Linie als Variablenamen zu verstehen. Dies gilt insbesondere für „GLN6“, da die Aminosäure GLN6 aufgrund ihrer Position relativ zu Bindungstasche theoretisch nicht an der Bindung beteiligt sein kann.

**Tabelle 7: Ergebnisse der multivariaten Regression für 305 K (20% DMSO)**

	TRP21 (NH1)	GLN20 (NH)	GLN20 ( $\gamma$ H)	GLY25 (NH)	“GLN6“ (NH)
$K_a$ [ $M^{-1}$ ]	917 $\pm$ 360	784 $\pm$ 143	831 $\pm$ 137	1352 $\pm$ 360	939 $\pm$ 226
$\delta_{\text{bound}}$ [ppm]	10.26 $\pm$ 0.001	9.944 $\pm$ 0.009	0.349 $\pm$ 0.016	9.238 $\pm$ 0.003	9.146 $\pm$ 0.007
$\delta_{\text{free}}$ [ppm]	10.21 $\pm$ 0.001	9.775 $\pm$ 0.003	0.675 $\pm$ 0.006	9.190 $\pm$ 0.001	9.043 $\pm$ 0.003
$\Delta G = -RT \ln K_a$ [kJ/mol]	-17.29 $\pm$ 1.1	-16.90 $\pm$ 0.4	-17.05 $\pm$ 0.4	-18.28 $\pm$ 0.7	-17.35 $\pm$ 0.8

Aus der NMR-Titration ergibt sich für Chitotriose in 20%iger DMSO-Lösung eine mittlere Bindungskonstante von  $965 \pm 245 M^{-1}$  und eine Freie Bindungsenthalpie von  $-17.4 \pm 0.6$  kJ/mol. Die Bindungskonstante in 20% DMSO ist damit bereits um eine Größenordnung gegenüber der Bindungskonstante in reinem Wasser reduziert [75]. Sie ist aber noch größer als die für den Hevein-Chitotriose-Komplex bestimmte, daher kann auch bei einer Temperaturerhöhung von einer ausreichend starken Bindung ausgegangen werden<sup>1</sup>.

### 5.3.6 Hevein/Chitotriose: Van't Hoff Analyse (20% DMSO)

#### Vorbereitung der Probe

1.7 mg des durch HPLC gereinigten Heveins wurden in 720  $\mu$ l Puffer (0.05 M Na-Phosphat, 0.1 M NaCl, 15% D<sub>2</sub>O, pH=5.9) und 180  $\mu$ l DMSO gelöst. 450  $\mu$ l wurden entnommen und darin 6 mg Chitotriose aufgelöst. Die Hevein-Konzentration wurde durch UV-Absorptionsmessung zu 0.212 mM bestimmt. D.h. die Heveinprobe war wahrscheinlich nicht ganz trocken, da theoretisch eine Konzentration von 0.39 mM zu erwarten gewesen wäre. Für die Auswertung wurde der Wert 0.212 mM verwendet.

#### Beschreibung des NMR-Experiments

Zur Ausgangsprobe wurden nacheinander 5  $\mu$ l, 5 $\mu$ l, 10 $\mu$ l, 20  $\mu$ l, 35  $\mu$ l, 75  $\mu$ l, 100  $\mu$ l und 100  $\mu$ l der Chitotriose-Lösung zugefügt und jeweils ein 1D-<sup>1</sup>H-NMR-Spektrum bei 300 K, 310 K und 315 K aufgenommen.

#### Ergebnisse

Die aus den Titrations-Kurven (Abb. X220 bis Abb. X222) berechneten Bindungskonstanten  $K_a$  sind in Tabelle 8 bis Tabelle 10 zusammengefasst. Während die aus den Bindungskonstanten  $K_a$  berechnete Freie Enthalpie für alle Temperaturen relativ gut reproduzierbar ist (Abb. 59), ergibt sich für die Enthalpie- und Entropiewerte aus der van't Hoff -Analyse (Abb. X223 bis Abb. X227) ein uneinheitliches Bild. Es geht zwar eindeutig daraus hervor, daß die Enthalpie negativ ist, das Vorzeichen der Entropie kann jedoch nicht sicher bestimmt werden (Tabelle 11). Unter der Annahme, daß die mittlere Enthalpie mit einem kleineren Fehler behaftet ist als die Entropie, läßt sich über den Umweg über die Freie Enthalpie trotzdem eine Abschätzung über das Vorzeichen der Entropie machen: Aus der van't Hoff Analyse ergibt sich eine mittlere Enthalpie  $\Delta H = -26.4$  kJ/mol und eine mittlere Freie Enthalpie  $\Delta G = -17.3$  kJ/mol. Aus der Gibbs-Helmholtz-Gleichung errechnet sich daraus bei 300K

1. Bei exothermen Reaktionen nimmt die Gleichgewichtskonstante bei Temperaturerhöhung ab.

eine Entropie von  $-30$  J/mol. Dies ist ein Hinweis, darauf, daß sich beim Übergang von reinem Wasser zu einer Wasser/DMSO-Mischung im Verhältnis 4:1 nicht nur die Enthalpie von  $-36.4$  kJ/mol auf  $-26.4$  kJ/mol reduziert hat, sondern sich auch der Entropieanteil des Systems bei der Komplexbildung weniger erniedrigt hat. Dies ist ein weiteres Beispiel für das als „Enthalpie-Entropie-Kompensation“ bezeichnete Phänomen [81,86].

**Tabelle 8: Ergebnisse der multivariaten Regression für 20% DMSO (300 K)**

	TRP21 (NH1)	GLN20 (NH1)	GLN20 ( $\gamma$ H)	GLY25 (NH)	“GLN6“ (NH)
$K_a$ [ $M^{-1}$ ]	394 $\pm$ 43	1198 $\pm$ 131	1236 $\pm$ 119	1368 $\pm$ 150	1089 $\pm$ 144
$\delta_{\text{bound}}$ [ppm]	10.30 $\pm$ 0.002	10.040 $\pm$ 0.003	0.303 $\pm$ 0.004	9.268 $\pm$ 0.001	9.144 $\pm$ 0.002
$\delta_{\text{free}}$ [ppm]	10.24 $\pm$ 0.001	9.879 $\pm$ 0.003	0.571 $\pm$ 0.005	9.207 $\pm$ 0.001	9.045 $\pm$ 0.002
$\Delta G = -RT \ln K$ [kJ/mol]	-14.90 $\pm$ 0.25	-17.68 $\pm$ 0.26	-17.75 $\pm$ 0.25	-18.01 $\pm$ 0.27	-17.44 $\pm$ 0.28

**Tabelle 9: Ergebnisse der multivariaten Regression für 20% DMSO (310 K)**

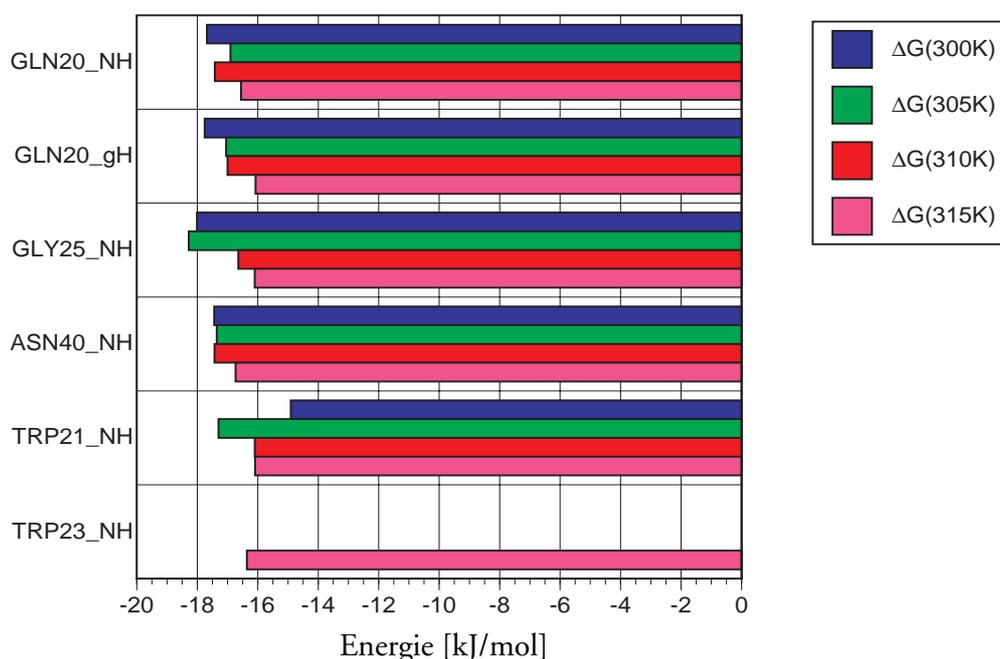
	TRP21 (NH1)	GLN20 (NH1)	GLN20 ( $\gamma$ H)	GLY25 (NH)	“GLN6“ (NH)
$K$ [ $M^{-1}$ ]	636 $\pm$ 75	1079 $\pm$ 66	911 $\pm$ 51	791 $\pm$ 30	1083 $\pm$ 63
$\delta_{\text{bound}}$ [ppm]	10.26 $\pm$ 0.001	9.977 $\pm$ 0.002	0.356 $\pm$ 0.003	9.236 $\pm$ 0.001	9.144 $\pm$ 0.001
$\delta_{\text{free}}$ [ppm]	10.20 $\pm$ 0.001	9.819 $\pm$ 0.002	0.656 $\pm$ 0.003	9.180 $\pm$ 0.001	9.056 $\pm$ 0.001
$\Delta G = -RT \ln K$ [kJ/mol]	-16.63 $\pm$ 0.31	-18.00 $\pm$ 0.15	-17.56 $\pm$ 0.14	-17.20 $\pm$ 0.10	-18.01 $\pm$ 0.15

**Tabelle 10: Ergebnisse der multivariaten Regression für 20% DMSO (315 K)**

	TRP23	TRP21	GLN20 (NH)	GLN20 ( $\gamma$ H)	GLY25 (NH)	“GLN6“ (NH)
$K$ [ $M^{-1}$ ]	706 $\pm$ 127	633 $\pm$ 82	762 $\pm$ 83	629 $\pm$ 81	636 $\pm$ 44	820 $\pm$ 47
$\delta_{\text{bound}}$ [ppm]	10.37 $\pm$ 0.001	10.24 $\pm$ 0.001	9.957 $\pm$ 0.003	0.383 $\pm$ 0.008	9.217 $\pm$ 0.001	9.144 $\pm$ 0.001
$\delta_{\text{free}}$ [ppm]	10.35 $\pm$ 0.001	10.18 $\pm$ 0.001	9.785 $\pm$ 0.003	0.681 $\pm$ 0.007	9.169 $\pm$ 0.001	9.065 $\pm$ 0.001
$\Delta G = -RT \ln K$ [kJ/mol]	-17.17 $\pm$ 0.45	-16.89 $\pm$ 0.35	-17.37 $\pm$ 0.31	-16.87 $\pm$ 0.35	-16.90 $\pm$ 0.18	-17.57 $\pm$ 0.15

**Tabelle 11: Enthalpie und Entropie für das Hevein-Chitotriose-System in 20% DMSO**

	$\Delta H^0$ [kJ mol <sup>-1</sup> ]	$\Delta S^0$ [J T <sup>-1</sup> mol <sup>-1</sup> ]
GLN20_NH	-16.3	-3.7
GLN20_γH	-30.4	42.5
GLY25_NH	-44.4	87.2
GLN6_NH	-11.1	-21.2
TRP21_NH1	-29.7	41.4

**Abb. 59:** Freie Bindungsenergie der Bindung von Chitotriose an Hevein (20% DMSO)

### 5.3.7 Hevein-Chitotriose: Bestimmung der Bindungskonstante in 50% DMSO (309 K)

#### Vorbereitung der Probe

Aus der Original-Probe (Abschnitt 5.3.3) wurden 225  $\mu$ l entnommen und mit 225  $\mu$ l Puffer (0.05 M Na-Phosphat, 0.1 M NaCl, 15% D<sub>2</sub>O, pH=5.9) und 450  $\mu$ l DMSO verdünnt. Zur Konzentrationsbestimmung wurden 20  $\mu$ l der Hevein-Probe mit 300  $\mu$ l H<sub>2</sub>O verdünnt (1:16) und ein UV-Spektrum aufgenommen. Die Absorption bei 280 nm betrug 0.230. Mit einem Extinktionskoeffizienten  $\epsilon(280\text{nm})=12660 \text{ M}^{-1} \text{ cm}^{-1}$  erhält man eine Heveinkonzentration der Probe von 0.29 mM.

Aus der 50%-DMSO-Probe wurden 450  $\mu$ l entnommen und darin 1.5 mg Chitotriose (= 5.3 mM) aufgelöst.

## Beschreibung des NMR-Experiments

Zur Ausgangsprobe wurden nacheinander 15  $\mu\text{l}$ , 15  $\mu\text{l}$ , 20  $\mu\text{l}$ , 30  $\mu\text{l}$ , 50  $\mu\text{l}$ , 70  $\mu\text{l}$ , 100  $\mu\text{l}$  und 117  $\mu\text{l}$  der Chitotriose-Lösung zugefügt und jeweils ein  $1\text{D-}^1\text{H-NMR}$ -Spektrum bei  $36^\circ\text{C}$  aufgenommen.

## Ergebnisse

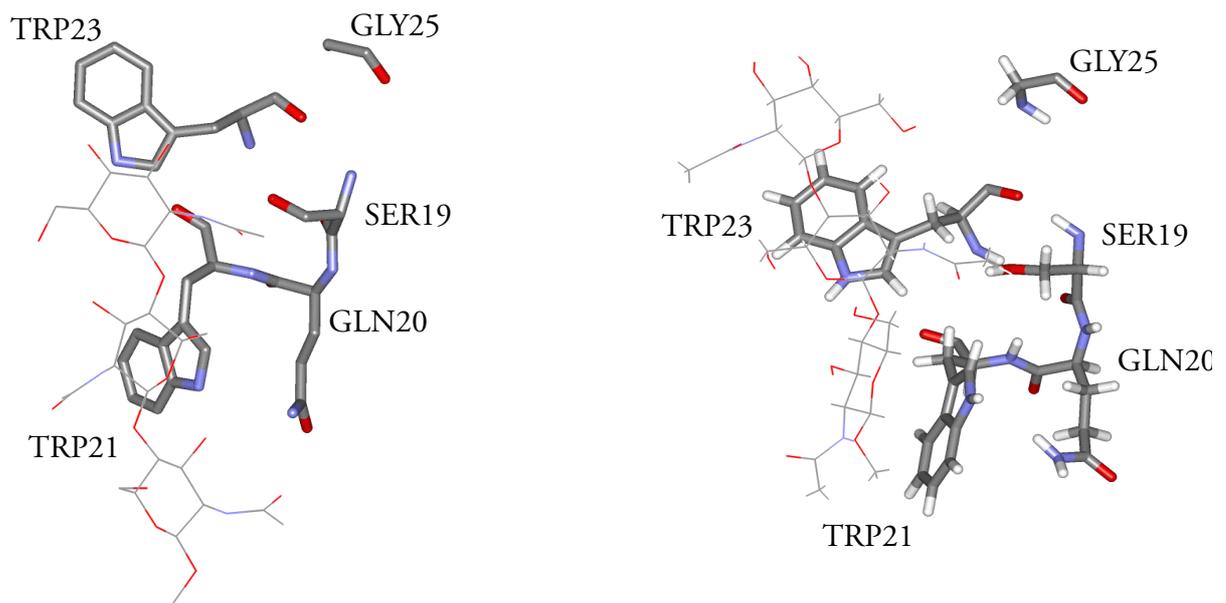
Die Spektren (Abb. X228) zeigen mit zunehmender Chitotriose-Konzentration eine Linienverbreiterung für das NH1-Proton von TRP23, eine Tieffeldverschiebung des Signal für das NH-Proton von GLN20, desweiteren eine Hochfeldverschiebung des NH1-Protons von TRP23. Eine Hochfeldverschiebung des  $\gamma$ -Protons von GLN20 kann wegen Signalüberlagerung nicht beobachtet werden. Eine Verschiebung des TRP21-Signals ist praktisch kaum noch zu sehen, die Linienverbreiterung ist zwar noch vorhanden, aber weniger ausgeprägt wie bei 20% DMSO. Aus der selektiven Verschiebung der Signalen geht jedoch eindeutig hervor, daß auch in 1:1 DMSO/Wasser-Gemischen immer noch eine Komplexbindung von Chitotriose an Hevein stattfindet. Die Auswertung der Titrations-Kurven (Abb. X229) ergibt eine mittlere Bindungskonstante von  $460\text{ M}^{-1}$  (Tabelle 12). Die auffällige Differenz in den Bindungskonstanten aus GLN20(NH) und GLY25(NH) einerseits ( $K_a=640 \pm 125\text{ M}^{-1}$ ) und TRP23 (NH1) und „GLN6“(NH) andererseits ( $K_a'=276 \pm 47\text{ M}^{-1}$ ) legt den Schluß nahe, daß sich die Bindungsaffinitäten zu den einzelnen Aminosäuren nicht gleichmäßig geändert haben, sondern daß eine teilweise Ablösung des Liganden stattgefunden hat.

**Tabelle 12: Ergebnisse der multivariaten Regression für 50% DMSO (309 K)**

	GLN20 (NH)	TRP23 (NH1)	GLY25 (NH)	„GLN6“ (NH)
$K_a\text{ [M}^{-1}\text{]}$	$523 \pm 131$	$231 \pm 11$	$765 \pm 120$	$322 \pm 84$
$\delta_{\text{bound}}[\text{ppm}]$	$9.967 \pm 0.008$	$10.41 \pm 0.0005$	$9.189 \pm 0.003$	$9.021 \pm 0.004$
$\delta_{\text{free}}[\text{ppm}]$	$9.877 \pm 0.002$	$10.62 \pm 0.0005$	$9.129 \pm 0.001$	$9.058 \pm 0.001$
$\Delta G = -RT \ln K$ [kJ/mol]	$-16.1 \pm 0.5$	$-14.0 \pm 0.1$	$-17.05 \pm 0.35$	$-14.83 \pm 0.6$

Sofern die Zuordnung GLY25 ( $\delta=9.19\text{ ppm}$ ) richtig ist<sup>1</sup>, wäre die hohe Bindungskonstante für GLY25(NH) ein Hinweis auf das Vorliegen des Strukturvorschlags B (Abschnitt 5.2.5), da nur dort eine Interaktion zwischen Hevein und Chitotriose möglich ist (vgl. Abb. 60).

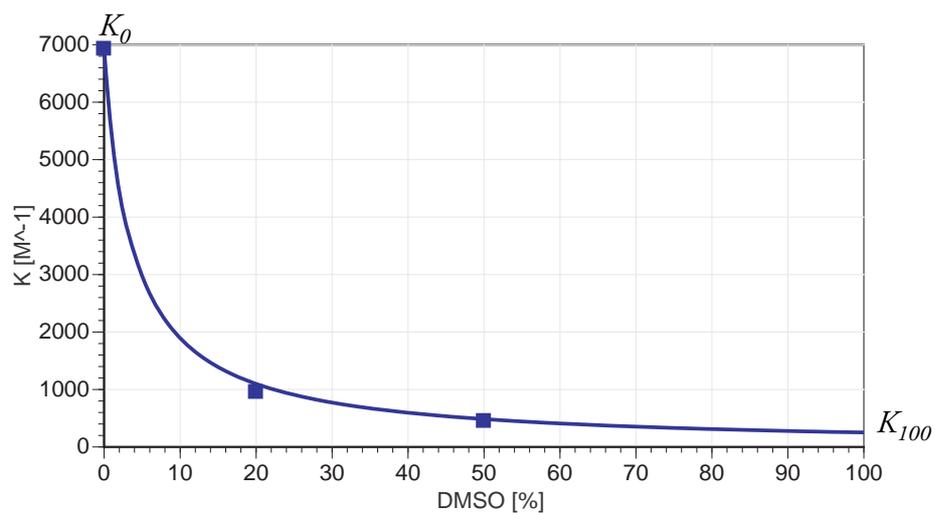
1. Eine sichere Zuordnung ist nur anhand von  $2\text{D-}^1\text{H-NMR}$ -Spektren möglich.



**Abb. 60:** Strukturvorschlag A und B für den Hevein/Chitotriose -Komplex

### 5.3.8 Änderung der Bindungskonstante mit zunehmender DMSO-Konzentration

Aus der Änderung der mittleren Bindungskonstanten  $K_a$  für 0% [75], 20% und 50% DMSO lässt sich auf eine Erniedrigung der Bindungskonstanten mit  $a/(b+x)$  schließen ( $a= 22825$ ;  $b= 3.3$ ;  $R^2=0.99$ ). Unter dieser Voraussetzung wäre in reinem DMSO immerhin noch eine Bindungskonstante von  $250 \text{ M}^{-1}$  zu erwarten.



**Abb. 61:** Änderung der Bindungskonstante des Hevein-Chitotriose-Komplexes bei Erhöhung der Konzentration von DMSO

### 5.3.9 Interpretation der Ergebnisse

Aus den  $^1\text{H}$ -NMR-Messungen in DMSO/Wasser-Mischungen ergibt sich, daß die Änderung des Lösungsmittels einen deutlichen Einfluß auf die Bindungsaffinität von Protein-Kohlenhydraten besitzt. Durch Zugabe von 20% DMSO erniedrigt sich die Bindungskonstante bereits um eine Größenordnung. Die weitere Zugabe von DMSO erniedrigt die Bindungskonstante nicht mehr in gleichem Maße (vgl. Abb. 61). Da durch qualitativen Vergleich eines 2D-NOESY-Spektrums einer Lösung von Hevein und Chitotriose in 50% DMSO mit einem Spektrum, das in reinem Wasser aufgenommen wurde, festgestellt wurde, daß alle strukturbestimmenden NOEs für Hevein noch vorhanden sind, kann davon ausgegangen werden, daß selbst in einer 1:1 DMSO/Wassermischung die Tertiärstruktur von Hevein im wesentlichen noch erhalten ist. Die Vermutung liegt nahe, daß es sich bei der beobachteten Erniedrigung der Bindungskonstanten hauptsächlich um einen Solvensreorganisationseffekt handelt. Der Anteil, den die Reorganisation des Lösungsmittels als Folge der Komplexbildung an der Freien Enthalpie hat, wird in der Literatur kontrovers diskutiert [84,233]. Daß die Reorganisation der Solvensstruktur jedoch einem deutlichen Einfluß auf die Komplexbildung hat, zeigt sich daran, daß schon die Verwendung von  $\text{D}_2\text{O}$  statt  $\text{H}_2\text{O}$  als Lösungsmittel die Enthalpie von Di- und Trisacchariden komplexiert an Concanavalin A um 0.9 bis 7.5 kJ/mol reduziert [234]. Da bei diesem Experiment davon ausgegangen werden kann, daß sowohl die Konformationen von Protein und Ligand nicht durch den H $\rightarrow$ D-Austausch beeinflußt werden, kann geschlossen werden, daß die negative Enthalpie von der Reorganisation der Solvens als Folge der Komplexbildung herrührt [84]. Es ist bekannt, daß die Entfernung von Wassermolekülen von einer Proteinoberfläche ein enthalpisch begünstigter Prozeß ist [81,90]. Zwar kann die Stärke einzelner H-Brücken in einer Solvenshülle größer sein als im freien Wasser, dafür ist jedoch die Anzahl der durch Solvataion gebrochener H-Brücken, bedingt durch die geometrischen Verhältnisse auf der Proteinoberfläche, ebenfalls größer. Ein Wassermolekül kann im freien Wasser insgesamt eine energetisch viel günstigere Anordnung annehmen, als in einer Solvenshülle. Dies wird auch durch MMC-Rechnungen bestätigt [235].

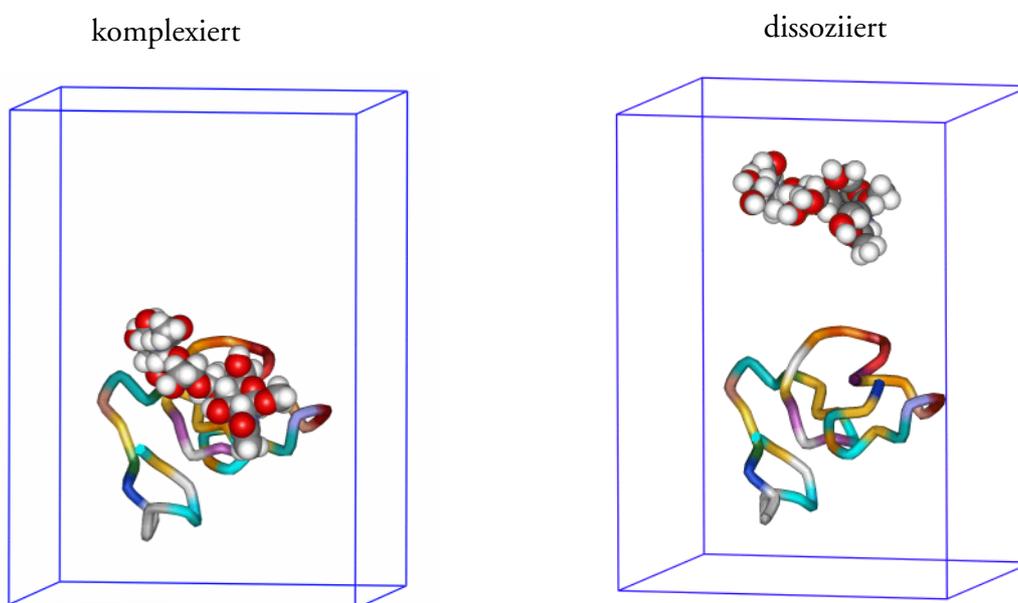
Geht man also davon aus, daß ein wesentlicher Teil der Bindungsaffinität dadurch verursacht wird, daß Wassermoleküle durch die Komplexbildung von einer Protein- oder Zuckeroberfläche ins Solvens entlassen werden und dort in eine besonders energiearmes Netzwerk von H-Brückenbindungen integriert werden, so läßt sich die gefundene Abhängigkeit der Bindungskonstanten von der DMSO-Konzentration (vgl. Abb. 61) wie folgt erklären: durch Zugabe von wenig DMSO wird die Wasserstruktur bereits deutlich gestört [236,237]. Die Ausbildung eines größeren Netzwerks von Wasserstoffbrückenbindungen ist nicht mehr möglich. Aus diesem Grund ist die Stabilisierung, die das Gesamtsystem durch Übergang von Wassermolekülen von der Proteinoberfläche ins freie Lösungsmittel erfährt, deutlich geringer. Durch die Zugabe von DMSO sind im Wasser viele „Mikroproteinoberflächen“ in Form von DMSO-Molekülen vorhanden und es ergibt sich für ein Wassermolekül energetisch kein wesentlicher Unterschied, ob es sich in einer Solvenshülle von einem Protein oder einem DMSO-Molekül befindet. Um den in Abb. 61 gezeigten Verlauf der Bindungsaffinität in Abhängigkeit von der DMSO-Konzentration weiter zu untermauern, wäre es sinnvoll die Bindungskonstante zusätzlich noch in einer 5%igen DMSO/Wasser-Mischung zu bestimmen. Würde sich der steile Abfall der Kurve bei niedriger DMSO-Konzentration bestätigen, wäre dies ein wichtiges Indiz für die Richtigkeit der Annahme, daß die Abnahme der Bindungsaffinität durch DMSO-Zugabe auf die in geringerem Maße freiwerdende Solvensreorganisationenthalpie zurückzuführen ist. Da die Bindungskonstante mit zunehmender DMSO-Konzentration gegen einen Grenzwert  $K_{100}$  konvergiert könnte aus der Differenz  $\Delta G(\text{reines Wasser}) - \Delta G(100\% \text{DMSO})$  der Einfluß des Wassers auf die Komplexbildungskonstante bestimmt werden. Voraussetzung hierfür ist jedoch, daß sich die Struktur des Proteins durch die Zugabe von DMSO nicht ändert. Dies kann für hohe Konzentrationen von DMSO nicht ausgeschlossen werden [213]. Bestimmt man die Bindungskonstanten nur für geringe Konzentrationen von DMSO so ist anzunehmen, daß das DMSO nur geringe Effekte auf den Kom-

plex, aber große Effekte auf die Solvensstruktur hat. Durch eine Serie von Messungen zwischen 0% und 50% DMSO könnten genug Datenpunkte erhalten werden, um durch einen Fit  $K_{100}$  und damit den „Wassereffekt“ zu bestimmen.

## 5.4 Zusammenfassung und Ausblick

Im Rahmen der theoretischen Analyse von Beiträgen zur Bindungsaffinität, die auf der direkten Wechselwirkung von Protein und Zucker beruhen, wurden durch eine detaillierte Interaktionsanalyse auf Residue-Basis die wesentlichen Wechselwirkungen zwischen Hevein und Chitobiose in Übereinstimmung mit den experimentellen Ergebnissen gefunden und nach ihrer Wichtigkeit geordnet. Weiterhin ist es gelungen, basierend auf Interaktionsanalysen, sinnvolle Strukturvorschläge für den Hevein-Chitotriose-Komplex herauszuarbeiten und im Vergleich zum Hevein-Chitobiose-Komplex zusätzliche günstige Wechselwirkungen mit dem Protein zu lokalisieren. Es ist daher wahrscheinlich, daß diese zusätzlichen Wechselwirkungen zur Erhöhung der Bindungsaffinität beim Übergang von Chitobiose zu Chitotriose signifikant beitragen.

Wie aus den  $^1\text{H}$ -NMR-Messungen klar hervorgeht, beruht ein wesentlicher Anteil der stabilisierenden Effekte jedoch nicht auf der direkten Wechselwirkung zwischen Protein und Ligand, sondern auf dem Enthalpieunterschied zwischen  $\text{H}_2\text{O}$ (Solvenschülle) und  $\text{H}_2\text{O}$ (freies Wasser). Der Solvenseinfluß läßt sich daher durch MD-Simulationen eines Protein-Kohlenhydrat-Systems ausschließlich im gebundenen Zustand nicht vollständig beschreiben. Zur Simulation von Solvenseffekten ist es daher notwendig jeweils eine MD-Simulation im gebundenen Zustand und eine entsprechende MD-Simulation im dissoziierten Zustand durchzuführen (vgl. Abb. 62) und anschließend die Unterschiede der Solvenseigenschaften zu analysieren. Dies war aufgrund der Größe des zu simulierenden Systems und wegen der zu analysierenden Datenmenge bis vor kurzem ein technisch nicht durchführbares Vorhaben. Mittlerweile sind jedoch durch den Einsatz von Parallelrechnern, wie auch vom sonstigen Hard-



**Abb. 62:** MD-Simulation von Chitotriose und Hevein im gebundenen und dissoziierten Zustand. (Das Wasser in der Solventbox ist der Übersichtlichkeit wegen nicht dargestellt).

wareumfeld, die technischen Voraussetzungen vorhanden um Simulationen dieser Größenordnung in akzeptabler Zeit durchzuführen. Das technische Problem der Datenanalyse wurde im Rahmen dieser Arbeit durch Entwicklung der Analyse-Software CAT2000 gelöst.

Im Prinzip sind damit die technischen Voraussetzungen geschaffen um Fragestellungen der Art: Welche Faktoren beeinflussen die Bindungsaffinität von Protein-Kohlenhydrat-Komplexen und wo muß man ansetzen, um die Affinität eines Liganden zu erhöhen? - auf einem simulationstechnischen Niveau zu bearbeiten, das bis vor kurzem noch als undurchführbar galt.

## 6 Zusammenfassung

Die Entschlüsselung der wesentlichen Faktoren, die dafür verantwortlich sind, daß ein Lektin spezifisch an ein bestimmtes Oligosaccharid bindet, bildet die Grundlage zur Entwicklung von hochaktiven Therapeutika auf Zucker-Basis. Im Rahmen dieser Arbeit wurden daher Simulations-Strategien und Analyse-Methoden entwickelt, die es ermöglichen, in effizienter Weise Konformationsanalysen von Oligosacchariden im freien und gebundenen Zustand durchzuführen. Durch eine detaillierte Wechselwirkungsanalyse auf Residue-Basis und durch die Untersuchung des Lösungsmiteleinflusses auf die Bindungsaffinität mittels NMR-Methoden ist es gelungen, wesentliche Aspekte der Komplexbildung herauszuarbeiten.

Hochtemperatur-Moleküldynamik (HTMD)-Simulationen bilden die Grundlage zur Berechnung der relativen Freien Energie einer Konformation im ungebundenen Zustand. Die Berechnung der Freien Gridgruppen-Energie (FGE) aus der relativen Population einer Konformation (Gridgruppe) im thermischen Gleichgewicht ist eine der zentralen Methoden der vorliegenden Arbeit. Voraussetzung der Methode ist eine Überprüfung, ob sich während der MD-Simulation das Konformationsgleichgewicht für die zu analysierenden Konformationen eingestellt hat. Hierzu wurden, basierend auf einem empirischen Modell, EQ-Faktoren definiert, welche das Erreichen des Konformationsgleichgewichts anzeigen. Über die Freie Gridgruppen-Energie erhält man aus den HTMD-Daten einen direkten Zugang zu den Rotationsprofilen der  $\Phi$ ,  $\Psi$  und  $\omega$ -Torsionswinkel eines Oligosaccharids. Weiterhin lassen sich aus den Daten einer einzigen HTMD-Simulation die  $\Phi/\Psi$ -Maps aller glycosidischen Bindungen berechnen. Am Beispiel des biologisch wichtigen Tetrasaccharids Sialyl-Lewis<sup>X</sup> wurde gezeigt, daß zum Erreichen des Konformationsgleichgewichts im  $\Phi/\Psi$ -Raum eine Simulationszeit von ca. 30 ns notwendig ist (Simulationstemperatur = 1000 K). Die hierfür benötigte CPU-Zeit liegt in der Größenordnung von einem Tag (IBM SP2 oder SGI O2).

Weiterhin konnte gezeigt werden, daß HTMD-Simulationen geeignet sind, um alle stabilen  $\Phi/\Psi$ -Konformationen von Oligosacchariden (lokale Minima) im n-dimensionalen Torsionshyperraum aufzuspüren. Voraussetzung hierfür ist eine leistungsfähige Analyse-Software, die sowohl die hierbei anfallenden großen Datenmengen (bis zu  $10^6$  Datenpunkte) verarbeiten kann, als auch die notwendigen Gruppierungs-Funktionen zur Verfügung stellt, um die stabilen Konformationen im Torsionshyperraum zu markieren und hinsichtlich ihrer Eigenschaften zu analysieren. Ein solches Analyse-Programm wurde im Rahmen dieser Arbeit entwickelt (CAT2000). Durch die Implementierung einer leistungsfähigen Script-Sprache (CSL) ist es möglich, grundlegende Konformationsanalysen vollständig zu automatisieren (High Performance Analysis) oder komplexe Analysen interaktiv durchzuführen. Die HTMD-Methode ist soweit ausgearbeitet, daß Konformationsanalysen von Oligosacchariden im Routinebetrieb von Servicepersonal durchgeführt werden können.

Nicht alle Aspekte des konformativen Verhaltens eines Oligosaccharids lassen sich durch HTMD-Simulationen analysieren. So lassen sich zum Beispiel die spezifischen Effekte, die durch Ausbildung von H-Brückenbindungen verursacht werden, nur bei Simulationstemperaturen  $< 500$  K untersuchen. Zur Überprüfung, ob eine Gleichgewichtseinstellung im  $\Phi/\Psi$ -Raum mit einer Simulationstemperatur von 300 K überhaupt zur Zeit technisch möglich ist, wurde eine MD-Simulation von  $\beta$ -Maltose mit einer Gesamtsimulationszeit von 2  $\mu$ s (Mikrosekunden!) durchgeführt (CPU Zeit  $\approx 1600$  h; IBM SP2). Zwar sind innerhalb der Simulationsperiode vier syn-anti-Übergänge erfolgt, was vorher bei 300 K-MD-Simulationen von Maltose noch nie beobachtet werden konnte, eine Einstellung des Konformationsgleichgewichts konnte jedoch nicht erreicht werden. Als Schlußfolgerung ergibt sich, daß Standard-MD-Simulationen bei 300 K nicht geeignet sind um Konformerübergänge im  $\Phi/\Psi$ -Raum zu untersuchen.

Am Beispiel von  $\beta$ -Maltose wurde weiterhin überprüft, wie lange es dauert, bis sich bei einer MD-Simulation bei Raumtemperatur (300 K) das Rotamerengleichgewicht für die OH-Gruppen bzw. die Hydroxymethylgruppen ( $\omega$ ) eingestellt hat. Ergebnis: Für eine einzelne sekundäre OH-Gruppe ist durchschnittlich eine Simulationsperiode von 500 ps notwendig, um den Gleichgewichtszustand zu erreichen. Die Hydroxymethylgruppen equilibrieren deutlich langsamer; erst nach 300-400 ns hat sich das Gleichgewicht eingestellt. Diese Zeitspanne ist nahezu zwei Größenordnungen länger, als die momentan üblichen Simulationszeiten ( $< 10$  ns), so daß nicht davon ausgegangen werden kann, daß gegenwärtig durch Routine-MD-Simulationen bei 300 K ein Konformationsgleichgewicht für die Hydroxymethylgruppen erreicht werden kann. Die Populationsverteilung der primären OH-Gruppen hängt von der Orientierung der Hydroxymethylgruppen ab, daher ist auch hier keine Gleichgewichtseinstellung innerhalb weniger Nanosekunden zu erwarten.

Um trotzdem eine Konformationsanalyse im  $\Phi/\Psi/\omega$ -Raum bei 300 K zu ermöglichen, wurde die GCMD-Methode entwickelt. Die Anwendung der GCMD-Methode bietet für Oligosaccharide mittlerer Größe (bis zum Pentasaccharid) Zugang zu den Eigenschaften der lokalen Minima im  $\Phi/\Psi/\omega$ -Raum unter gleichzeitiger Berücksichtigung der Rotamerengleichgewichte der OH-Gruppen bei 300 K. Diese Methode kann daher in Kombination mit NMR-Methoden wichtige Beiträge zur Strukturaufklärung von Oligosacchariden liefern.

Die Durchführung von MD-Simulationen bei verschiedenen Simulationstemperaturen ermöglicht es aus den mittleren Lebensdauern der Konformationen deren effektive Übergangsbarrieren durch Arrhenius-Analysen zu bestimmen. Hierzu wurde die technische Umsetzung entwickelt und in CAT2000 implementiert. Weiterhin wurden Methoden zum Aufspüren von spezifischen Solvenseffekten entwickelt, so daß insgesamt ein ganzer Pool von Methoden zur Verfügung steht, um die thermodynamischen und kinetischen Eigenschaften eines Oligosaccharids auch in Lösung zu analysieren.

Im Rahmen der theoretischen Analyse von enthalpischen Einzelbeiträgen zur Bindungsaffinität, die auf der direkten Wechselwirkung von Protein und Zucker beruhen, wurden durch eine detaillierte Interaktionsanalyse auf Residue-Basis die wesentlichen Wechselwirkungen zwischen Hevein und Chitobiose in Übereinstimmung mit den experimentellen Ergebnissen gefunden und nach ihrer Wichtigkeit eingeordnet. Weiterhin ist es gelungen, basierend auf Interaktionsanalysen, sinnvolle Strukturvorschläge für den Hevein-Chitotriose-Komplex herauszuarbeiten und im Vergleich zum Hevein-Chitobiose-Komplex zusätzliche günstige Wechselwirkungen mit dem Protein zu lokalisieren. Es ist daher wahrscheinlich, daß diese zusätzlichen Wechselwirkungen zur Erhöhung der Bindungsaffinität beim Übergang von Chitobiose zu Chitotriose signifikant beitragen. Der Einfluß des Lösungsmittels auf die Wechselwirkungsenergie wurde ansatzweise untersucht.

Eine umfangreichere Analyse von Solvenseffekten war bis vor kurzem aufgrund der Größe des zu simulierenden Systems und wegen der zu analysierenden Datenmenge technisch nicht durchführbar. Mittlerweile sind jedoch durch den Einsatz von Parallelrechnern, wie auch vom sonstigen Hardwareumfeld, die technischen Voraussetzungen vorhanden um Simulationen dieser Größenordnung in akzeptabler Zeit durchzuführen. Das technische Problem der Datenanalyse wurde im Rahmen dieser Arbeit durch Entwicklung der Analyse-Software CAT2000 gelöst.

Die Durchführung von Interaktionsanalysen auf Residue-Basis ist mit CAT2000 routinemäßig möglich. Die Arbeitszeit, die zur Durchführung der Analyse notwendig ist, konnte wesentlich verkürzt werden (vorher: ca. eine Woche; jetzt: eine Stunde). Damit sind die Voraussetzungen zum Einsatz der Methode im Rahmen eines rationalen Drug Designs geschaffen.

---

Der Einfluß des Lösungsmittels auf die Bindungsaffinität wurde schließlich durch NMR-Methoden untersucht. Dazu wurde die Bindungskonstante des Hevein-Chitotriose-Komplexes in DMSO/Wasser-Mischungen durch NMR-Titrationsen bestimmt. Durch eine van't Hoff Analyse wurde der enthalpische und entropische Anteil der Freien Enthalpie bestimmt.

Bereits in einer 1:4 Mischung von DMSO und Wasser nimmt die Bindungskonstante um eine Größenordnung ab. Eine weitere Erhöhung der DMSO-Konzentration auf 50% bewirkt nur noch eine geringe zusätzliche Erniedrigung der Bindungsaffinität. Dies legt den Schluß nahe, daß eine intakte Wasserstruktur eine wichtige Voraussetzung für eine hohe Bindungsaffinität ist.

Durch die van't Hoff-Analyse konnte gezeigt werden, daß die Komplexbildung zwischen Hevein und Chitotriose in 20%iger DMSO-Lösung, wie bei den meisten Kohlenhydrat-Protein-Komplexen, enthalpisch begünstigt ist.

## 7 Anhang

### 7.1 Gruppierungs- und Clustering-Methoden von CAT2000

#### *Template-Grouping-Methode*

Bei dieser Methode wird eine virtuelle  $\Phi/\Psi$ -Konformation („Template“) und eine maximale RMS-Toleranz ( $r_{\max}$ ) vorgegeben. Alle Frames, deren  $\Phi/\Psi$ -Kombinationen um maximal  $r_{\max}$  von der  $\Phi/\Psi$ -Kombination des Templates abweichen, werden derselben Gruppe wie das Template zugeordnet. Kann ein Frame aufgrund des RMS-Kriteriums mehreren Templates zugeordnet werden, so entscheidet der kleinere RMS-Wert. Die „Template-Grouping“-Methode hat den Vorteil, daß sie Gruppierungen in Koordinatensystemen beliebiger Dimensionalität erlaubt und daß sie im Gegensatz zu automatischen Cluster- oder Gruppierungsmethoden unabhängig von den vorhandenen Daten eine reproduzierbare Gruppierung sicherstellt. Sie ist damit für mechanistische Untersuchungen die „Gruppierungsmethode der Wahl“.

#### *Hand-Over-Hand (HOH)*

Der Algorithmus arbeitet analog der Template-Grouping Methode mit einer vorgegebenen RMS-Toleranz  $r_{\max}$ . Der Unterschied besteht darin, daß jede neu in die Gruppe aufgenommene Struktur selbst Template für den nächsten Suchzyklus wird. Der Algorithmus „hangelt“ sich so von Punkt zu Punkt bis innerhalb des Suchradius keine neuen Frames mehr gefunden werden. Die Methode funktioniert nur, wenn die Cluster deutlich voneinander separiert sind ( $> r_{\max}$ ). Anwendung findet die Methode zum Clustern von Minimumstrukturen.

#### *Variable-RMS-Clustering (VRMS) und Fixed-RMS-Clustering (FRMS)*

Diese Methode arbeitet nach folgendem Prinzip: Um das Häufigkeitsmaximum zu finden, werden ausgehend von einem  $\Phi/\Psi$ -Wertepaar („Zentrum“), alle Strukturen, deren  $\Phi/\Psi$ -Werte sich maximal um einen RMS  $r_{\text{init}}$  davon unterscheiden, derselben temporären Gruppe zugeordnet. Für alle Frames dieser temporären Gruppe wird der Mittelwert der  $\Phi/\Psi$ -Werte berechnet. Unterscheidet sich der Mittelwert um mehr als  $0.1 r_{\text{init}}$  vom ursprünglichen  $\Phi/\Psi$ -Zentrum, so wird der berechnete Mittelwert das neue „virtuelle“  $\Phi/\Psi$ -Zentrum und ausgehend davon wird eine neue temporäre Gruppe definiert. Fällt das  $\Phi/\Psi$ -Zentrum schließlich mit dem  $\Phi/\Psi$ -Mittelwert zusammen, ist das Populationsmaximum gefunden und der RMS-Radius wird solange vergrößert, wie das Zentrum stabil bleibt. Der RMS-Radius darf auf maximal den doppelten Wert von  $r_{\text{init}}$  anwachsen. Sind Zentrum und optimaler RMS schließlich ermittelt, so wird die erste finale Gruppe definiert. Anschließend versucht der Algorithmus für die restlichen Frames weitere Gruppen zu finden. Entscheidend für den Erfolg des Clusterings ist bei diesem Algorithmus die richtige Wahl von  $r_{\text{init}}$ . Die einfachere Variante ist die Fixed-RMS-Clustering (FRMS) Methode. Bei dieser wird  $r_{\text{init}}$  nicht optimiert.

#### *Populationsgrid-basiertes-Fixed-RMS-Clustering (PFRMS)*

Hierbei werden die Zentren maximaler lokaler Population (lokale Minima) im n-dimensionalen Hyperraum basierend auf einem n-dimensionalen Histogramm gefunden. Die zu einer Gruppe gehörenden Frames werden ausgehend von den lokalen Minima durch einen (Toleranz-) RMS-Radius analog wie bei FRMS beschrieben zugeordnet.

Eine detailliertere Beschreibung der in CAT2000 implementierten Funktionen findet sich im CAT Referenzhandbuch (Informationen sind erhältlich unter [www.martin-frank.de](http://www.martin-frank.de)).

## 7.2 Experimentelle Bestimmung des Rotamerengleichgewichts

Zur Analyse des Rotamerengleichgewichts stehen im wesentlichen stehen zwei experimentelle Methoden zur Auswahl: Messung der optischen Rotation und NMR-Messungen [212].

Die Methode der Wahl ist die Messung der vicinalen Kopplungskonstanten  $J_{H5,H6R}$  und  $J_{H5,H6S}$ . Problematisch ist dabei oftmals die Zuordnung der Signale für H6R und H5R. Das Rotamerengleichgewicht läßt sich empirisch aus einer Karplus-analogen Beziehung zwischen Kopplungskonstante und Torsionswinkel  $H6x-C6-C5-H5$  ( $x=R,S$ ) berechnen (Gleichung 46 bis 51) [212].

$$1.3 P_{gg} + 2.7 P_{gt} + 11.7 P_{tg} = J_{H5,H6S} \quad (46)$$

$$1.3 P_{gg} + 11.5 P_{gt} + 5.8 P_{tg} = J_{H5,H6R} \quad (47)$$

$$P_{gg} + P_{gt} + P_{tg} = 1 \quad (48)$$

umgeformt ergibt sich:

$$P_{gg} = -0.0902 J_{H5,H6R} - 0.0571 J_{H5,H6S} + 1.1915 \quad (49)$$

$$P_{gt} = 0.1042 J_{H5,H6R} - 0.0451 J_{H5,H6S} - 0.0768 \quad (50)$$

$$P_{tg} = -0.0140 J_{H5,H6R} + 0.1022 J_{H5,H6S} - 0.1146 \quad (51)$$

## 7.3 Tabellen

**Tabelle 13: Berechnung der Destabilisierungsenergie durch 1,2-gauche und 1,3-diaxiale Wechselwirkung nach Ref. [2] S.24 (Werte in kJ/mol)**

	Glc (gg)	Glc (gt)	Glc (tg)	Gal (gg)	Gal (gt)	Gal (tg)
1,2-gauche (O6-O5)	+1.89	+1.89		+1.89	+1.89	
1,2-gauche (O6-C4)	+1.89		+1.89	+1.89		+1.89
1,3-syn (O6-O4)			+6.3	+6.3		
1,3-syn (O6-H4)	+1.89					
1,3-syn (H6-O4)	+1.89	+1.89			+1.89	+1.89
Energie [kJ/mol]	7.56	3.78	8.19	10.08	3.78	3.78
$\Delta E$	3.78	0	4.41	6.3	0	0

**Tabelle 14: Rotationsbarrieren**

Molekül	[kJ/mol]	[kcal/mol]	Ref.
CH <sub>3</sub> -CH <sub>3</sub>	12.2	2.88	A
CH <sub>3</sub> -CH <sub>2</sub> CH <sub>3</sub>	14.2	3.4	A
CH <sub>3</sub> -CH(CH <sub>3</sub> ) <sub>2</sub>	16.3	3.9	A
CH <sub>3</sub> -C(CH <sub>3</sub> ) <sub>3</sub>	19.6	4.7	A
CH <sub>3</sub> CH <sub>2</sub> -CH <sub>2</sub> CH <sub>3</sub>	19.2/14.2*	4.6 /3.4*	A
CH <sub>3</sub> -OH	4.47	1.07	A
CH <sub>3</sub> -CH <sub>2</sub> OH	13.9	3.3	B
CH <sub>3</sub> -OCH <sub>3</sub>	11.3	2.7	A

Entnommen aus: A [208], S.107; B [203], S. 107

\* Bei Butan gibt es zwei verschiedene Maxima

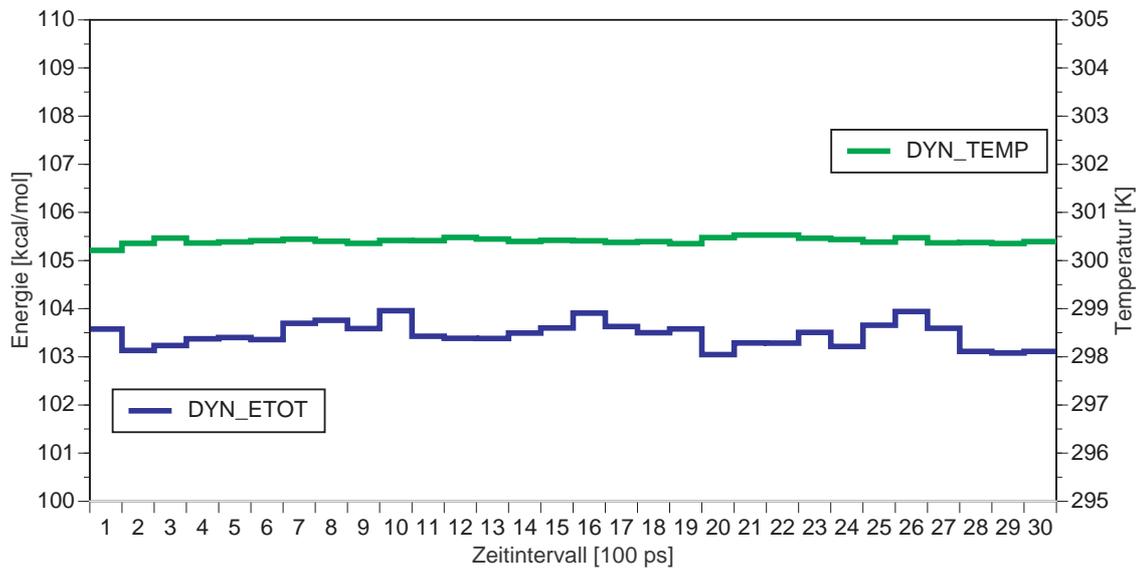
Originalliteratur: [206], [207]

Techniken: Mikrowellen-Spektroskopie, Electronic Diffraction, Ultraschallabsorption, und Infrarotspektroskopie

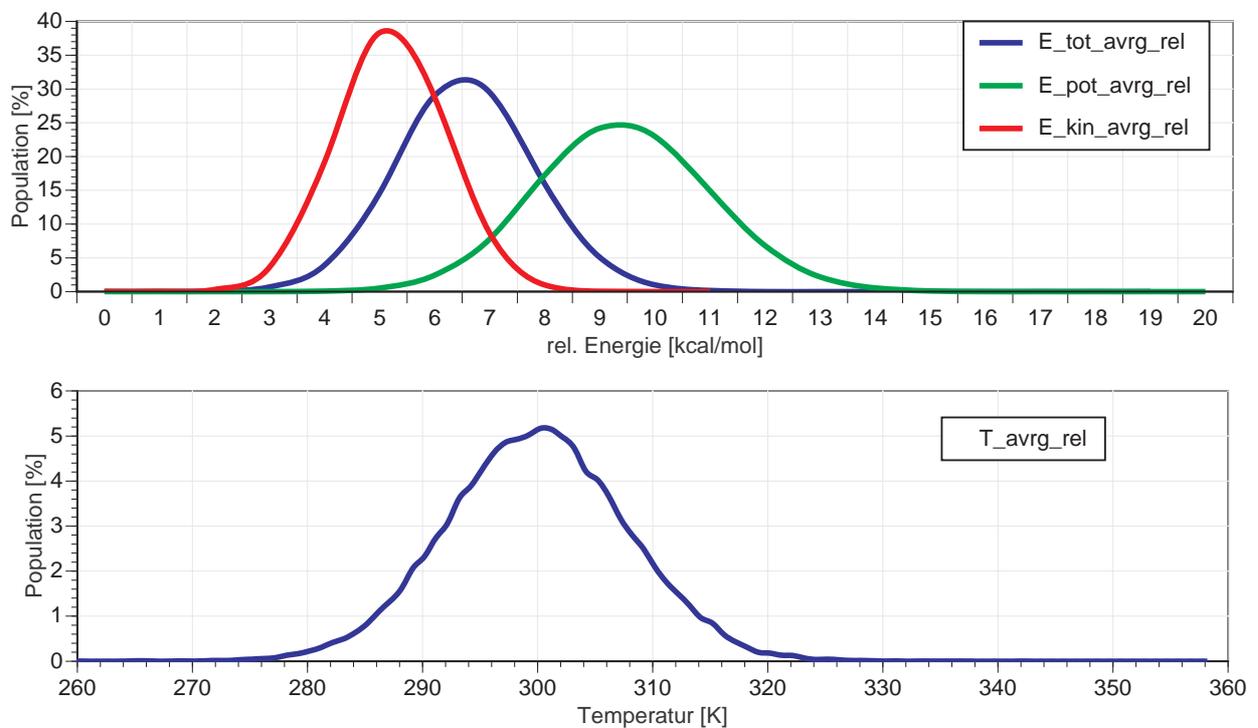
**Tabelle 15: Partialladung ausgewählter Atome**

Atom	CVFF	CFF91	AMBER	RHF (Glc) (gemittelt)	DFT (Glc) (gemittelt)
O4	-0.380	-0.537	-0.650	-0.680	-0.572
OH4	0.350	0.424	0.400	0.360	0.326
O5	-0.300	-0.227	-0.400	-0.689	-0.517
O6	-0.380	-0.537	-0.566	-0.655	-0.548
OH6	0.350	0.424	0.400	0.351	0.317

## 8 Abbildungen (X)<sup>1</sup>

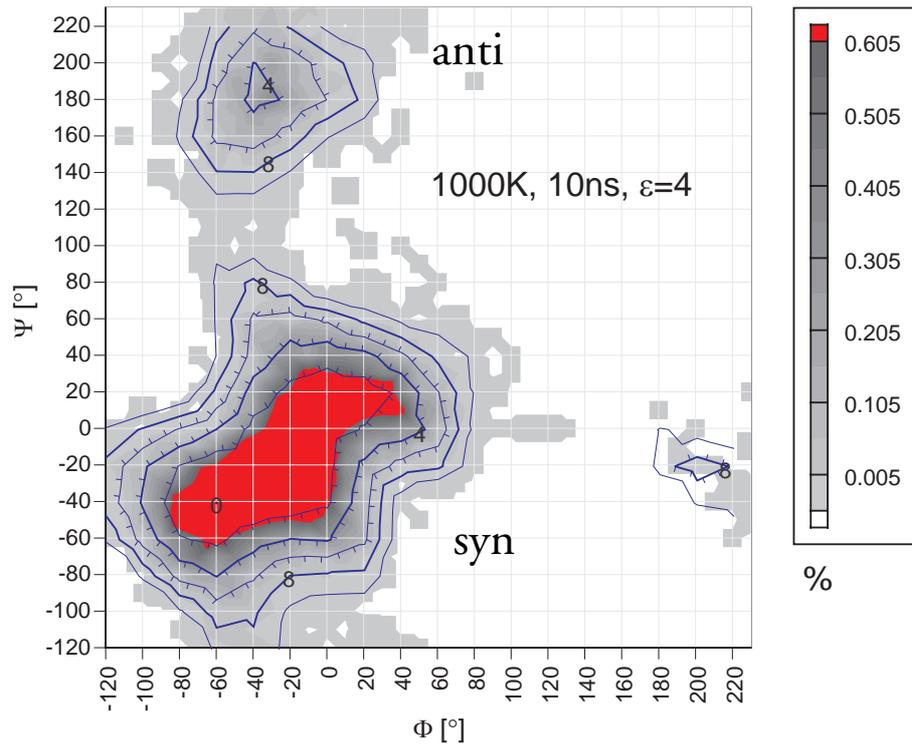


**Abb. X1:** Änderung der mittleren Energie und Temperatur in Abhängigkeit von der Simulationszeit. Die Werte wurden jeweils über eine Zeitperiode von 100 ps gemittelt. (MD-Simulation von  $\beta$ -Maltose: CVFF, 300 K,  $\varepsilon = 4$ ,  $t_s = 3$ ns,  $\Delta t_s = 50$ fs)

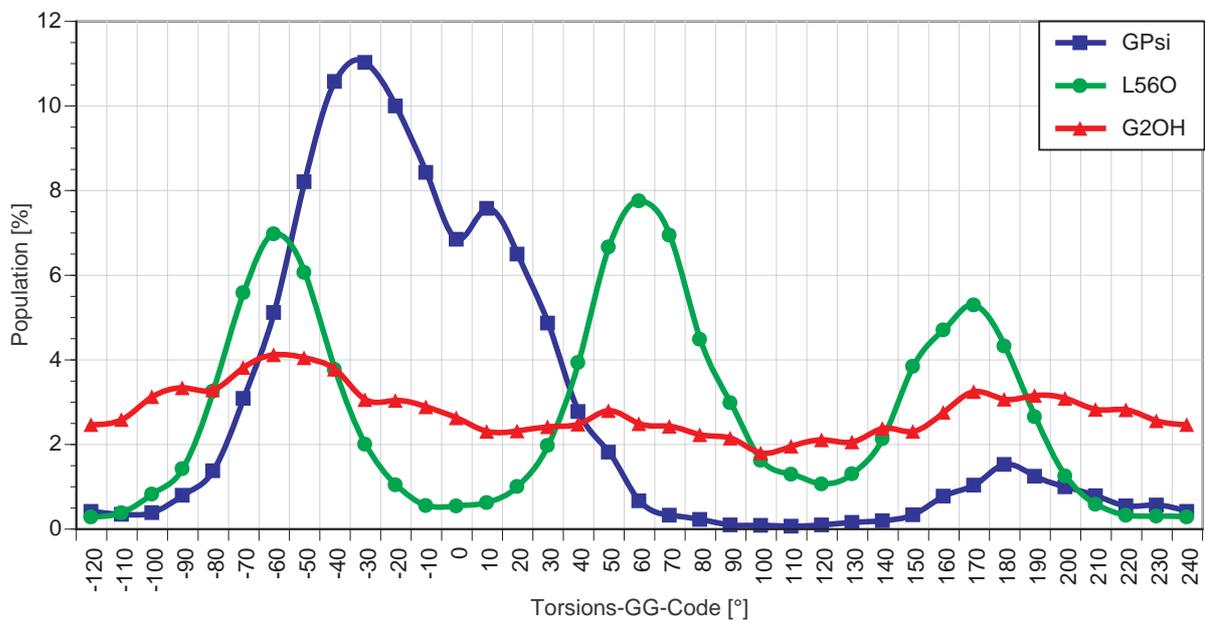


**Abb. X2:** Energie- und Temperaturverteilung (MD-Simulation von  $\beta$ -Maltose: CVFF,  $\varepsilon = 4$ , 300 K,  $t_s = 3$ ns,  $\Delta t_s = 50$ fs)

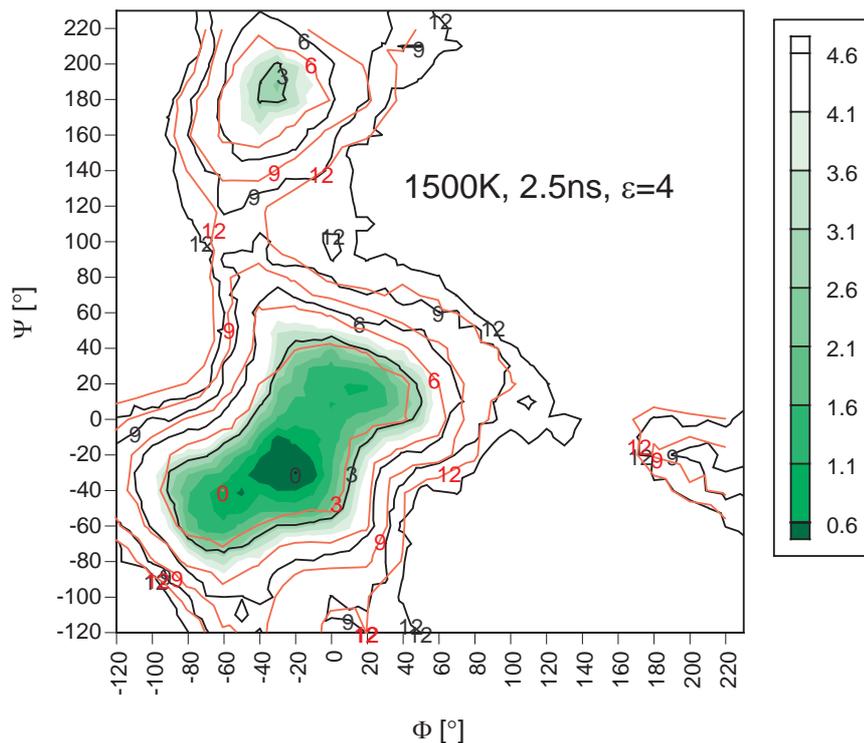
1. Die Reihenfolge der Abbildungen richtet sich im Wesentlichen nach der Reihenfolge der Verweise im Textteil.



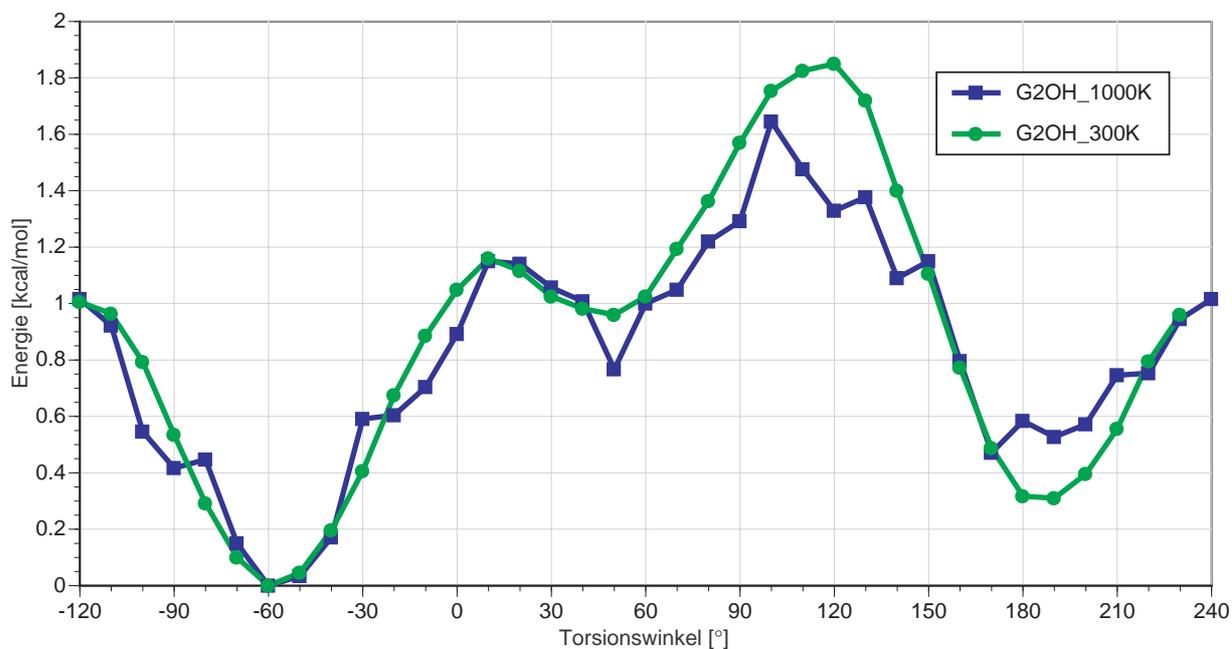
**Abb. X3:** 2D-Häufigkeitsverteilung für  $\Phi/\Psi$  bei 1000 K ( $\beta$ -Maltose, CVFF,  $\epsilon = 4$ ).  
Contour-(Linien-)Plot: Adiabatic Map (vgl. Abschnitt 3.3.1).



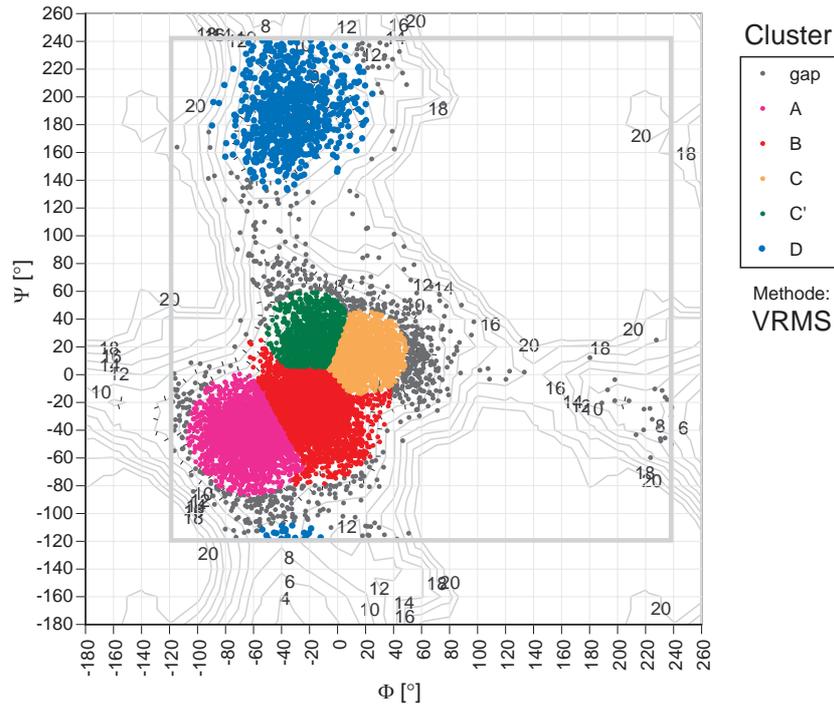
**Abb. X4:** Häufigkeitsverteilung der glycosidischen (GPhi), der Hydroxymethyl- (L56O) und der OH-Torsionswinkel bei 1000 K ( $\beta$ -Maltose, CVFF,  $\epsilon = 4$ ,  $t_s = 10$  ns,  $\Delta t_s = 1$  ps)



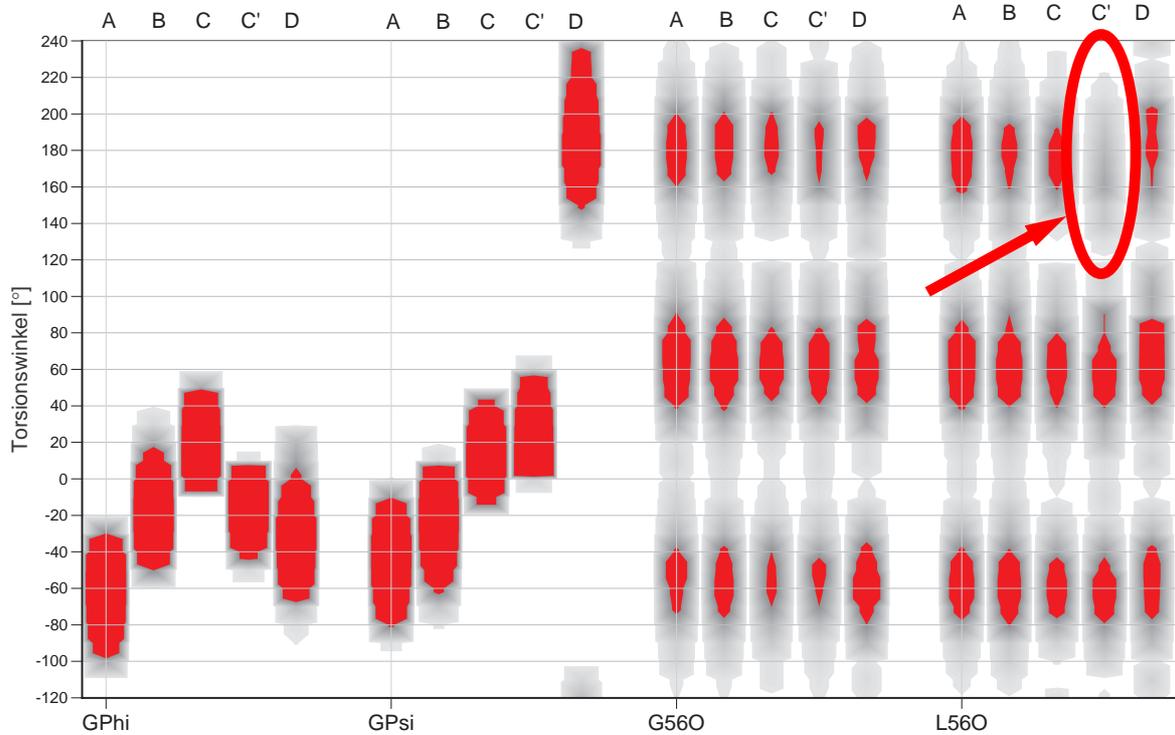
**Abb. X5:**  $\Phi / \Psi$ -Energiekarte aus MD-Daten (MD-Simulation von  $\beta$ -Maltose: CVFF,  $\Delta t_s = 50$ fs) (rote Linien: Adiabatic Map, vgl. Abschnitt 3.3.1)



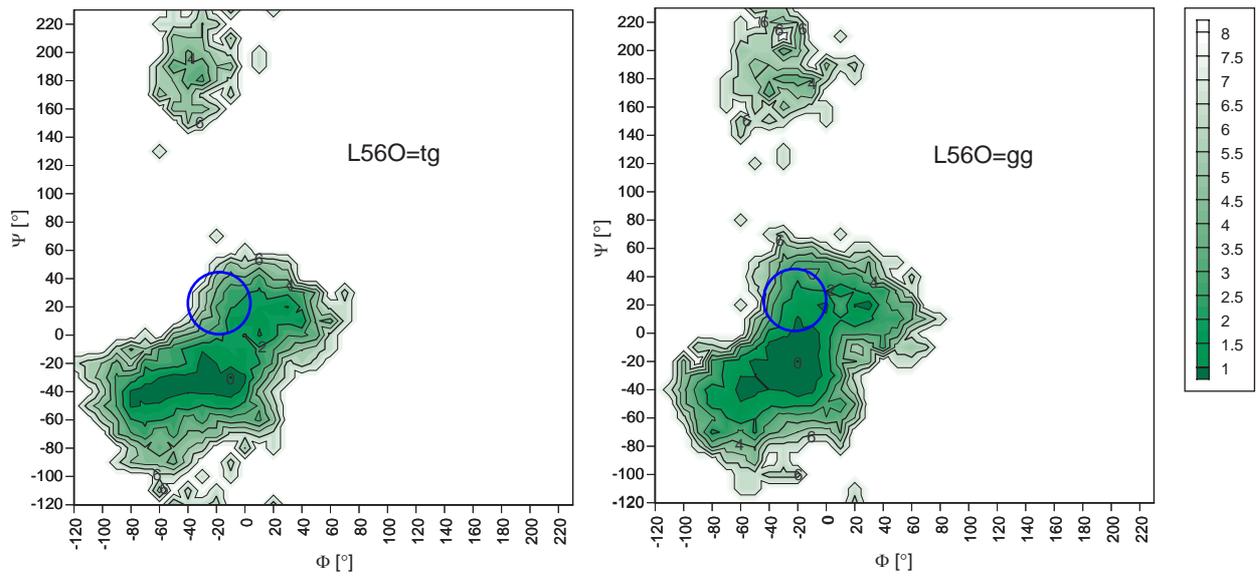
**Abb. X6:** FGE-Profil eines OH-Torsionswinkels aus einer MD-Simulation bei 1000 K und bei 300 K ( $\beta$ -Maltose, G2OH, CVFF,  $\epsilon = 4$ ,  $t_s = 10$  ns)



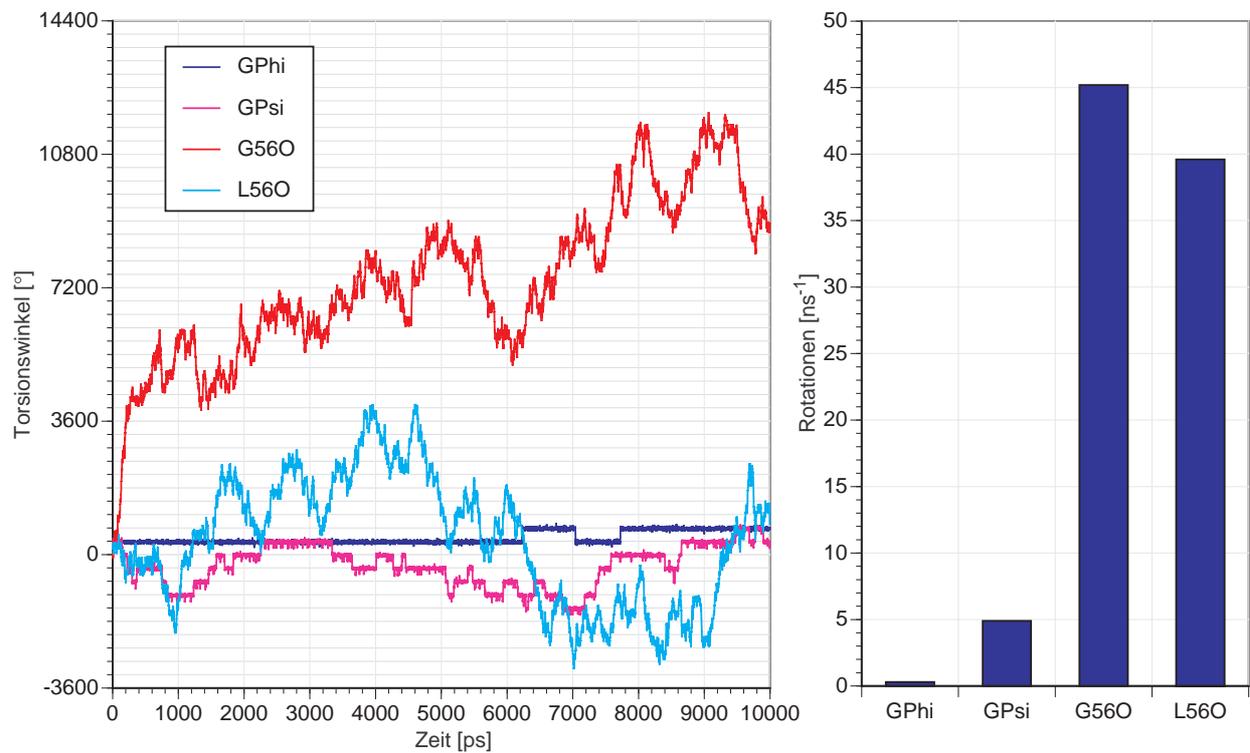
**Abb. X7:** Auto-Clustering der 1000 K-Trajektorien-Daten im  $\Phi/\Psi$ -Raum ( $\beta$ -Maltose, CVFF,  $\varepsilon = 4$ ). „Variable-RMS-Clustering“-Methode (Initial-RMS=30°). Contour-(Linien-)Plot: Adiabatic Map aus einem  $\Phi/\Psi/\omega_1/\omega_2$ -Grid-Search (vgl. Abschnitt 3.3.1).



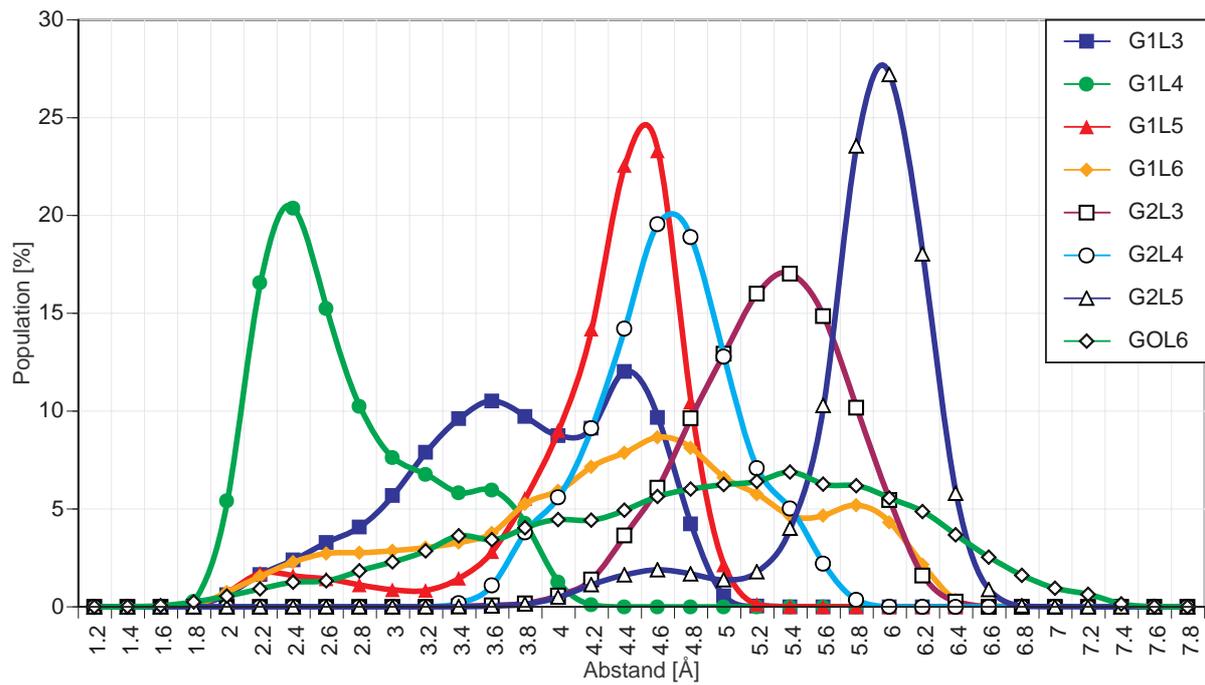
**Abb. X8:** Häufigkeitsverteilung für  $\Phi$ ,  $\Psi$  und Hydroxymethylgruppen bei 1000 K aufgeteilt nach  $\Phi/\Psi$ -Cluster (A, B, C, C', D) ( $\beta$ -Maltose, CVFF,  $\varepsilon = 4$ ). Für Cluster C' ist L56O = tg nicht stabil (siehe Pfeil)



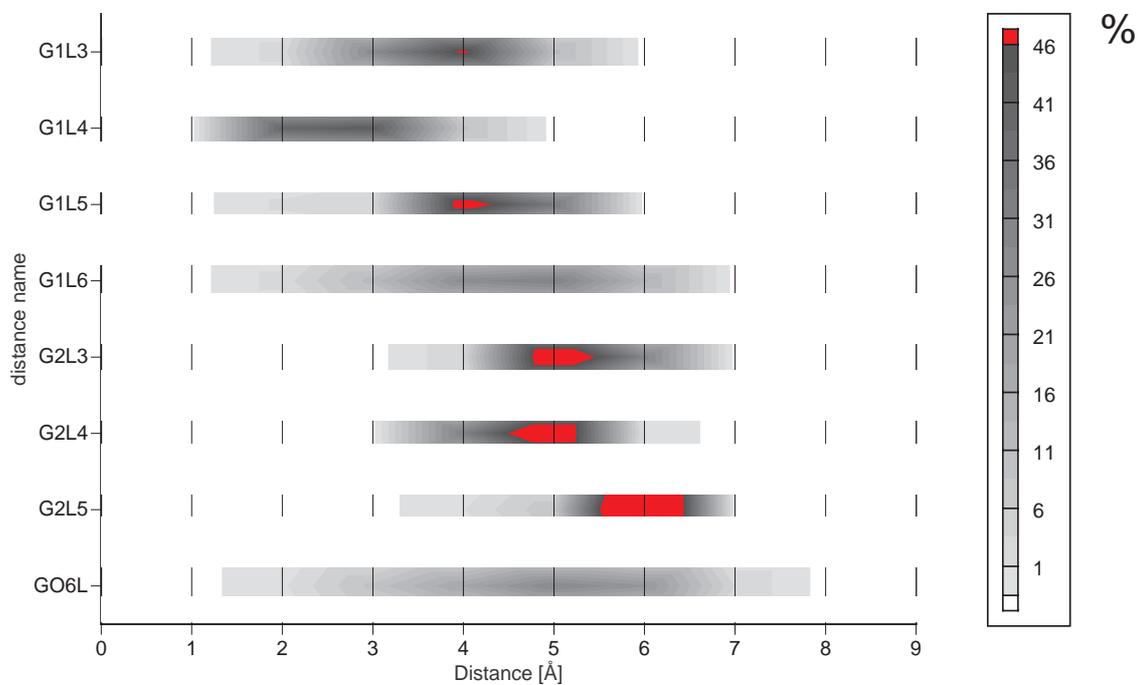
**Abb. X9:** FGE-Map für  $\Phi/\Psi$  für verschiedene Einstellungen von L56O ( $\beta$ -Maltose, CVFE,  $\varepsilon = 4$ ).



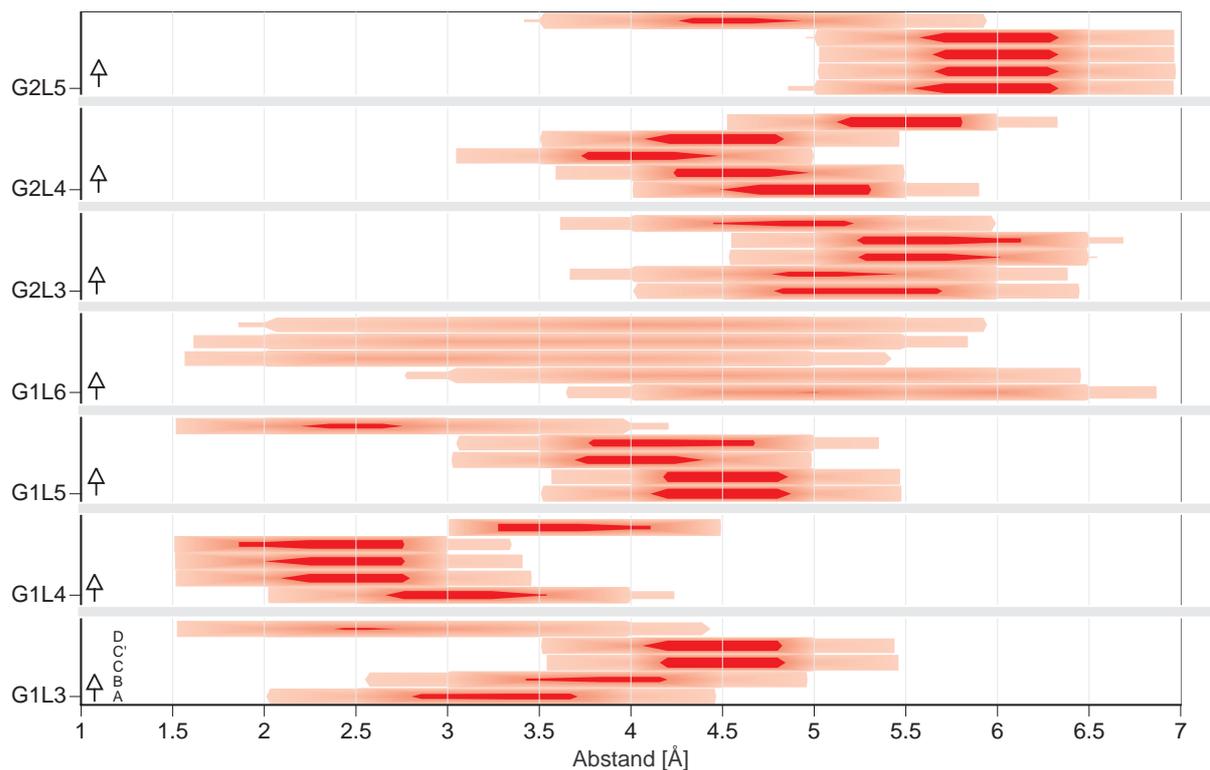
**Abb. X10:** Rotationstrajektorie für  $\Phi$ ,  $\Psi$  und Hydroxymethylgruppen bei 1000 K ( $\beta$ -Maltose, CVFE,  $\varepsilon = 4$ ,  $t_s = 10$  ns).



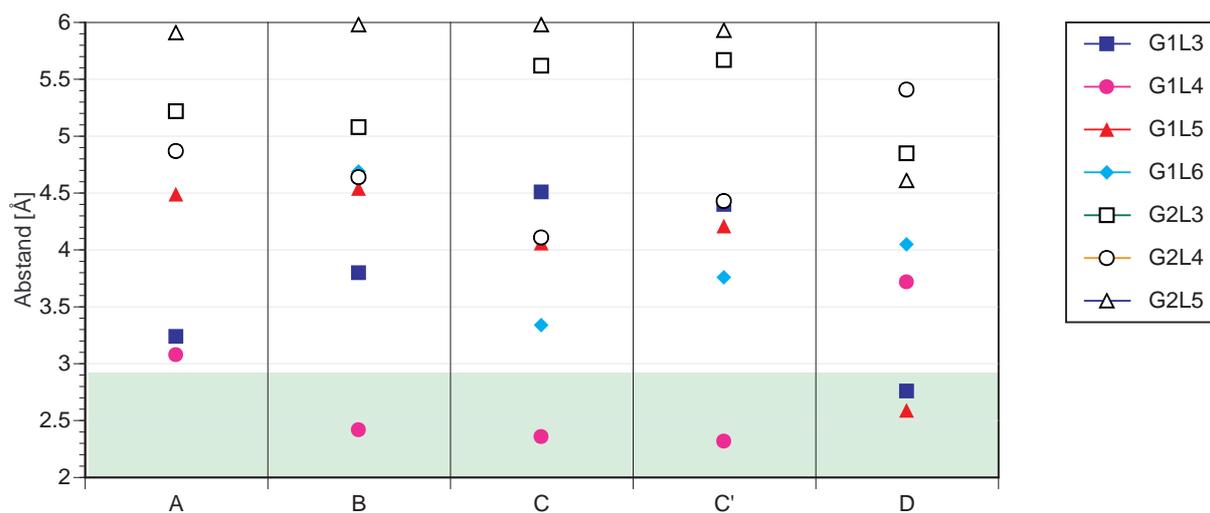
**Abb. X11:** Häufigkeitsverteilung der H-H-Abstände bei 1000 K ( $\beta$ -Maltose, CVFF,  $\varepsilon = 4$ ,  $t_s = 10$  ns).



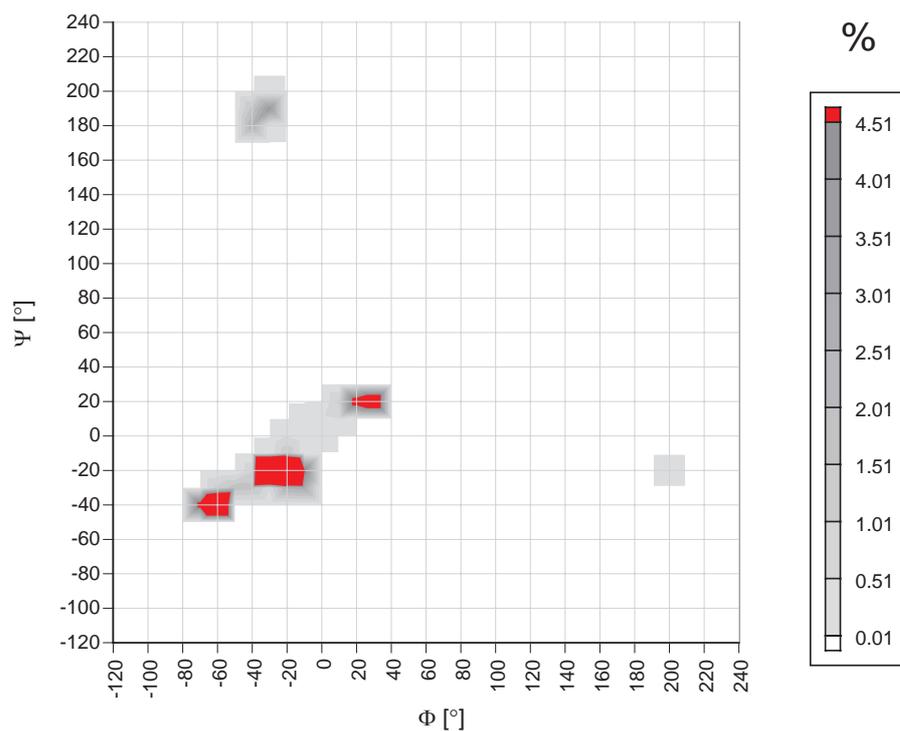
**Abb. X12:** Häufigkeitsverteilung der H-H-Abstände bei 1000 K als „Pseudo-3D“-Histogramm ( $\beta$ -Maltose, CVFF,  $\varepsilon = 4$ ,  $t_s = 10$  ns).



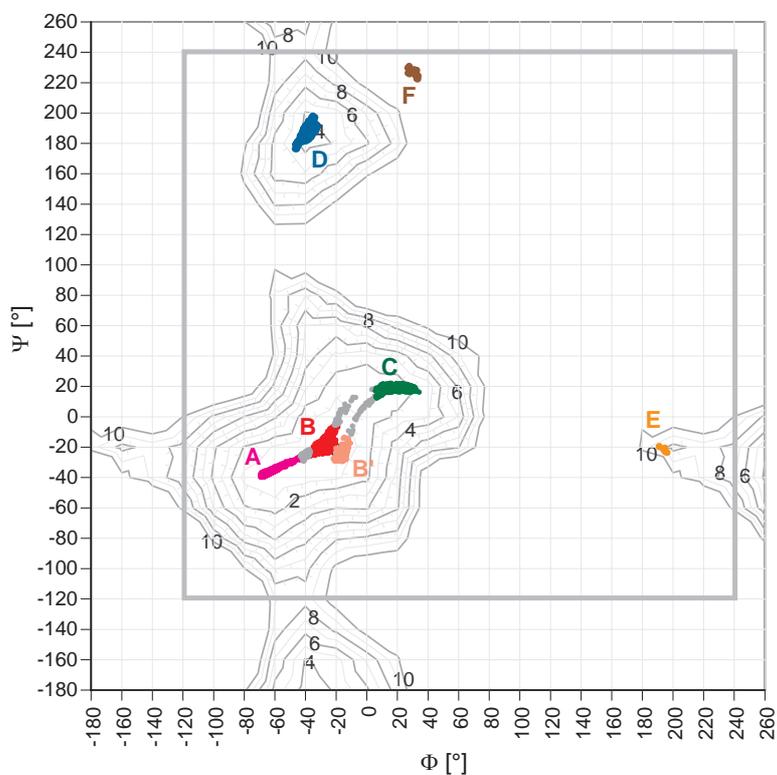
**Abb. X13:** Häufigkeitsverteilung der H-H-Abstände bei 1000 K als „Pseudo-3D“-Histogramm aufgeteilt nach  $\Phi/\Psi$ -Clustern (A, B, C, C', D) ( $\beta$ -Maltose, CVFF,  $\varepsilon = 4$ ,  $t_s = 10$  ns).



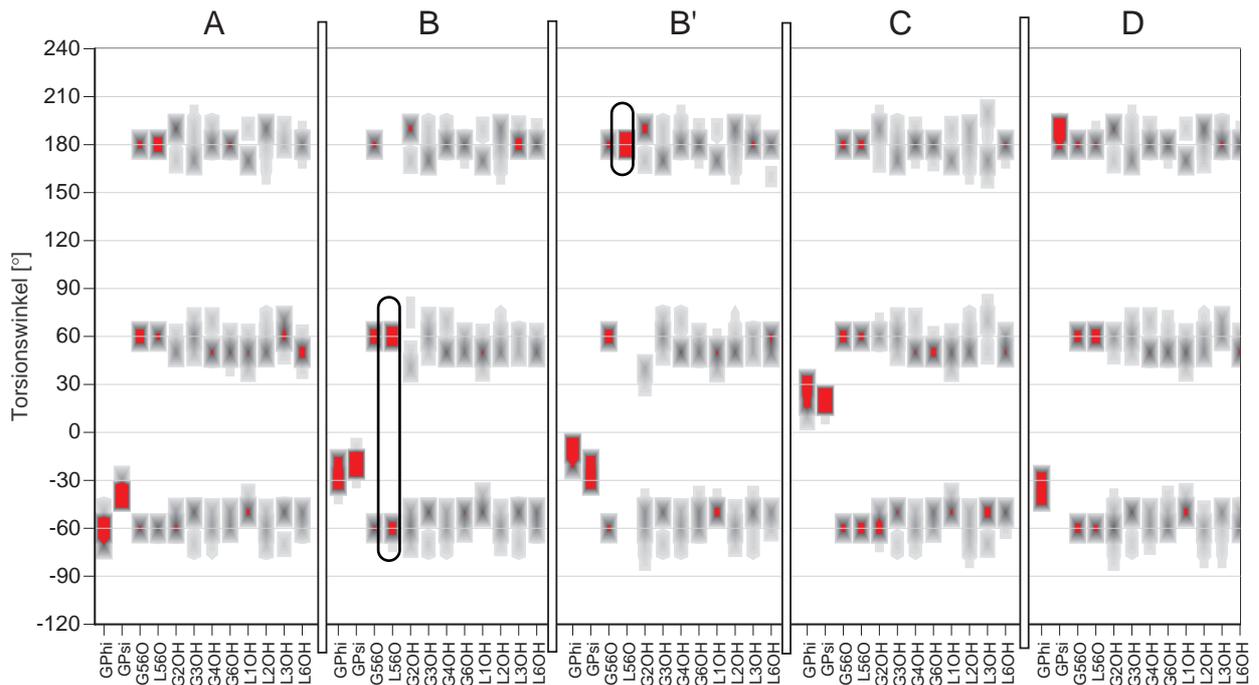
**Abb. X14:**  $\langle r^{-6} \rangle$ -gewichteter Mittelwert der H-H-Abstände bei 1000 K aufgeteilt nach  $\Phi/\Psi$ -Clustern (A, B, C, C', D) ( $\beta$ -Maltose, CVFF,  $\varepsilon = 4$ ,  $t_s = 10$  ns).



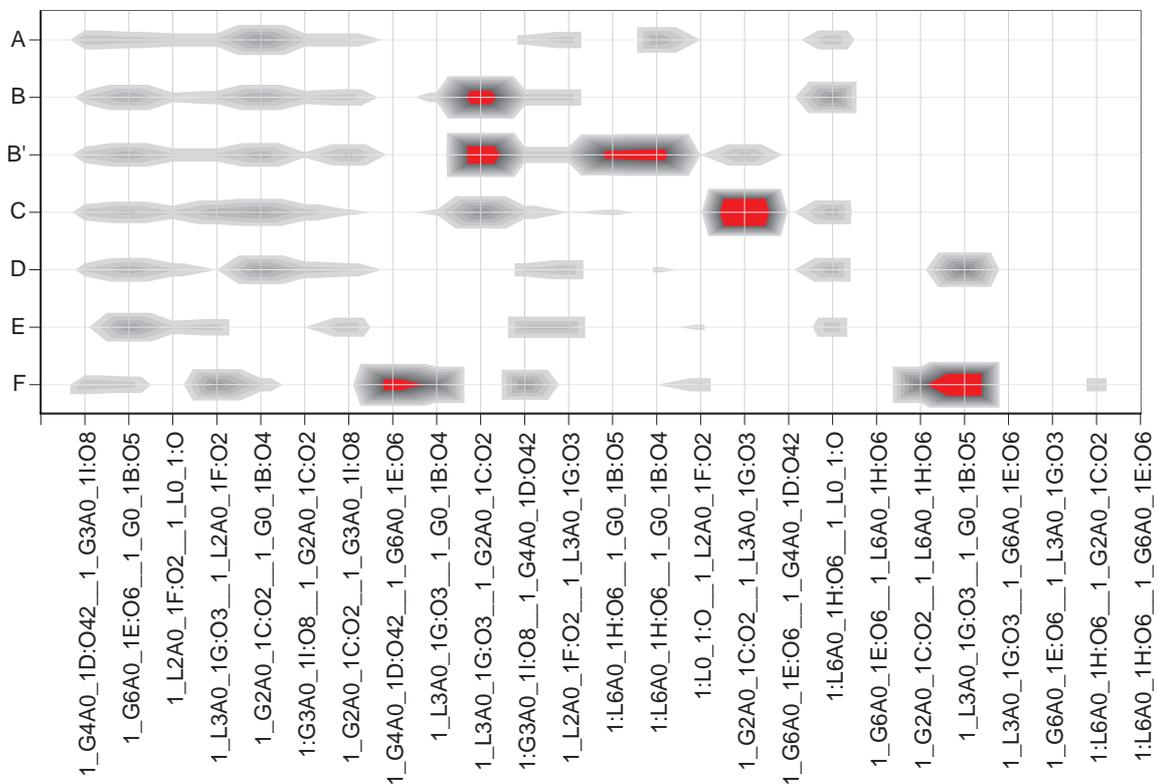
**Abb. X15:** 2D-Häufigkeitsverteilung der  $\Phi/\Psi$  -Werte der minimierten 10000 HTMD-Strukturen („Conformational Clustering“) aus einer 1000K-MD-Simulation ( $\beta$ -Maltose, CVFF,  $\varepsilon = 4$ ).



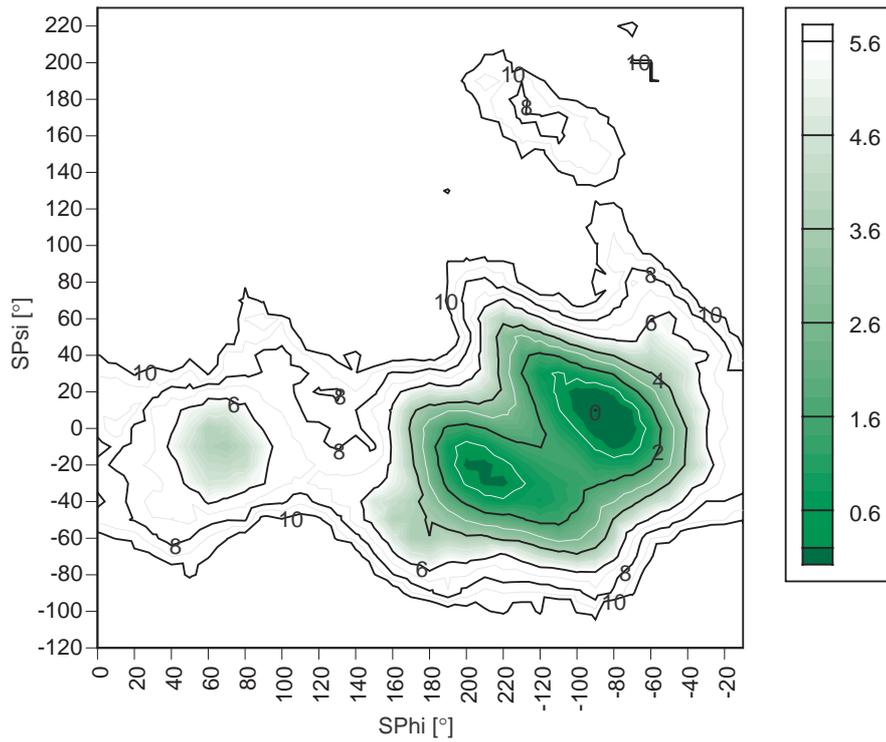
**Abb. X16:** Clustering der minimierten Strukturen der 1000K-MD-Simulation ( $\beta$ -Maltose, CVFF,  $\varepsilon = 4$ , 10000 Strukturen). Clustering-Algorithmus: HOH und FRMS (siehe Abschnitt 7.1)



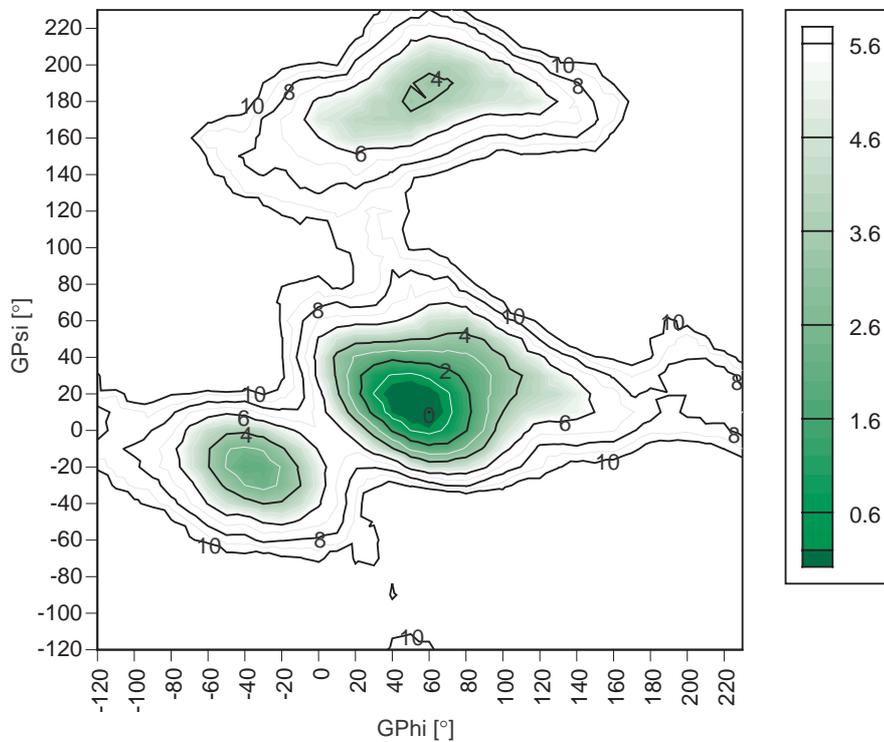
**Abb. X17:** Torsionswinkelverteilung für minimierte Strukturen der 1000K-MD-Simulation aufgeteilt nach  $\Phi/\Psi$ -Clustern (A, B, B', C, D), vgl. Abb. X16 ( $\beta$ -Maltose, CVFF,  $\varepsilon = 4$ ). Unterschied zwischen B und B': siehe Markierung



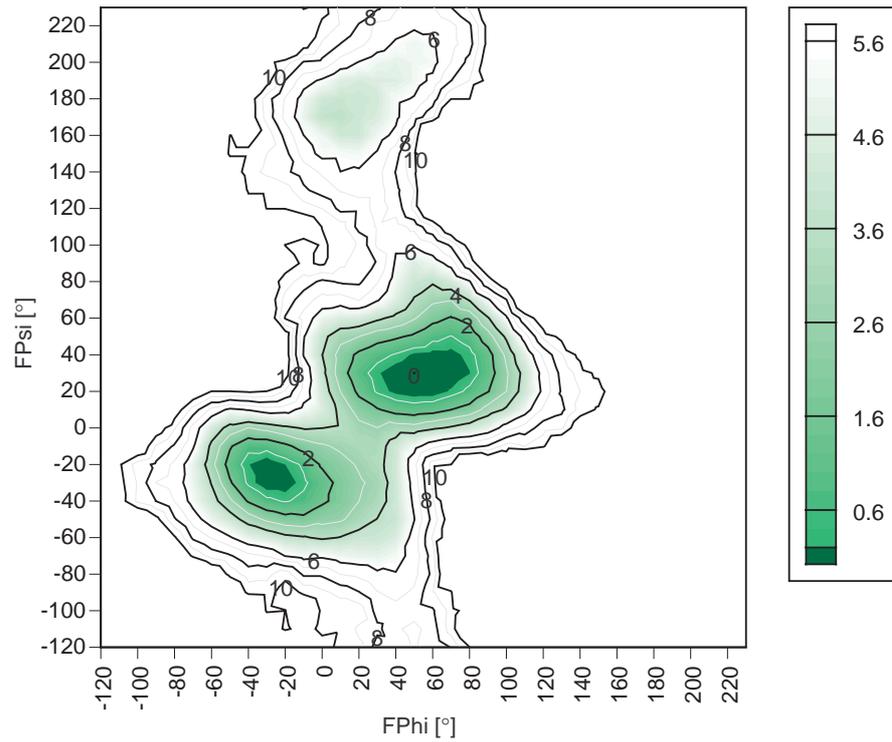
**Abb. X18:** H-Brücken pro Gruppe für minimierte Strukturen der 1000K-MD-Simulation ( $\beta$ -Maltose, CVFF,  $\varepsilon = 4$ ).



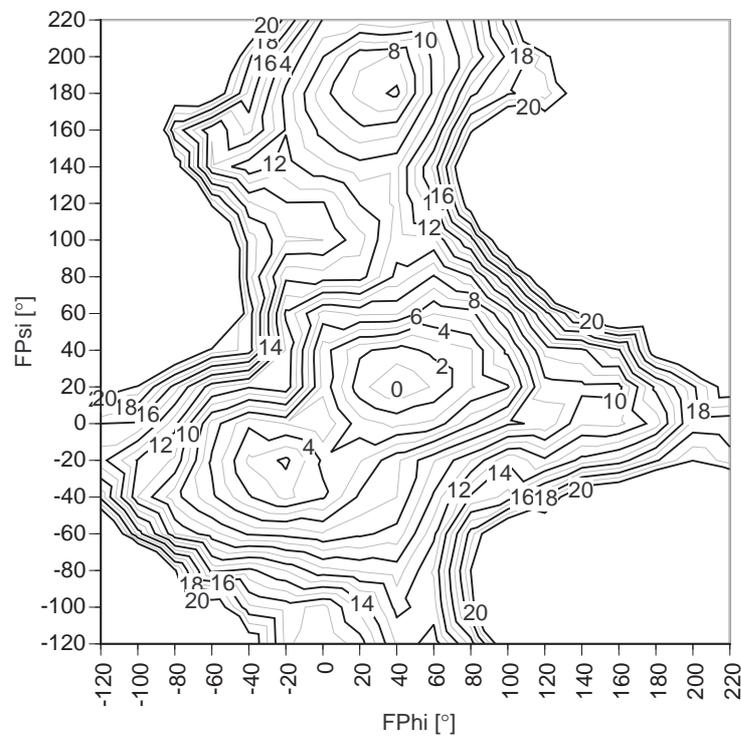
**Abb. X19:** FGE-Map für Neu5Acα(2-3)Gal (Sialyl-Lewis<sup>X</sup>, CVFF,  $\epsilon = 80$ , 1000 K,  $t_s = 100$  ns).



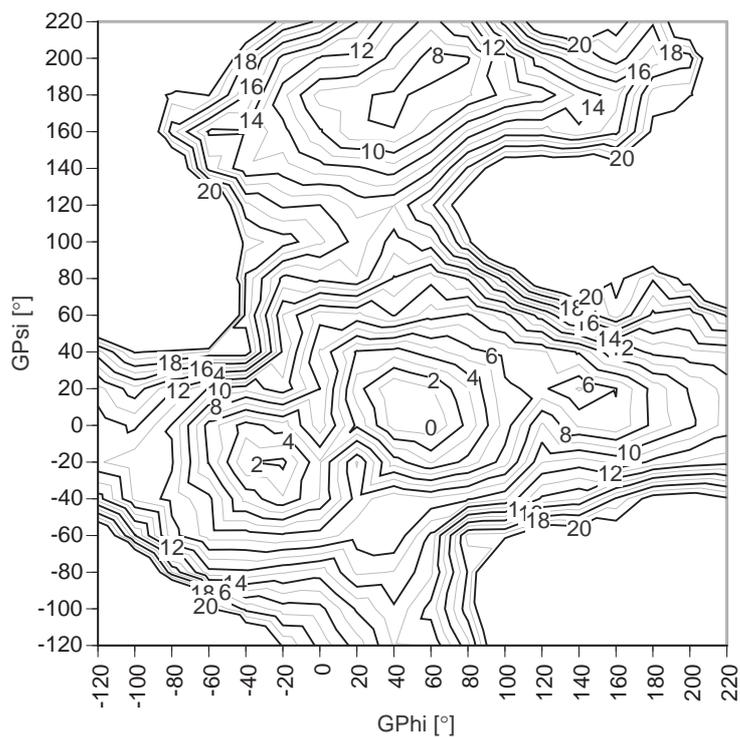
**Abb. X20:** FGE-Map für Galβ(1-4)GlcNAc (Sialyl-Lewis<sup>X</sup>, CVFF,  $\epsilon = 80$ , 1000 K,  $t_s = 100$  ns).



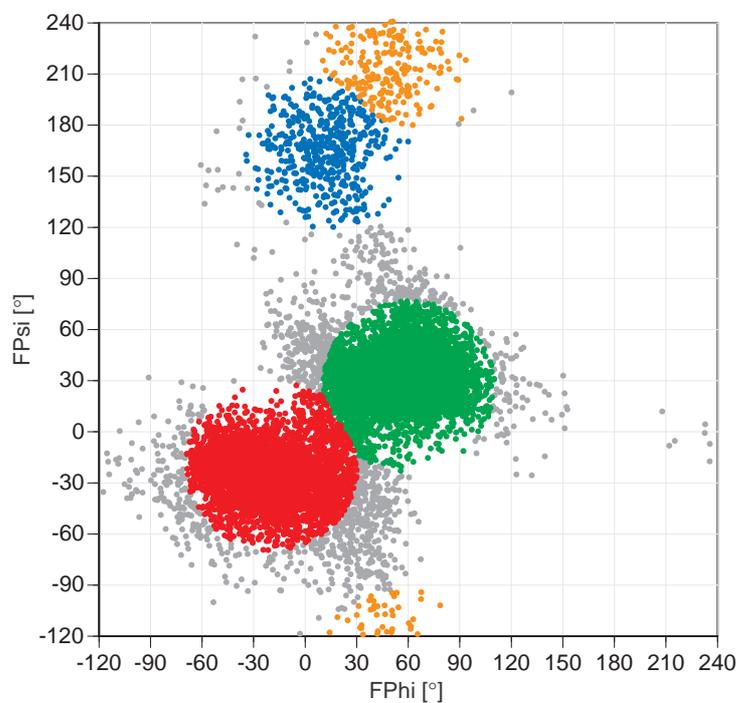
**Abb. X21:** FGE-Map für  $\text{Fuc}\alpha(1-3)\text{GlcNAc}$  (Sialyl-Lewis<sup>X</sup>, CVFF,  $\epsilon = 80$ , 1000 K,  $t_s = 100$  ns).



**Abb. X22:** Adiabatic Map für  $\text{Fuc}\alpha(1-3)\text{GlcNAc}$  (Sialyl-Lewis<sup>X</sup>, CVFF,  $\epsilon = 80$ )  
Zweistufiger Grid-Search (GPhi/GPsi; FPhi/FPsi): Beschreibung siehe Abschnitt 3.3.1



**Abb. X23:** Adiabatic Map für Gal $\beta$ (1-4)GlcNAc (Sialyl-Lewis<sup>X</sup>, CVFF,  $\epsilon = 80$ )  
Zweistufiger Grid-Search (FPhi/FPsi, GPhi/GPsi): Beschreibung siehe Abschnitt 3.3.1



**Abb. X24:** PFRMS-Clustering (RMS-Toleranz = 50°) für Fuca $\alpha$ (1-3)GlcNAc (Sialyl-Lewis<sup>X</sup>, CVFF,  
 $\epsilon = 80$ , 1000 K,  $\tau_s = 10$  ns)

Template Grouping: Neu5Ac(2-3)Gal

SPP-Code	SPhi	SPsi	RMS-Toleranz	%
1	-70	5	30	21.33
2	-160	-20	30	20.47
4	70	-10	30	2.82
5	-110	170	30	0.63

Template Grouping: Gal $\beta$ (1-4)GlcNAc

GPP-Code	GPhi	GPsi	RMS-Toleranz	%
1	50	10	30	34.58
2	-40	-20	30	11.73
3	70	190	30	6.93
4	210	20	30	1.04

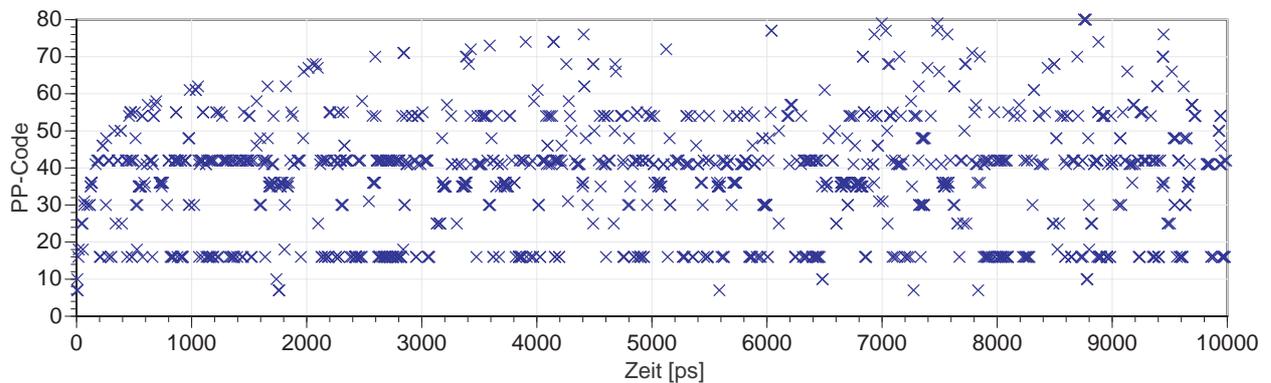
Template Grouping: Fuc $\alpha$ (1-3)GlcNAc

FPP-Code	FPhi	FPsi	RMS-Toleranz	%
1	50	30	30	28.29
2	-30	-25	30	23.78
3	35	185	30	2.71

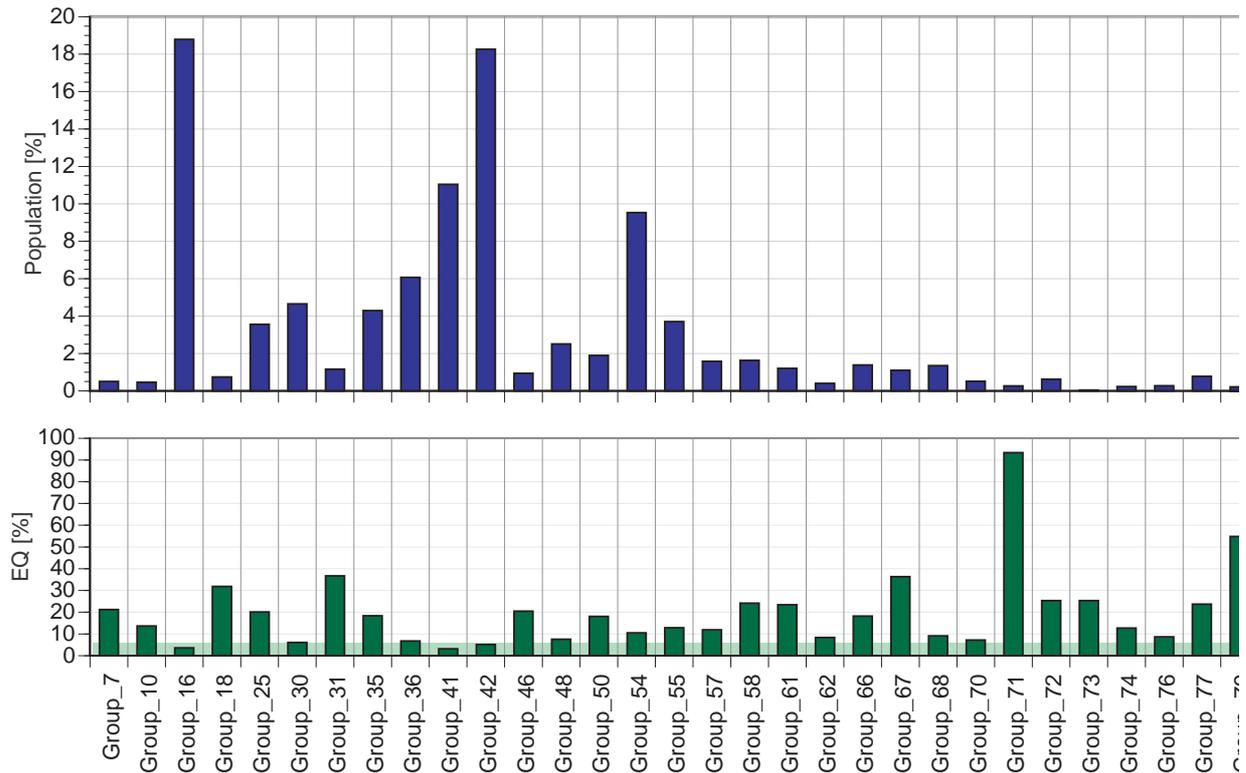
Kombination: SPP / GPP / FPP

PP-Code	7	10	16	18	25	30	31	35	36	41	42	46	48	50	54	55	57	58	61	62	66	67	68	70	71	72	73	74	76	77	79	80
SPP	4	5	1	1	1	1	2	1	2	1	2	1	2	2	2	4	4	2	1	2	1	1	2	2	4	5	2	5	1	4	4	5
GPP	2	2	1	4	3	2	3	2	2	1	1	1	2	3	1	1	1	1	2	4	3	3	3	4	2	1	2	1	4	3	3	2
FPP	2	2	1	1	2	1	3	2	2	2	1	3	1	2	2	1	2	3	3	1	1	3	1	2	1	2	3	1	2	3	1	1

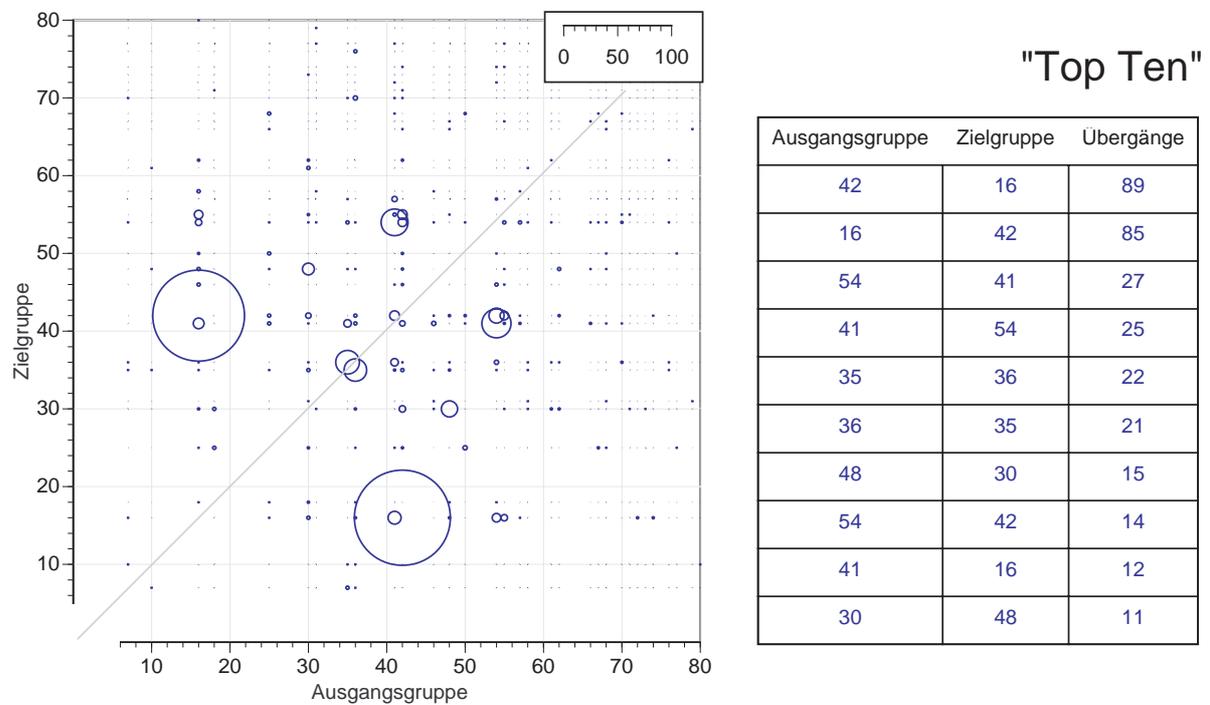
**Abb. X25:** Gruppierung im sechsdimensionalen  $\Phi/\Psi$ -Hyperraum (Sialyl-Lewis<sup>X</sup>, CVFF,  $\varepsilon = 80$ , 1000 K,  $t_s = 10$  ns). (Anm.: SPP-Gruppe 3 existiert in dieser Auswertung nicht)



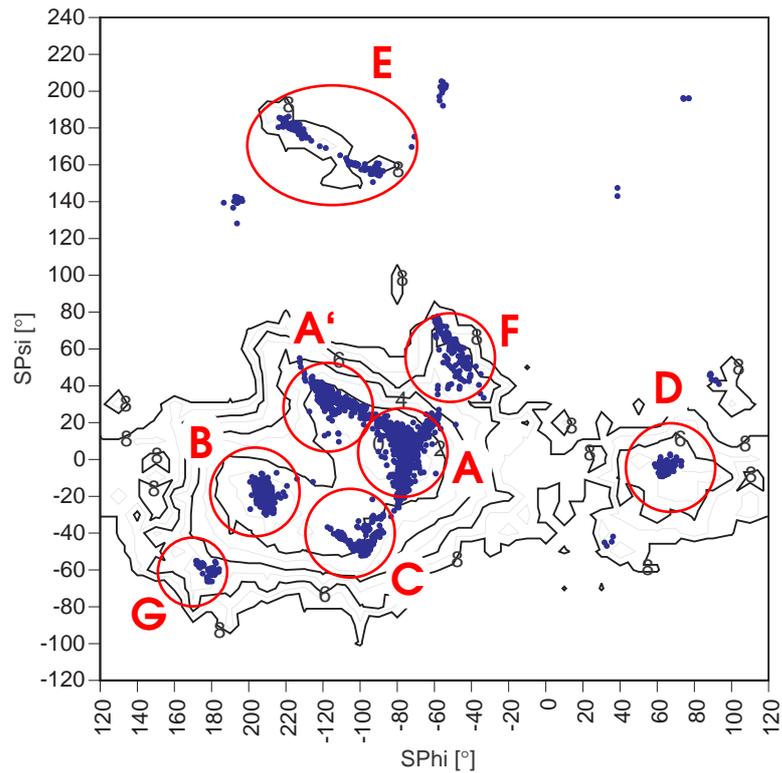
**Abb. X26:** Conformational-Search: Trajektorie der PP-Gruppen (Sialyl-Lewis<sup>X</sup>, CVFF,  $\varepsilon = 80$ , 1000 K,  $t_s = 10$  ns). Die Daten der PP-Gruppen, die in der S-, G- oder F-Dimension einen Sub-PP-Code von „Null“ enthalten, sind nicht dargestellt.



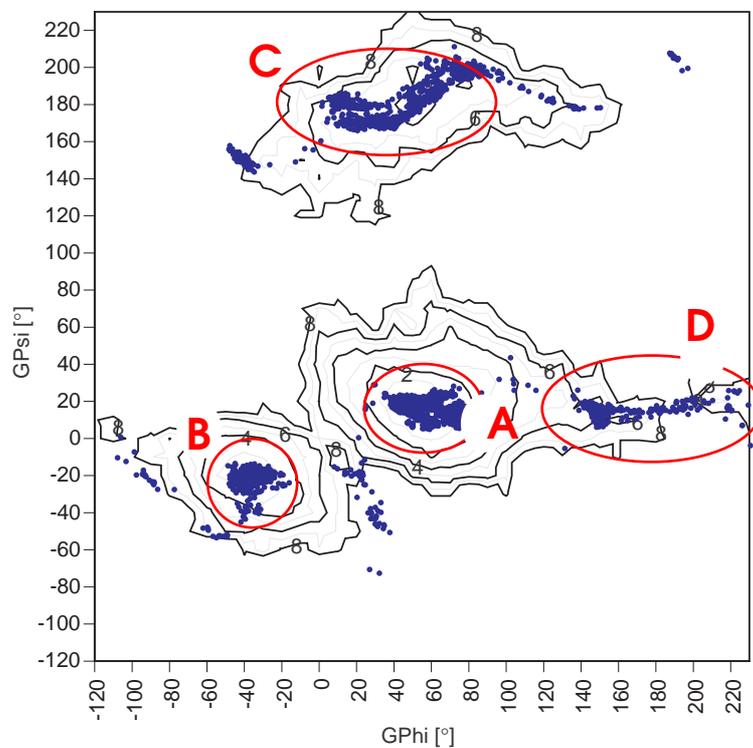
**Abb. X27:** Conformational-Search: Populations- und Gleichgewichtsanalyse der PP-Gruppen (Sialyl-Lewis<sup>X</sup>, CVFF,  $\varepsilon = 80$ , 1000 K,  $t_s = 10$  ns).



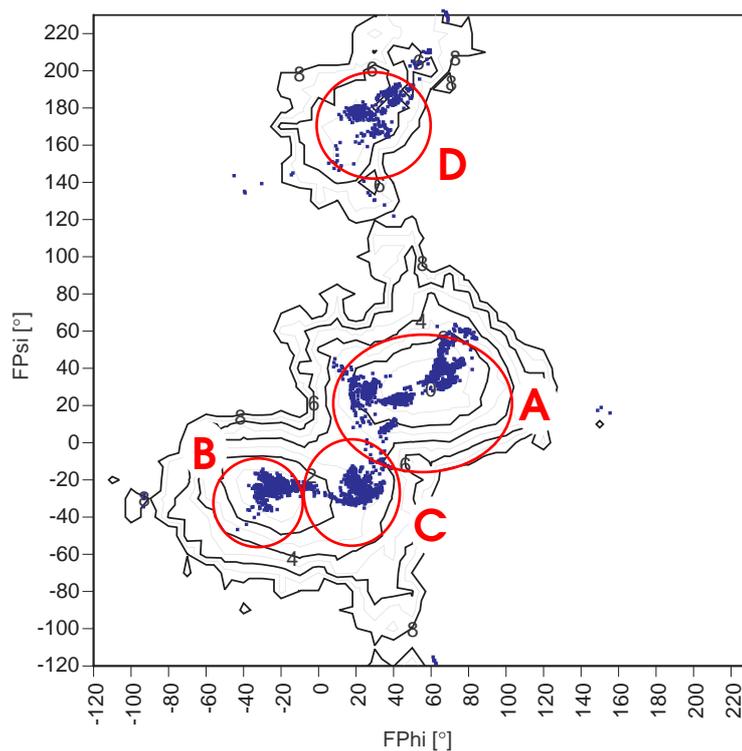
**Abb. X28:** Übergangsmatrix für PP-Gruppen-Übergänge (Sialyl-Lewis<sup>X</sup>, CVFF,  $\varepsilon = 80$ , 1000 K,  $t_s = 10$  ns).



**Abb. X29:** SPhi/SPsi-Werte der minimierten Strukturen aus einer 1000K-MD-Simulation (Sialyl-Lewis<sup>X</sup>, CVFF,  $\epsilon = 80$ ,  $t_s = 10$  ns, 10000 Strukturen)



**Abb. X30:** GPhi/GPsi-Werte der minimierten Strukturen aus einer 1000K-MD-Simulation (Sialyl-Lewis<sup>X</sup>, CVFF,  $\epsilon = 80$ ,  $t_s = 10$  ns, 10000 Strukturen).



**Abb. X31:** FPhi/FPsi-Werte der minimierten Strukturen aus einer 1000K-MD-Simulation (Sialyl-Lewis<sup>X</sup>, CVFF,  $\epsilon = 80$ ,  $t_s = 10$  ns, 10000 Strukturen)

Template Grouping: Gal $\beta$ (1-4)GlcNAc

GPP Code	GPhi	GPsi	Tolerance	Percent
1	50	10	30	36.7
2	-40	-20	30	10.41
3	70	190	30	6.48
4	210	20	30	1.27

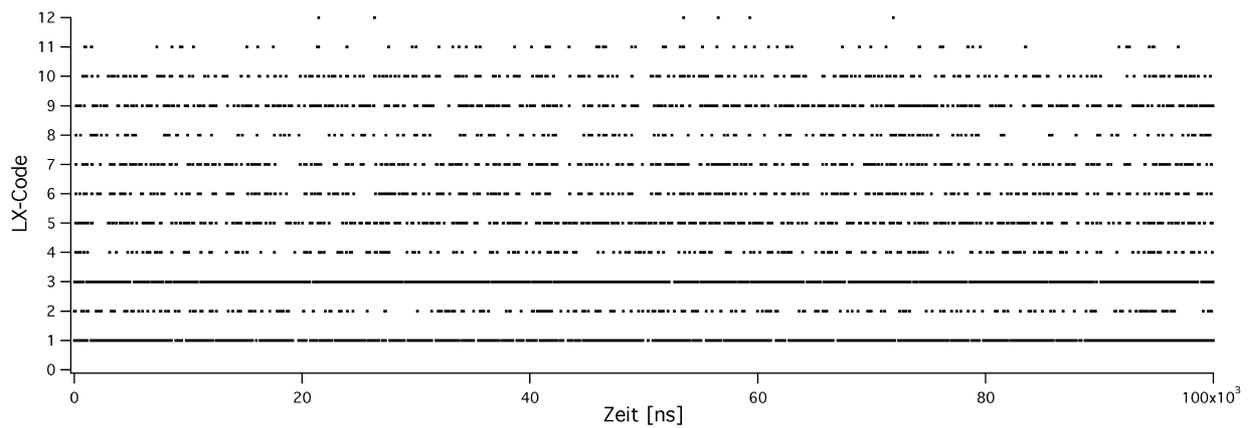
Template Grouping: Fuc $\alpha$ (1-3)GlcNAc

FPP Code	FPhi	FPsi	Tolerance	Percent
1	50	30	30	29.64
2	-30	-25	30	22.22
3	35	185	30	3.11

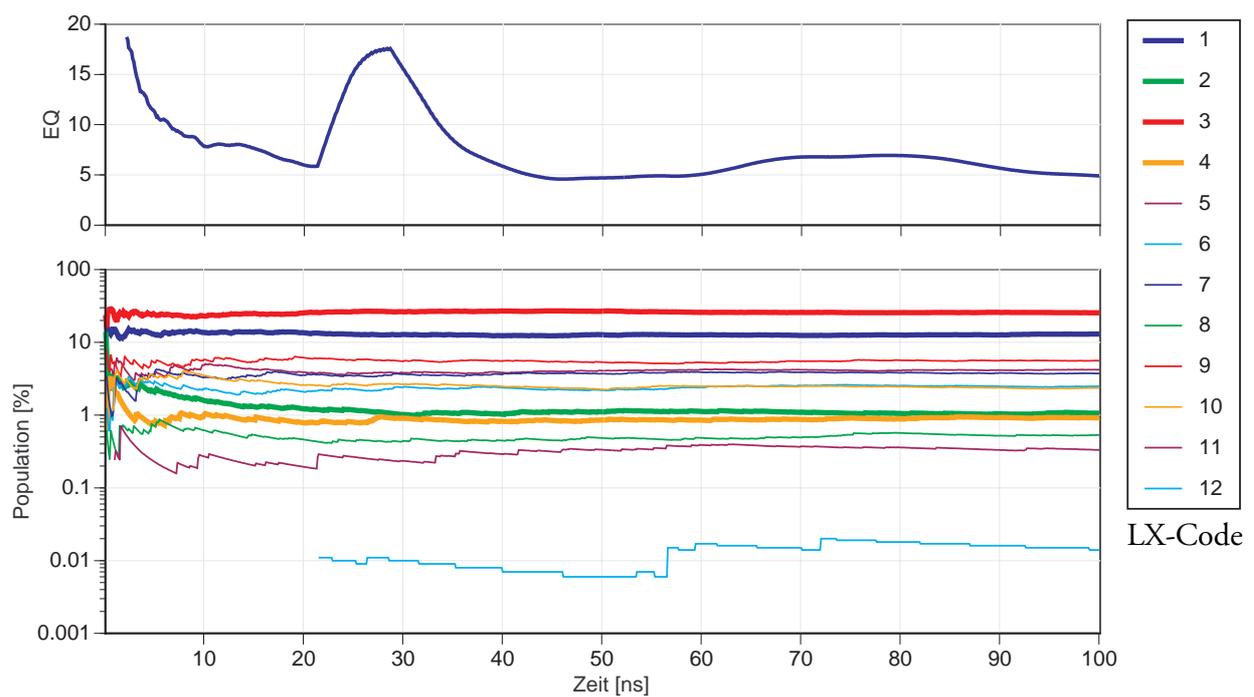
Kombination: GPP / FPP

LX Code	1	2	3	4	5	6	7	8	9	10	11	12
GPP	1	1	1	4	2	3	3	4	2	3	2	4
FPP	2	3	1	1	1	1	2	2	2	3	3	3
%	19	1	51	1	6	2	2	1	14	2	0	0
FGE	1.9	7.2	0	7.1	4.1	6.5	6	8	2.6	6.8	9.8	15.2
E <sub>pot</sub>	1.9	7.1	0.0	7.0	3.8	6.5	6.0	7.8	1.8	6.9	6.5	19.2

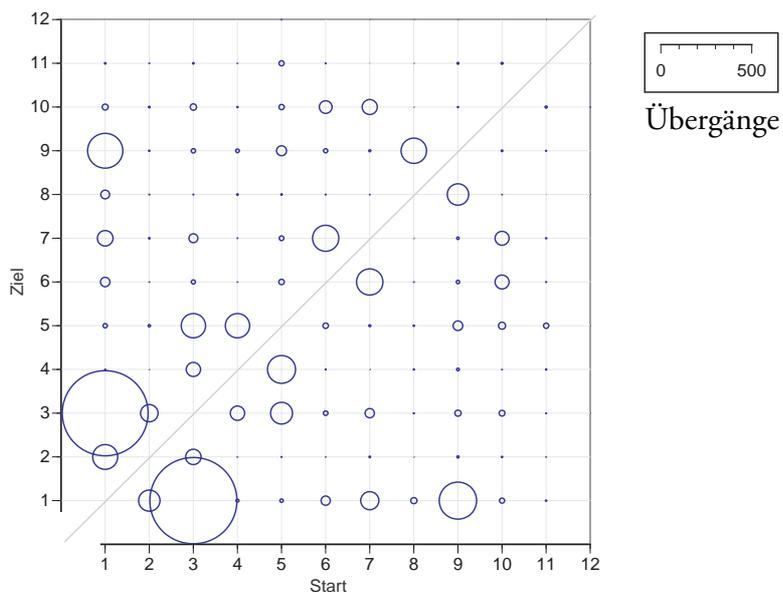
**Abb. X32:** Gruppierung für den Lewis<sup>X</sup>-Teil (Sialyl-Lewis<sup>X</sup>, CVFF,  $\epsilon = 80$ , 1000 K,  $t_s = 100$  ns).



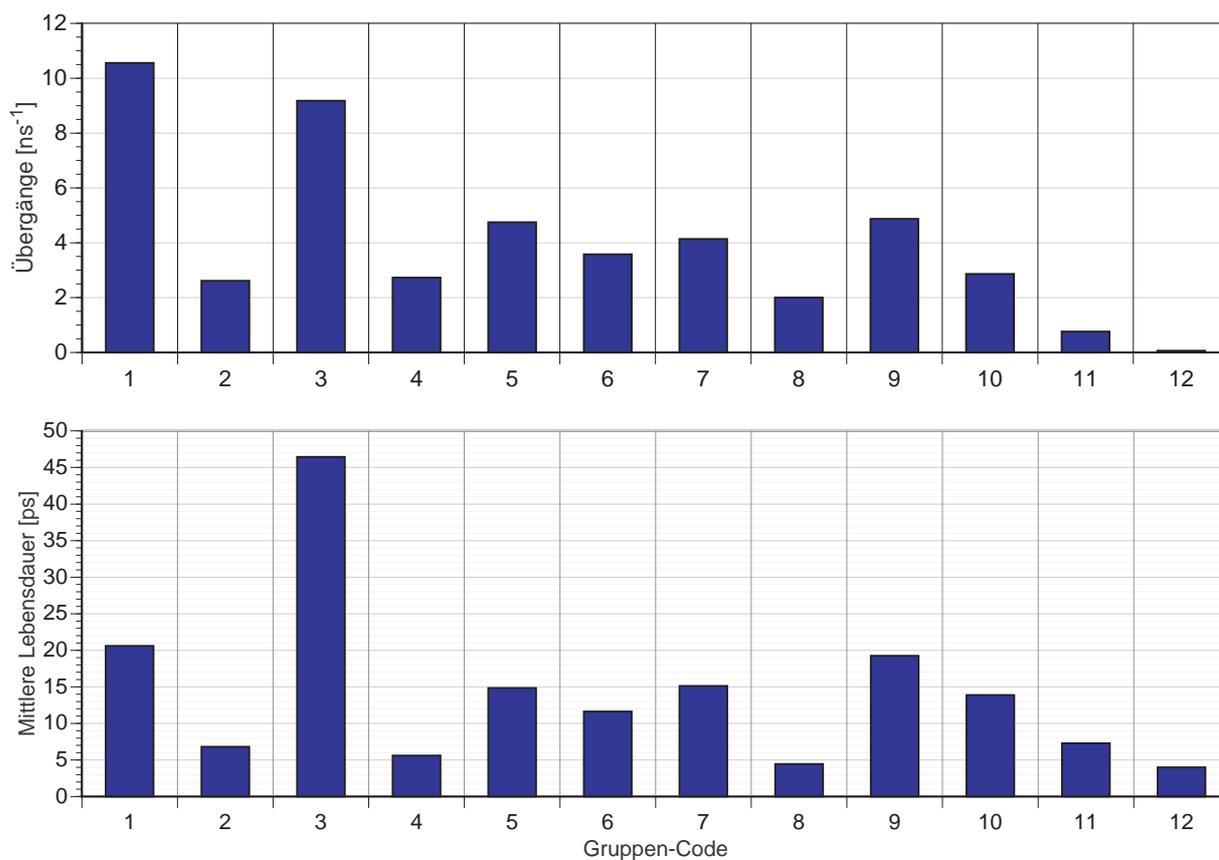
**Abb. X33:** Conformational-Search: Trajektorie der LX-Gruppen (Sialyl-Lewis<sup>X</sup>, CVFF,  $\epsilon = 80$ , 1000 K,  $t_s = 100$  ns).



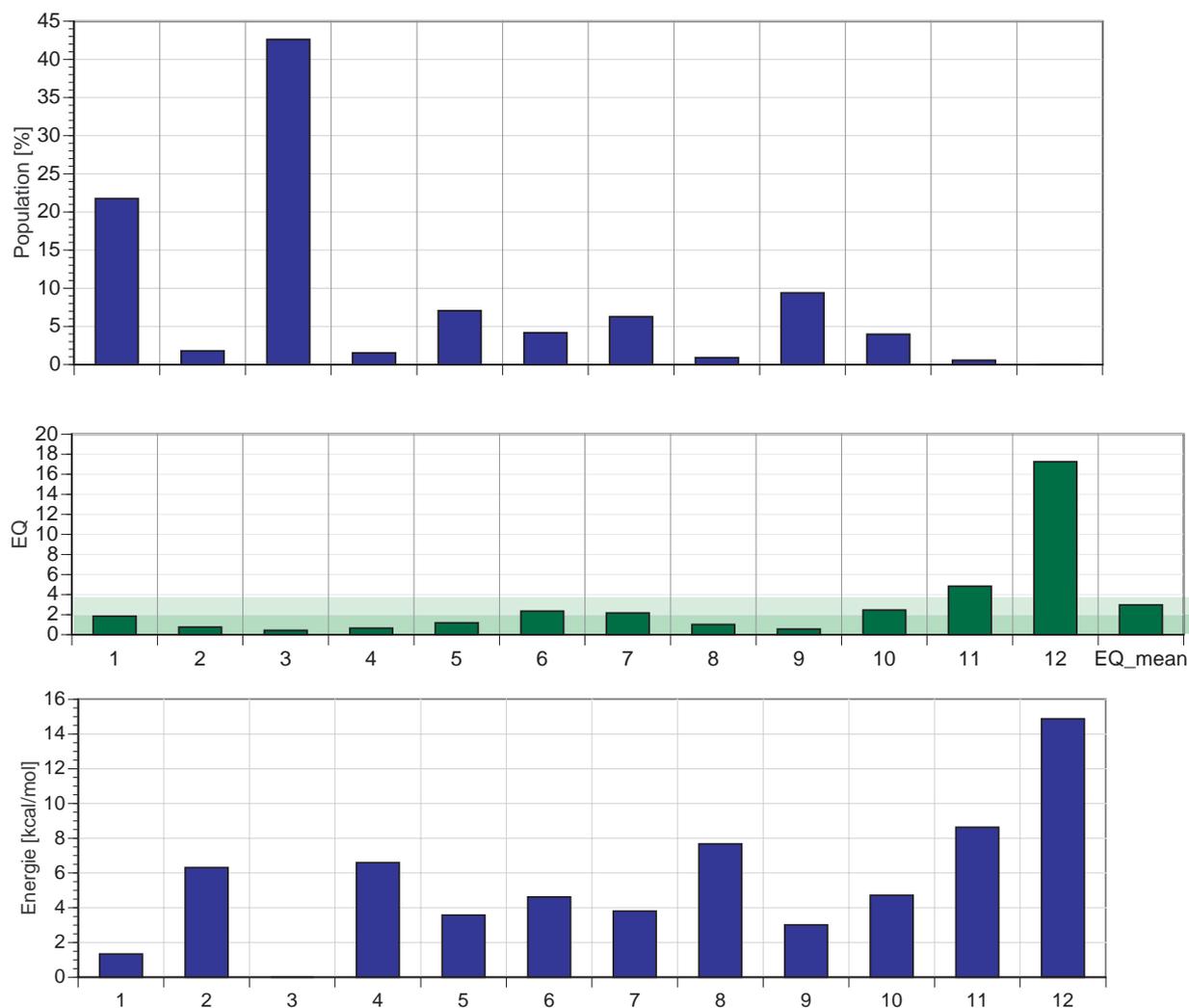
**Abb. X34:** Einstellung des Konformationsgleichgewichts des Lewis<sup>X</sup>-Teils (Sialyl-Lewis<sup>X</sup>, CVFF,  $\epsilon = 80$ , 1000 K,  $t_s = 100$  ns).



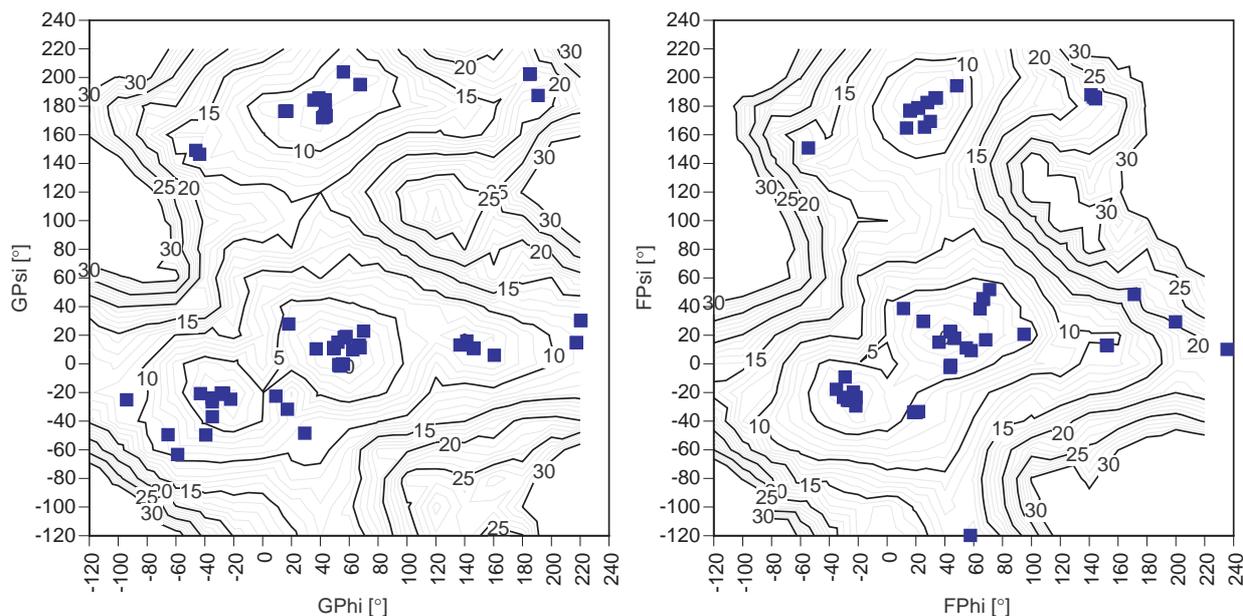
**Abb. X35:** Übergangsmatrix für die LX-Gruppen (Sialyl-Lewis<sup>X</sup>, CVFF,  $\epsilon = 80$ , 1000 K,  $t_s = 100$  ns).



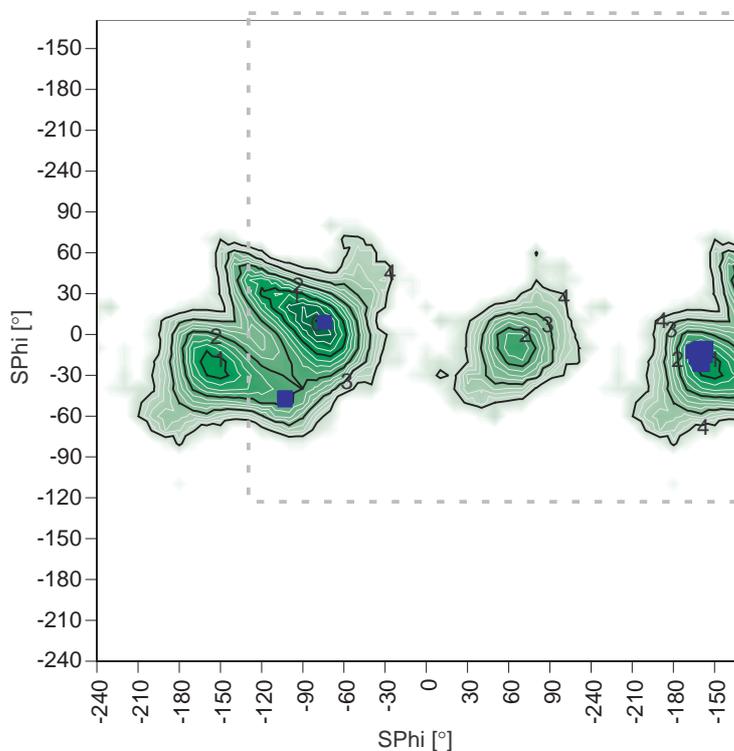
**Abb. X36:** Konformationsanalyse des Lewis<sup>X</sup>-Teils: Zahl der beobachteten Konformationsübergänge pro Nanosekunde absoluter Simulationszeit und mittlere Lebensdauer der LX-Gruppen (Sialyl-Lewis<sup>X</sup>, CVFF,  $\epsilon = 80$ , 1000 K,  $t_s = 100$  ns).



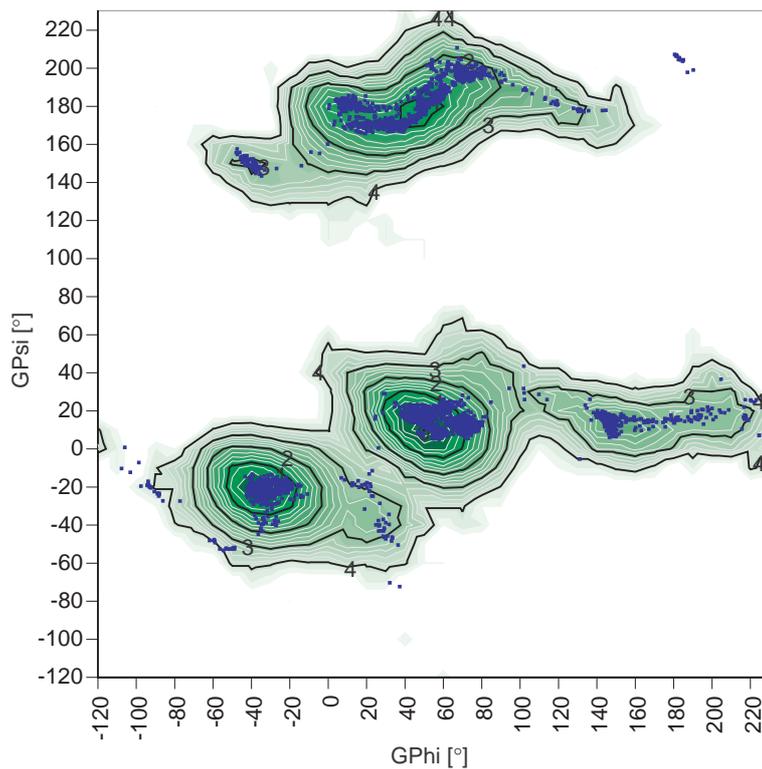
**Abb. X37:** Konformationsanalyse des Lewis<sup>X</sup>-Teils: Konformer-Population, Gleichgewichtseinstellung und Freie Energie der LX-Gruppen (1 bis 12) (Sialyl-Lewis<sup>X</sup>, CVFF,  $\epsilon = 80$ , 1000 K,  $t_s = 100$  ns).



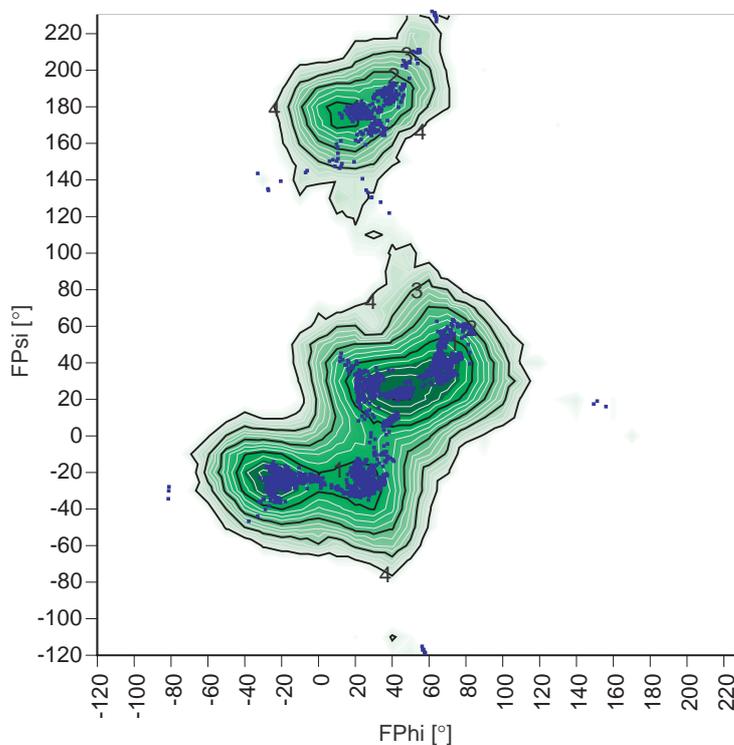
**Abb. X38:** Startstrukturen zur Analyse des Lewis<sup>X</sup>-Teils von Sialyl-Lewis<sup>X</sup> mit Hilfe der GCMD-Methode. Contourlinien: Adiabatic Map aus zweistufigem Grid-Search (Beschreibung siehe Abschnitt 3.3.1)



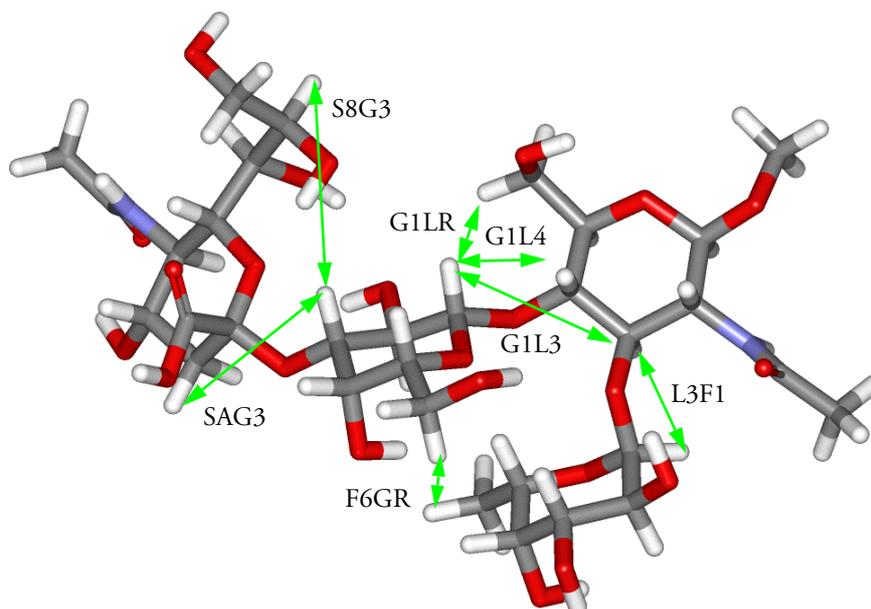
**Abb. X39:** SPhi/SPsi-Werte der Startstrukturen zur Analyse des Lewis<sup>X</sup>-Teils von Sialyl-Lewis<sup>X</sup> mit Hilfe der GCMD-Methode (nach dem Grid-Search / Conformational Clustering). Contourlinien: FGE-Map aus der sich anschließenden MD-Simulation



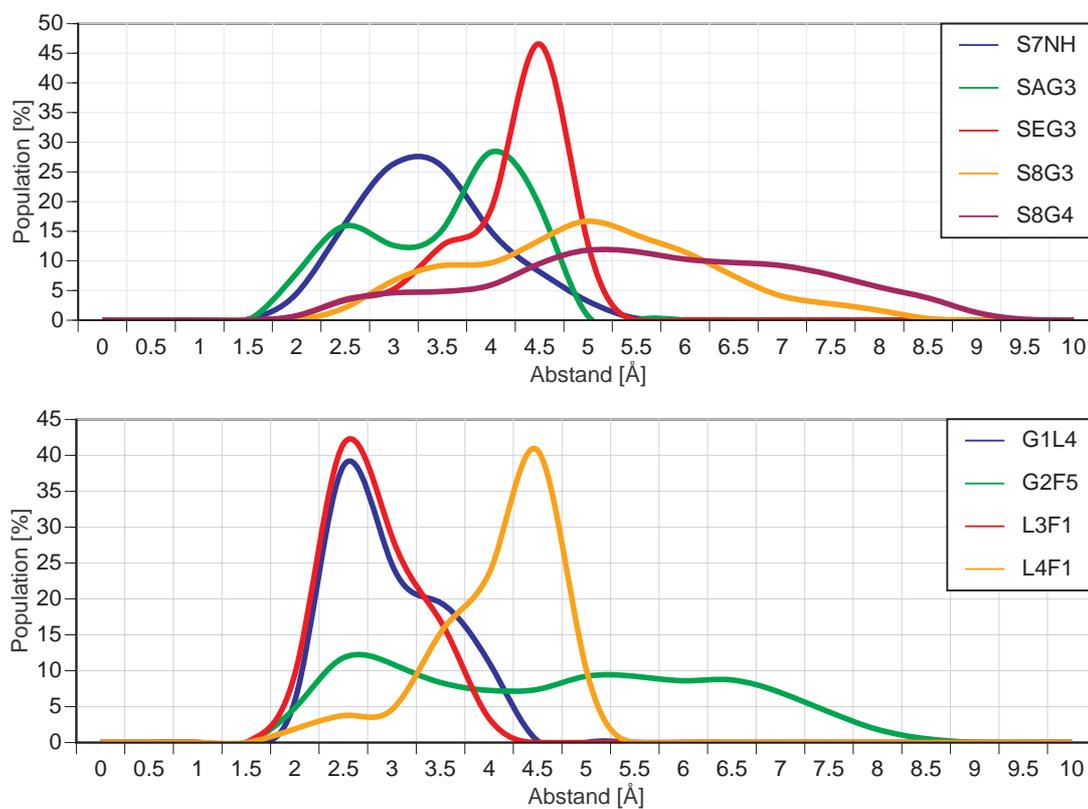
**Abb. X40:** Clustering in der GPhi/GPsi-Ebene (Sialyl-Lewis<sup>X</sup>, CVFF,  $\epsilon = 80$ , 1000 K,  $t_s = 10$  ns).  
Contourlinien: lokal gültige FGE-Map aus der GCMD-Simulation (300 K,  $t_s = 7.6$  ns)



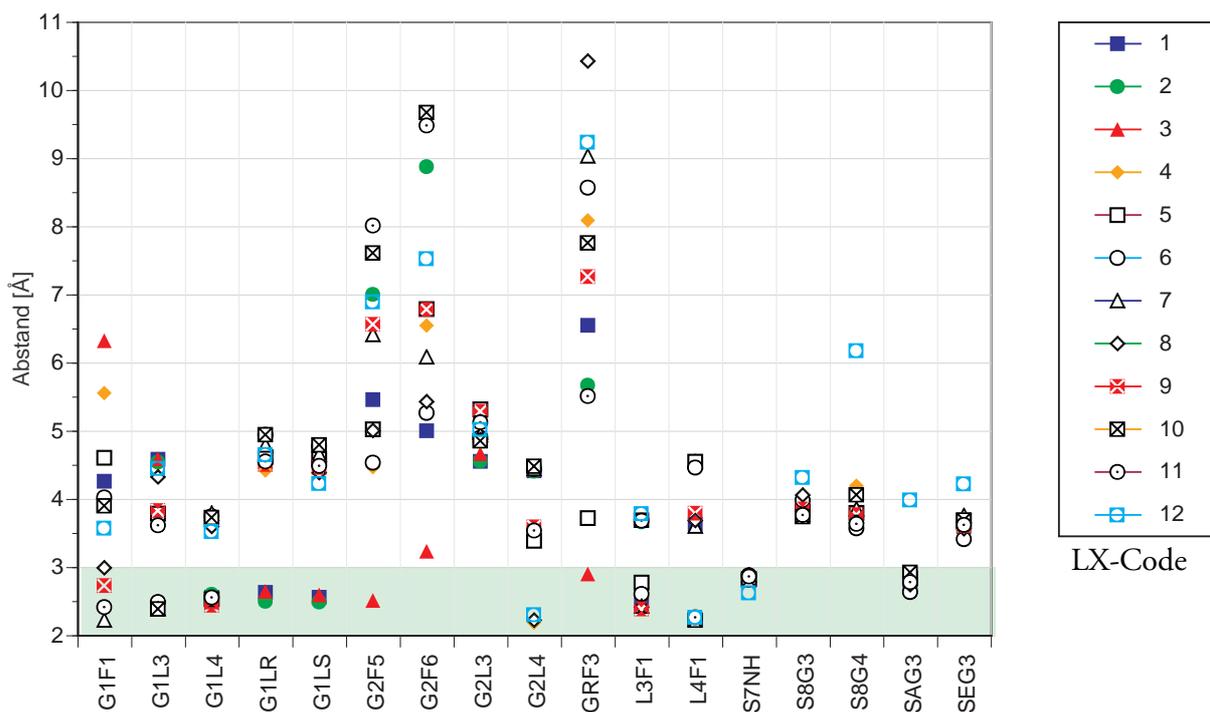
**Abb. X41:** Clustering in der FPhi/FPsi-Ebene (Sialyl-Lewis<sup>X</sup>, CVFF,  $\epsilon = 80$ , 1000 K,  $t_s = 10$  ns).  
Contourlinien: lokal gültige FGE-Map aus der GCMD-Simulation (300 K,  $t_s = 7.6$  ns)



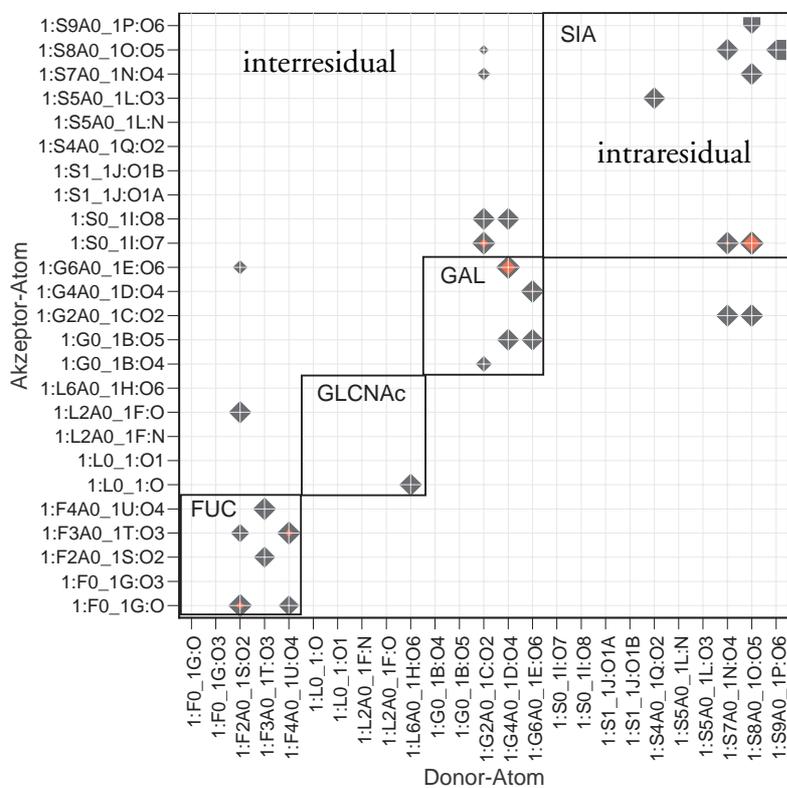
**Abb. X42:** Definition ausgewählter Interproton-Abstände für Sialyl-Lewis<sup>X</sup>.



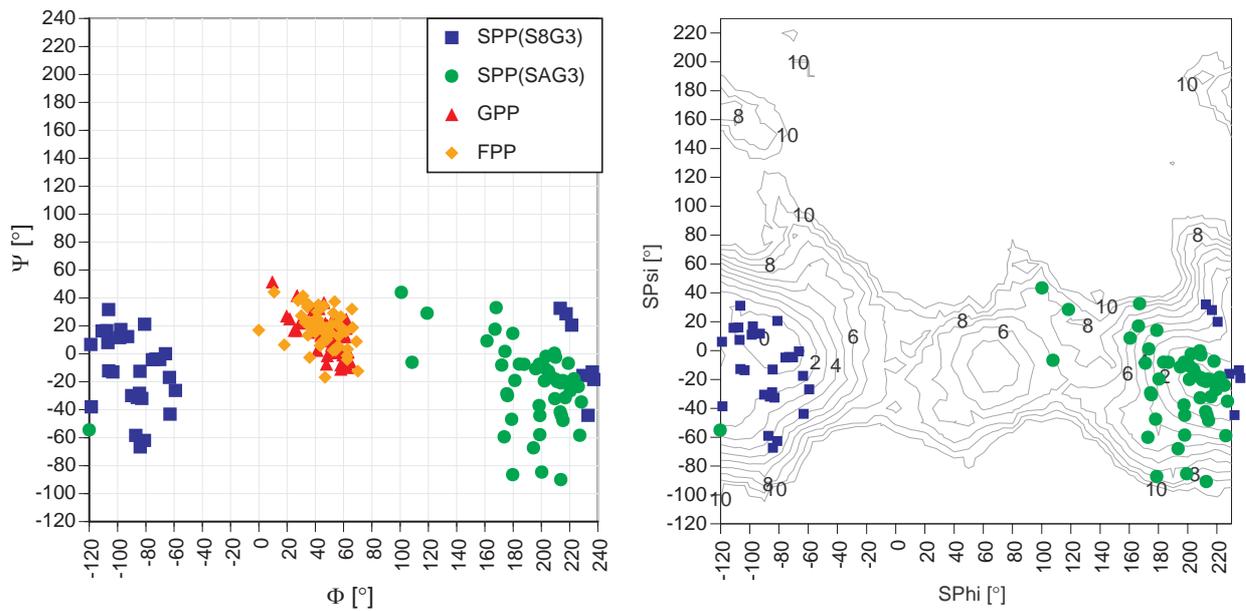
**Abb. X43:** Häufigkeitsverteilung ausgewählter Interproton-Abstände für Sialyl-Lewis<sup>X</sup> (CVFF,  $\epsilon = 80$ , 1000 K,  $t_s = 100$  ns).



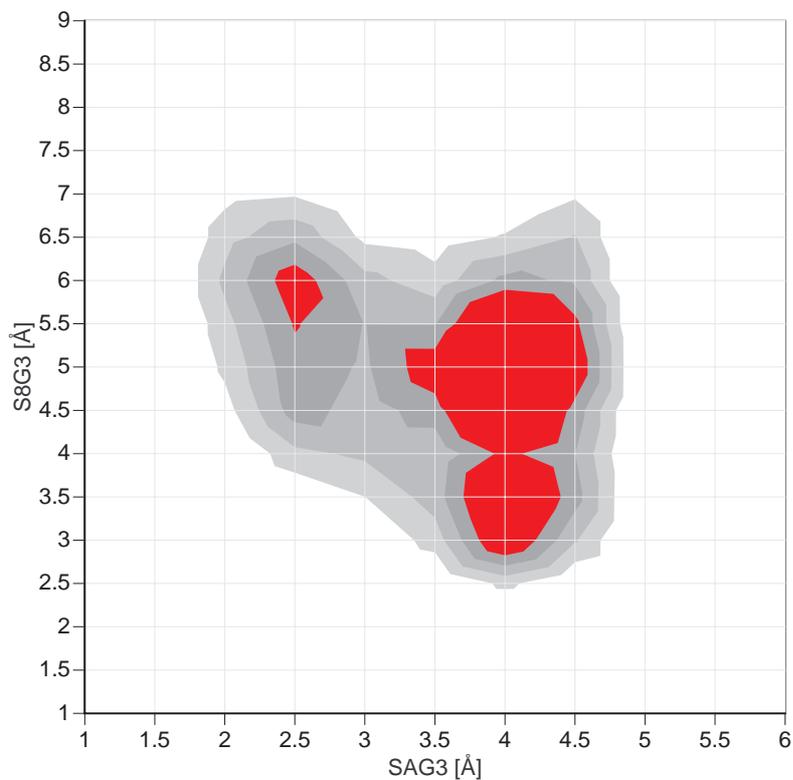
**Abb. X44:**  $\langle r^{-6} \rangle$ -gewichtete mittlere Interproton-Abstände aufgeteilt nach LX-Gruppen (Sialyl-Lewis<sup>X</sup>, CVFF,  $\epsilon = 4$ , 1000 K,  $t_s = 100$  ns). (vgl. Abb. X32)



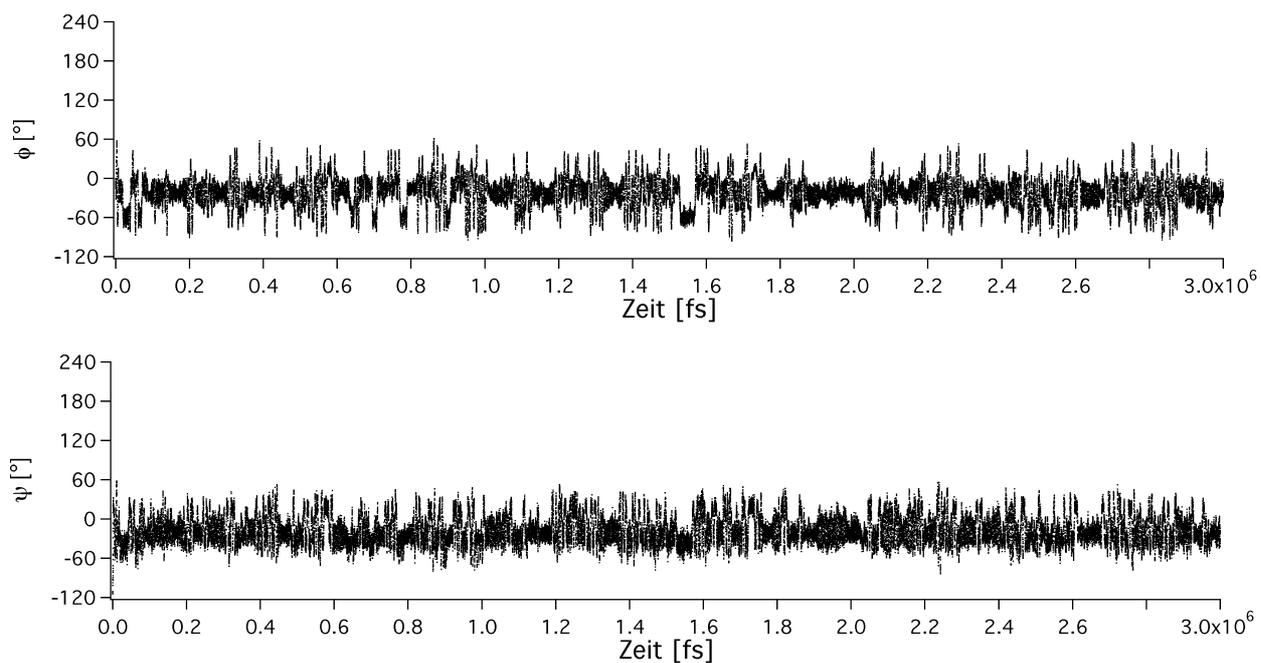
**Abb. X45:** H-Brücken-Matrix (Sialyl-Lewis<sup>X</sup>, CVFF,  $\epsilon = 80$ , 300 K,  $t_s = 7.6$  ns, GCMD-Methode).



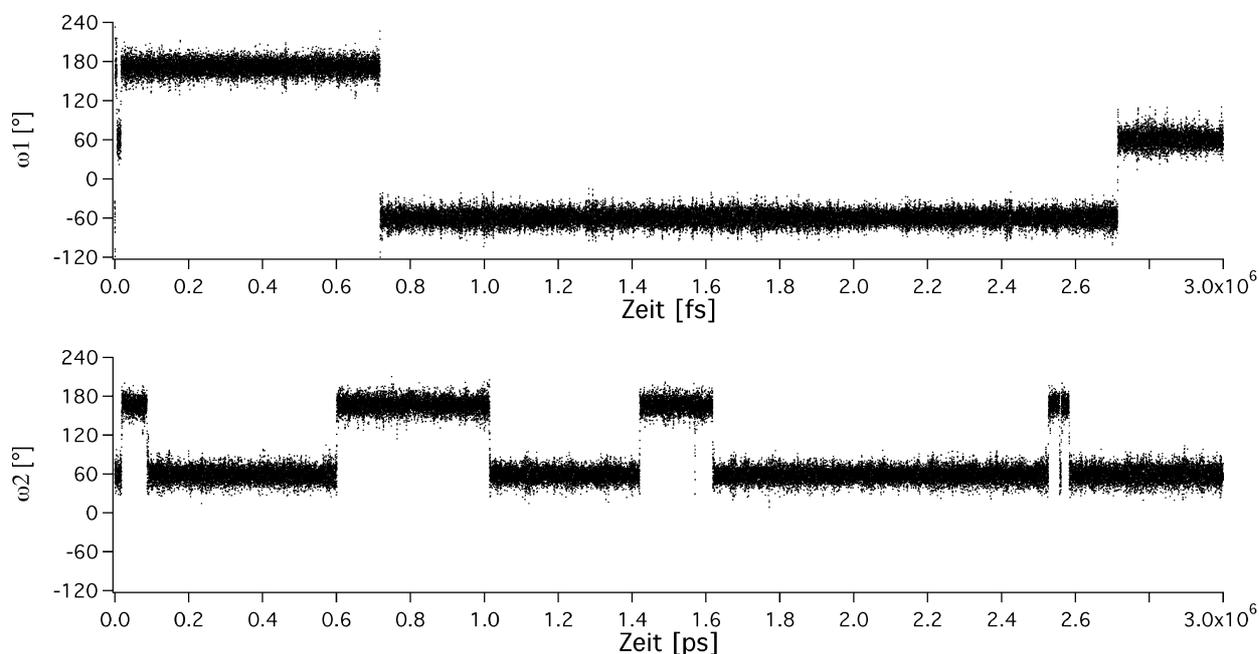
**Abb. X46:** Extraktion von allen Frames aus dem MD-Ensemble die einer Liste von NOE-Randbedingungen genügen (Sialyl-Lewis<sup>X</sup>, CVFF,  $\epsilon = 80$ , 1000 K,  $t_s = 100$  ns).



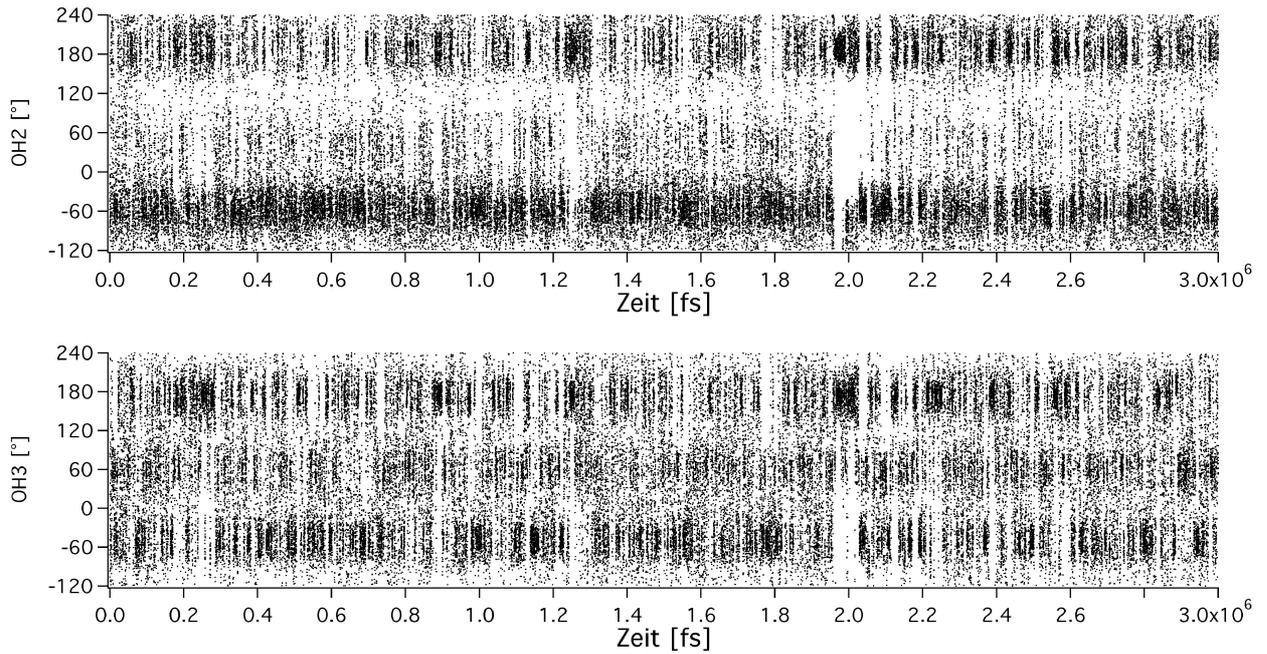
**Abb. X47:** Häufigkeitsverteilung in Abhängigkeit von den H-H-Abständen Sia(H8)-Gal(H3) und Sia(H3<sub>ax</sub>)-Gal(H3) (Sialyl-Lewis<sup>X</sup>, CVFF,  $\epsilon = 80$ , 1000 K,  $t_s = 100$  ns).



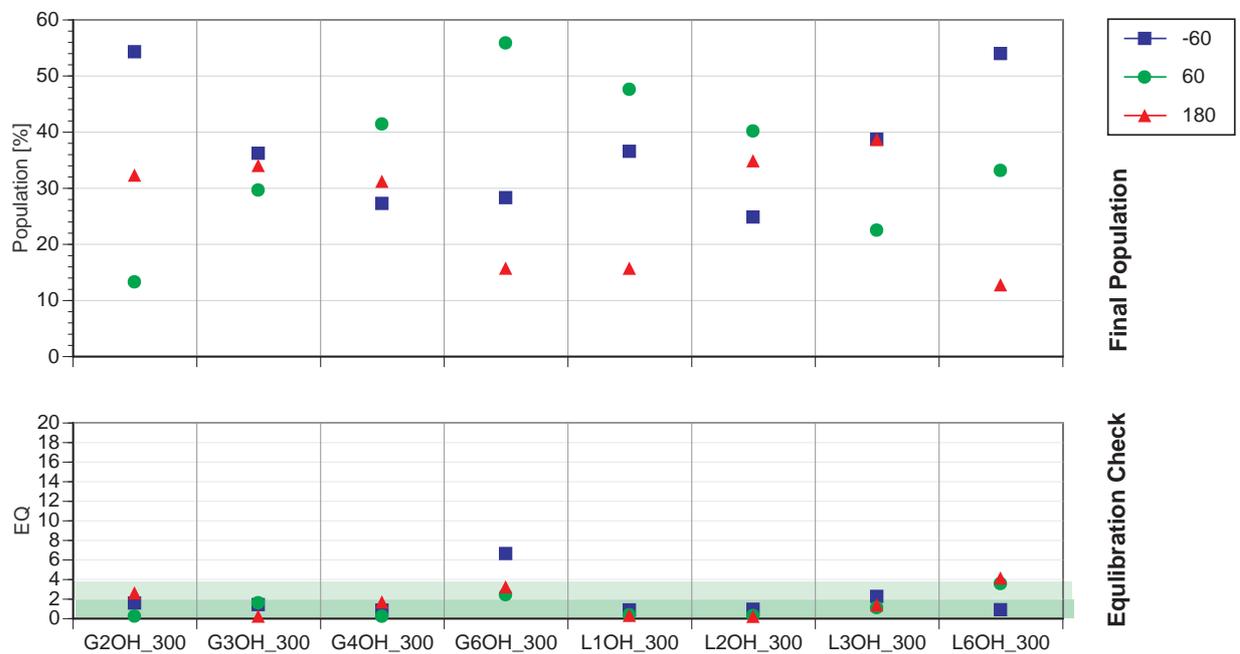
**Abb. X48:**  $\Phi$ - und  $\Psi$ -Trajektorie einer MD-Simulation von Maltose (CVFF,  $\varepsilon = 4$ , 300K,  $t_s = 3$  ns,  $\Delta t_s = 50$  fs)



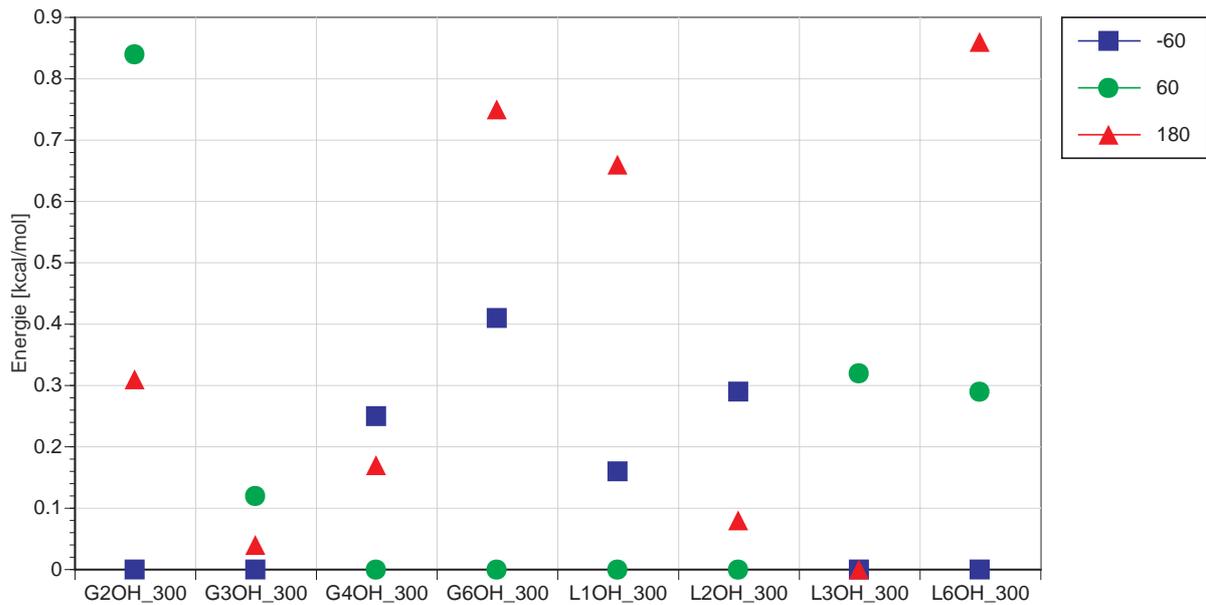
**Abb. X49:**  $\omega_1$ (G56O)- und  $\omega_2$ (L56O)-Trajektorie einer MD-Simulation von Maltose (CVFF,  $\varepsilon = 4$ , 300 K,  $t_s = 3$  ns,  $\Delta t_s = 50$  fs)



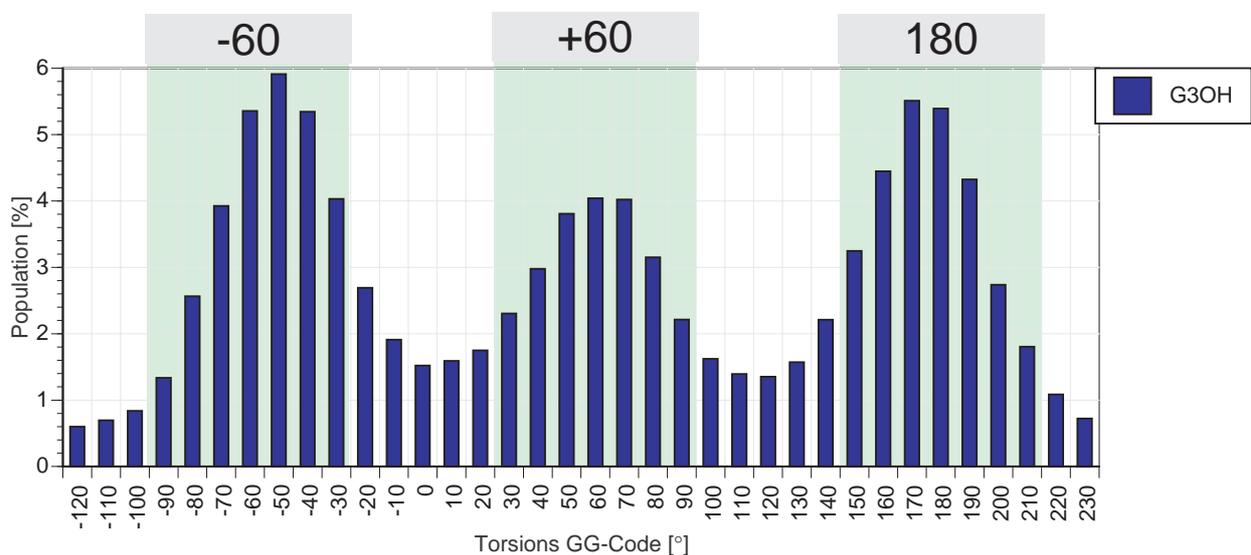
**Abb. X50:** Trajektorien von ausgewählten OH-Gruppen einer MD-Simulation von Maltose (CVFF,  $\epsilon = 4$ , 300 K,  $t_s = 3$  ns,  $\Delta t_s = 50$  fs)



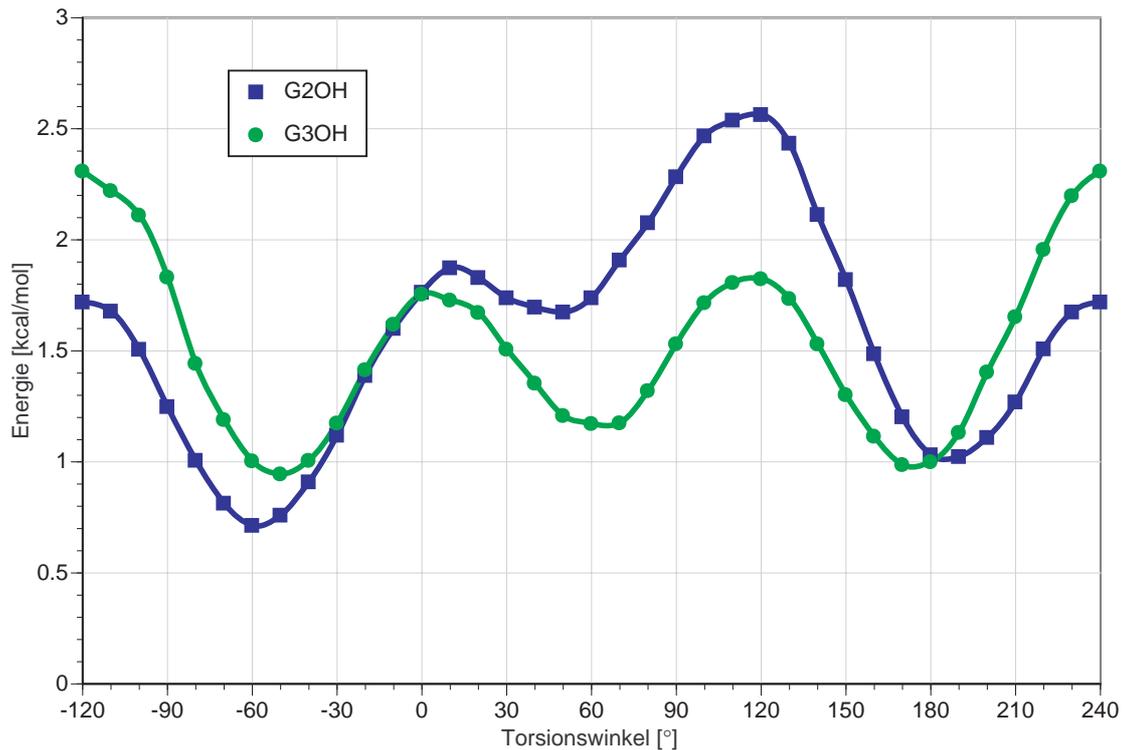
**Abb. X51:** Population der OH-Rotameren (Gridgruppen) und Gleichgewichtseinstellung (EQ3) bei 300K (MD-Simulation von Maltose: CVFF,  $\epsilon = 4$ , 300 K,  $t_s = 3$  ns,  $\Delta t_s = 50$  fs)



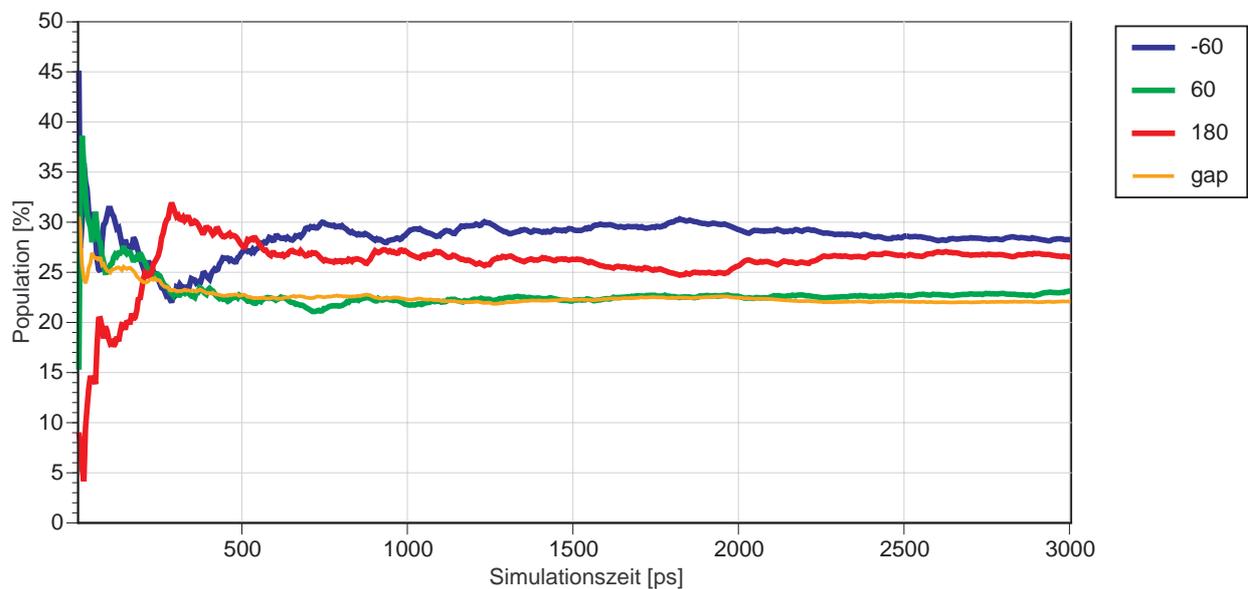
**Abb. X52:** Freie Gridgruppen-Energie der OH-Rotameren (MD-Simulation von Maltose: CVFF,  $\epsilon = 4$ , 300 K,  $t_s = 3$  ns,  $\Delta t_s = 50$  fs) (Anmerkung: Es ist kein direkter Vergleich der Energiewerte verschiedener OH-Gruppen möglich)



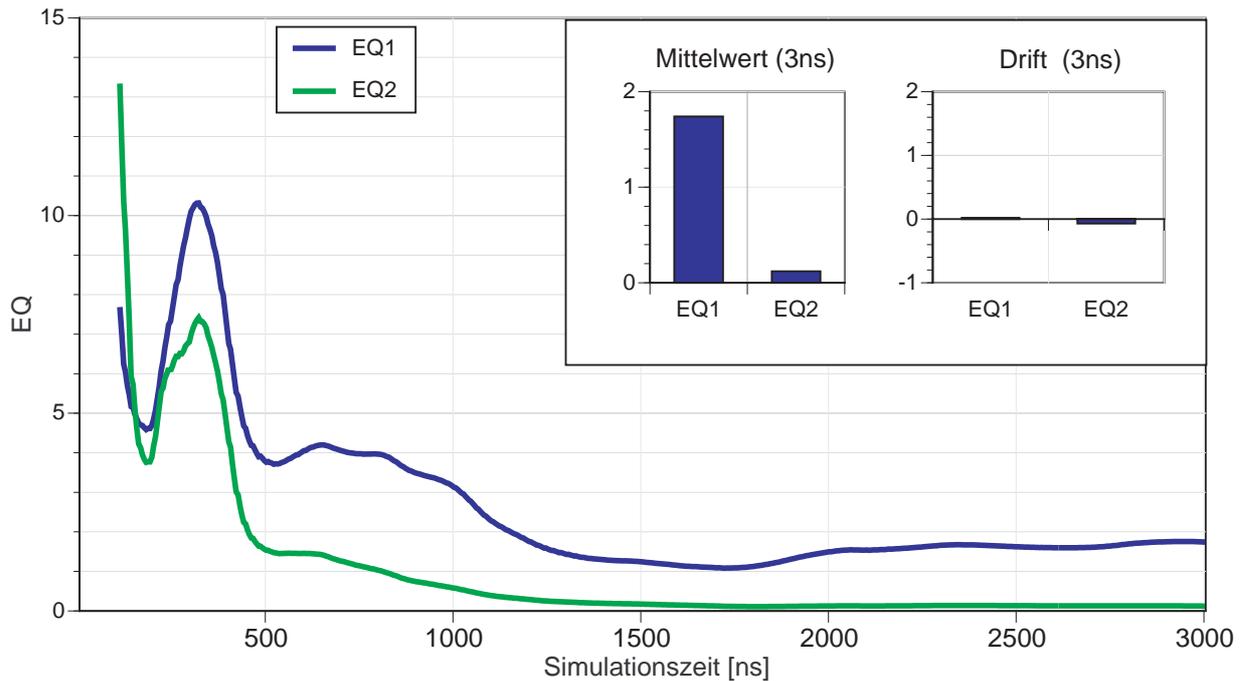
**Abb. X53:** Häufigkeitsverteilung und Definition der Rotameren für G3OH (MD-Simulation von Maltose: CVFF,  $\epsilon = 4$ , 300 K,  $t_s = 3$  ns,  $\Delta t_s = 50$  fs)



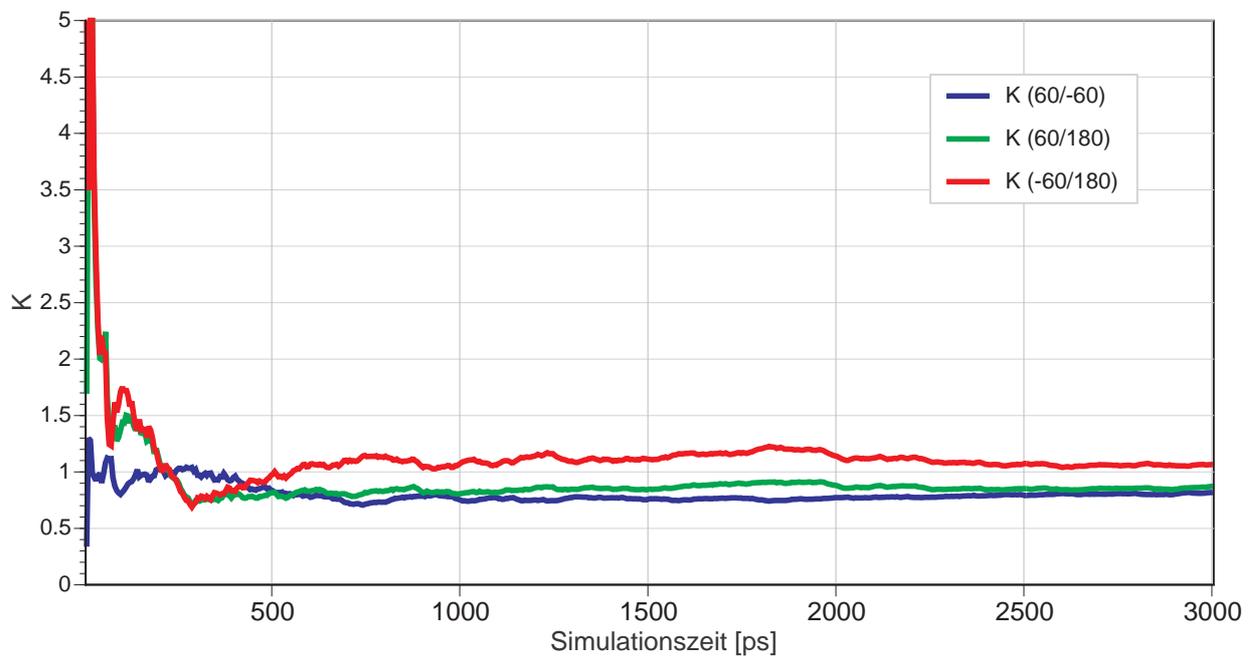
**Abb. X54:** Freie Gridgruppen-Energie für G2OH und G3OH (MD-Simulation von Maltose: CVFE,  $\epsilon = 4$ , 300 K,  $t_s = 3$  ns,  $\Delta t_s = 50$  fs)



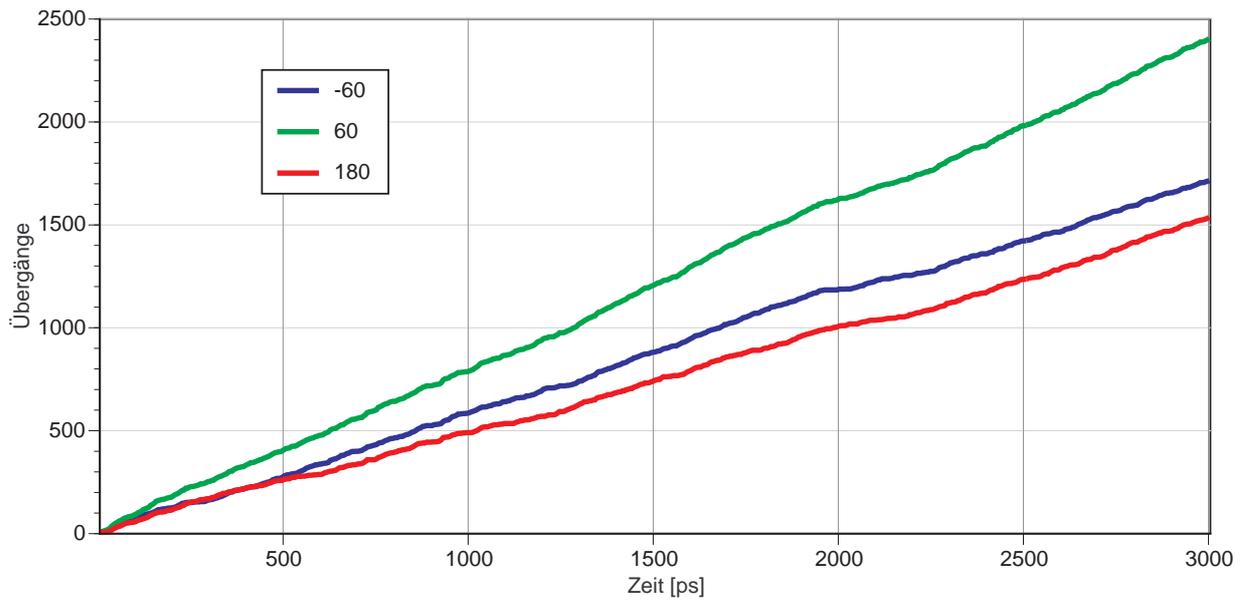
**Abb. X55:** Rotamerenverteilung für G3OH in Abhängigkeit von der Simulationszeit (MD-Simulation von Maltose: CVFE,  $\epsilon = 4$ , 300 K,  $t_s = 3$  ns,  $\Delta t_s = 50$  fs)



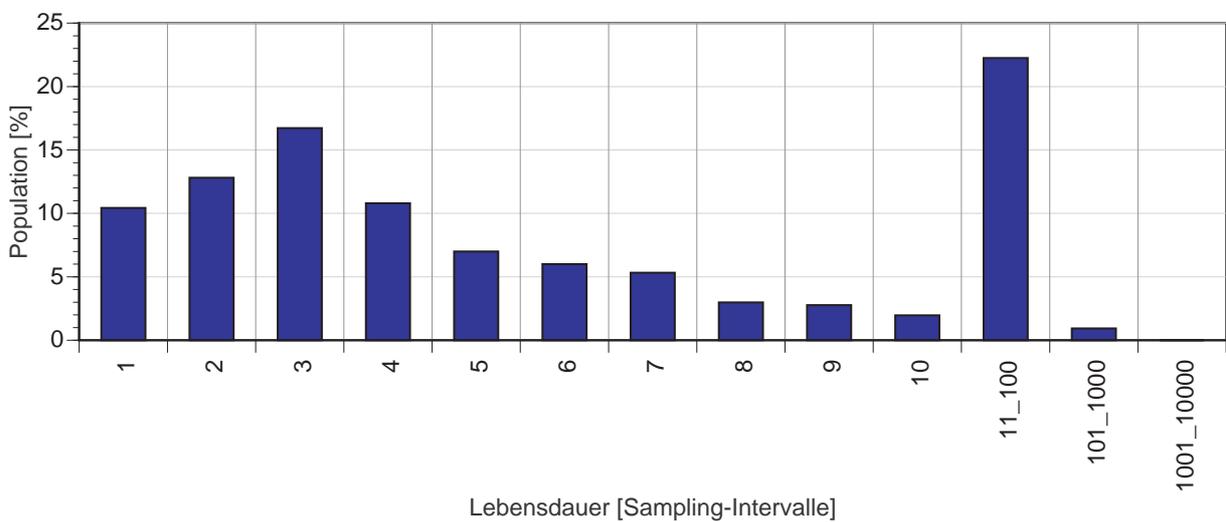
**Abb. X56:** Rotamerengleichgewichtseinstellung für G3OH in Abhängigkeit von der Simulationszeit (MD-Simulation von Maltose: CVFF,  $\epsilon = 4$ , 300K,  $t_s = 3$  ns,  $\Delta t_s = 50$  fs)



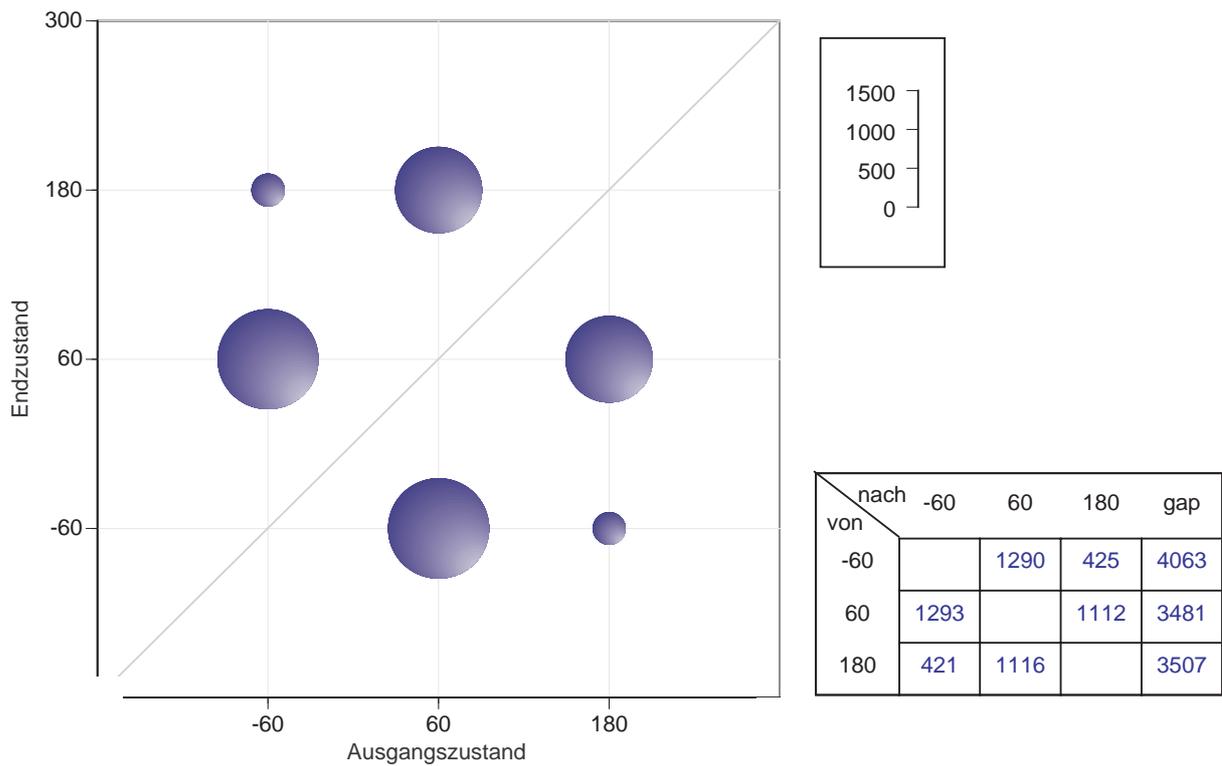
**Abb. X57:** Änderung der Gleichgewichtskonstanten für das Rotamerengleichgewicht von G3OH in Abhängigkeit von der Simulationszeit (MD-Simulation von Maltose: CVFF,  $\epsilon = 4$ , 300 K,  $t_s = 3$  ns,  $\Delta t_s = 50$  fs)



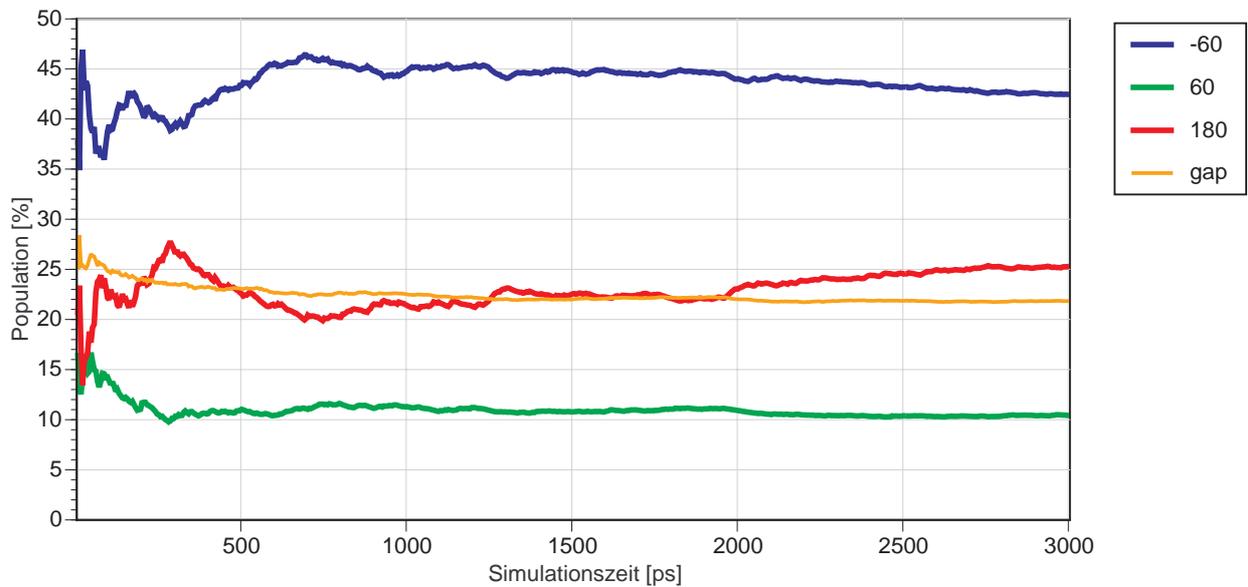
**Abb. X58:** Zahl der Rotamerenübergänge für G3OH in Abhängigkeit von der Simulationszeit (MD-Simulation von Maltose: CVFF,  $\varepsilon = 4$ , 300 K,  $t_s = 3$  ns,  $\Delta t_s = 50$  fs)



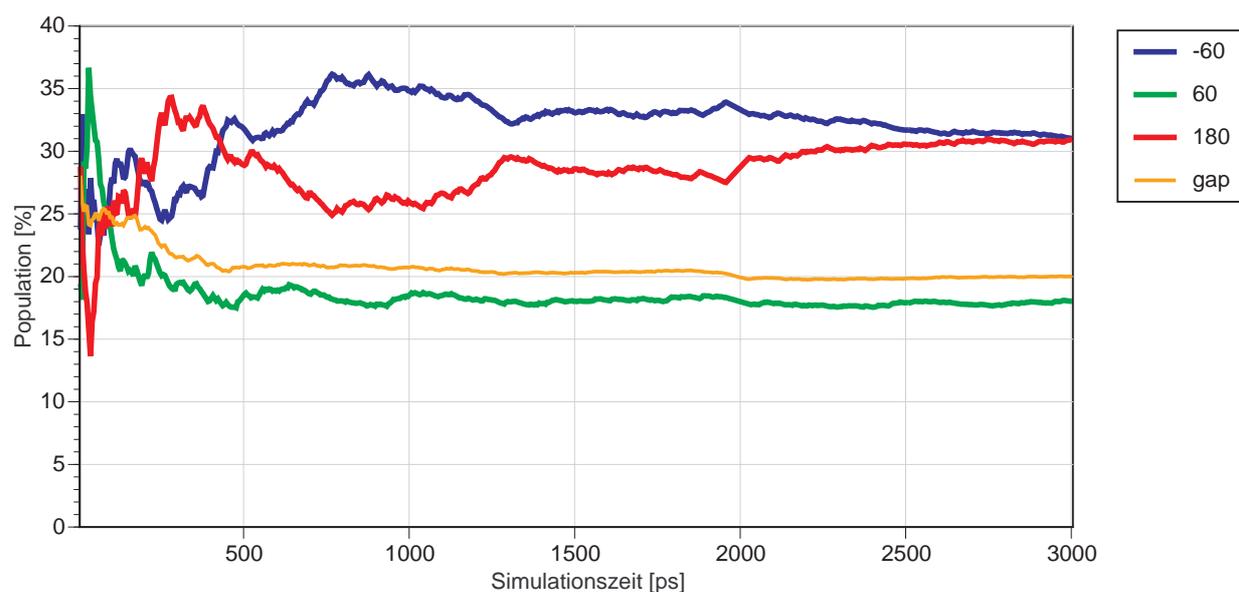
**Abb. X59:** Verteilung der Rotameren-Lebensdauern für G3OH (MD-Simulation von Maltose: CVFF,  $\varepsilon = 4$ , 300 K,  $t_s = 3$  ns,  $\Delta t_s = 50$  fs)



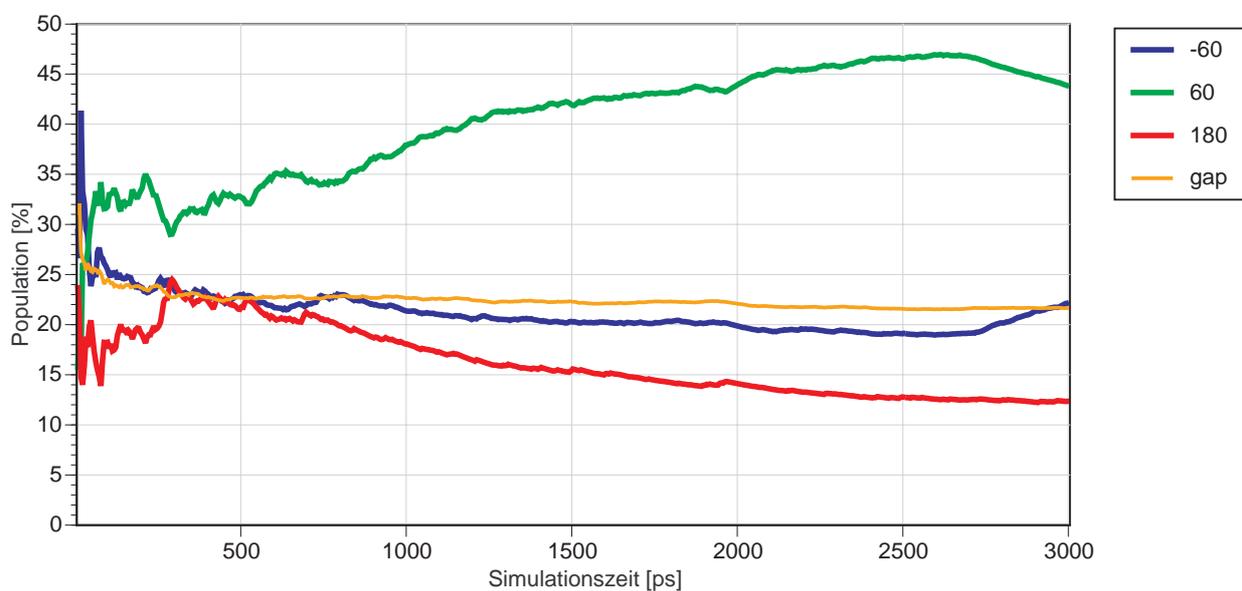
**Abb. X60:** Übergangsmatrix für G3OH (MD-Simulation von Maltose: CVFF,  $\varepsilon = 4$ , 300 K,  $t_s = 3$  ns,  $\Delta t_s = 50$  fs)



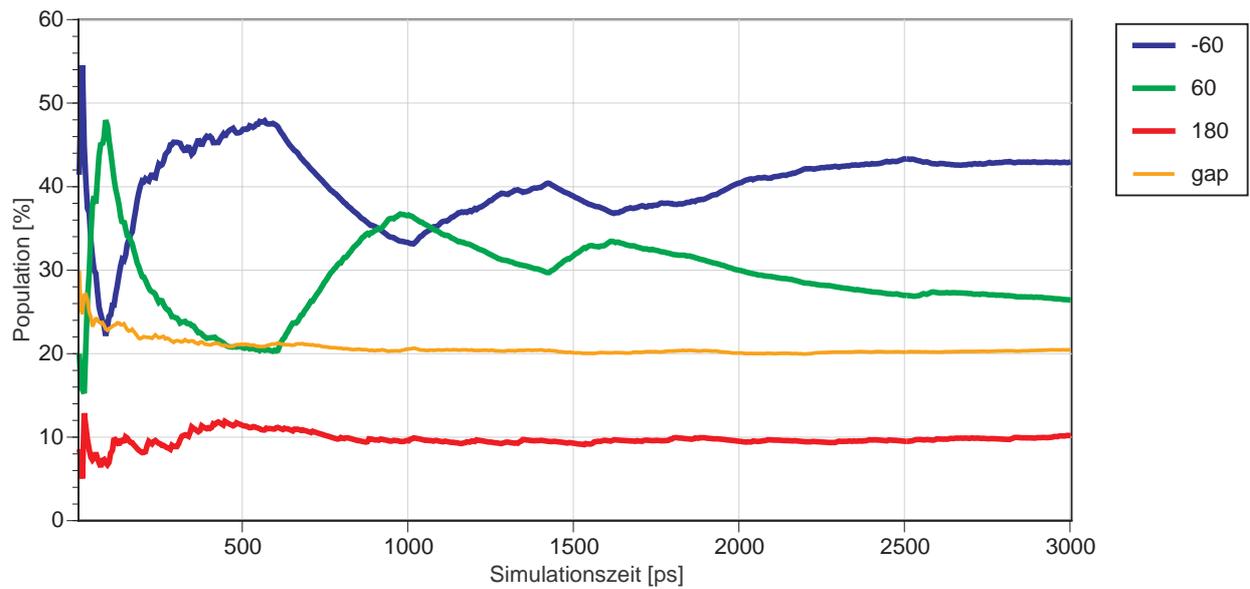
**Abb. X61:** Rotamerenverteilung für G2OH in Abhängigkeit von der Simulationszeit (MD-Simulation von Maltose: CVFF,  $\varepsilon = 4$ , 300 K,  $t_s = 3$  ns,  $\Delta t_s = 50$  fs)



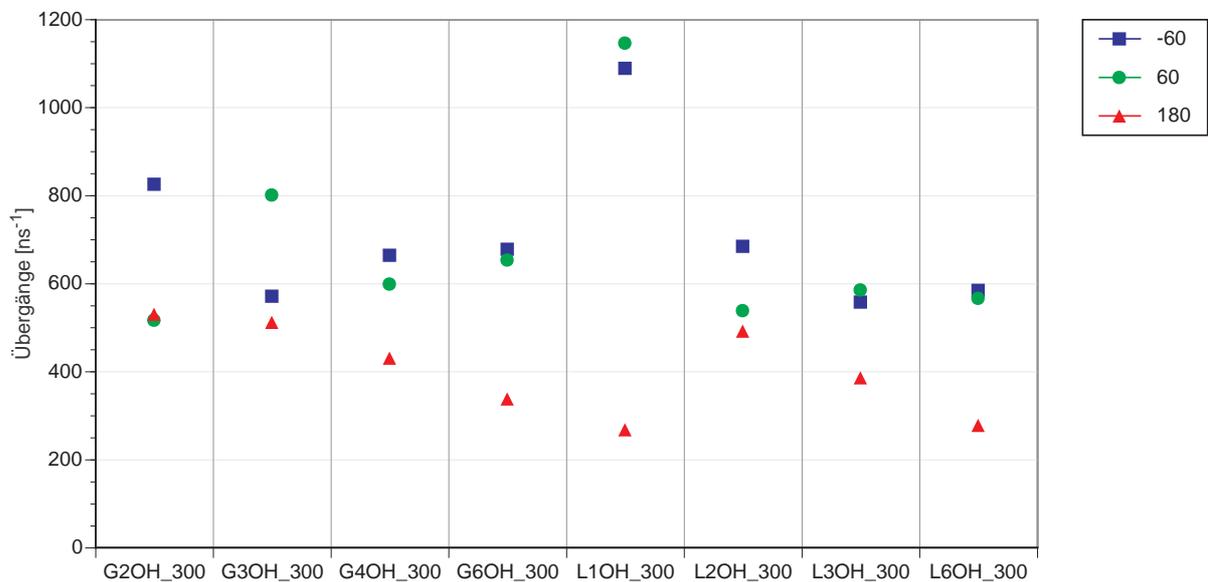
**Abb. X62:** Rotamerenverteilung für L3OH in Abhängigkeit von der Simulationszeit (MD-Simulation von Maltose: CVFF,  $\epsilon = 4$ , 300 K,  $t_s = 3$  ns,  $\Delta t_s = 50$  fs)



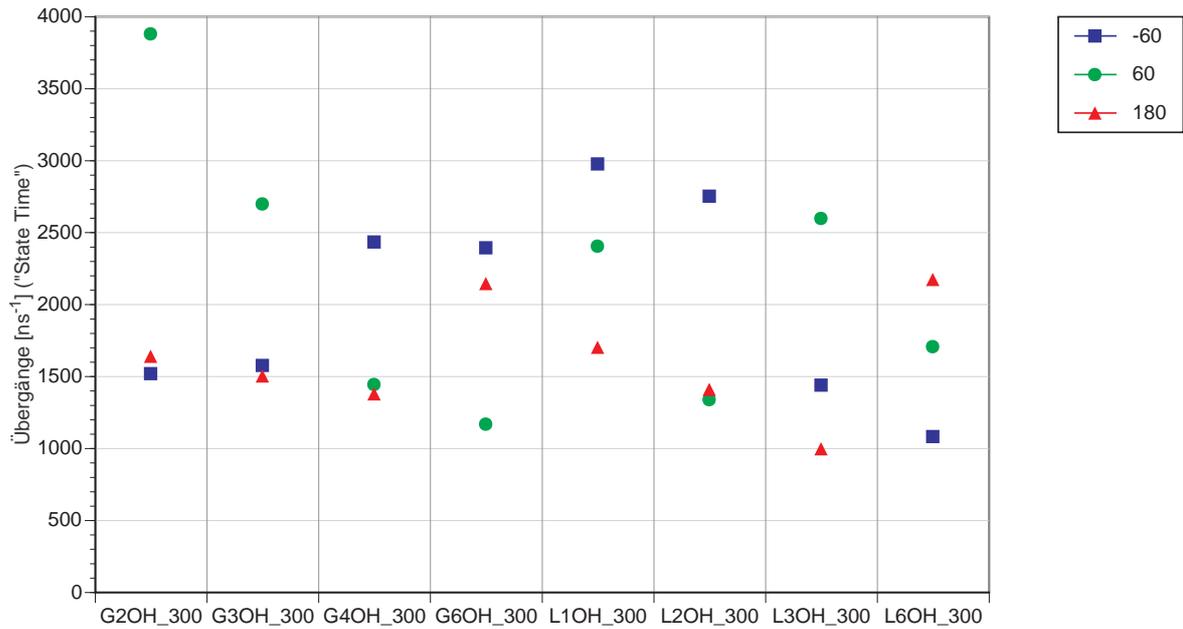
**Abb. X63:** Rotamerenverteilung für G6OH in Abhängigkeit von der Simulationszeit (MD-Simulation von Maltose: CVFF,  $\epsilon = 4$ , 300 K,  $t_s = 3$  ns,  $\Delta t_s = 50$  fs)



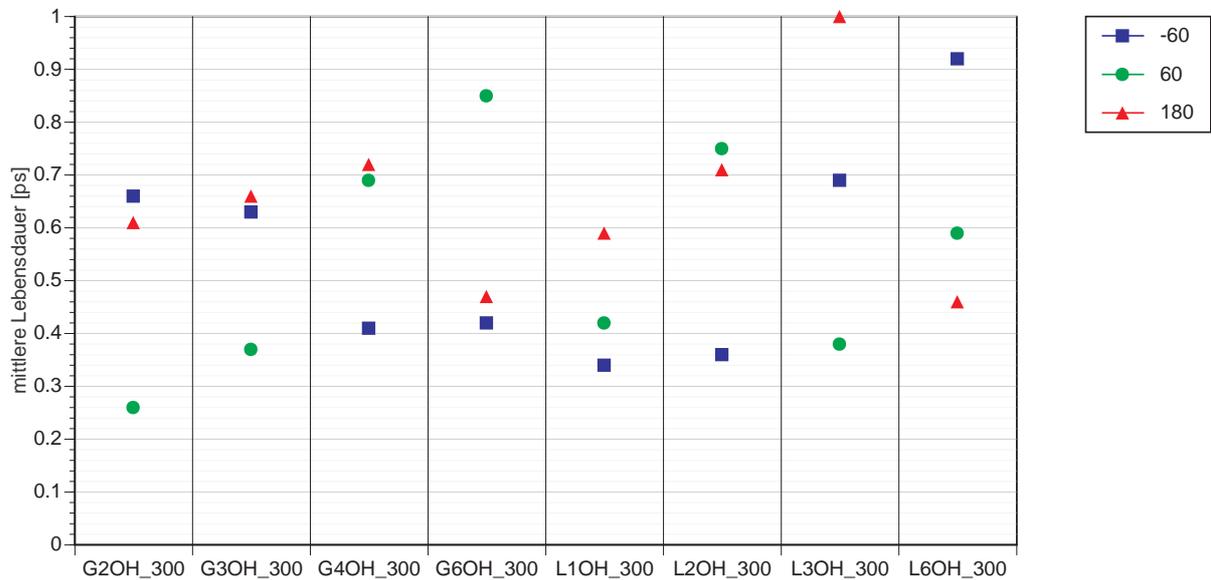
**Abb. X64:** Rotamerenverteilung für L6OH in Abhängigkeit von der Simulationszeit (MD-Simulation von Maltose: CVFF,  $\epsilon = 4$ , 300 K,  $t_s = 3$  ns,  $\Delta t_s = 50$  fs)



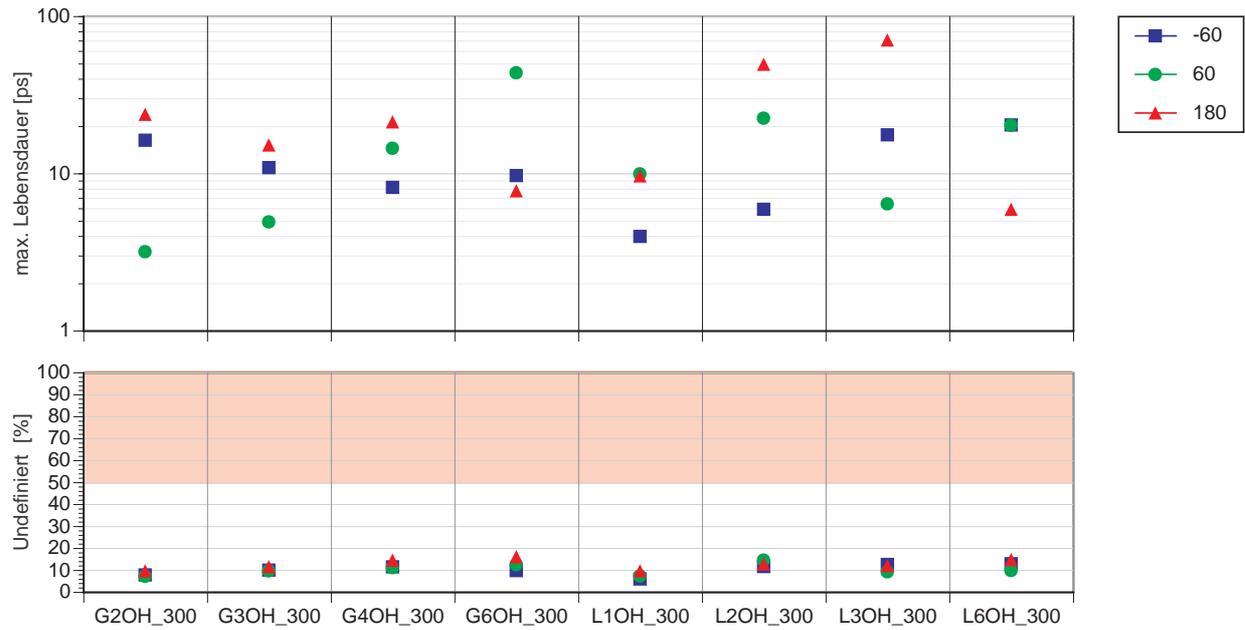
**Abb. X65:** Absolute Anzahl der Übergänge in/aus einem Rotamerenzustand pro Nanosekunde Simulationszeit (MD-Simulation von Maltose: CVFF,  $\epsilon = 4$ , 300 K,  $t_s = 3$  ns,  $\Delta t_s = 50$  fs)



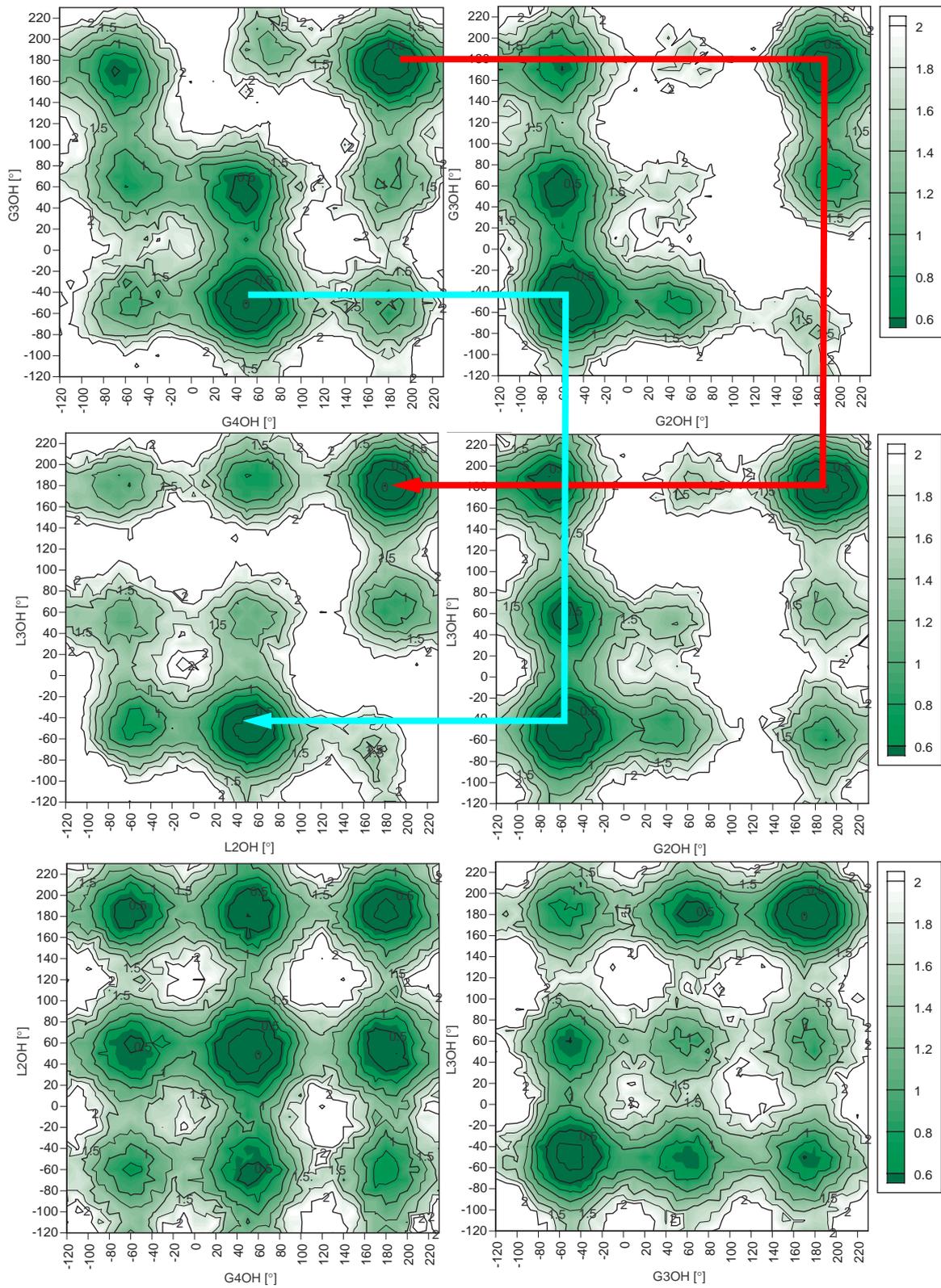
**Abb. X66:** Anzahl der Übergänge in/aus einem Rotamerenzustand pro Nanosekunde „State Time“ (MD-Simulation von Maltose: CVFF,  $\varepsilon = 4$ , 300 K,  $t_s = 3$  ns,  $\Delta t_s = 50$  fs)



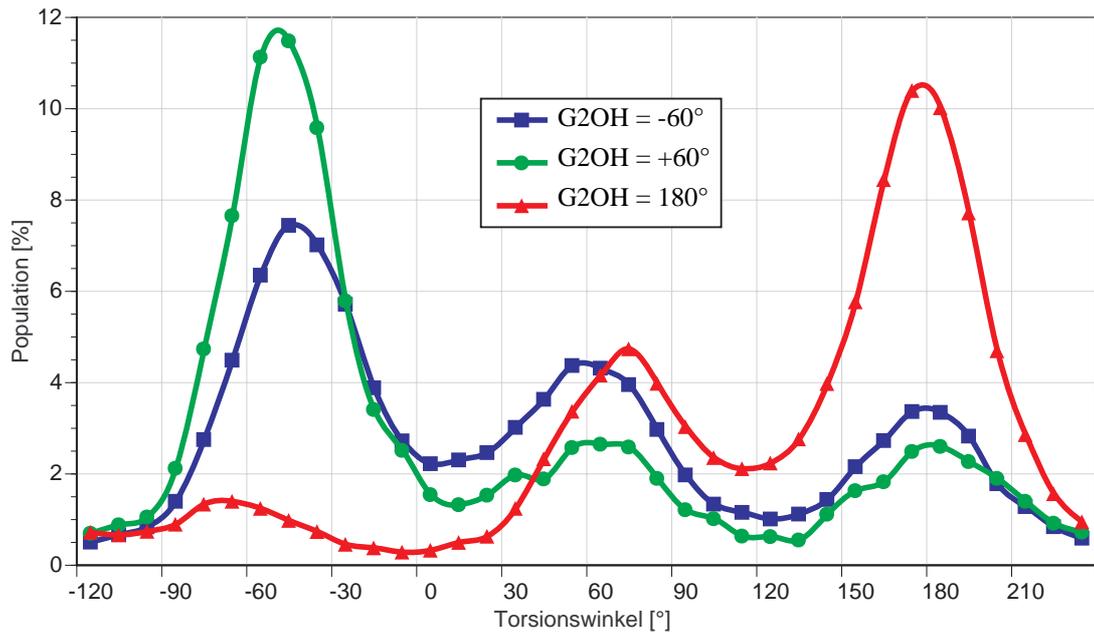
**Abb. X67:** Mittlere Lebensdauer eines Rotamerenzustands (MD-Simulation von Maltose: CVFF,  $\varepsilon = 4$ , 300 K,  $t_s = 3$  ns,  $\Delta t_s = 50$  fs)



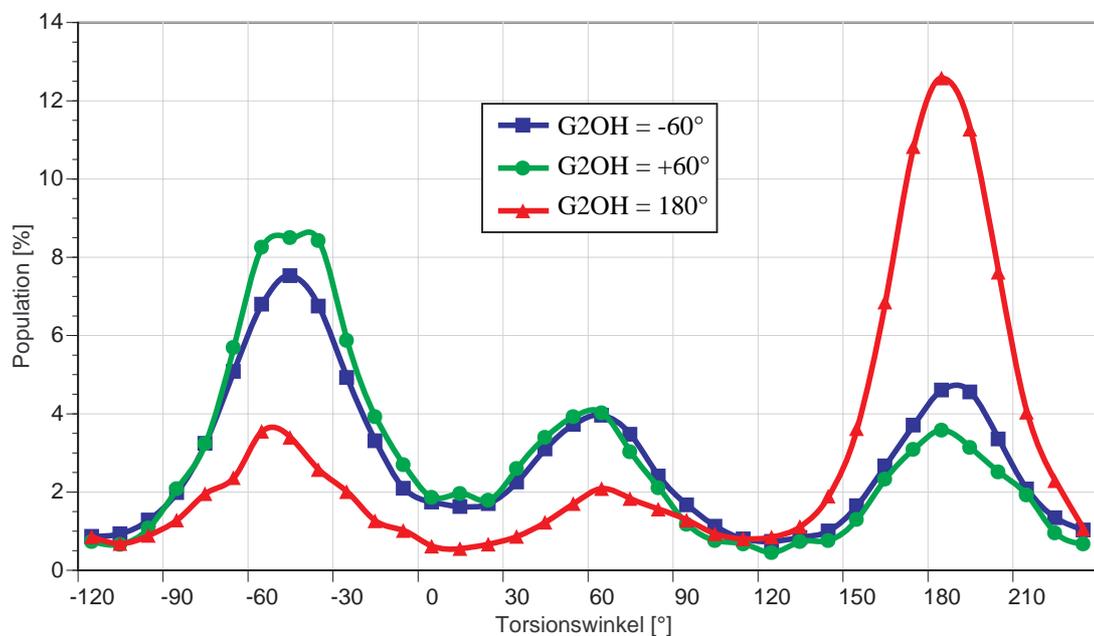
**Abb. X68:** Maximale Lebensdauer eines Rotamerenzustands und Prozentsatz der Lebensdauern die kleiner oder gleich dem Sampling-Intervall sind („Undefinierte Lebensdauern“) (MD-Simulation von Maltose: CVFF,  $\epsilon = 4$ , 300 K,  $t_s = 3$  ns,  $\Delta t_s = 50$  fs)



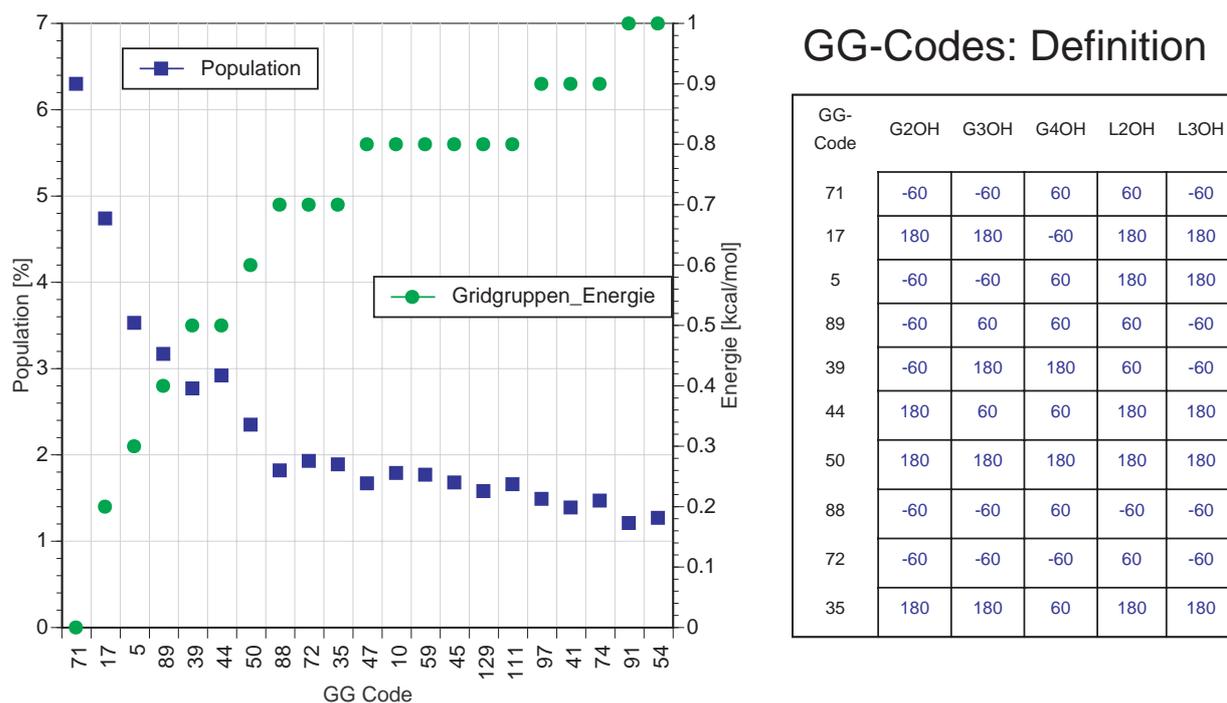
**Abb. X69:** Fernwirkungen von Rotamereneinstellungen der sekundären OH-Gruppen ( $\beta$ -Maltose, CVFF,  $\epsilon = 4$ , 300 K,  $t_s = 3$  ns,  $\Delta t_s = 50$  fs)



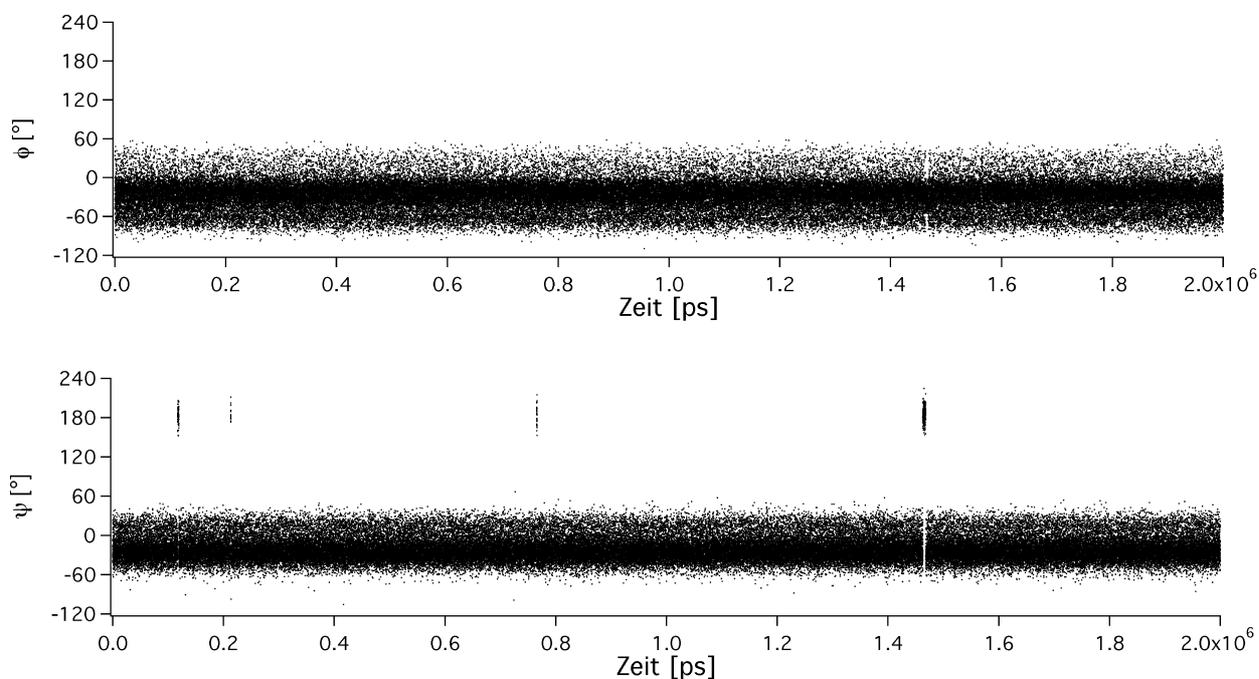
**Abb. X70:** Abhängigkeit der Rotamerenpopulation benachbarter OH-Gruppen von der Orientierung von G2OH ( $\beta$ -Maltose, Torsionswinkel G3OH, CVFF,  $\epsilon = 4$ , 300 K,  $t_s = 3$  ns)



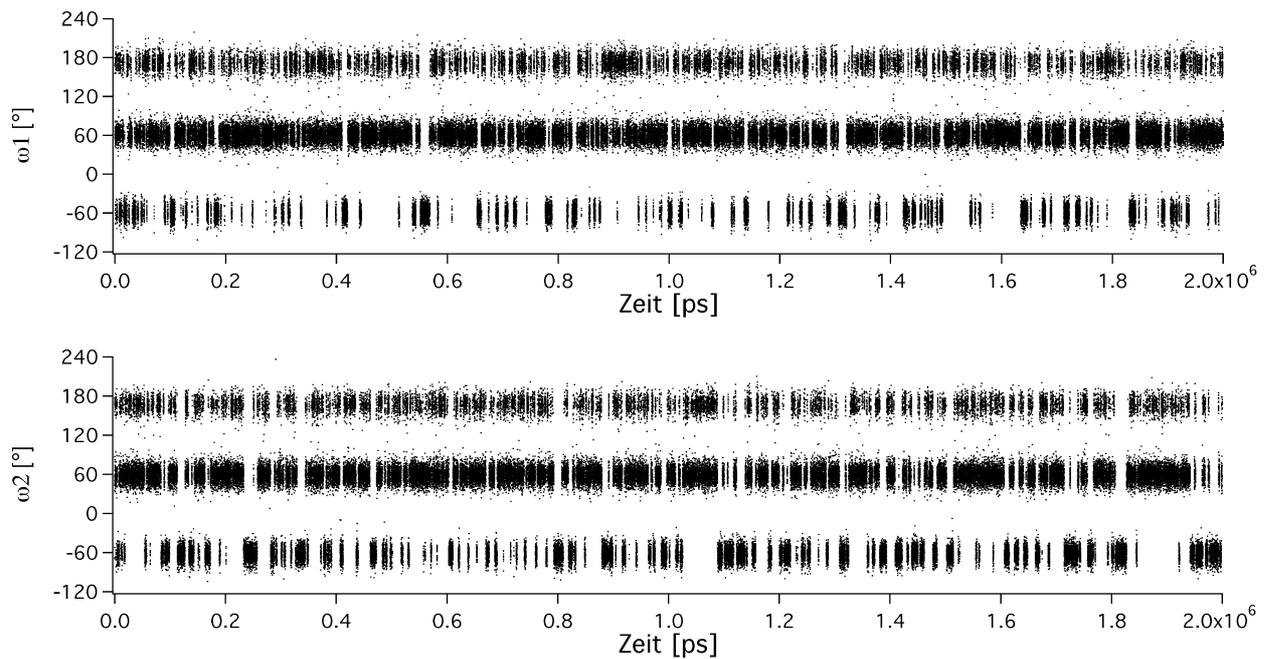
**Abb. X71:** Abhängigkeit der Rotamerenpopulation benachbarter OH-Gruppen von der Orientierung von G2OH ( $\beta$ -Maltose, Torsionswinkel L3OH, CVFF,  $\epsilon = 4$ , 300 K,  $t_s = 3$  ns)



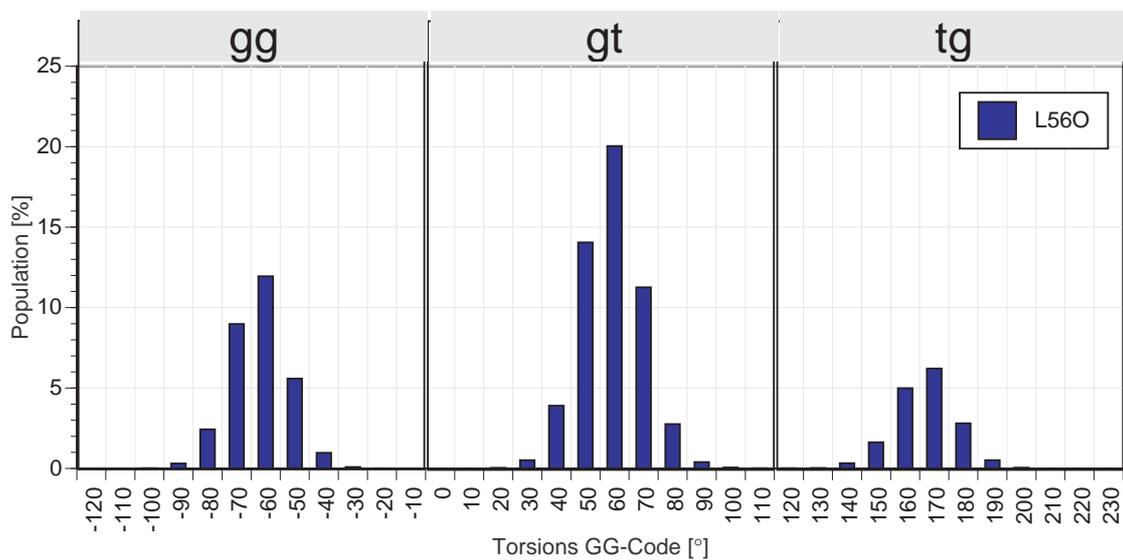
**Abb. X72:** Kombinierte Rotamerengruppen der Torsionswinkel G2OH, G3OH, G4OH, L2OH und L3OH zu neuen 5D-Gridgruppen ( $\beta$ -Maltose, CVFF,  $\epsilon = 4$ , 300K,  $t_s = 1$  ns,  $\Delta t_s = 20$  fs)



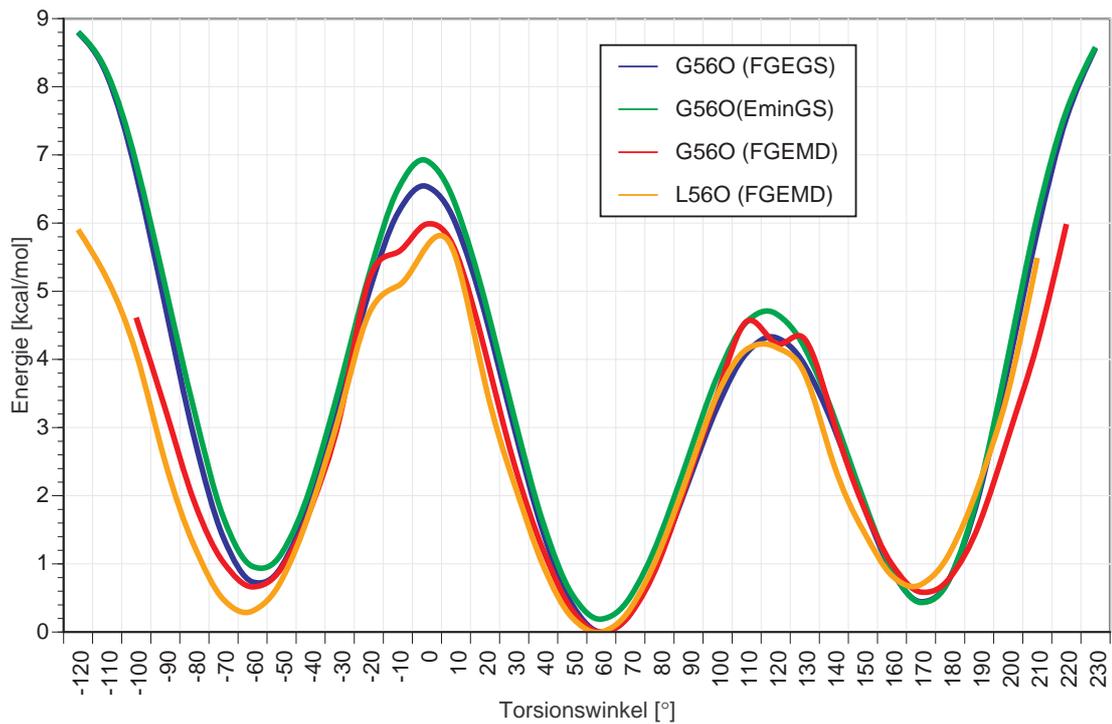
**Abb. X73:**  $\Phi$ - und  $\Psi$ -Trajektorie der Langzeitsimulation von  $\beta$ -Maltose (CVFF,  $\epsilon = 4$ , 300 K,  $t_s = 2$   $\mu$ s,  $\Delta t_s = 20$  ps)



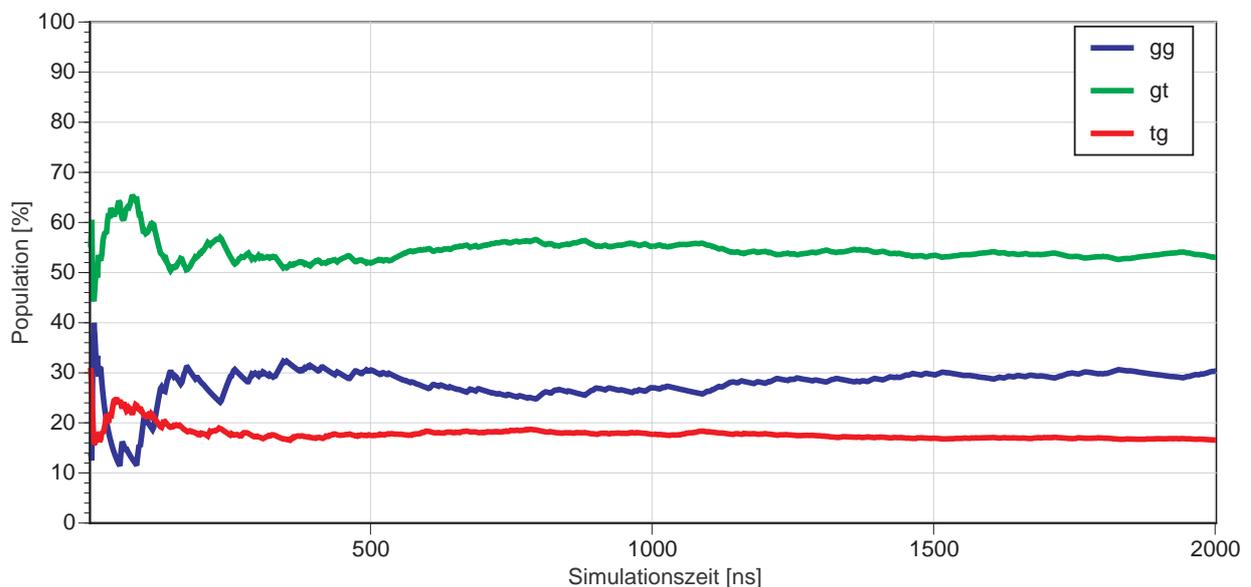
**Abb. X74:**  $\omega_1$ (G56O)- und  $\omega_2$ (L56O)-Trajektorie der Langzeitsimulation von  $\beta$ -Maltose (CVFF,  $\epsilon = 4$ , 300 K,  $t_s = 2 \mu\text{s}$ ,  $\Delta t_s = 20 \text{ ps}$ )



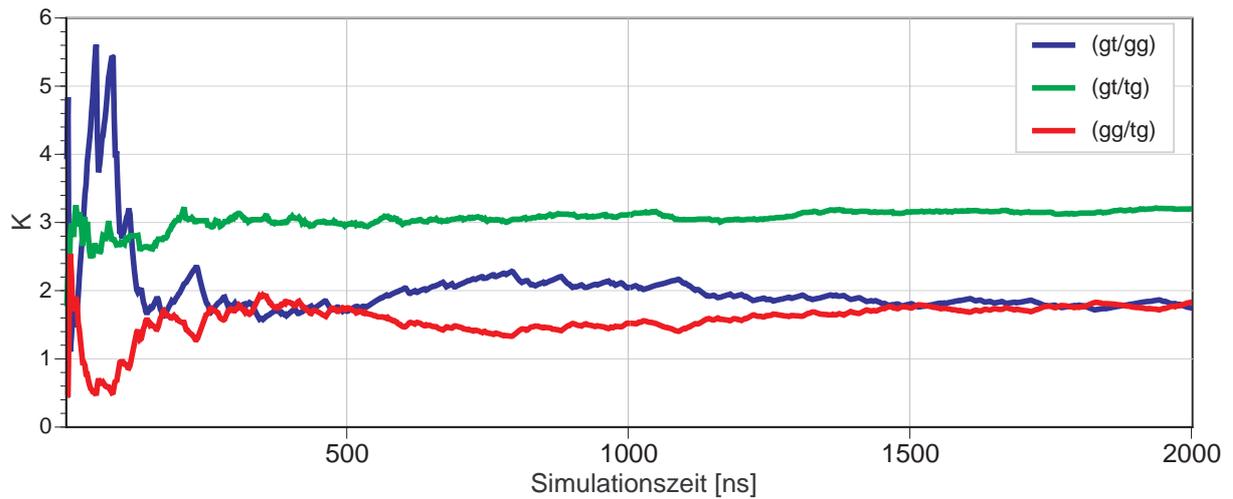
**Abb. X75:** Häufigkeitsverteilung des Torsionswinkels einer Hydroxymethylgruppe und Einteilung in Konformationen („gg“ [-115,-5], „gt“ [5,115], „tg“ [125,235]) ( $\beta$ -Maltose, L56O, CVFF,  $\epsilon = 4$ , 300 K,  $t_s = 2 \mu\text{s}$ ,  $\Delta t_s = 20 \text{ ps}$ )



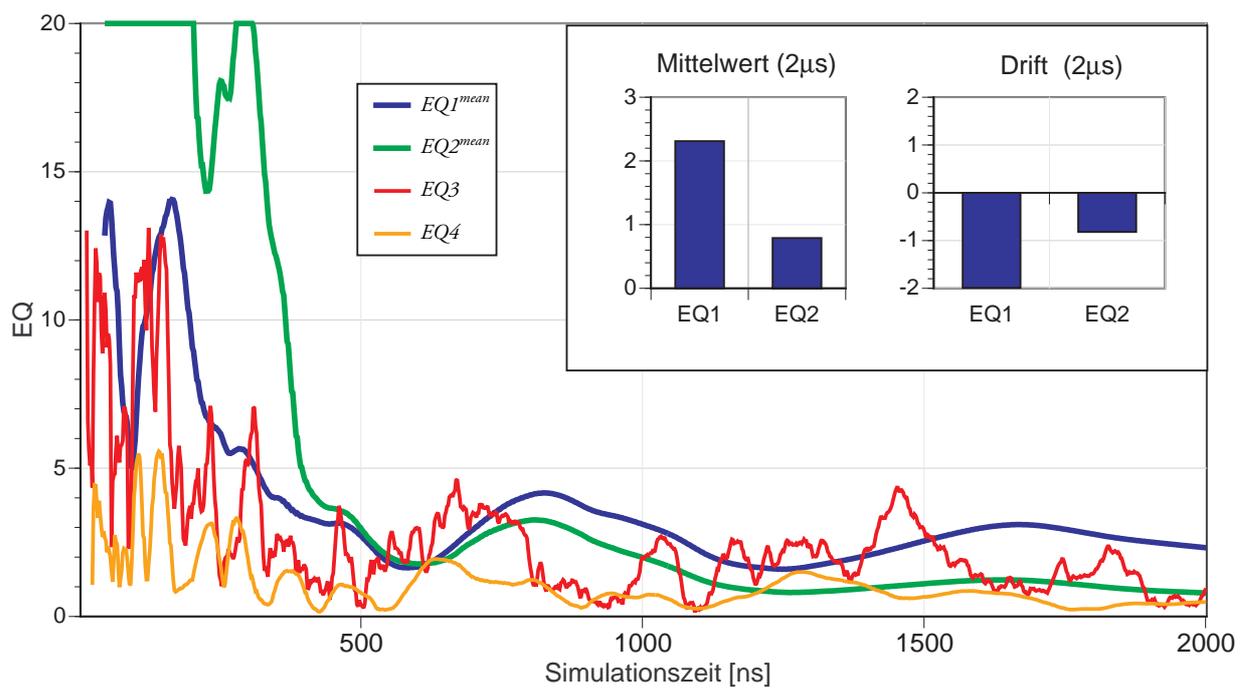
**Abb. X76:** Vergleich des Rotationsprofils für die Hydroxymethylgruppen berechnet aus MD-Daten und aus Grid-Search-Daten ( $\beta$ -Maltose, CVFF,  $\epsilon = 4$ , 300K,  $t_s = 2 \mu\text{s}$ ,  $\Delta t_s = 20 \text{ ps}$ ) (FGECS = Freie Gridgruppen-Energie aus einem Grid Search (vgl. Abschnitt 4.3.2); FGEMD = Freie Gridgruppen-Energie aus einer MD-Simulation)



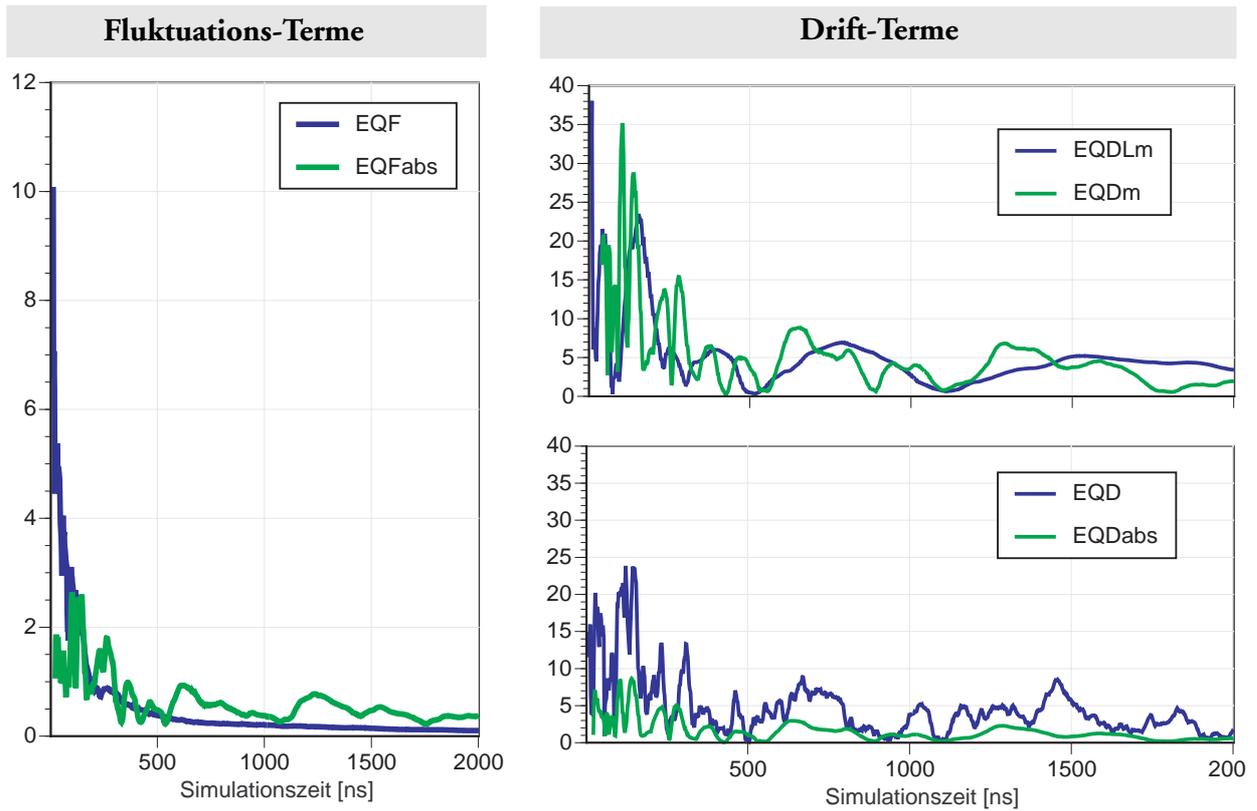
**Abb. X77:** Population der Hydroxymethyl-Rotameren „gg“ [-115,-5], „gt“ [5,115] und „tg“ [125,235] in Abhängigkeit von der Simulationszeit ( $\beta$ -Maltose, Torsionswinkel L56O, CVFF,  $\epsilon = 4$ , 300 K,  $t_s = 2 \mu\text{s}$ ,  $\Delta t_s = 20 \text{ ps}$ )



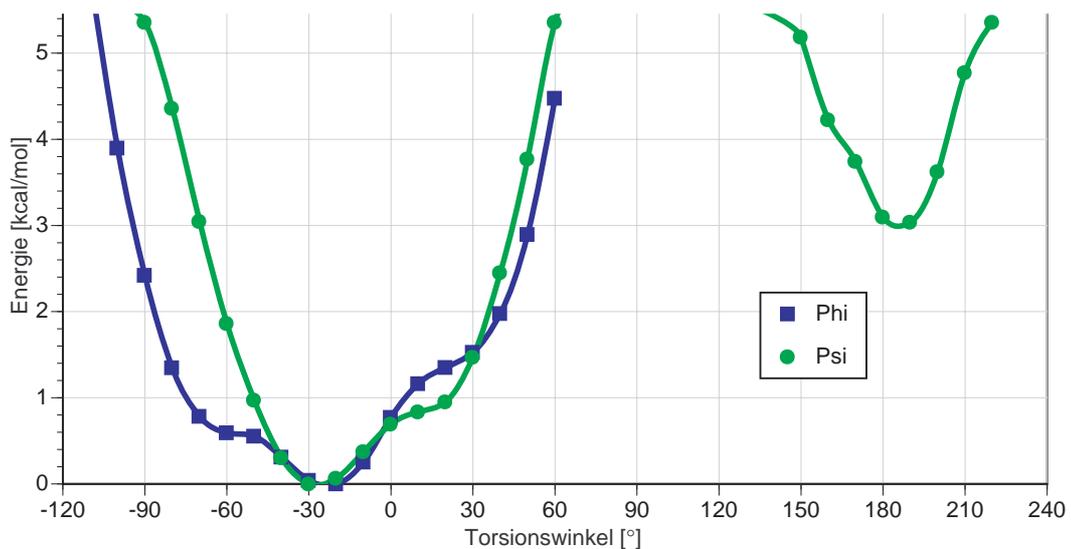
**Abb. X78:** Änderung der Gleichgewichtskonstanten  $K$  eines Rotamerengleichgewichts in Abhängigkeit von der Simulationszeit ( $\beta$ -Maltose, Torsionswinkel L56O, CVFF,  $\varepsilon = 4$ , 300 K,  $t_s = 2 \mu\text{s}$ ,  $\Delta t_s = 20 \text{ ps}$ )



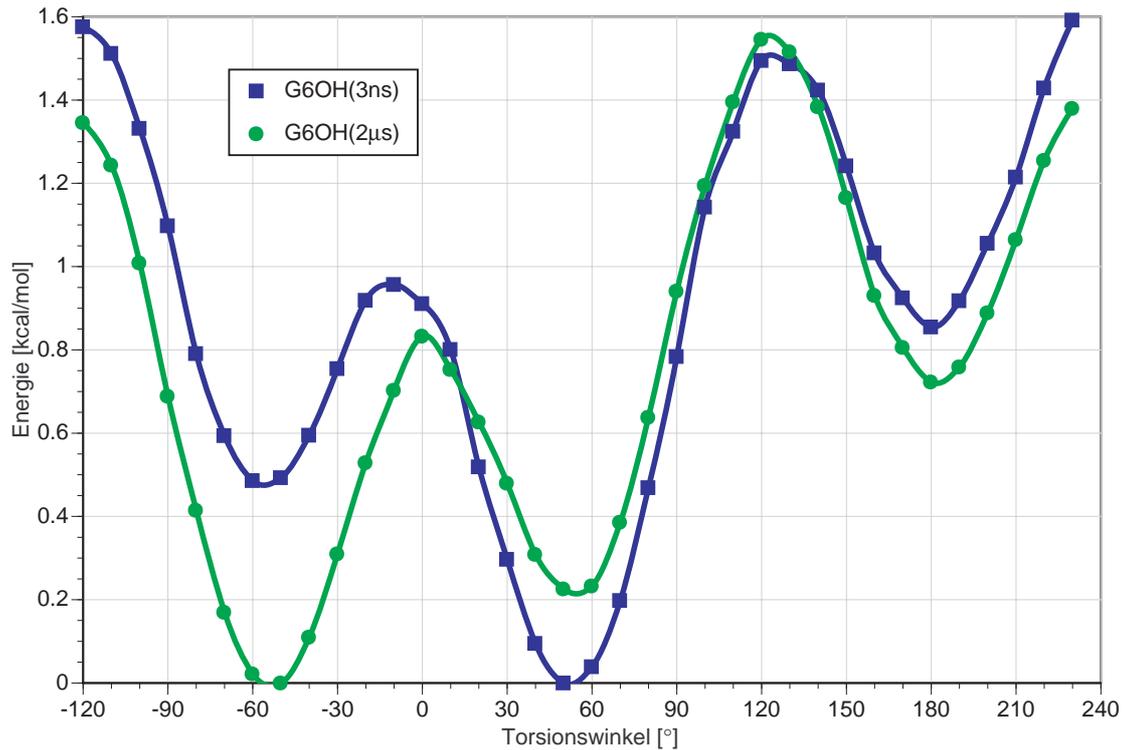
**Abb. X79:** Änderung der Gleichgewichtsparameter EQ eines Rotamerengleichgewichts in Abhängigkeit von der Simulationszeit ( $\beta$ -Maltose, Torsionswinkel L56O, CVFF,  $\varepsilon = 4$ , 300 K,  $t_s = 2 \mu\text{s}$ )



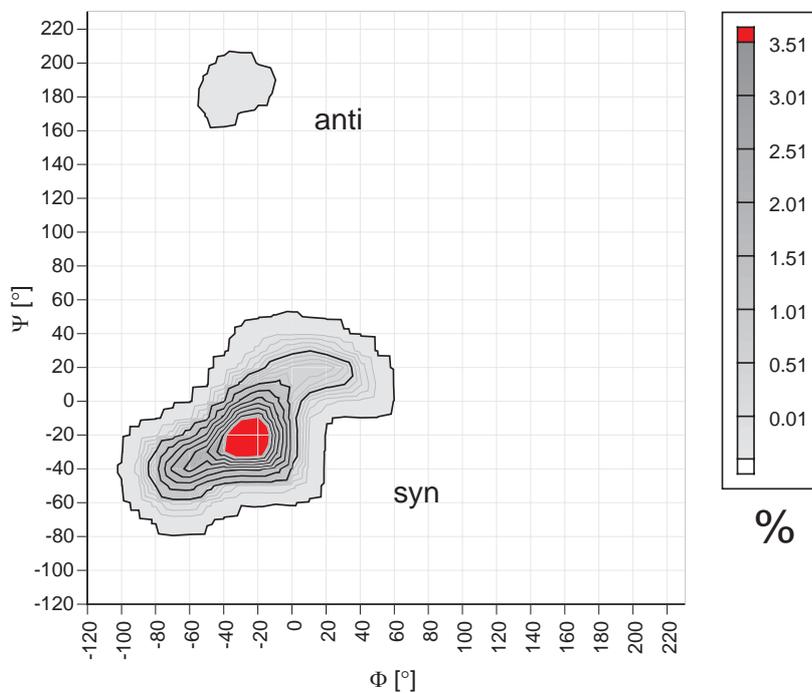
**Abb. X80:** Fluktuations- und Drift-Terme eines Rotamerengleichgewichts in Abhängigkeit von der Simulationszeit ( $\beta$ -Maltose, Torsionswinkel L56O, CVFF,  $\epsilon = 4$ , 300 K,  $t_s = 2 \mu\text{s}$ )



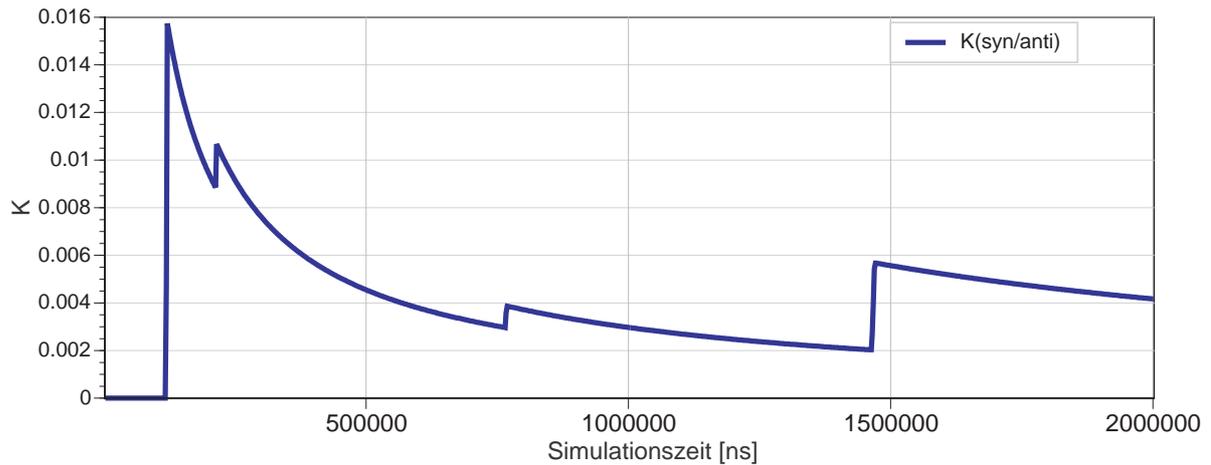
**Abb. X81:** Freie Gridgruppen-Energie für  $\Phi$  und  $\Psi$  (MD-Simulation von  $\beta$ -Maltose: CVFF,  $\epsilon = 4$ , 300 K,  $t_s = 2 \mu\text{s}$ ,  $\Delta t_s = 20 \text{ ps}$ )



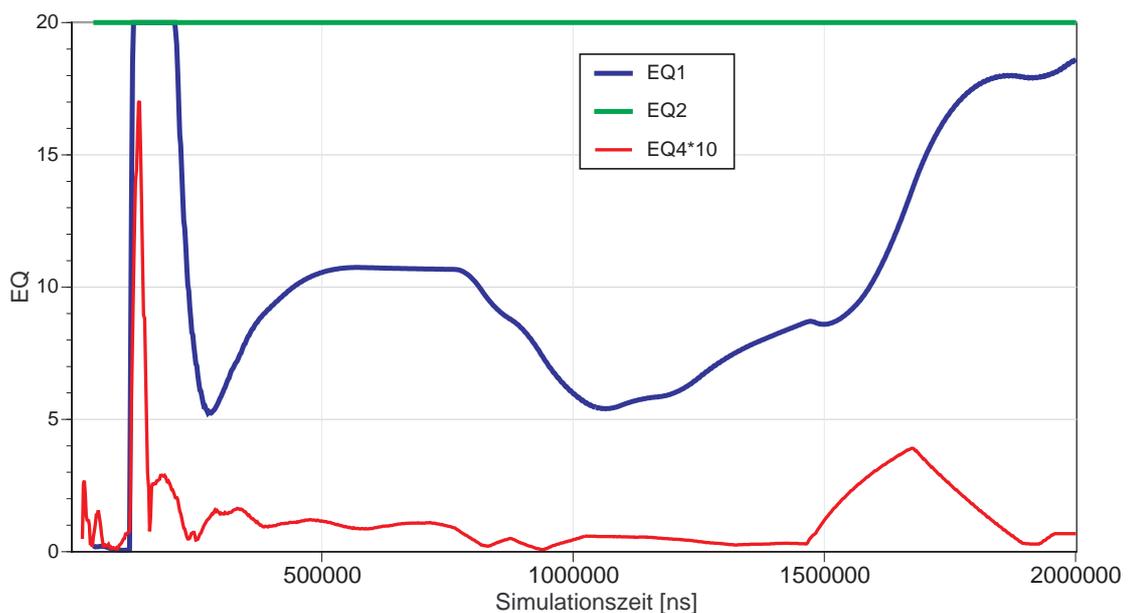
**Abb. X82:** Vergleich des Rotationsprofils der Freien Gridgruppen-Energie für die primäre OH-Gruppe G6OH mit (2  $\mu$ s) und ohne (3 ns) eingestelltem Hydroxymethylgleichgewicht ( $\beta$ -Maltose, CVFF,  $\epsilon = 4$ , 300 K)



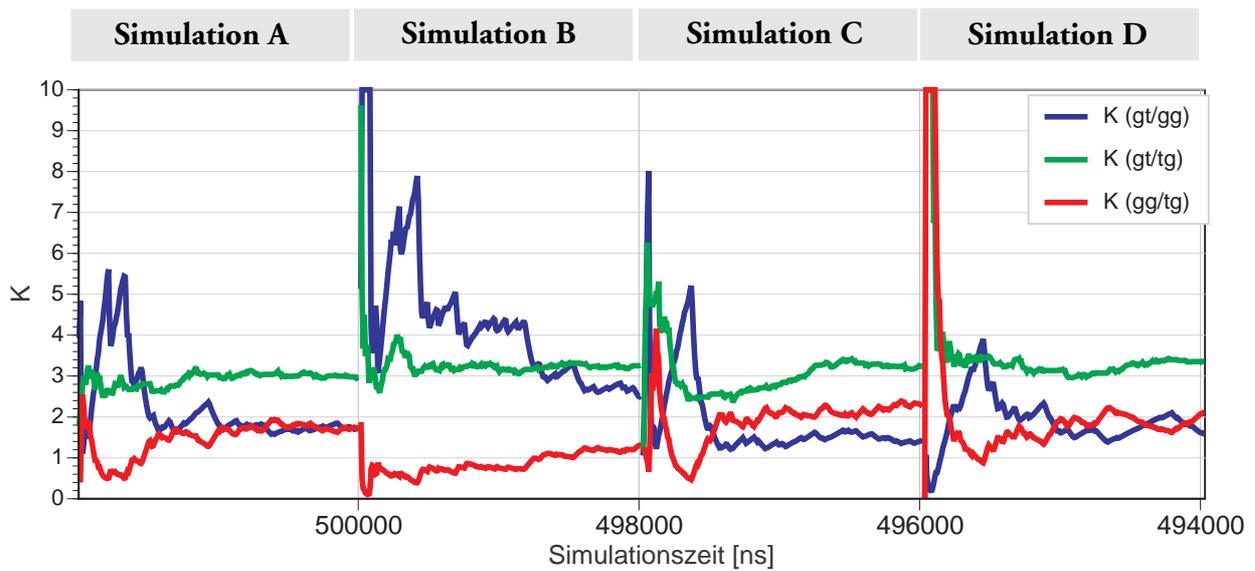
**Abb. X83:**  $\Phi / \Psi$  -Häufigkeitsverteilung (MD-Simulation von  $\beta$ -Maltose: CVFF,  $\epsilon = 4$ , 300 K,  $t_s = 2 \mu$ s,  $\Delta t_s = 20$  ps)



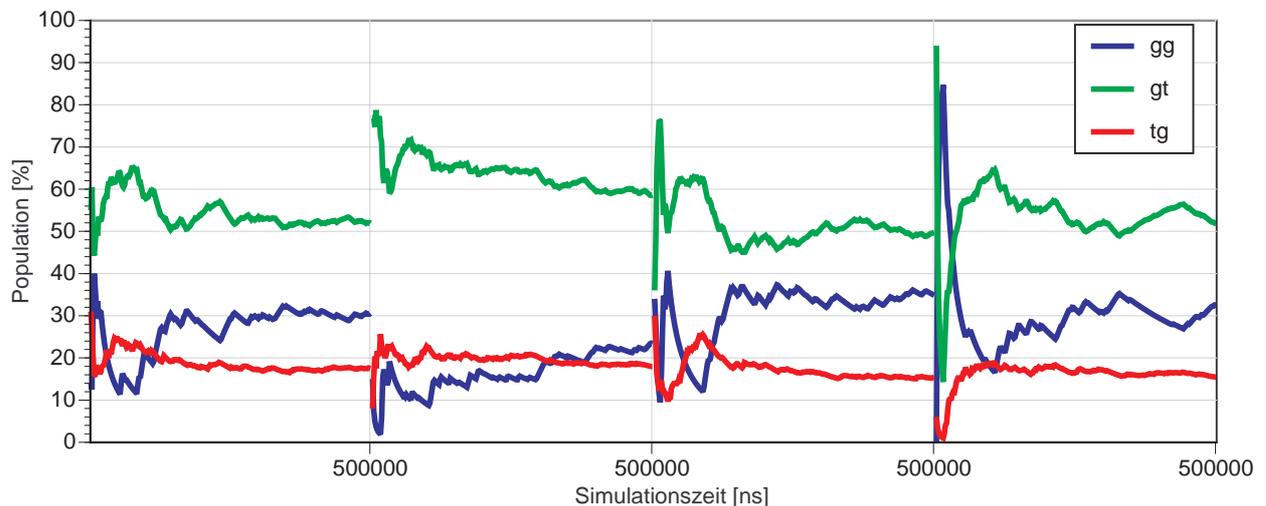
**Abb. X84:** Änderung der Gleichgewichtskonstanten des syn/anti-Gleichgewichts in Abhängigkeit von der Simulationszeit (Maltose, Torsionswinkel  $\Phi / \Psi$ , CVFF,  $\varepsilon = 4$ , 300 K,  $t_s = 2 \mu\text{s}$ ,  $\Delta t_s = 20 \text{ ps}$ )



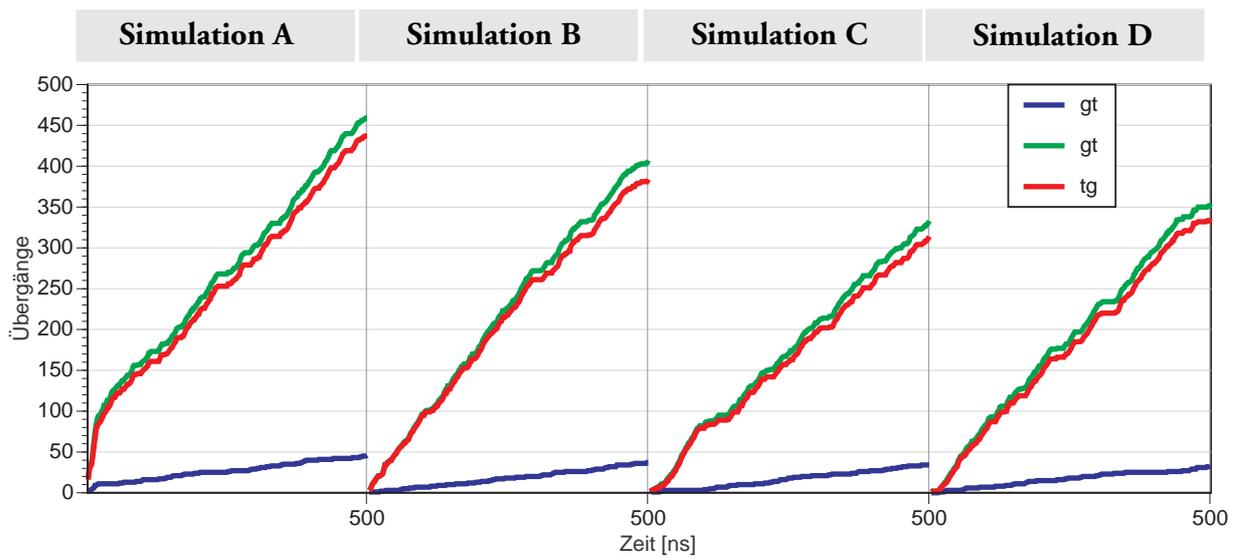
**Abb. X85:** Änderung der Gleichgewichtsparameter EQ eines syn/anti-Gleichgewichts in Abhängigkeit von der Simulationszeit (Maltose, Torsionswinkel  $\Phi / \Psi$ , CVFF,  $\varepsilon = 4$ , 300 K,  $t_s = 2 \mu\text{s}$ ,  $\Delta t_s = 20 \text{ ps}$ ). EQ2: alle Werte  $> 20$  wurden gleich 20 gesetzt.



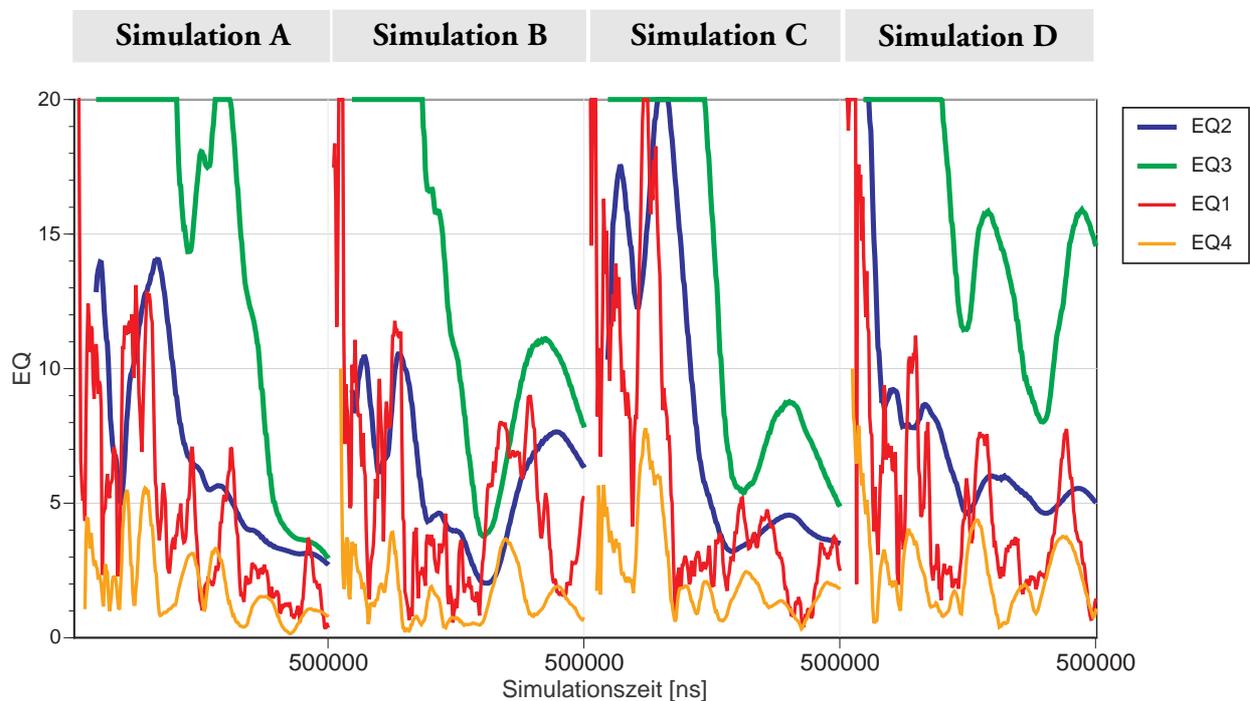
**Abb. X86:** Änderung der Gleichgewichtskonstanten  $K$  eines Rotamerengleichgewichts mit der Simulationszeit für vier Simulationen à 500 ns bei 300K (Maltose, Torsionswinkel L56O, CVFF,  $\epsilon = 4$ )



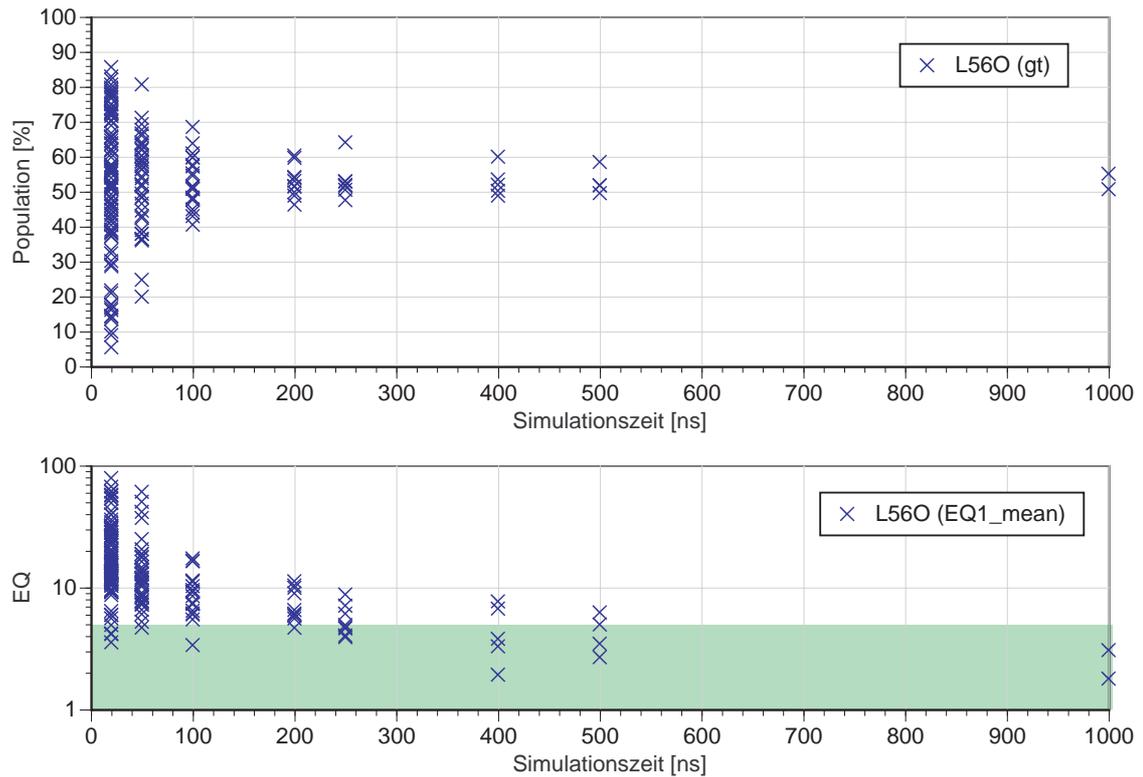
**Abb. X87:** Population der Rotameren („gg“ [-115,-5], „gt“ [5,115], „tg“ [125,235]) in Abhängigkeit von der Simulationszeit für vier Simulationen à 500 ns bei 300K (Maltose, Torsionswinkel L56O, CVFF,  $\epsilon = 4$ )



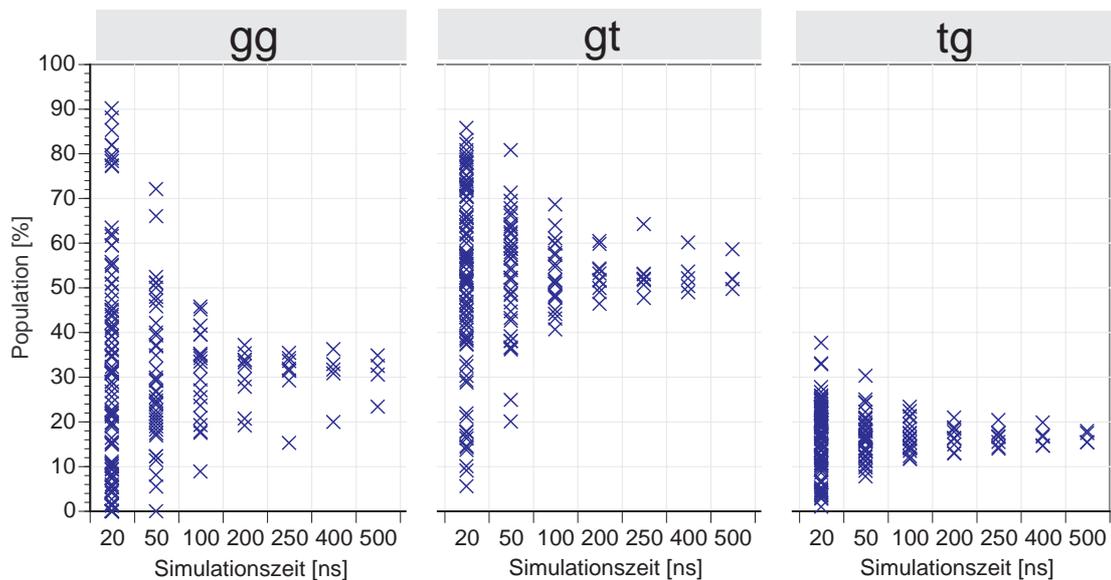
**Abb. X88:** Zahl der Konformationsübergänge in Abhängigkeit von der Simulationszeit (gezählt wurden die Übergänge aus einer Konformation heraus)  
MD Simulation von Maltose, Torsionswinkel L56O, CVFF,  $\epsilon = 4$ , 300K, jeweils 500 ns



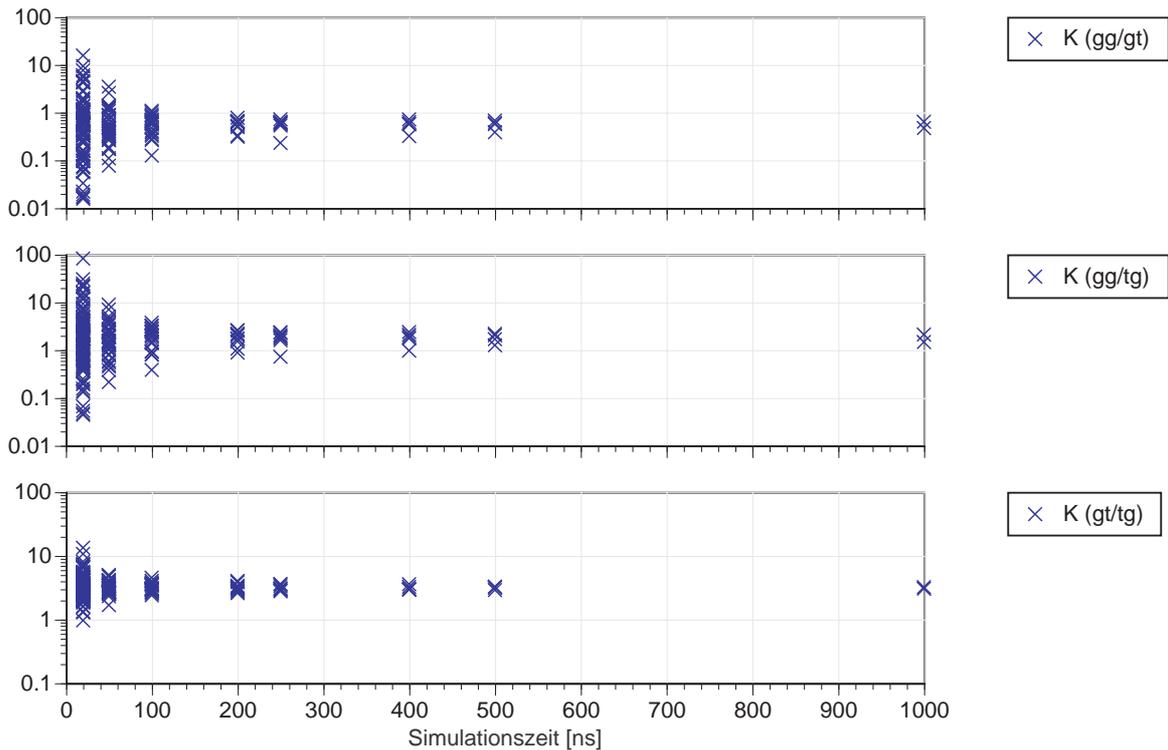
**Abb. X89:** Änderung der Gleichgewichtsparameter EQ in Abhängigkeit von der Simulationszeit (MD Simulation von Maltose, Torsionswinkel L56O, CVFF,  $\epsilon = 4$ , 300 K,  $t_s = 4 \times 500$  ns,  $\Delta t_s = 20$  ps)



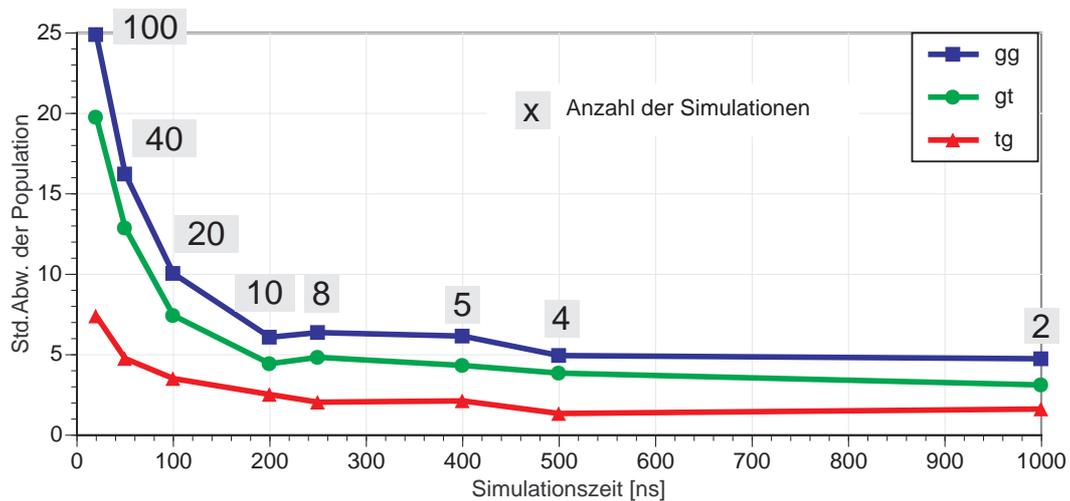
**Abb. X90:** Reproduzierbarkeit der Rotamerenpopulation in Abhängigkeit von der Simulationszeit Gleichgewichtsparameter EQ1 (Maltose, Torsionswinkel L56O, CVFF,  $\varepsilon = 4$ , 300K)



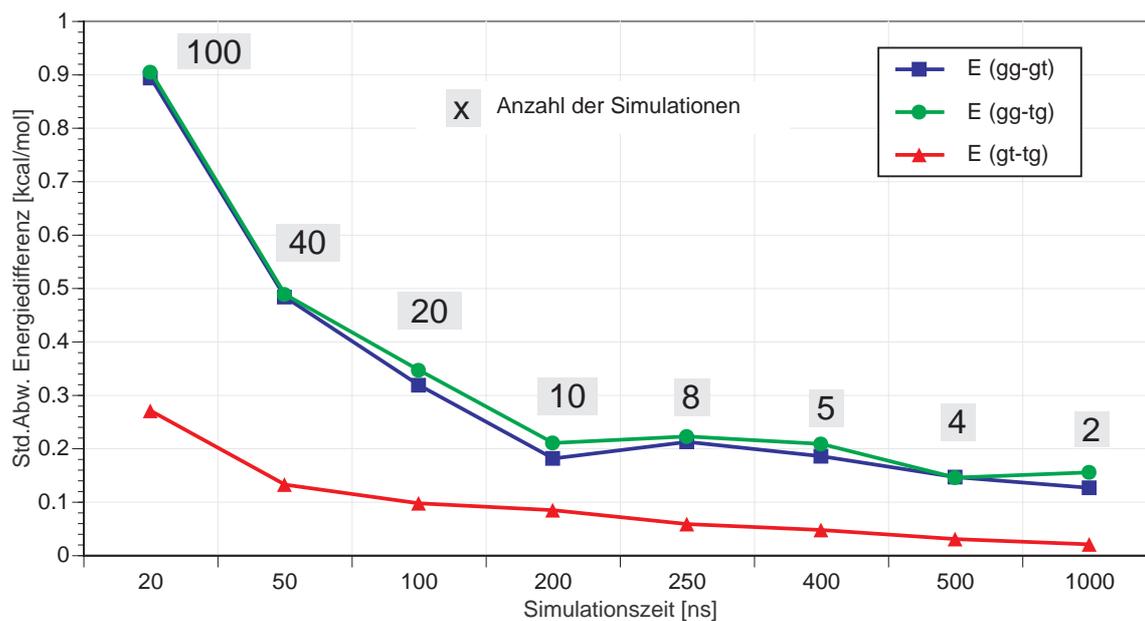
**Abb. X91:** Reproduzierbarkeit der Rotamerenpopulation in Abhängigkeit von der Simulationszeit (Maltose, Torsionswinkel L56O, CVFF,  $\varepsilon = 4$ , 300K)



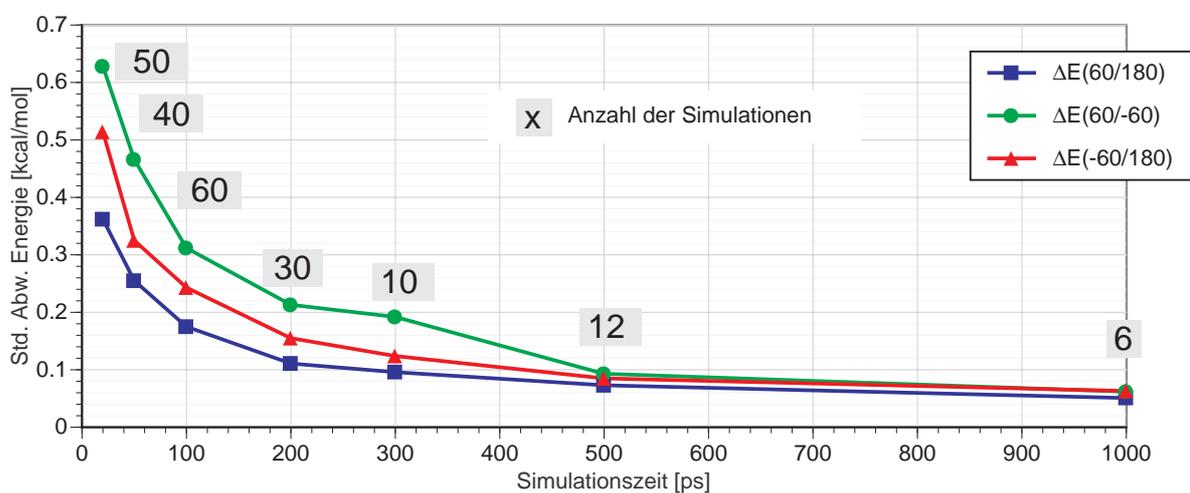
**Abb. X92:** Reproduzierbarkeit der Gleichgewichtskonstanten in Abhängigkeit von der Simulationszeit (Maltose, Torsionswinkel L56O, CVFF,  $\varepsilon = 4$ , 300K)



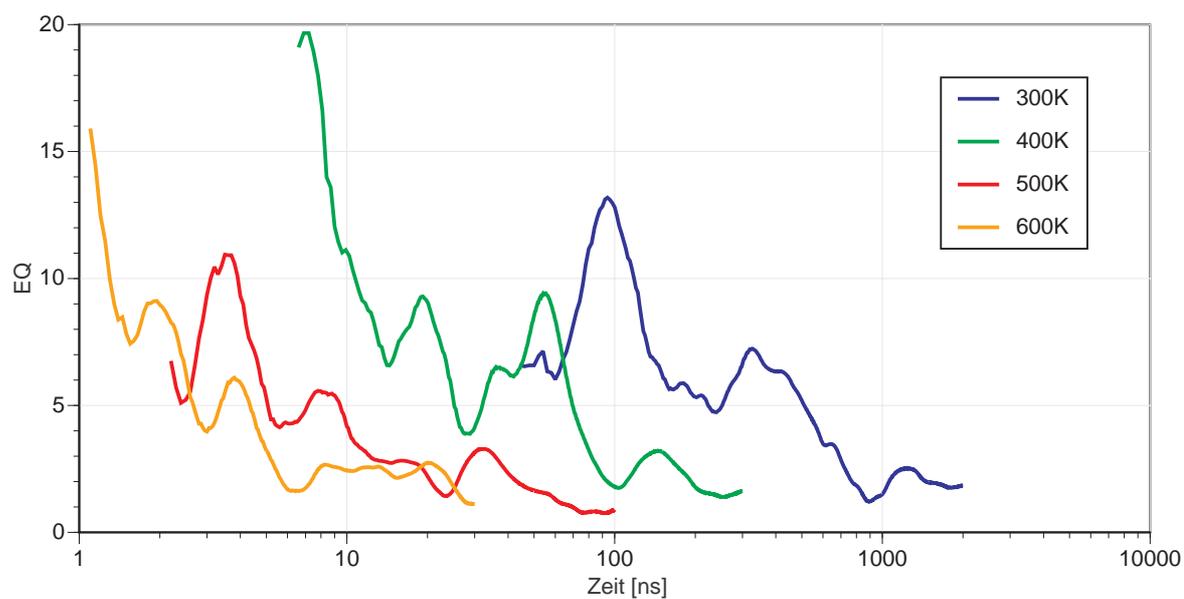
**Abb. X93:** Standardabweichung der Rotamerenpopulation in Abhängigkeit von der Simulationszeit (Maltose, Torsionswinkel L56O, CVFF,  $\varepsilon = 4$ , 300 K)



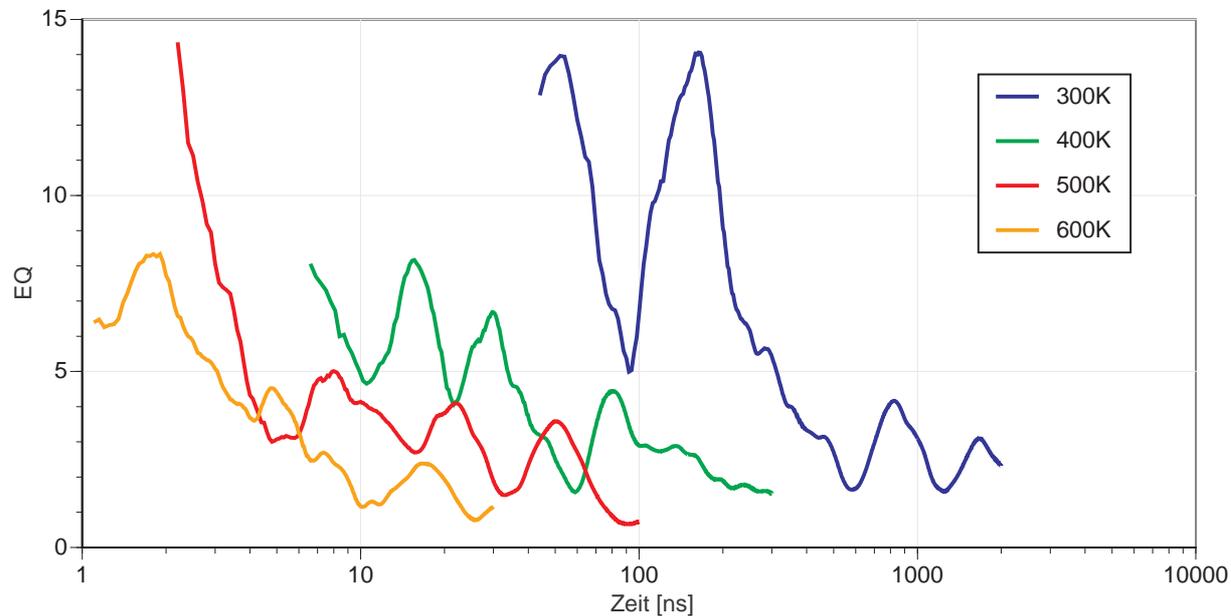
**Abb. X94:** Standardabweichung der Energiedifferenz berechnet aus den Gleichgewichtskonstanten in Abhängigkeit von der Simulationszeit (Maltose, Torsionswinkel L56O, CVFF,  $\epsilon = 4$ , 300 K)



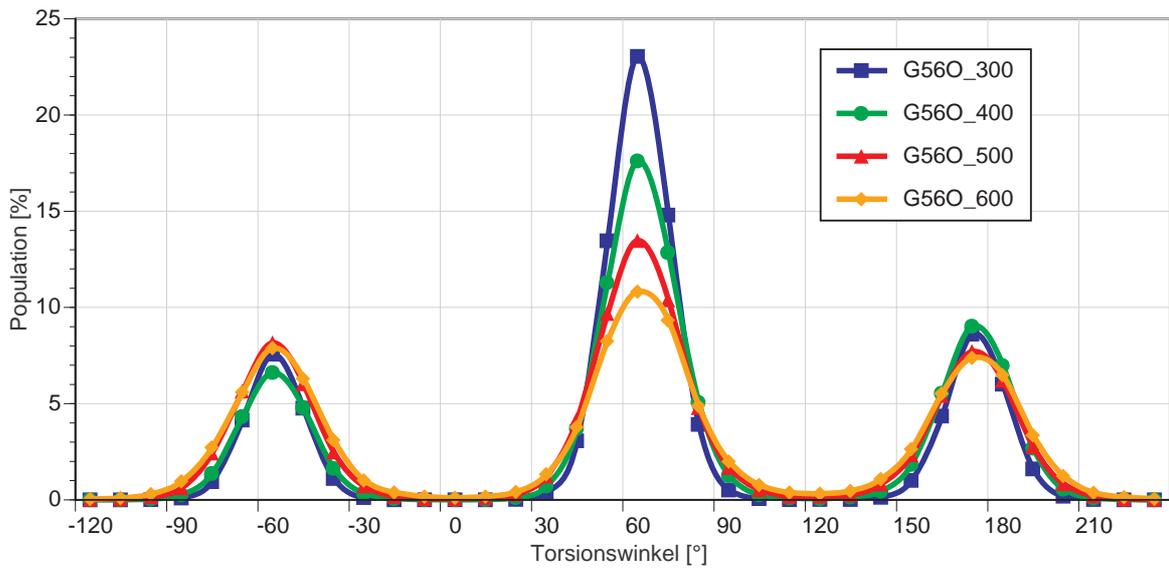
**Abb. X95:** Standardabweichung der Energiedifferenz berechnet aus den Gleichgewichtskonstanten in Abhängigkeit von der Simulationszeit (Maltose, Torsionswinkel G3OH, CVFF,  $\epsilon = 4$ , 300 K)



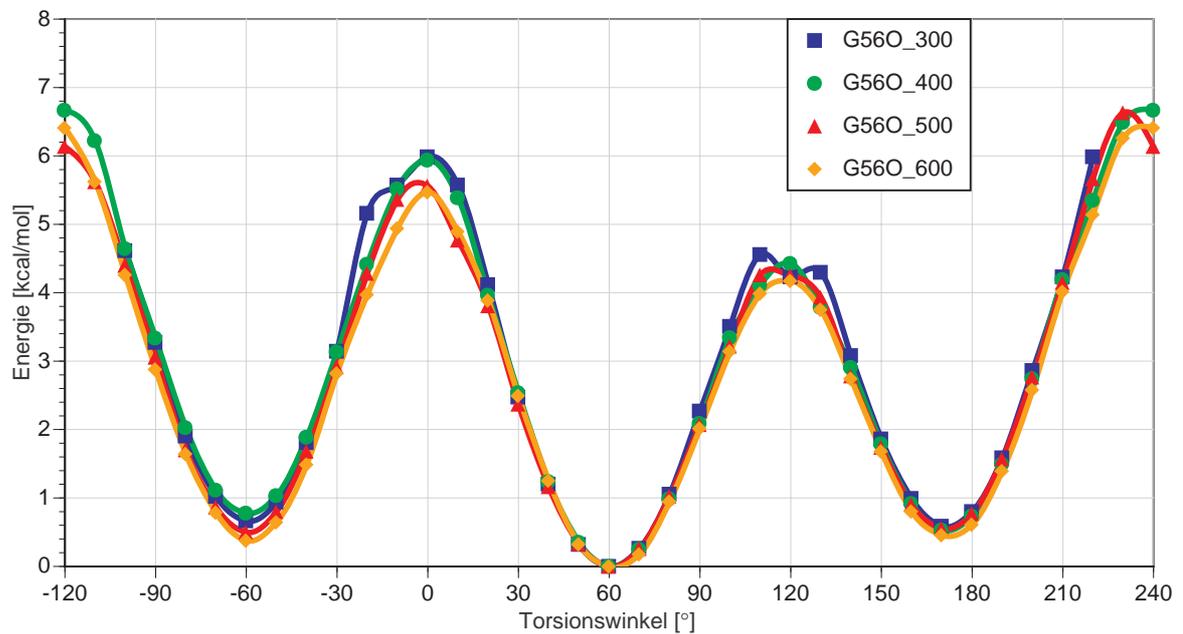
**Abb. X96:** Abhängigkeit der benötigten Simulationszeit zur Einstellung des Rotamerengleichgewichts für die Hydroxymethylgruppen von der Simulationstemperatur (Maltose, G56O, CVFF,  $\epsilon = 4$ )



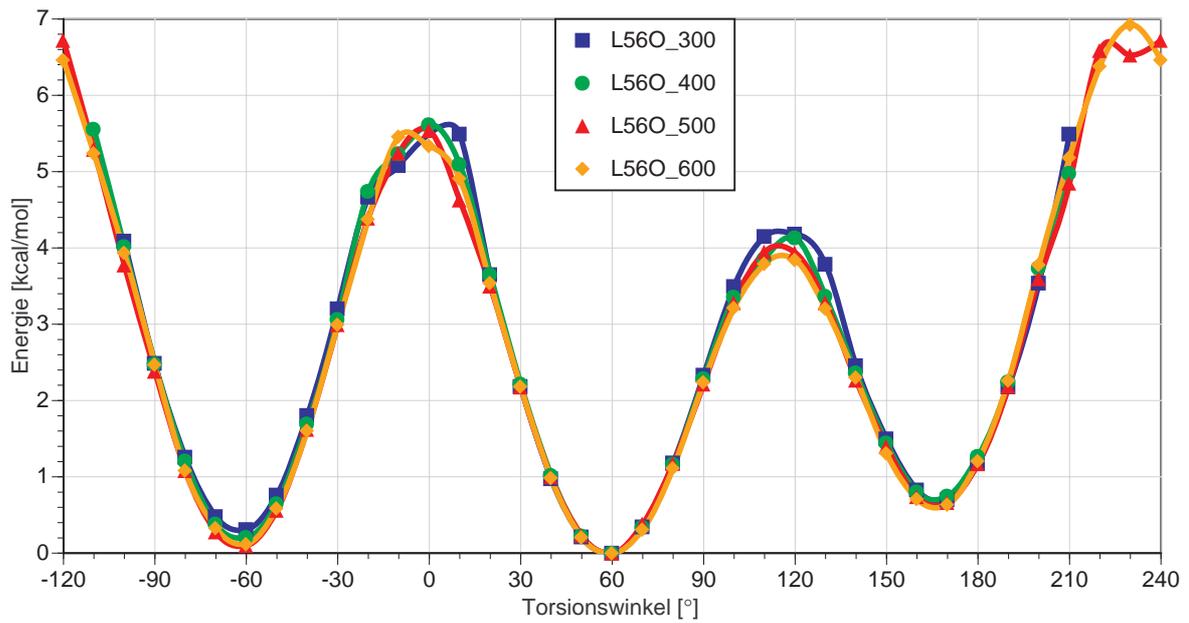
**Abb. X97:** Abhängigkeit der benötigten Simulationszeit zur Einstellung des Rotamerengleichgewichts für die Hydroxymethylgruppen von der Simulationstemperatur (Maltose, L56O, CVFF,  $\epsilon = 4$ )



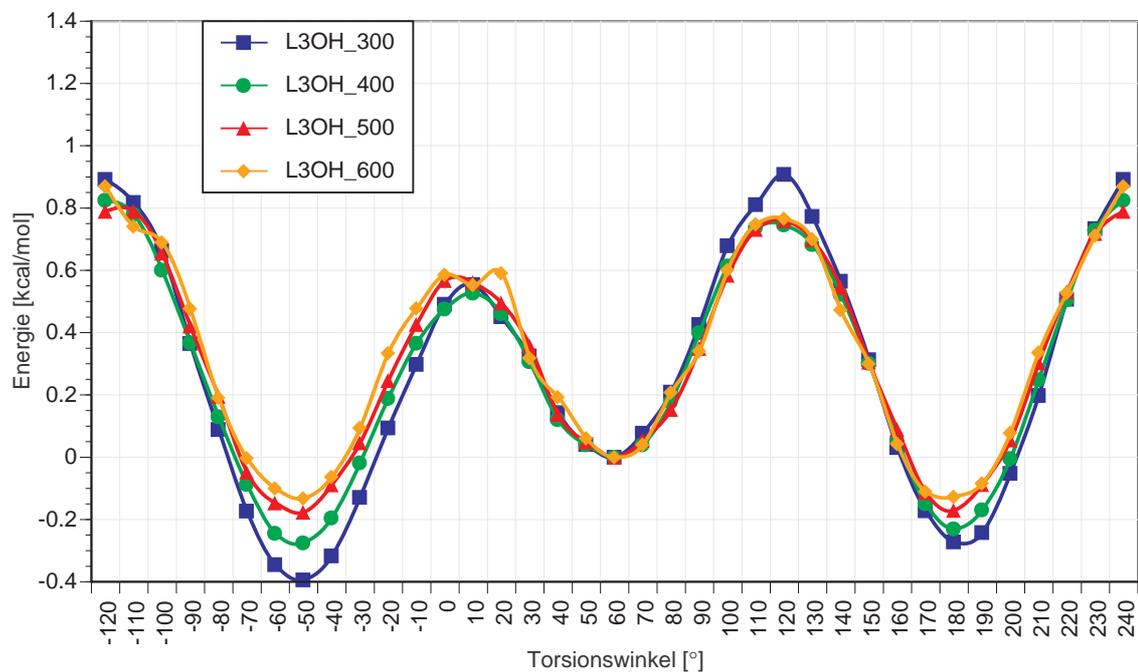
**Abb. X98:** Abhängigkeit der Häufigkeitsverteilung von der Simulationstemperatur (Maltose, G56O, CVFF,  $\epsilon = 4$ )



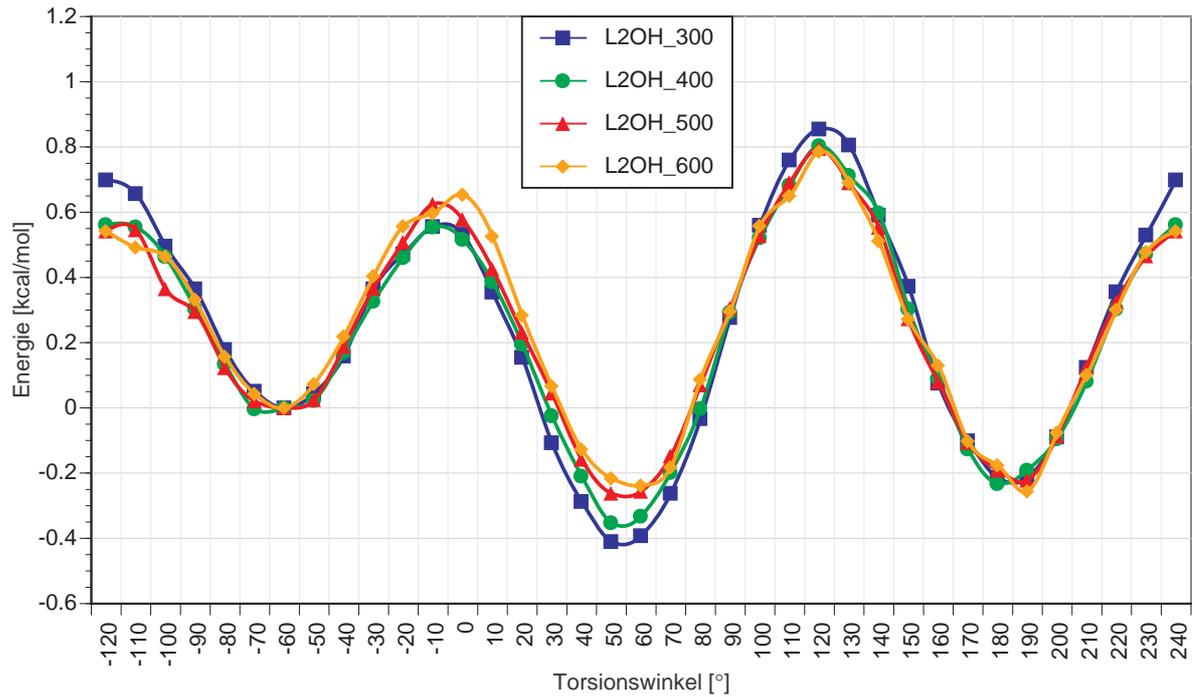
**Abb. X99:** Abhängigkeit des Energieprofils von der Simulationstemperatur (Maltose, G56O, CVFF,  $\epsilon = 4$ )



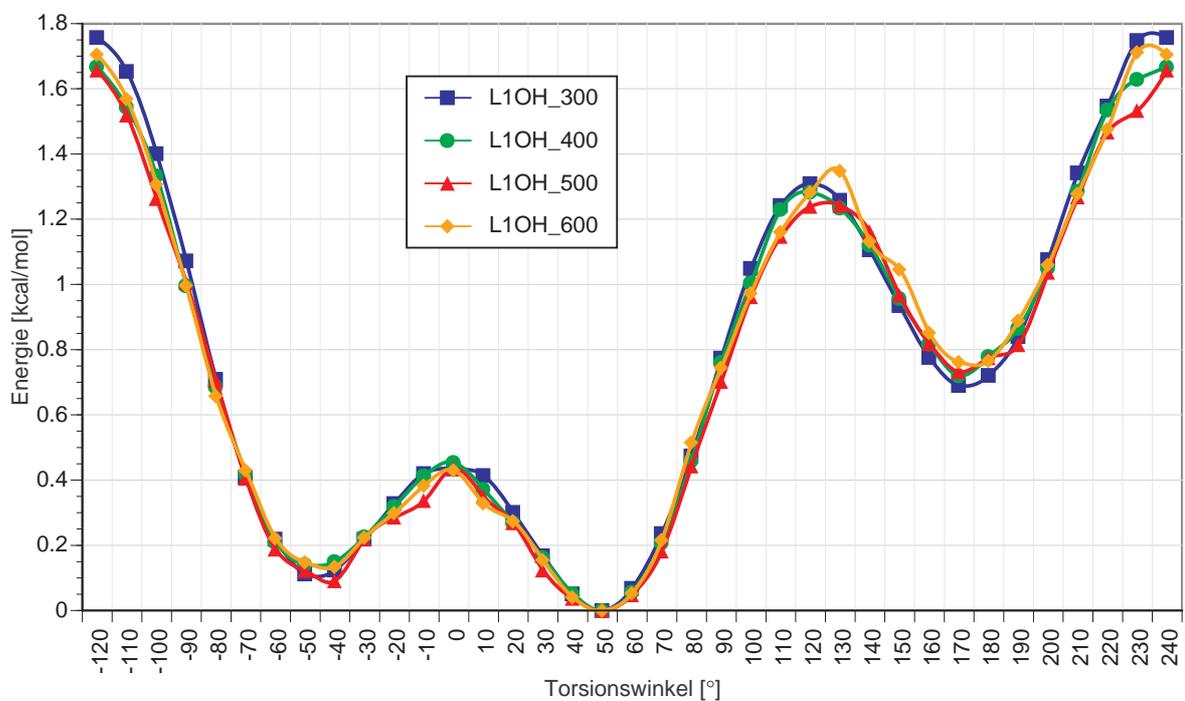
**Abb. X100:** Abhängigkeit des Energieprofils von der Simulationstemperatur (Maltose, L56O, CVFF,  $\epsilon = 4$ )



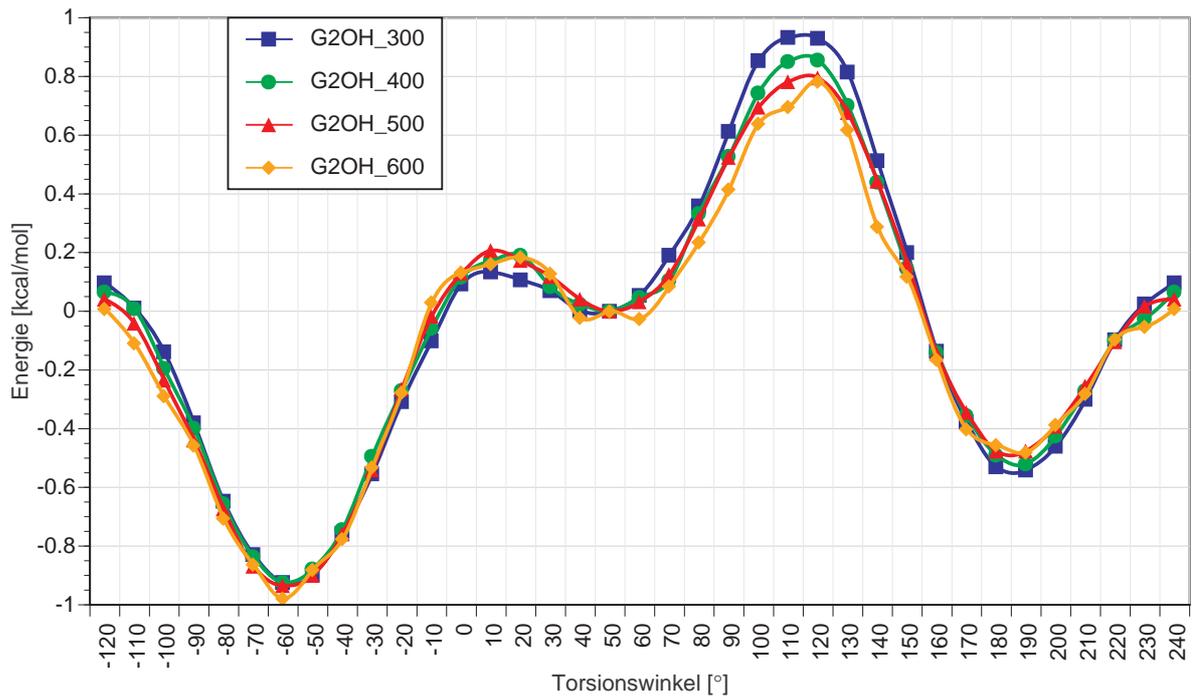
**Abb. X101:** Abhängigkeit des Energieprofils von der Simulationstemperatur (Maltose, L3OH, CVFF,  $\epsilon = 4$ ). Normierung der Energie auf L3OH = +60°.



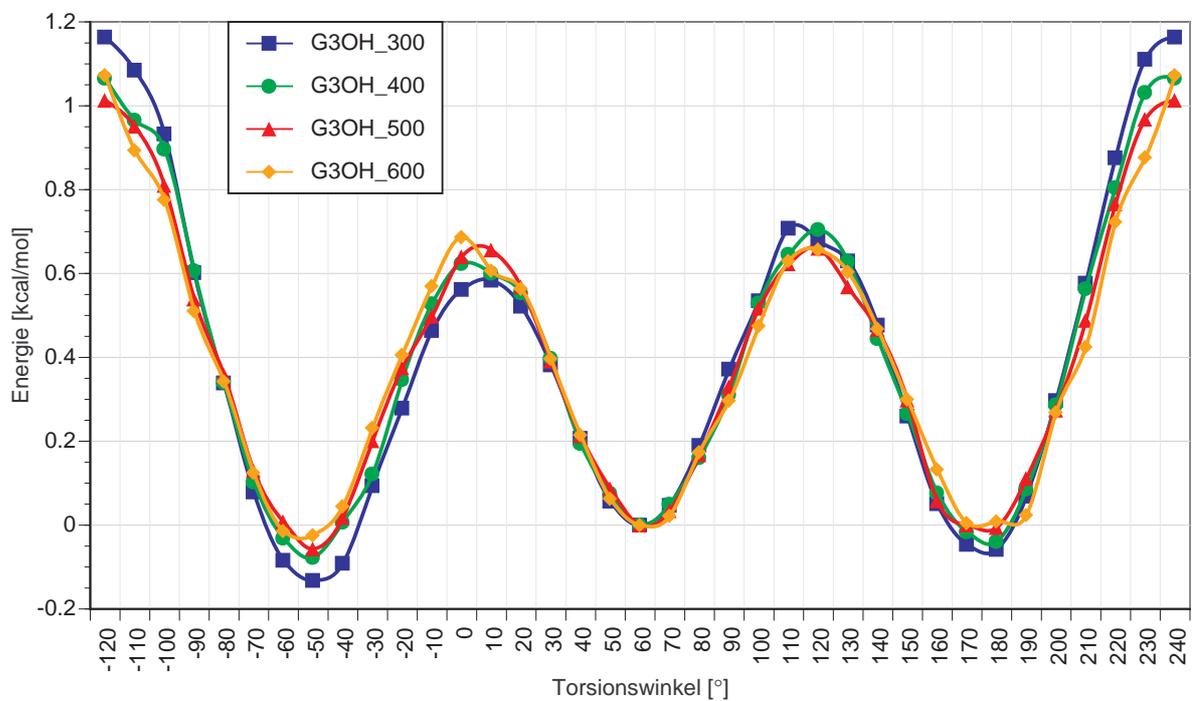
**Abb. X102:** Abhängigkeit des Energieprofils von der Simulationstemperatur (Maltose, L2OH, CVFF,  $\epsilon = 4$ ). Normierung der Energie auf L2OH =  $-60^\circ$ .



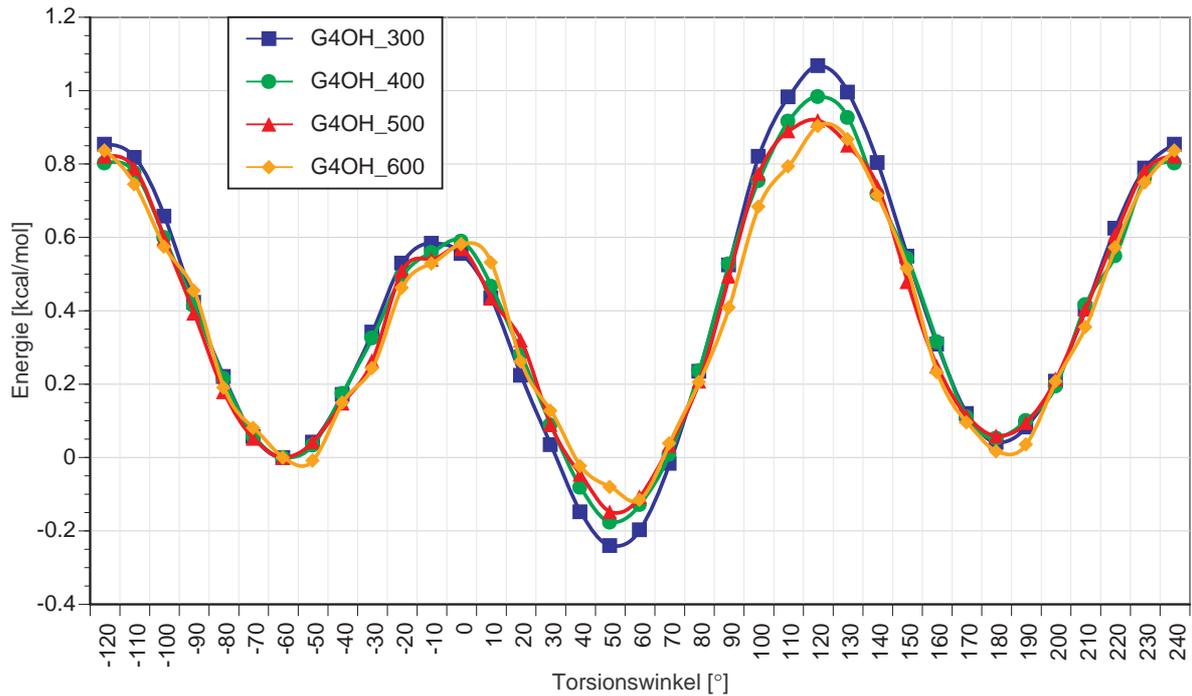
**Abb. X103:** Abhängigkeit des Energieprofils von der Simulationstemperatur ( $\beta$ -Maltose, L1OH, CVFF,  $\epsilon = 4$ )



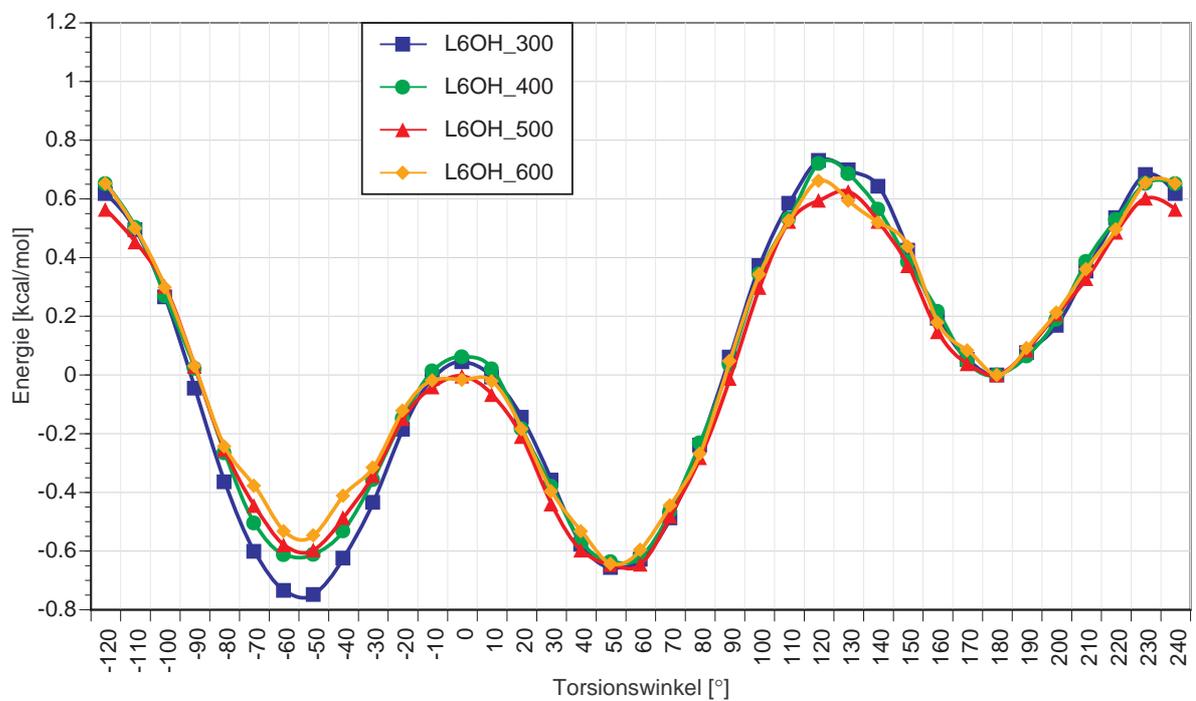
**Abb. X104:** Abhängigkeit des Energieprofils von der Simulationstemperatur (Maltose, G2OH, CVFF,  $\epsilon = 4$ ). Normierung der Energie auf G2OH = +60°.



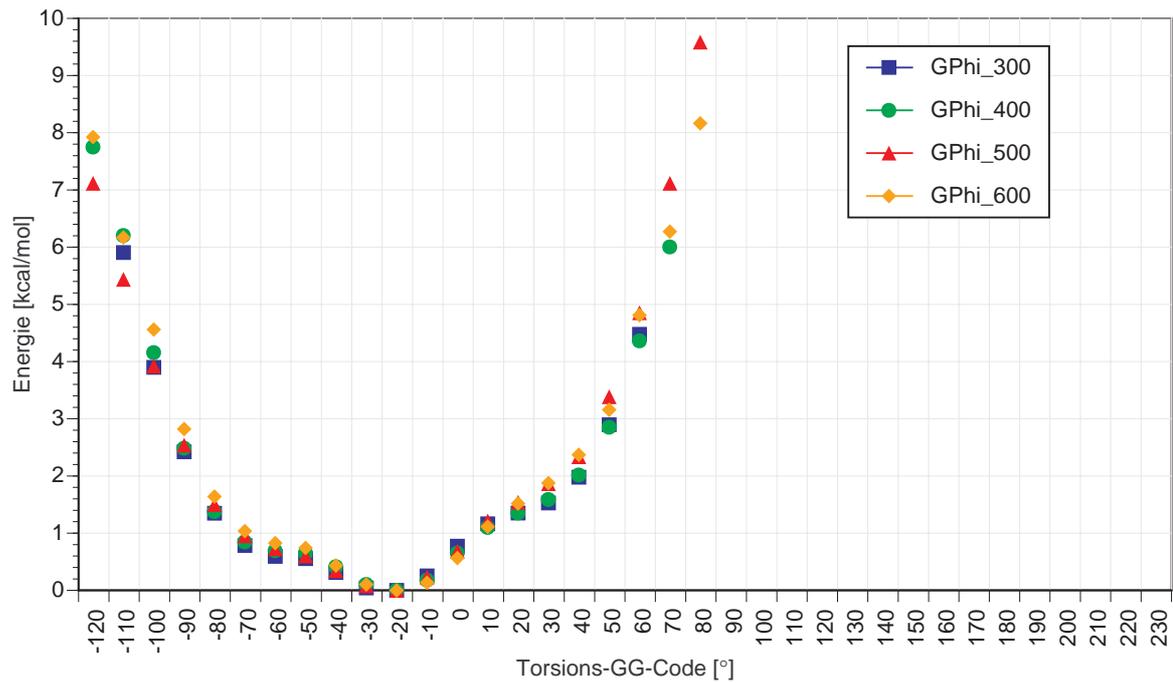
**Abb. X105:** Abhängigkeit des Energieprofils von der Simulationstemperatur (Maltose, G3OH, CVFF,  $\epsilon = 4$ ). Normierung der Energie auf G3OH = +60°.



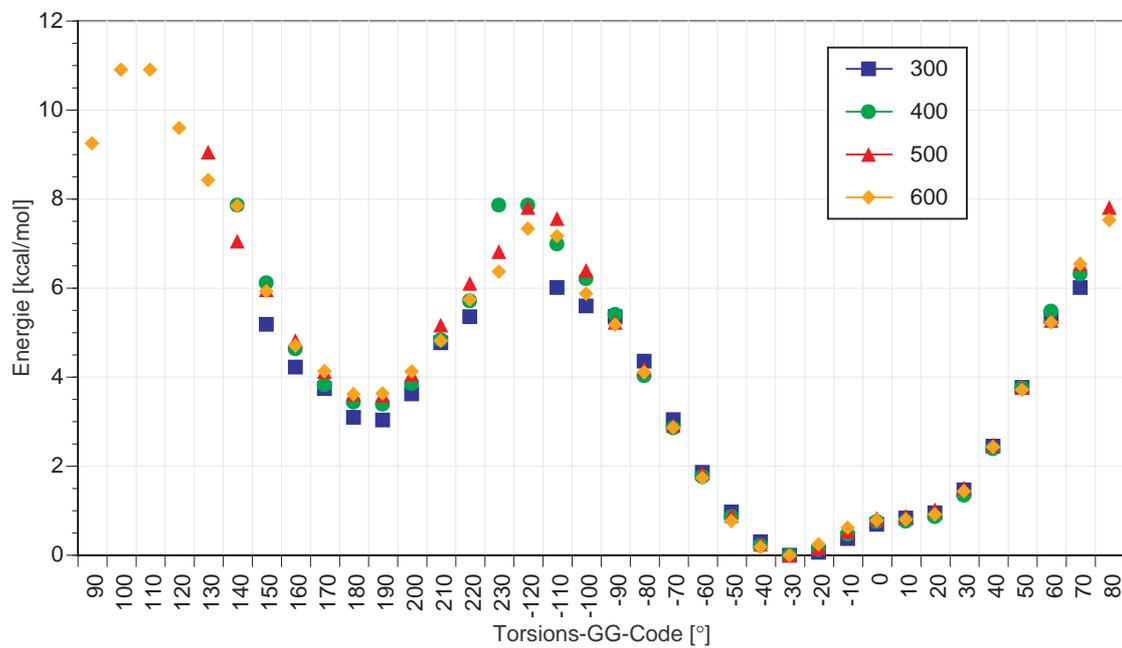
**Abb. X106:** Abhängigkeit des Energieprofils von der Simulationstemperatur (Maltose, G4OH, CVFF,  $\epsilon = 4$ ). Normierung der Energie auf G4OH =  $-60^\circ$ .



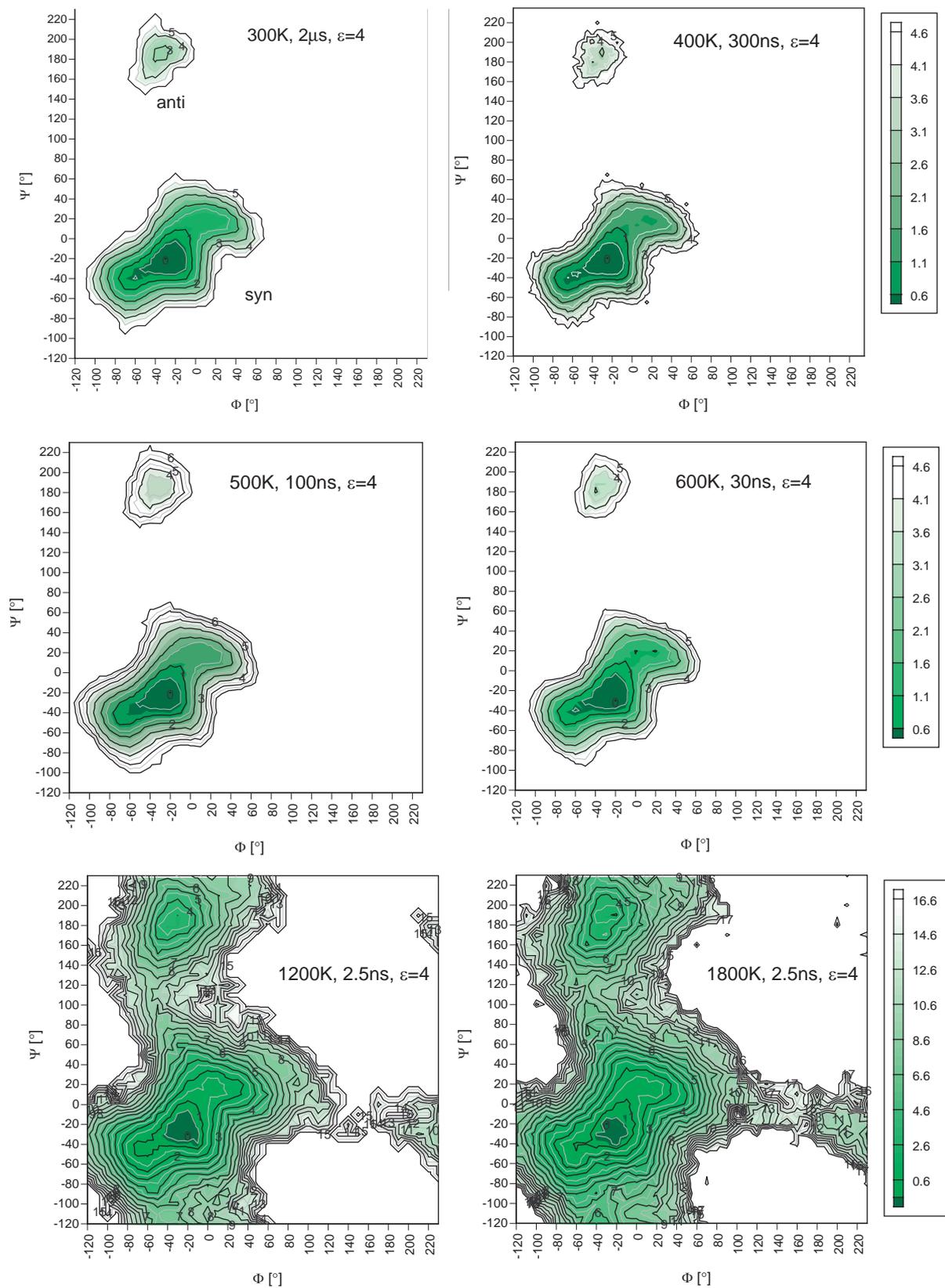
**Abb. X107:** Abhängigkeit des Energieprofils von der Simulationstemperatur (Maltose, L6OH, CVFF,  $\epsilon = 4$ ). Normierung der Energie auf L6OH =  $+180^\circ$ .



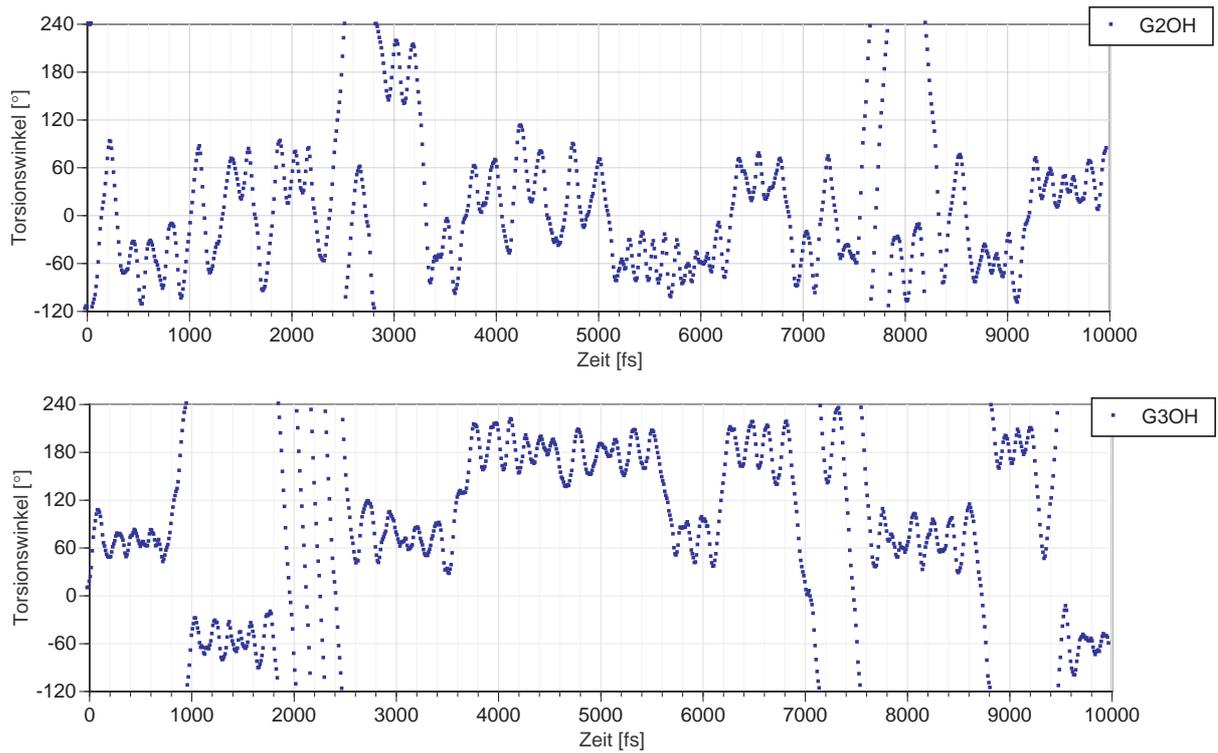
**Abb. X108:** Abhängigkeit des Energieprofils von der Simulationstemperatur (Maltose,  $\Phi$ , CVFF,  $\epsilon = 4$ )



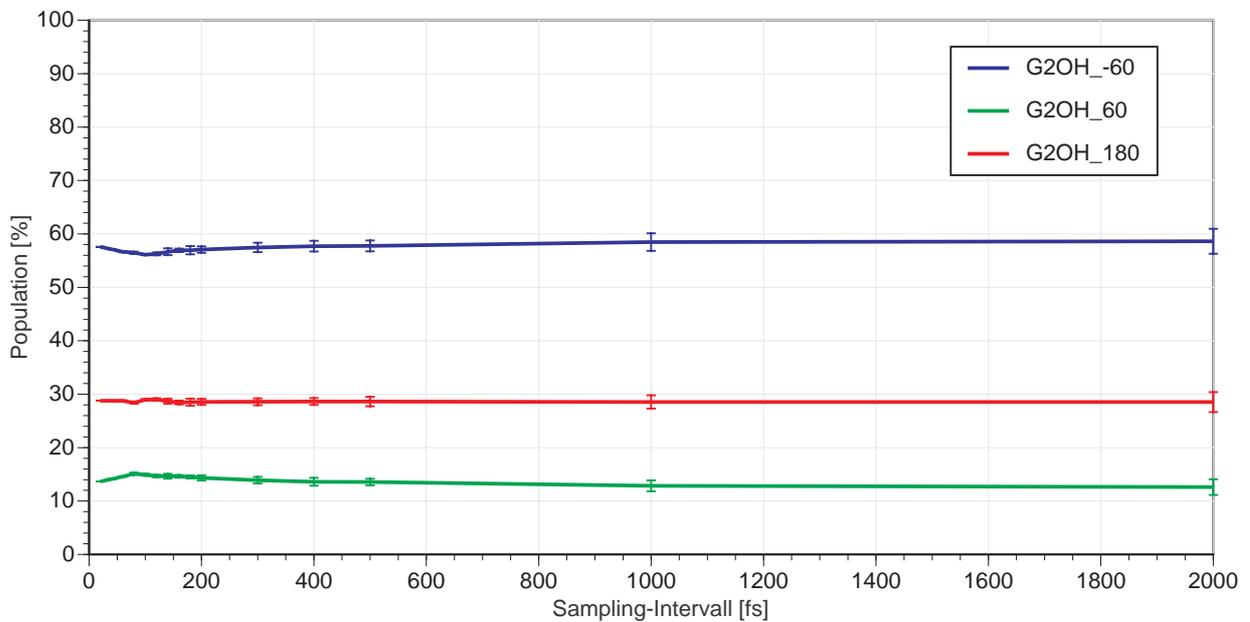
**Abb. X109:** Abhängigkeit des Energieprofils von der Simulationstemperatur (Maltose,  $\Psi$ , CVFF,  $\epsilon = 4$ )



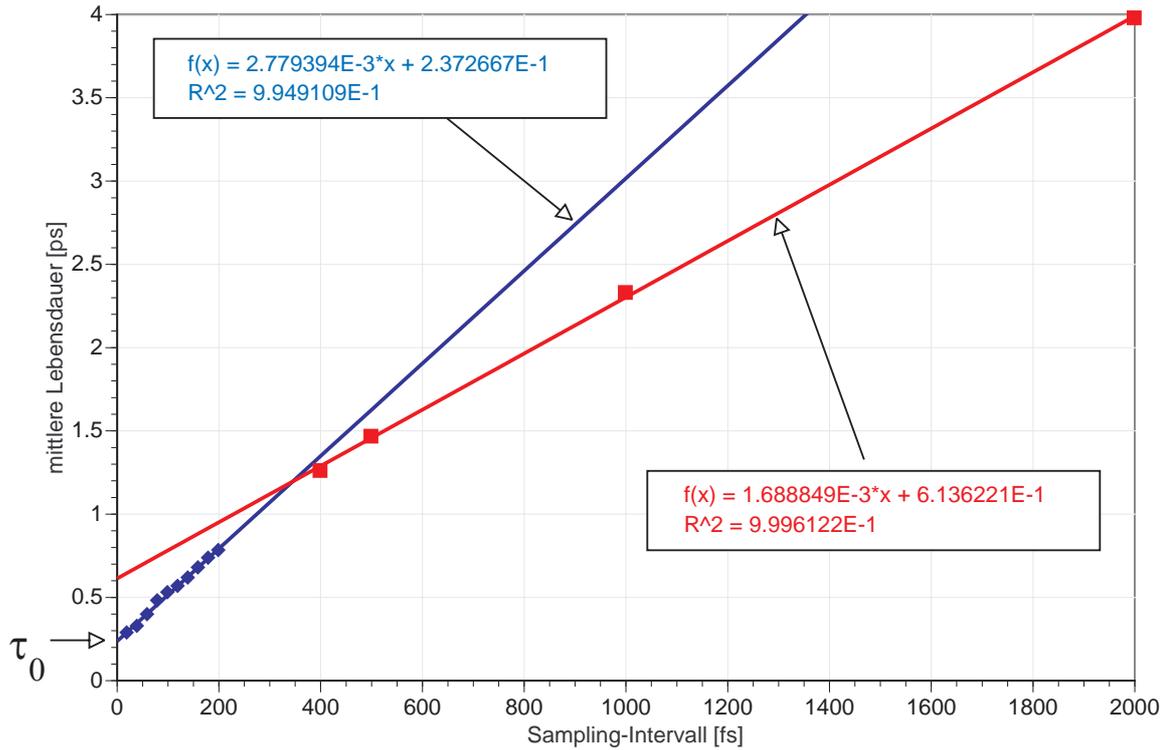
**Abb. X110:**  $\Phi / \Psi$ -FGE-Maps (MD-Simulation von Maltose: CVFF,  $\epsilon = 4$ )



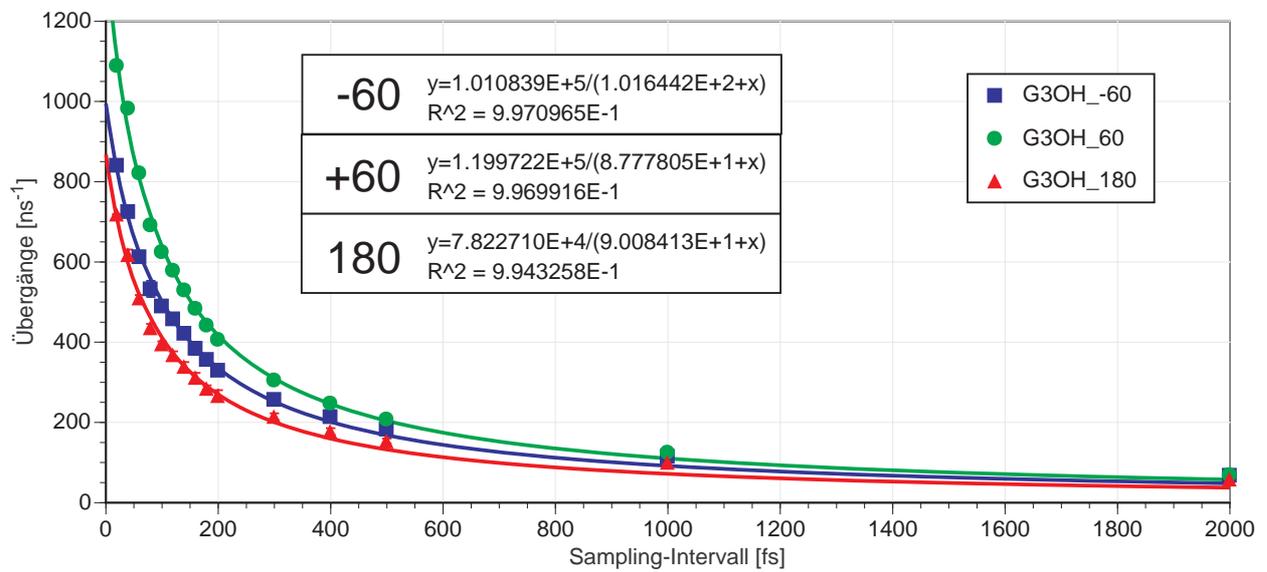
**Abb. X111:** Ausschnitt aus einer Trajektorie von G2OH und G3OH (Maltose, CVFF,  $\varepsilon = 4$ , 300K,  $t_s = 10$  ps,  $\Delta t_s = 10$  fs)



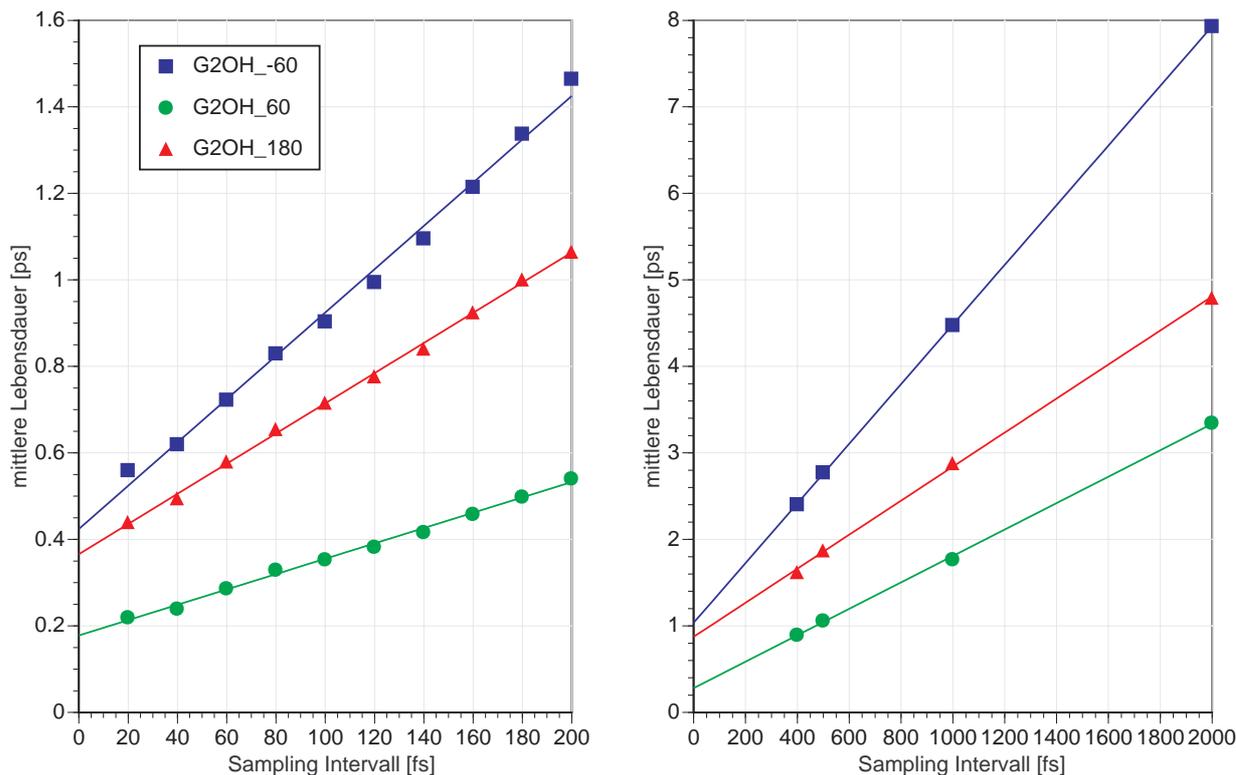
**Abb. X112:** Simulation der Abhängigkeit der Rotameren-Population vom Sampling-Intervall (MD-Simulation von Maltose: CVFF,  $\varepsilon = 4$ , 300 K,  $t_s = 1$  ns,  $\Delta t_s = 20$  fs)



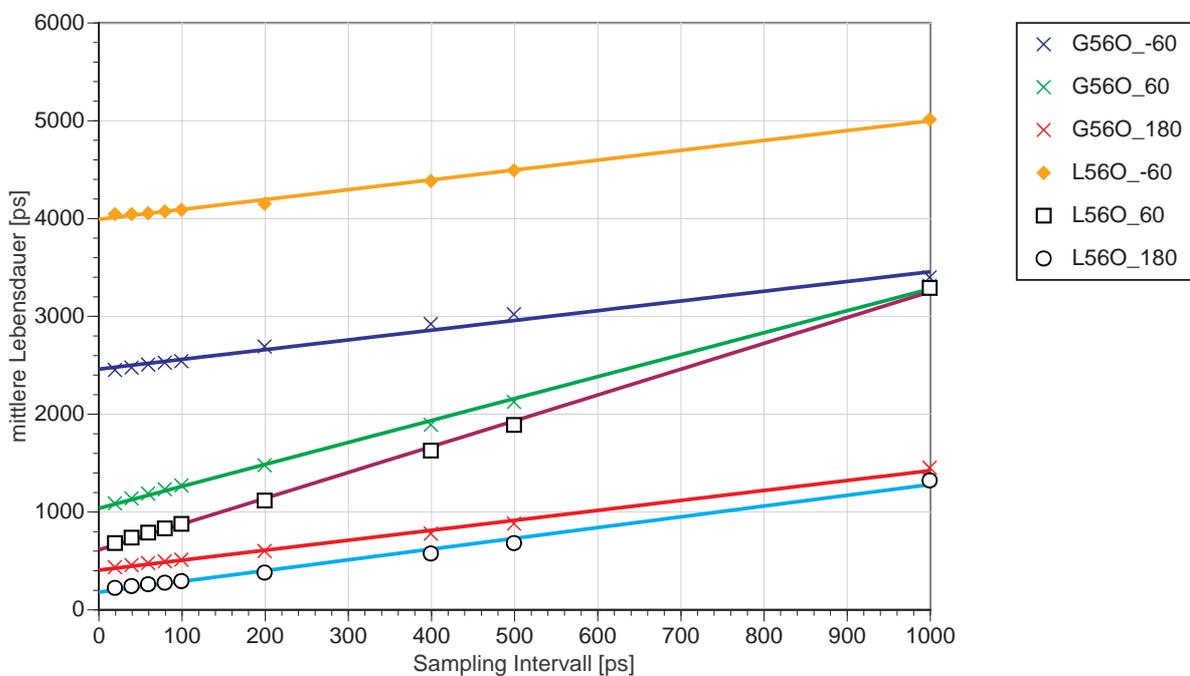
**Abb. X113:** Simulation der Abhängigkeit der mittleren Lebensdauer vom Sampling-Intervall (MD-Simulation von Maltose: CVFF,  $\varepsilon = 4$ , 300 K,  $t_s = 1$  ns,  $\Delta t_s = 20$  fs)



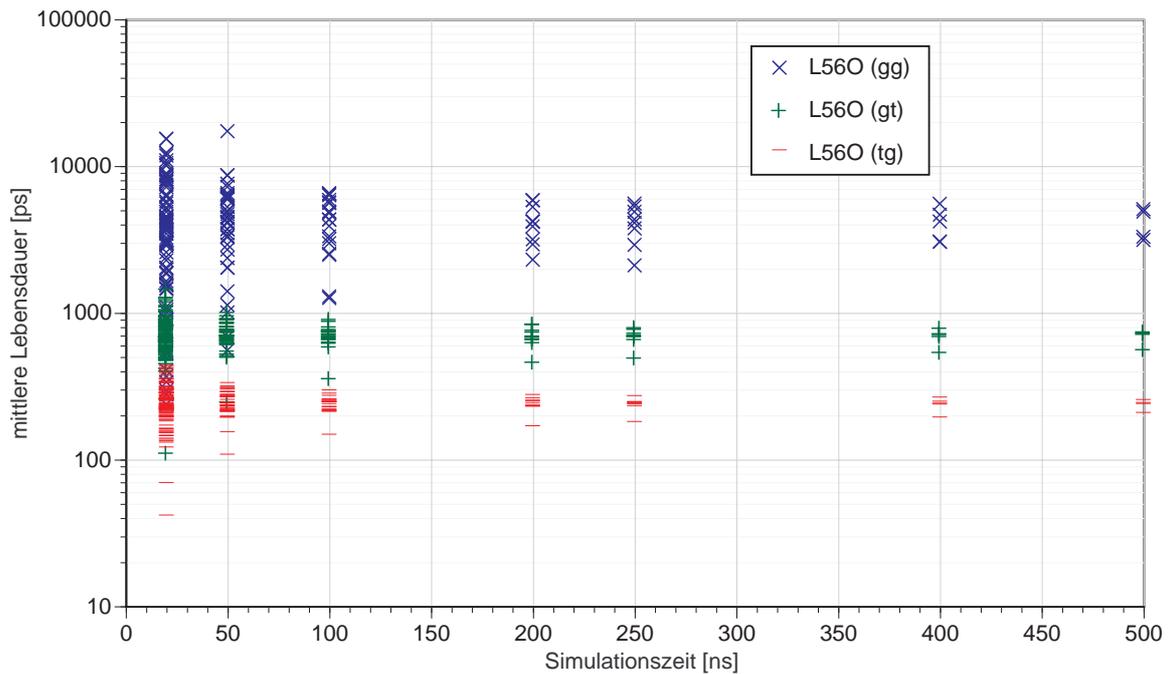
**Abb. X114:** Simulation der Abhängigkeit der Zahl der Übergänge vom Sampling-Intervall (MD-Simulation von Maltose: CVFF,  $\varepsilon = 4$ , 300 K,  $t_s = 1$  ns,  $\Delta t_s = 20$  fs)



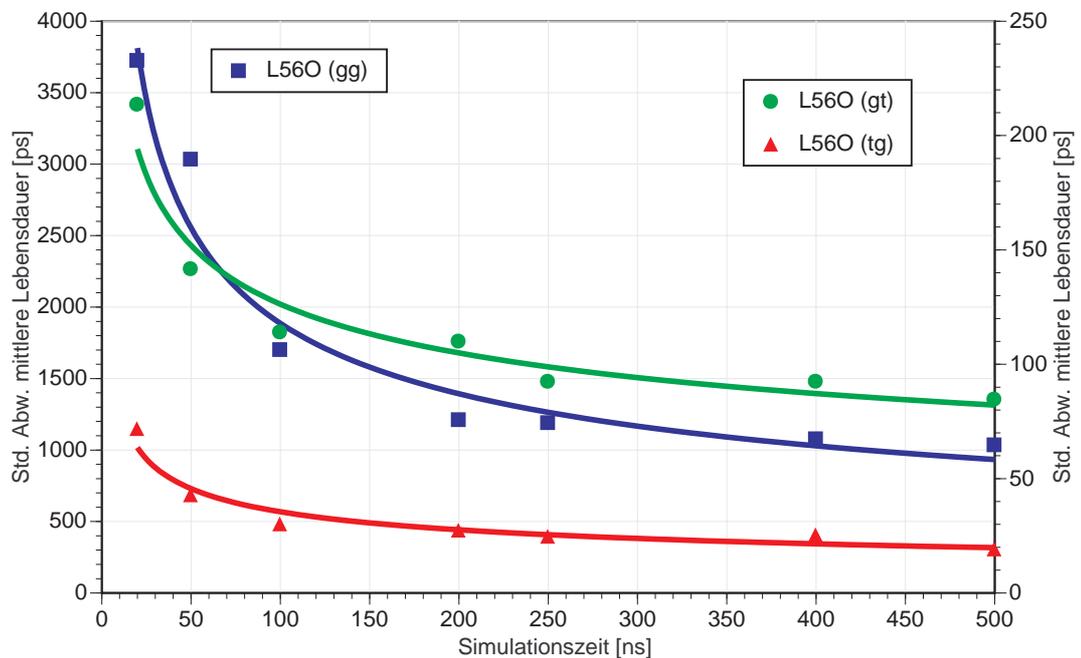
**Abb. X115:** Simulation der Abhängigkeit der mittleren Lebensdauer vom Sampling-Intervall (MD-Simulation von Maltose: CVFF,  $\varepsilon = 4$ , 300 K,  $t_s = 1$  ns,  $\Delta t_s = 20$  fs)



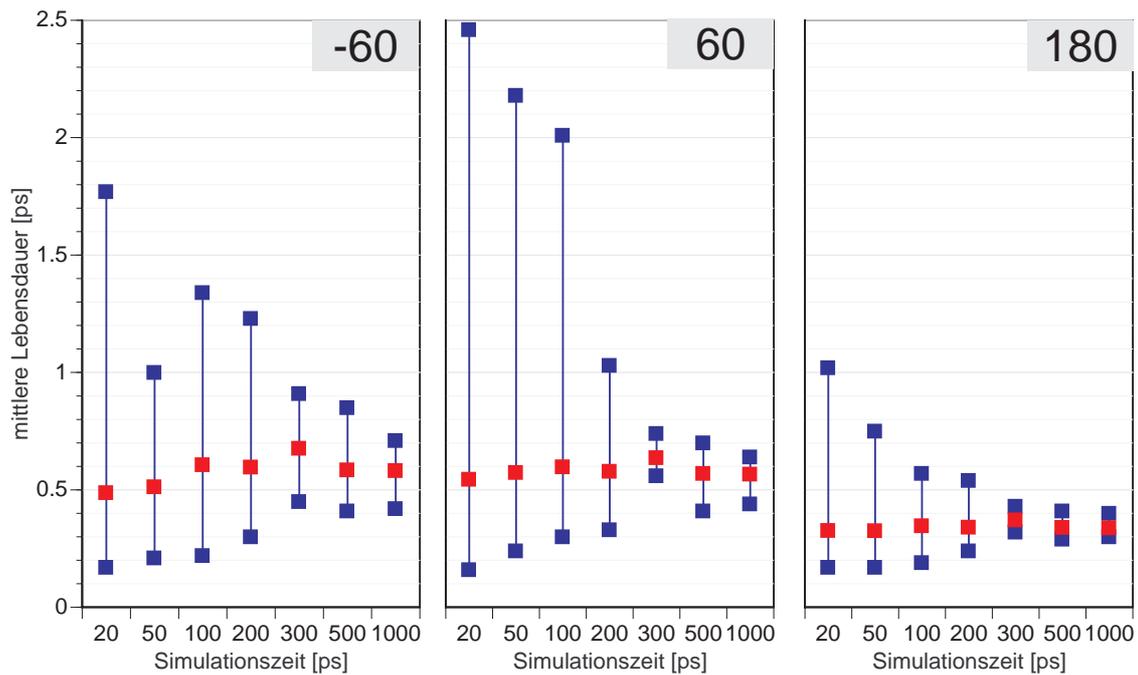
**Abb. X116:** Simulation der Abhängigkeit der mittleren Lebensdauer vom Sampling-Intervall (MD-Simulation von Maltose: CVFF,  $\varepsilon = 4$ , 300 K,  $t_s = 2$   $\mu$ s,  $\Delta t_s = 20$  ps)



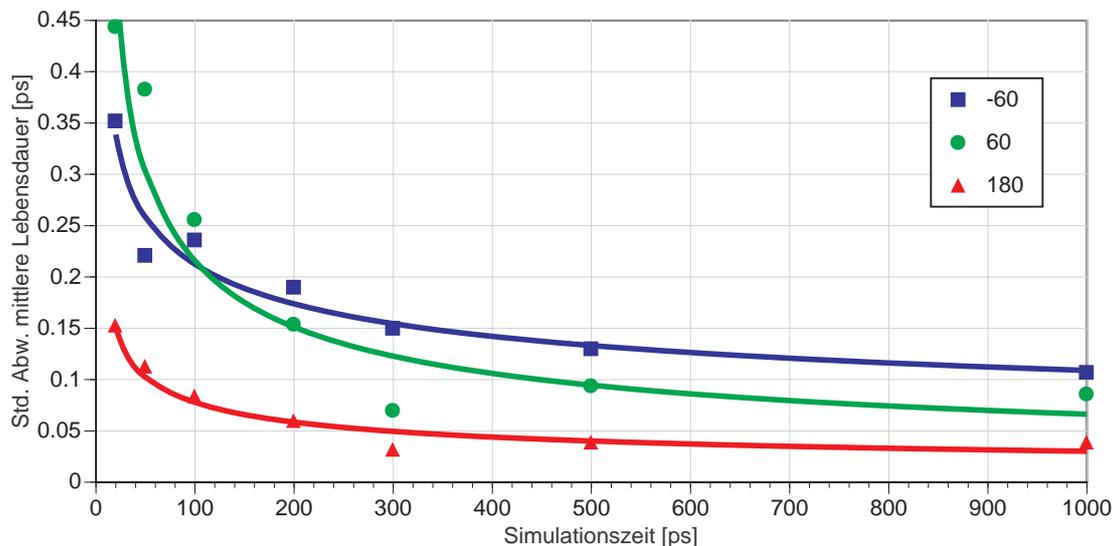
**Abb. X117:** Reproduzierbarkeit der mittleren Lebensdauern in Abhängigkeit von der Simulationszeit (Maltose, Torsionswinkel L56O, CVFF,  $\varepsilon = 4$ , 300 K)



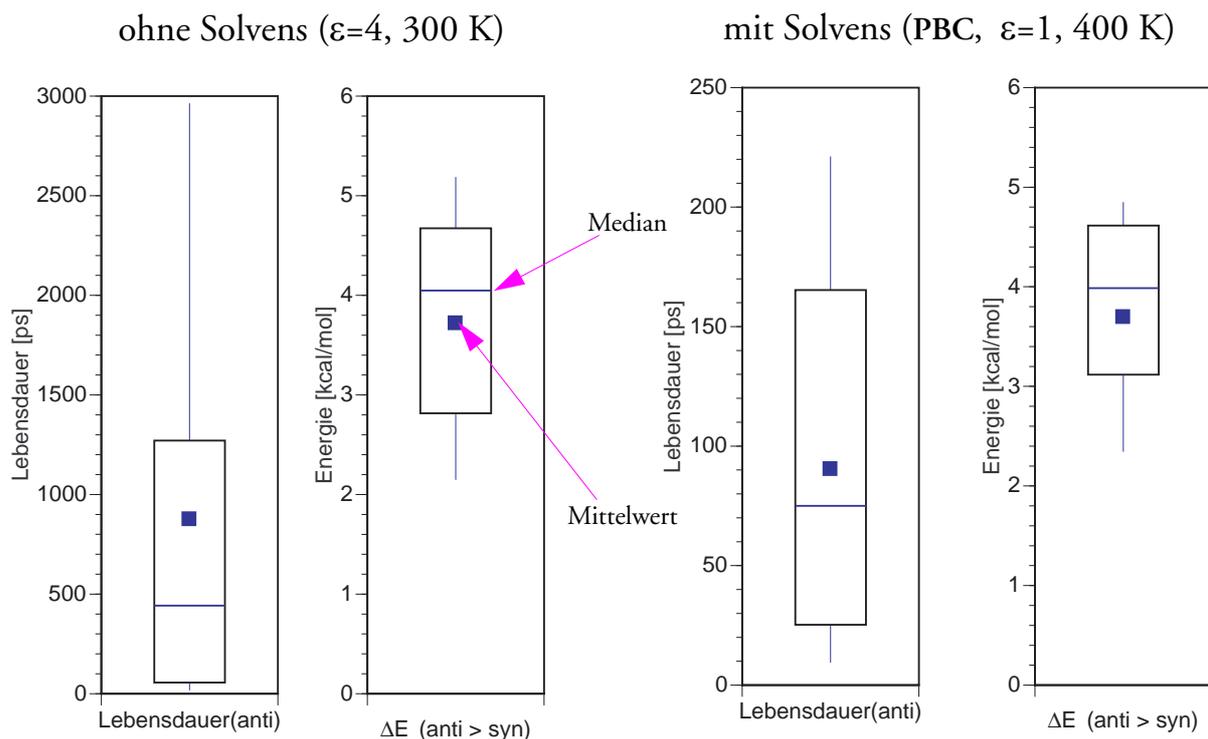
**Abb. X118:** Reproduzierbarkeit der mittleren Lebensdauern in Abhängigkeit von der Simulationszeit (Maltose, Torsionswinkel L56O, CVFF,  $\varepsilon = 4$ , 300 K)



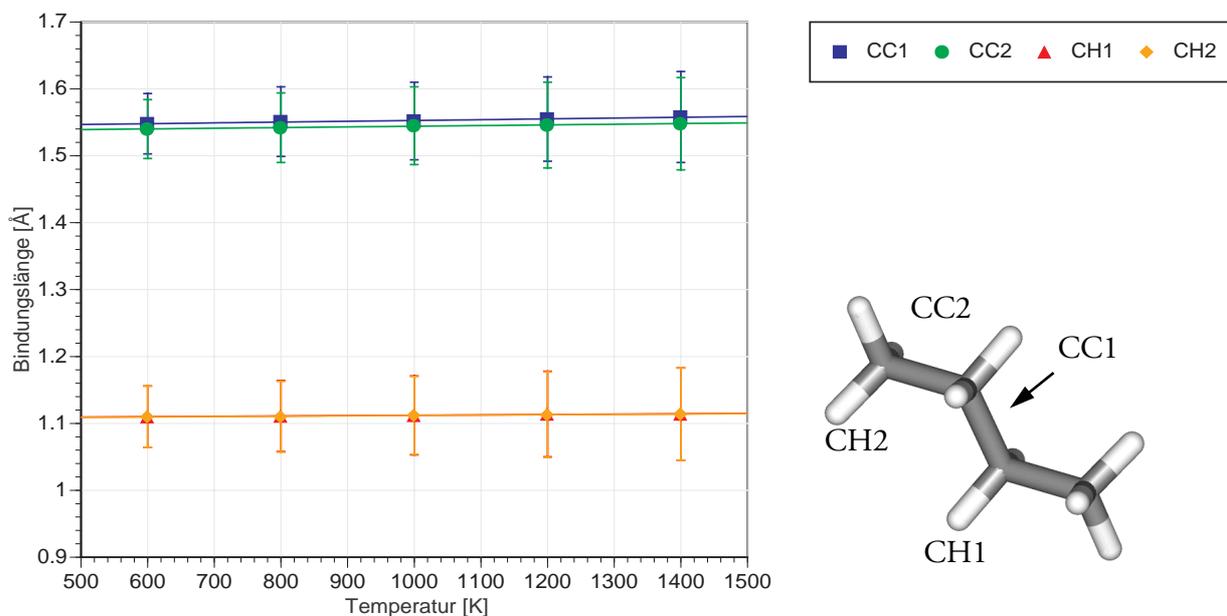
**Abb. X119:** Reproduzierbarkeit der mittleren Lebensdauern in Abhängigkeit von der Simulationszeit (Maltose, Torsionswinkel G3OH, CVFF,  $\epsilon = 4$ , 300K). Min-Max-Mittelwert-Plot.



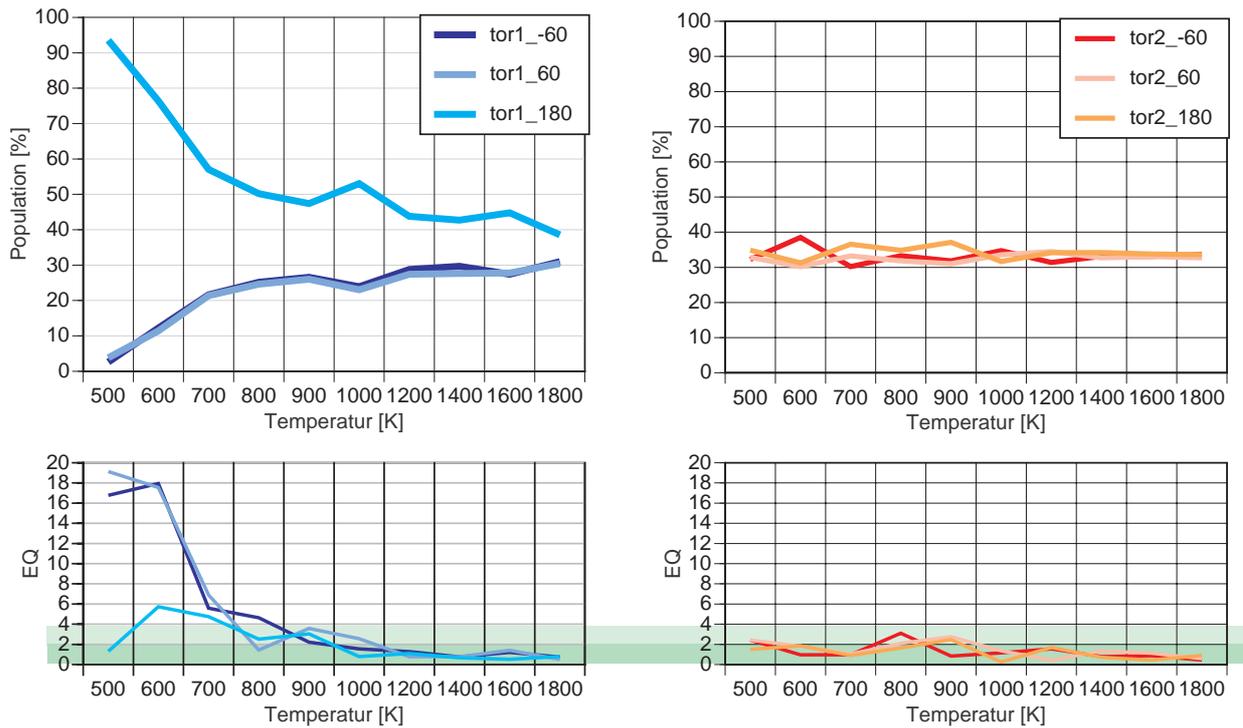
**Abb. X120:** Reproduzierbarkeit der mittleren Lebensdauern in Abhängigkeit von der Simulationszeit (Maltose, Torsionswinkel G3OH, CVFF,  $\epsilon = 4$ , 300 K)



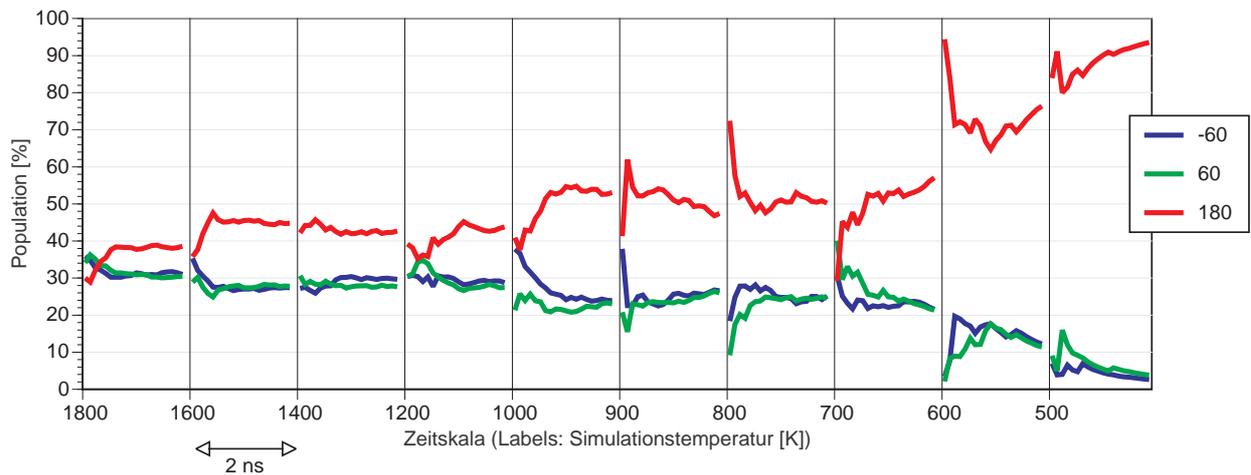
**Abb. X121:** Bestimmung der mittleren Lebensdauer der anti-Konformation von Maltose bei 300 K durch „Multiple-Start“-MD-Simulationen (CVFF,  $\epsilon = 4$ ,  $20x t_s = 3$  ns und  $20x t_s = 5$  ns) und bei 400 K (CVFF, PBC  $20\text{\AA} \times 20\text{\AA} \times 20\text{\AA}$ ,  $\epsilon = 1$ ,  $10x t_s = 250$  ps und  $10x t_s = 500$  ps). Boxplots: 50% der Werte liegen innerhalb der Box



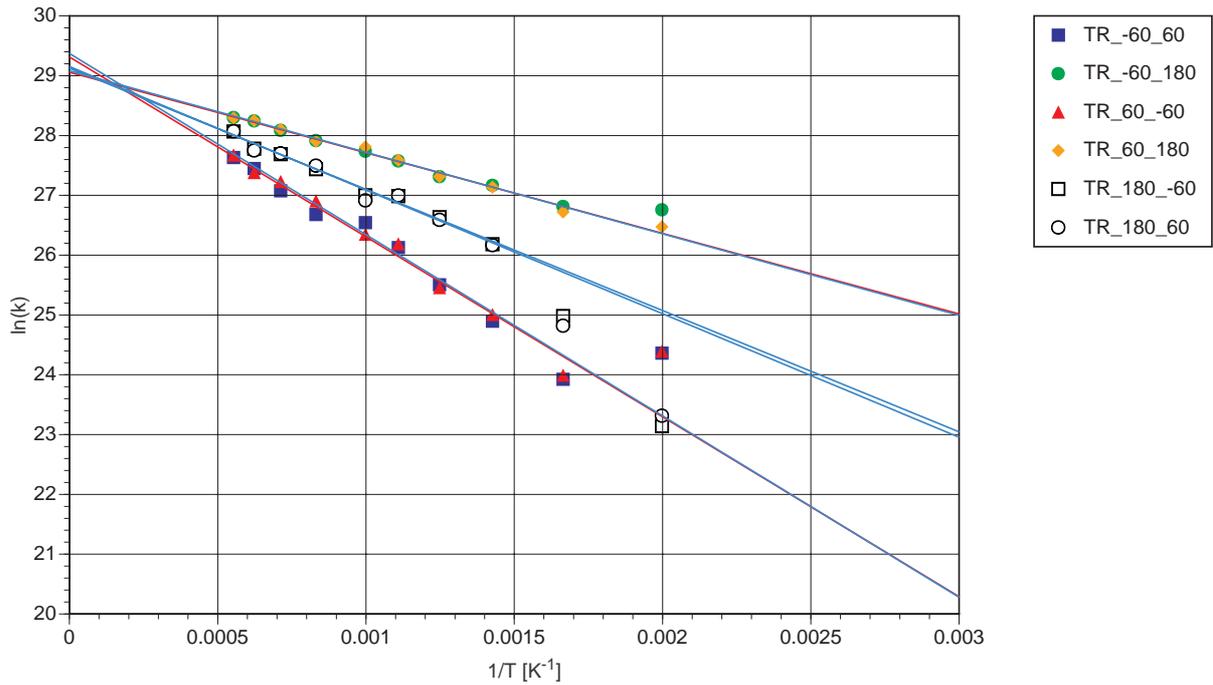
**Abb. X122:** Abhängigkeit der Bindungslänge von der Simulationstemperatur (n-Butan, CVFF,  $\epsilon = 10$ )



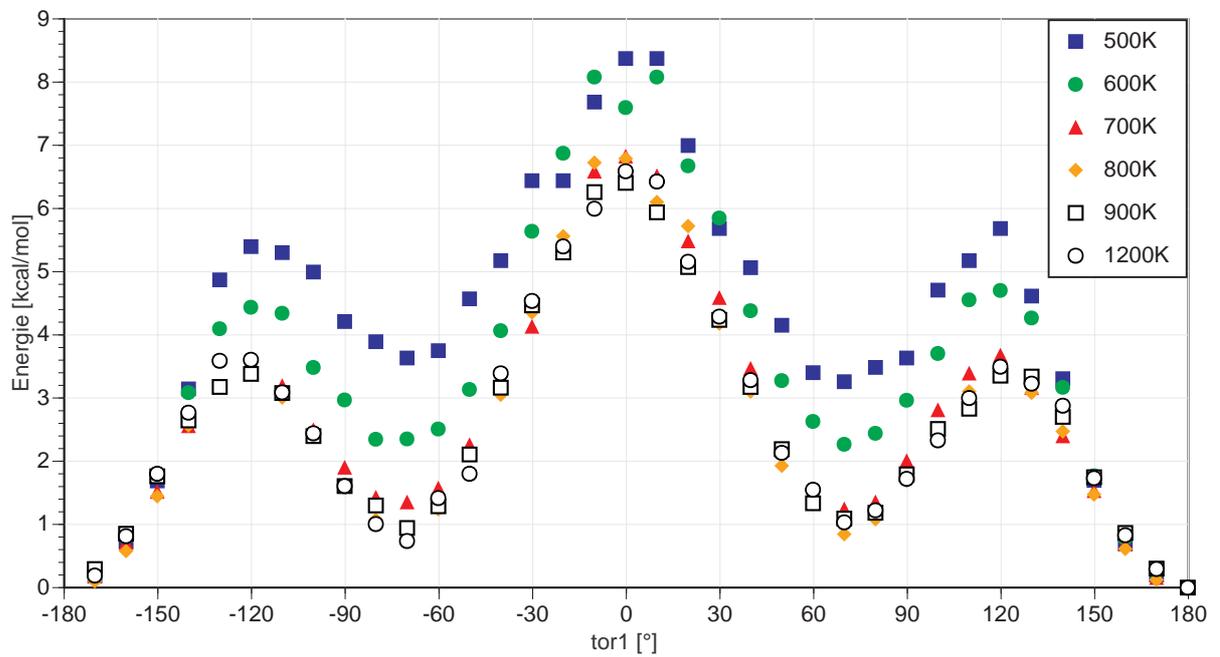
**Abb. X123:** Arrhenius-Analyse von n-Butan: Überprüfung des Konformationsgleichgewichts (CVFF,  $\varepsilon = 10$ ,  $t_s$  (pro Temperatur) = 2 ns)



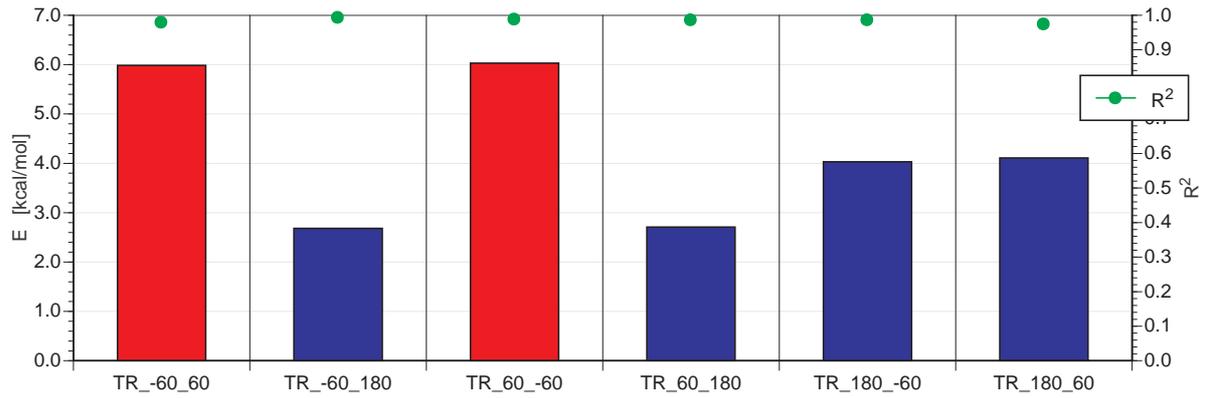
**Abb. X124:** Arrhenius-Analyse von n-Butan: Einstellung des Rotamerengleichgewichts in Abhängigkeit von der Simulationszeit und der Simulationstemperatur (CVFF,  $\varepsilon = 10$ )



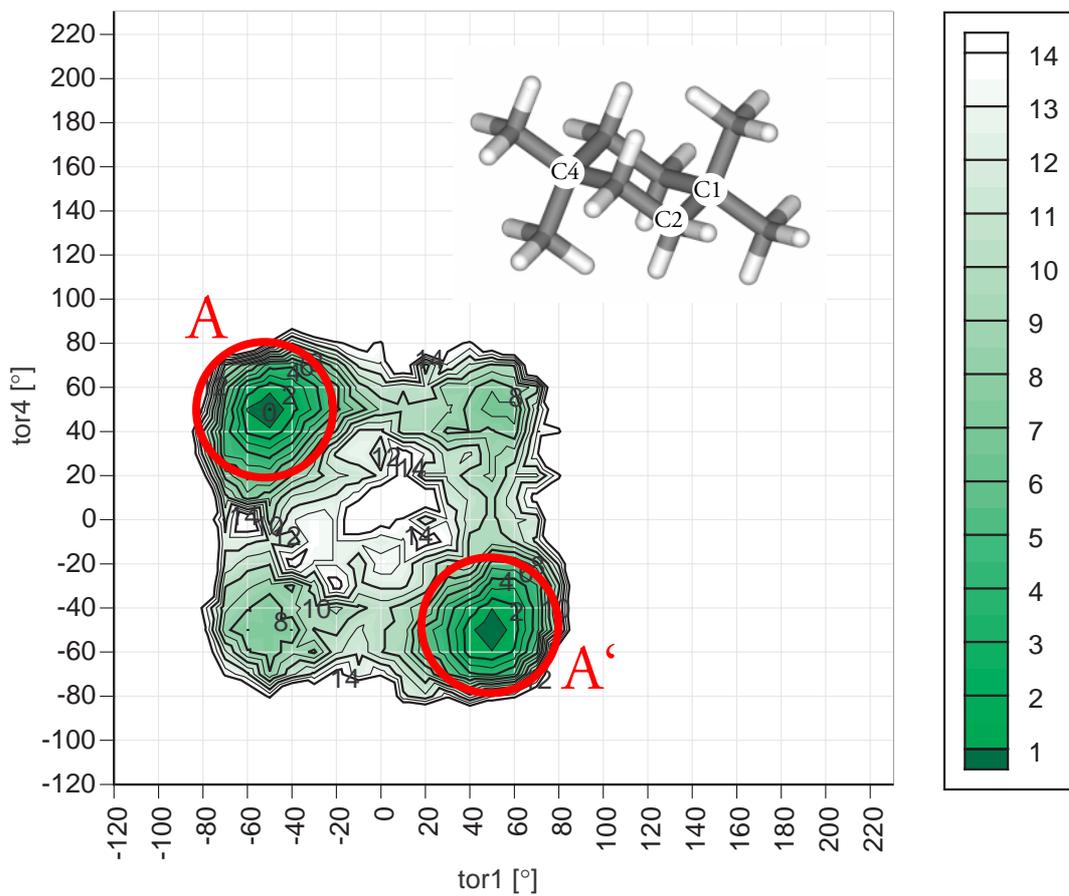
**Abb. X125:** Arrhenius-Analyse von n-Butan: Bestimmung der Rotationsbarrieren von „tor1“ (CVFF,  $\epsilon = 10$ )



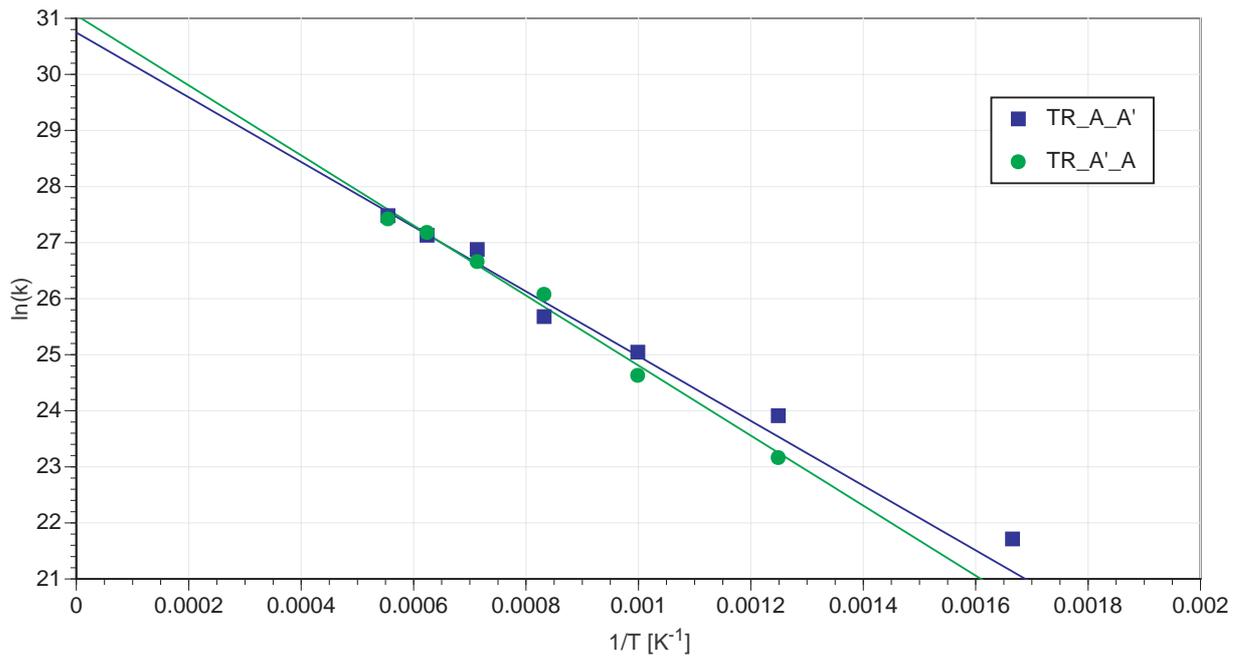
**Abb. X126:** FGE-Profile von „tor1“ in Abhängigkeit von der Simulationstemperatur (n-Butan, CVFF,  $\epsilon = 10$ )



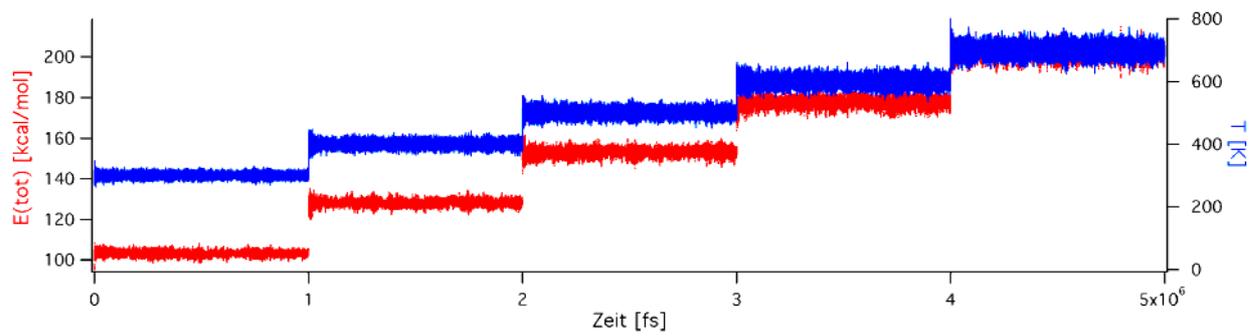
**Abb. X127:** Arrhenius-Analyse: Übergangsbarrieren (n-Butan, CVFF,  $\epsilon = 10$ )



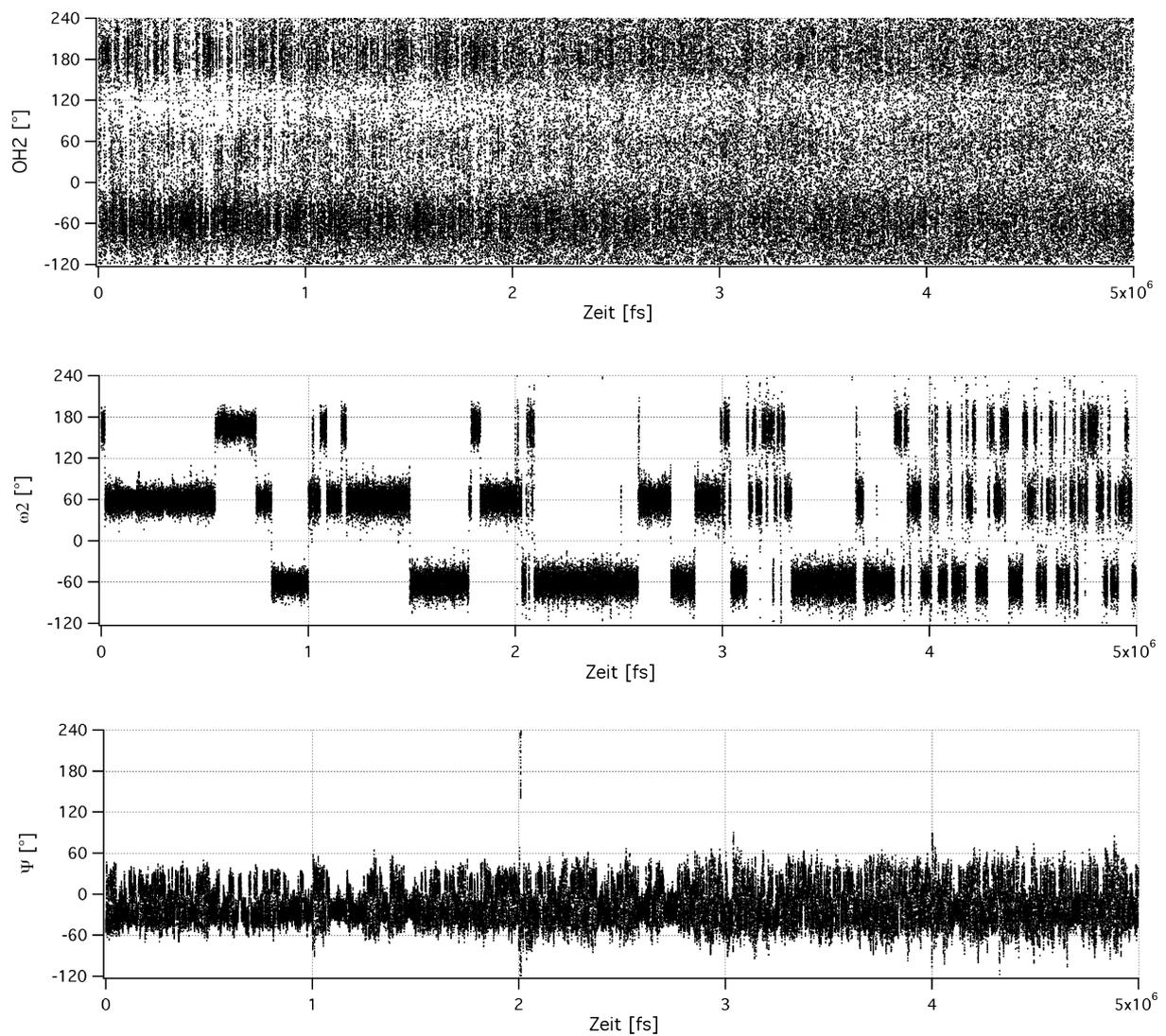
**Abb. X128:** FGE-Map (1,1,4,4-Tetramethylcyclohexan, CVFF,  $\epsilon = 4$ , 1600 K,  $t_s = 3$  ns)  
 tor1 = C1-C2-C3-C4, tor4 = C4-C5-C6-C1



**Abb. X129:** Arrhenius-Analyse: Bestimmung der Ringinversionsbarriere von 1,1,4,4,-Tetramethylcyclohexan (CVFF,  $\varepsilon = 4$ )

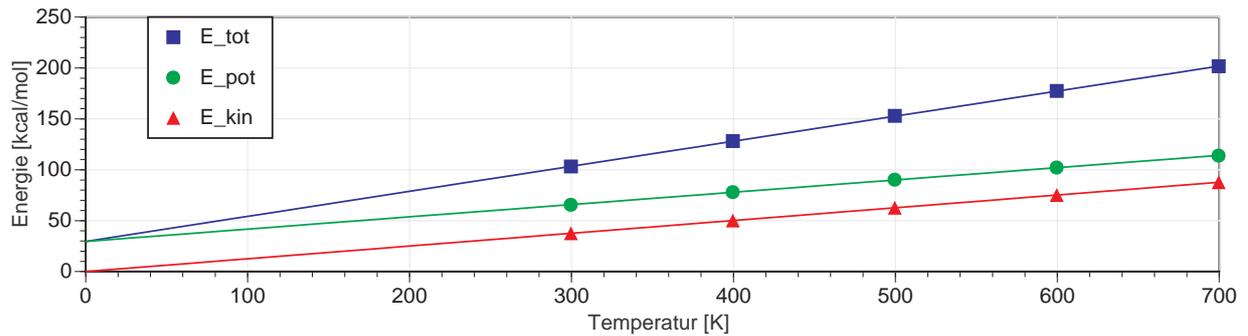


**Abb. X130:** Arrhenius-Analyse: Simulated Heating - Trajektorien von Temperatur und potentieller Energie (Maltose, CVFF,  $\varepsilon = 4$ ,  $t_s = 1$  ns / Temperatur,  $\Delta t_s = 50$  fs)

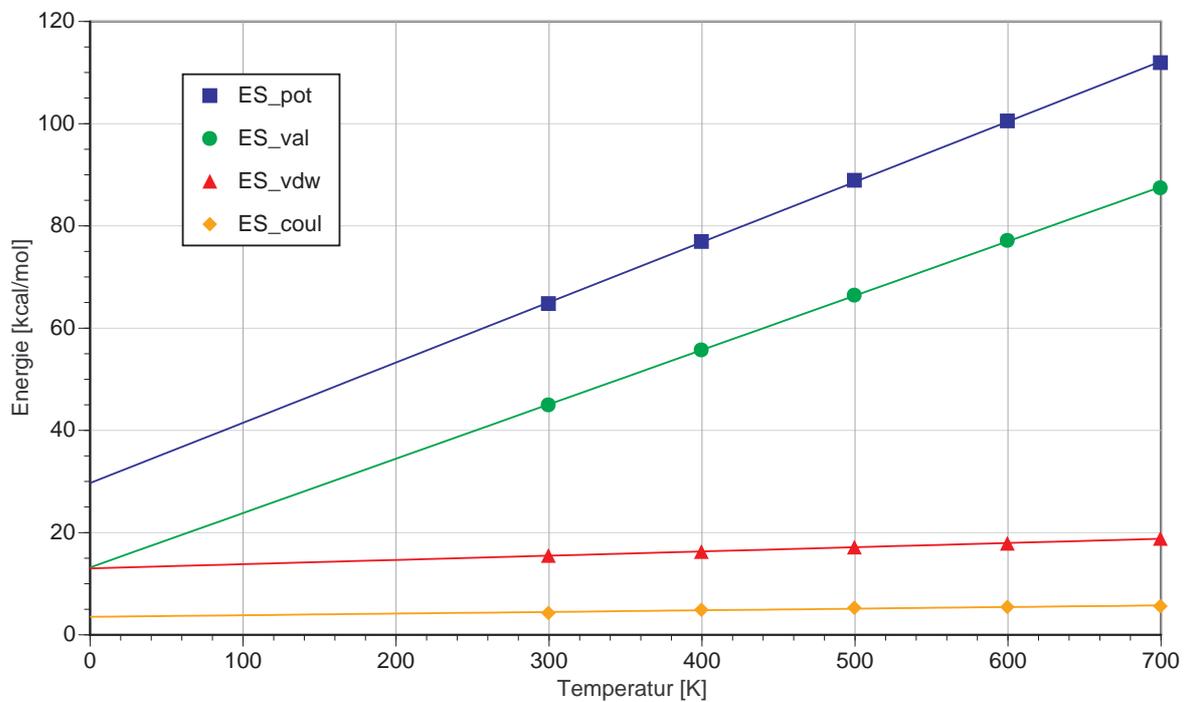


**Abb. X131:** Arrhenius-Analyse: Simulated Heating - Trajektorien von G2OH, L56O und Psi (Maltose, CVFF,  $\varepsilon = 4$ ,  $t_s = 1$  ns / Temperatur,  $\Delta t_s = 50$  fs)

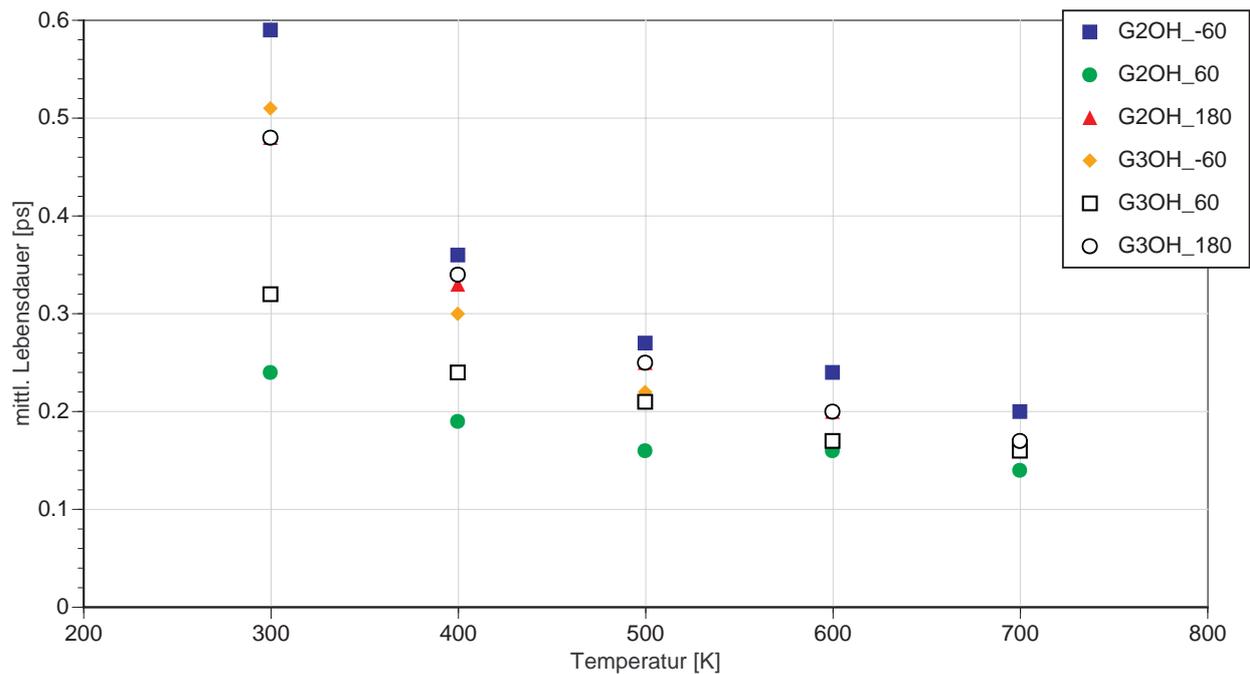
	E_tot	+/-	E_pot	+/-	E_kin	+/-	T	+/-
300	103.208	1.27	65.602	2.958	37.606	2.968	300.38	23.711
400	128.163	1.647	78.027	4.004	50.136	4.017	400.469	32.088
500	152.943	2.035	90.275	4.955	62.667	4.971	500.565	39.706
600	177.412	2.386	102.215	5.955	75.197	5.975	600.645	47.729
700	201.673	2.731	113.944	6.931	87.728	6.952	700.742	55.532



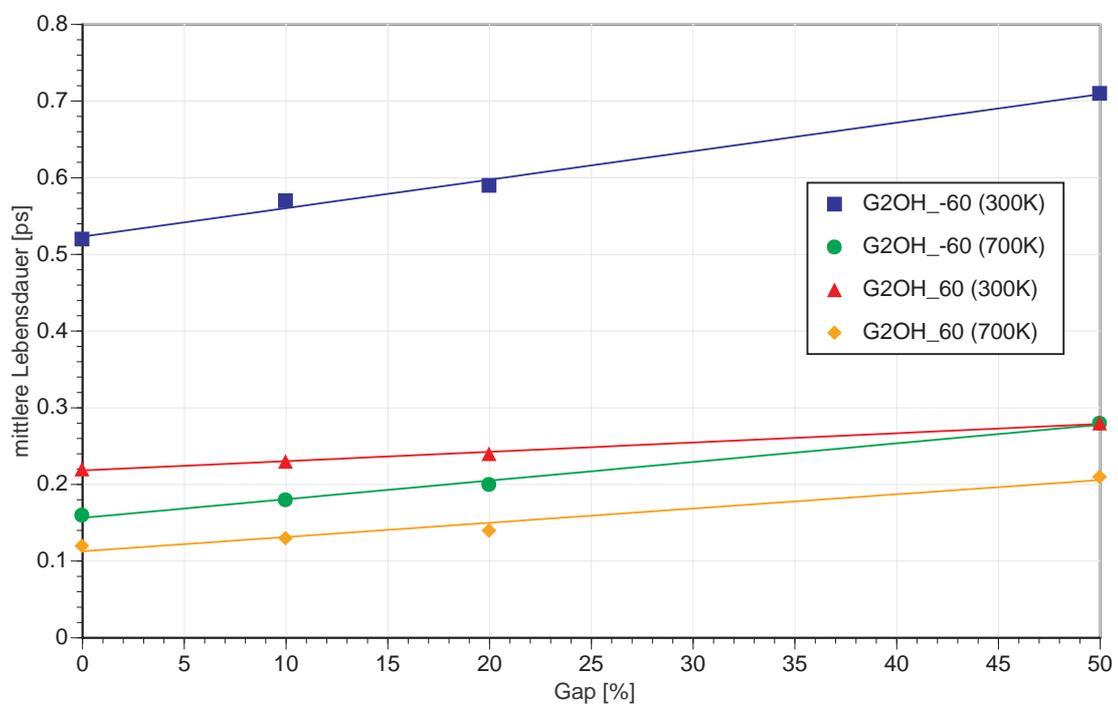
**Abb. X132:** Arrhenius-Analyse: Abhängigkeit der Systemenergie von der Temperatur (Maltose, CVFF,  $\epsilon = 4$ ,  $t_s = 1$  ns / Temperatur,  $\Delta t_s = 50$  fs)



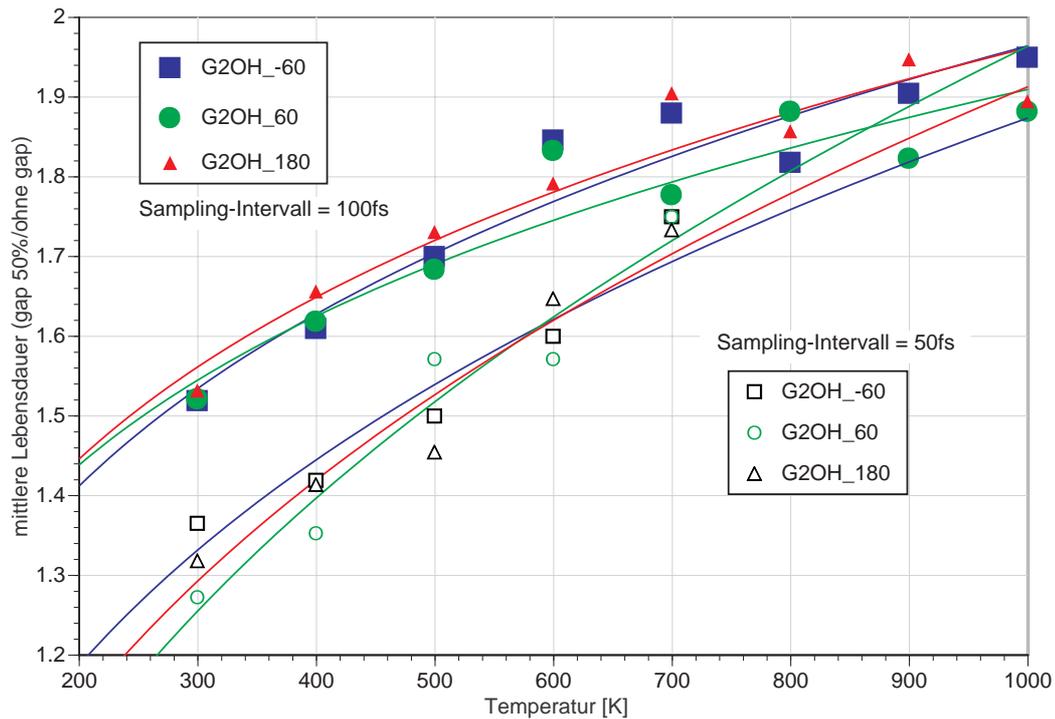
**Abb. X133:** Arrhenius-Analyse: Abhängigkeit der potentiellen Energie von der Temperatur (Maltose, CVFF,  $\epsilon = 4$ ,  $t_s = 1$  ns / Temperatur,  $\Delta t_s = 50$  fs)



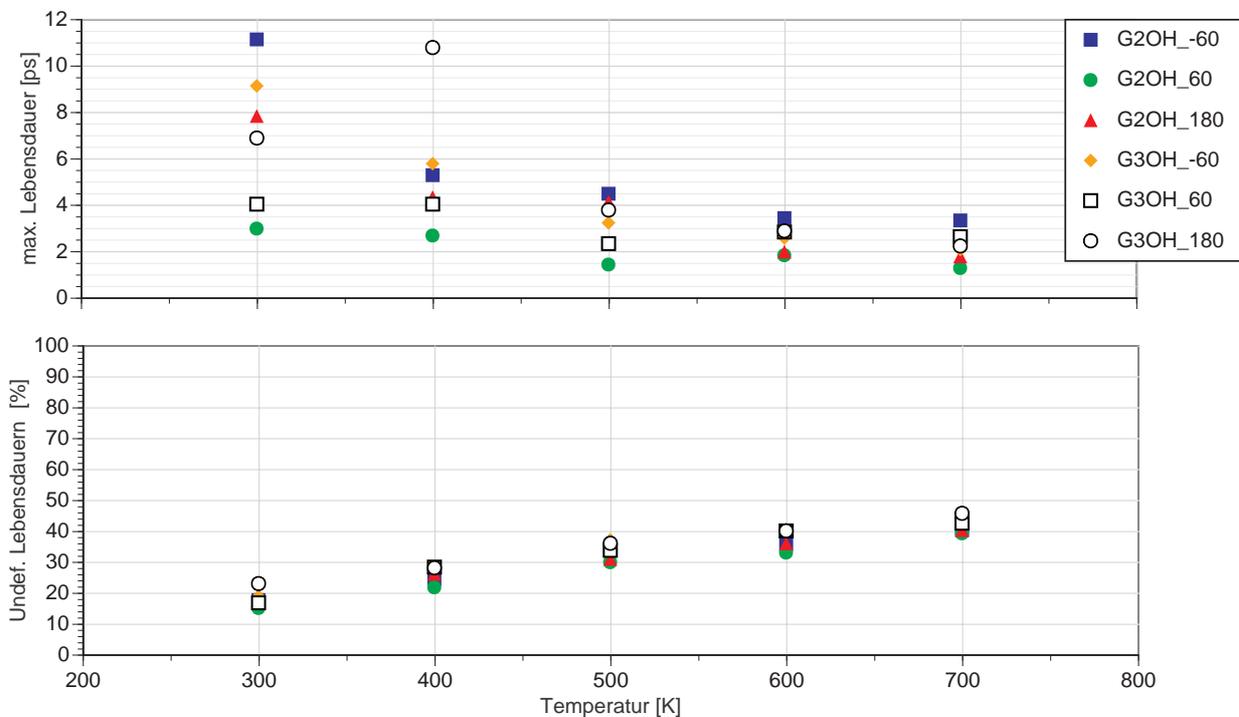
**Abb. X134:** Arrhenius-Analyse: Abhängigkeit der mittleren Lebensdauer ausgewählter Rotamere von der Temperatur (Maltose, CVFF,  $\varepsilon = 4$ ,  $t_s = 1$  ns / Temperatur,  $\Delta t_s = 50$  fs, Gap = 20%)



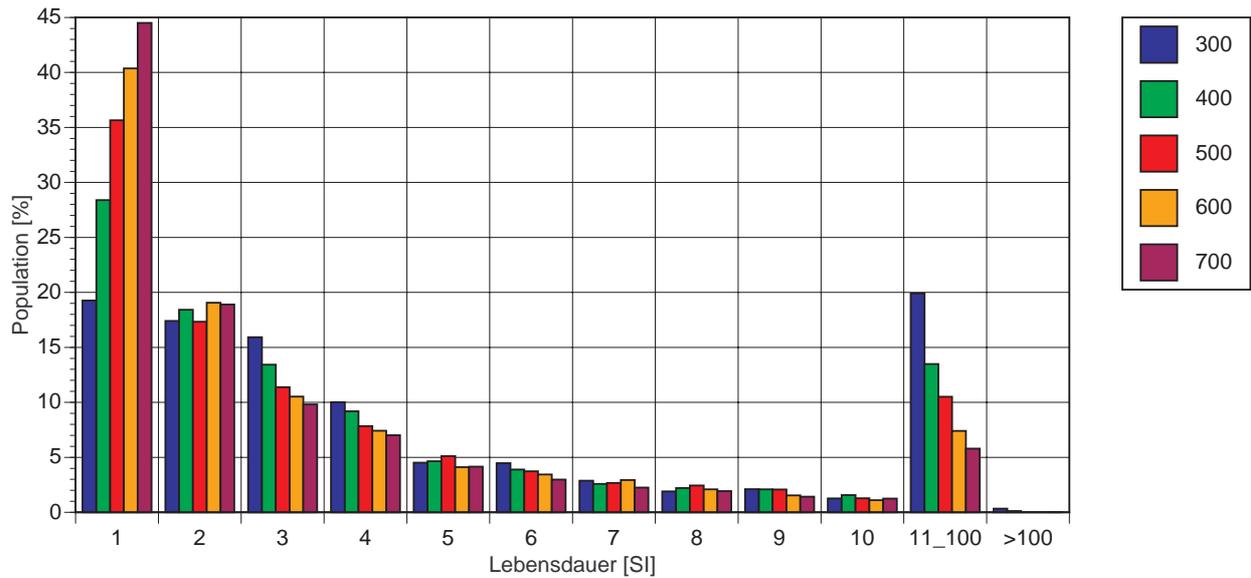
**Abb. X135:** Arrhenius-Analyse: Einfluß des Gaps auf die mittlere Lebensdauer (Maltose, G2OH, CVFF,  $\varepsilon = 4$ ,  $t_s = 1$  ns / Temperatur,  $\Delta t_s = 50$  fs)



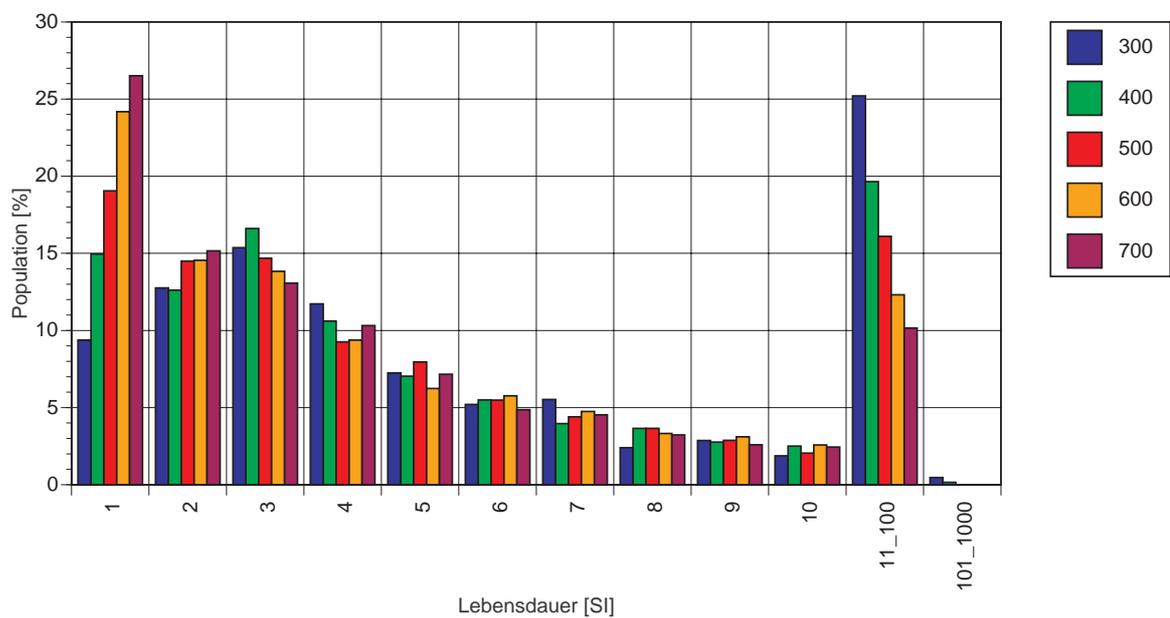
**Abb. X136:** Arrhenius-Analyse: Einfluß des Gaps auf die mittlere Lebensdauer in Abhängigkeit von der Temperatur (Maltose, G2OH, CVFF,  $\varepsilon = 4$ ,  $t_s = 1 \text{ ns} / \text{Temperatur}$ ) Kurven: Powerfit /Delta-graph



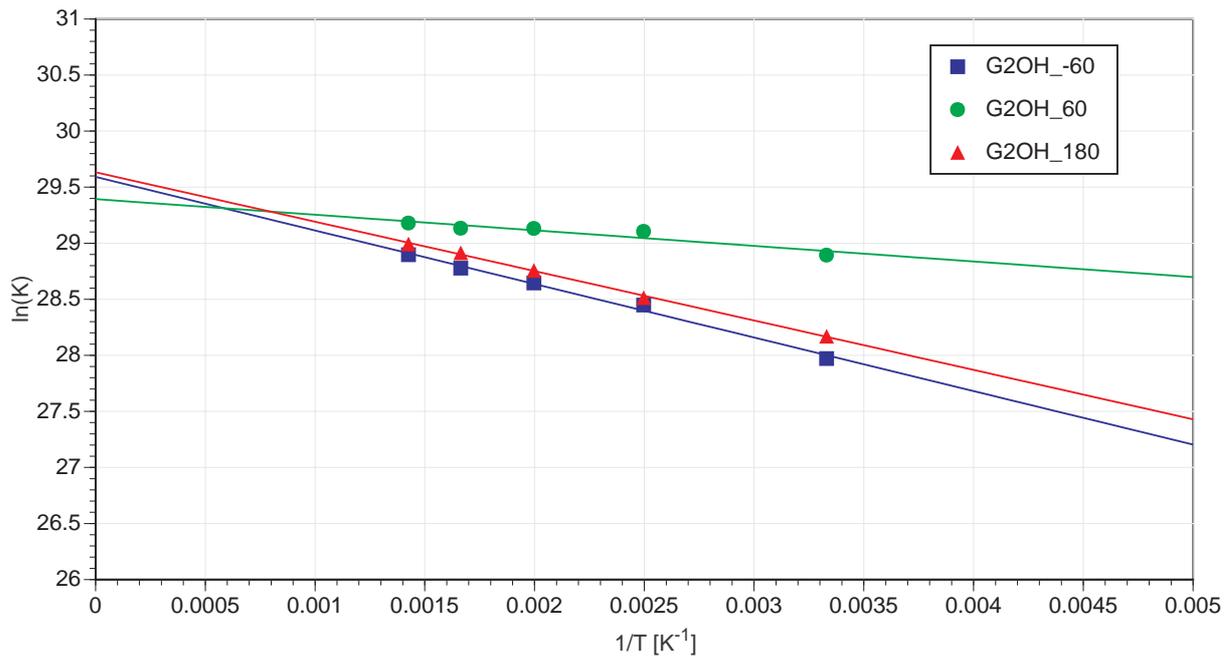
**Abb. X137:** Arrhenius-Analyse: Temperaturabhängigkeit der maximalen Lebensdauer und des Anteils der Lebensdauern die kleiner oder gleich den Sampling-Intervall sind („undefinierte Lebensdauern“) (Maltose, CVFF,  $\varepsilon = 4$ ,  $t_s = 1 \text{ ns} / \text{Temperatur}$ ,  $\Delta t_s = 50 \text{ fs}$ )



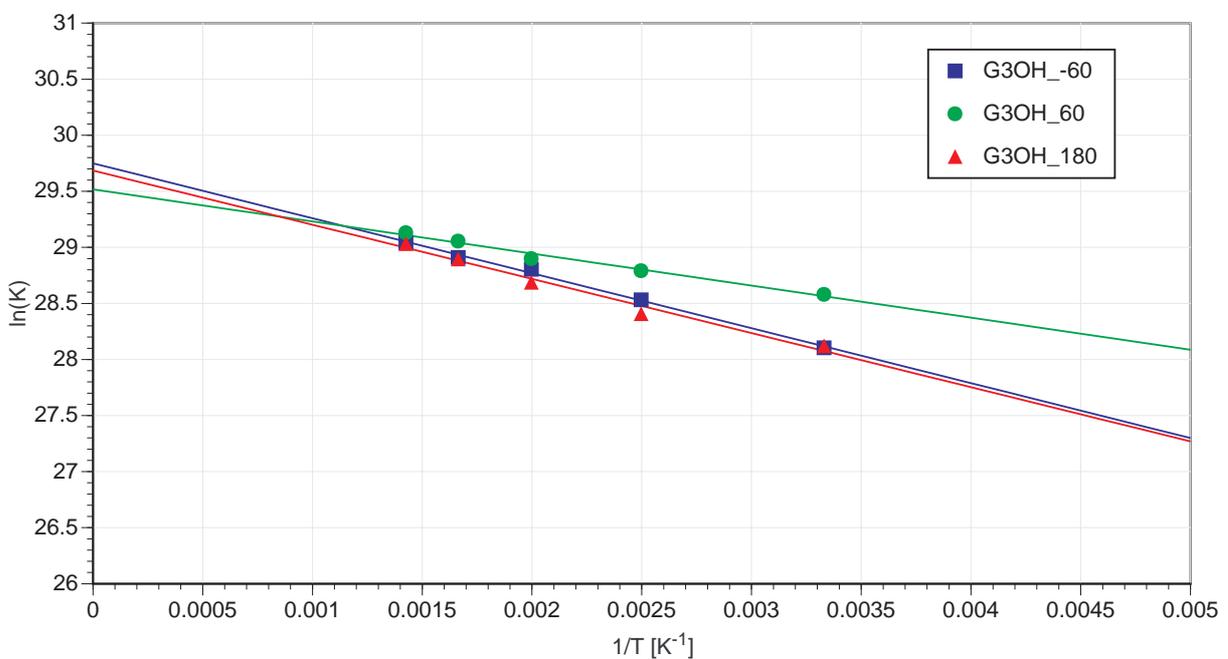
**Abb. X138:** Arrhenius-Analyse: Abhängigkeit der Lebensdauernverteilung von der Temperatur (Maltose, G3OH, CVFF,  $\varepsilon = 4$ ,  $t_s = 1$  ns / Temperatur,  $\Delta t_s = 50$  fs, Gap = 20%)



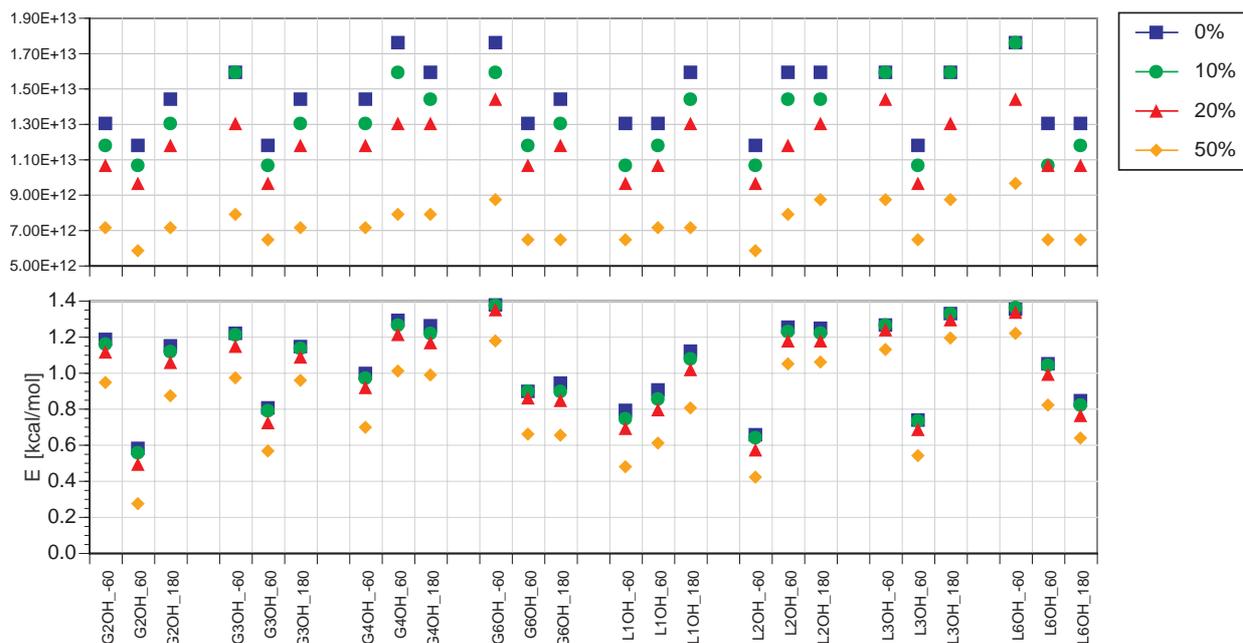
**Abb. X139:** Arrhenius-Analyse: Abhängigkeit der Lebensdauernverteilung von der Temperatur (Maltose, G3OH, CVFF,  $\varepsilon = 4$ ,  $t_s = 1$  ns / Temperatur,  $\Delta t_s = 50$  fs, Gap = 50%)



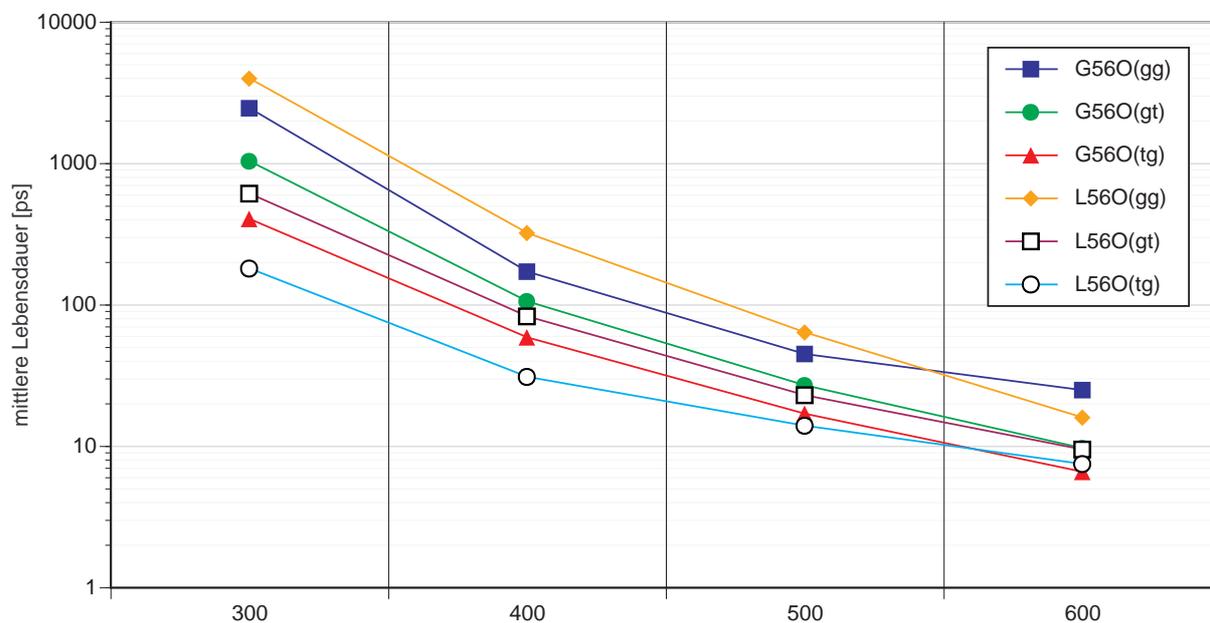
**Abb. X140:** Arrhenius-Analyse für die Rotameren von G2OH (Maltose, CVFF,  $\varepsilon = 4$ ,  $t_s = 1$  ns / Temperatur,  $\Delta t_s = 50$  fs, Gap = 50%)



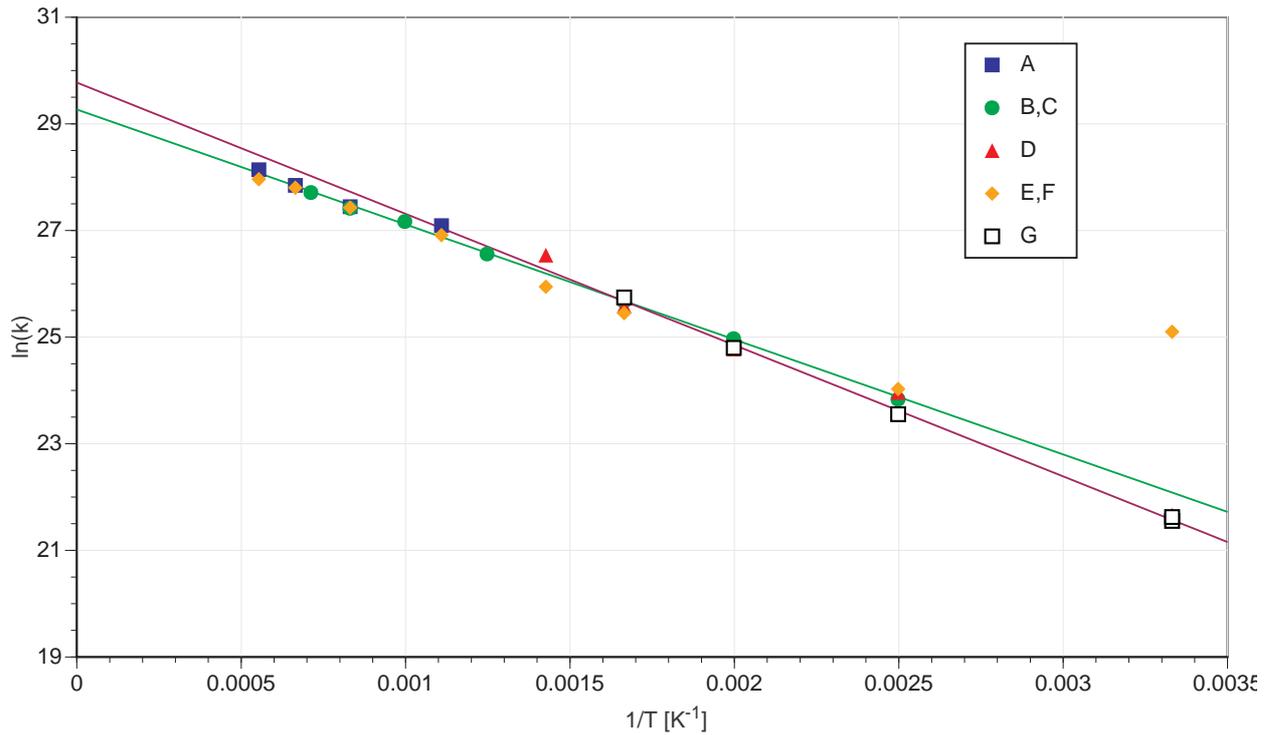
**Abb. X141:** Arrhenius-Analyse für die Rotameren von G3OH (Maltose, CVFF,  $\varepsilon = 4$ ,  $t_s = 1$  ns / Temperatur,  $\Delta t_s = 50$  fs, Gap = 50%)



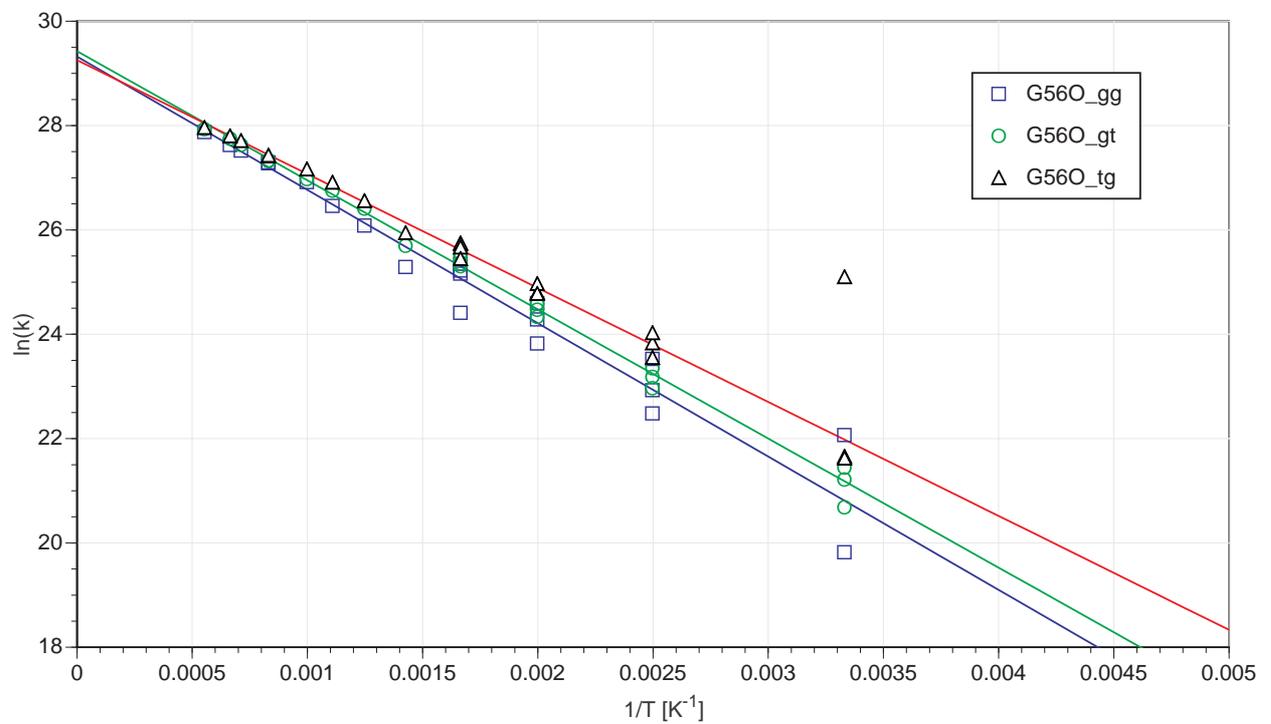
**Abb. X142:** Ergebnisse der Arrhenius-Analyse für die OH-Rotameren (Maltose, CVFF,  $\epsilon = 4$ ,  $T = 300$  K bis 700 K (+100 K),  $t_s = 1$  ns / Temperatur,  $\Delta t_s = 50$  fs)



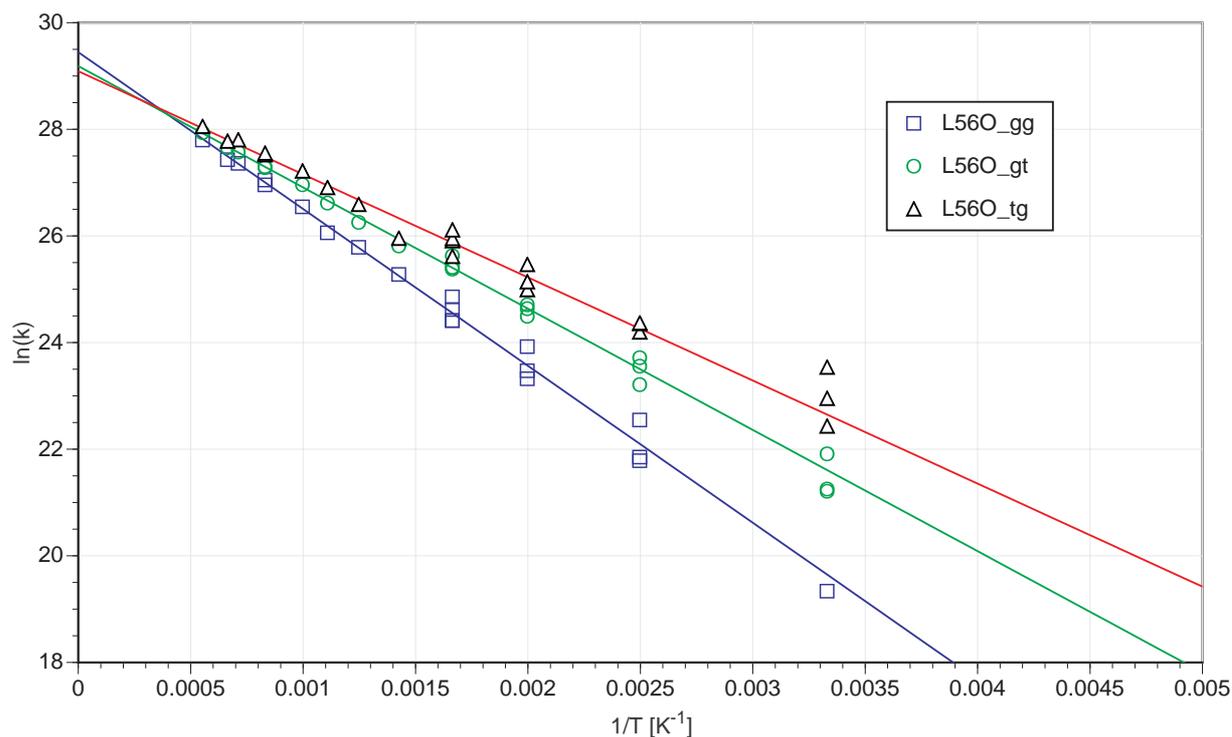
**Abb. X143:** Abhängigkeit der mittleren Lebensdauern von der Simulationstemperatur (Maltose, Hydroxymethylgruppen, CVFF,  $\epsilon = 4$ ) Simulationszeiten:  $t_s(300$  K) = 2  $\mu$ s;  $t_s(400$  K) = 300 ns;  $t_s(500$  K) = 100 ns;  $t_s(600$  K) = 30 ns



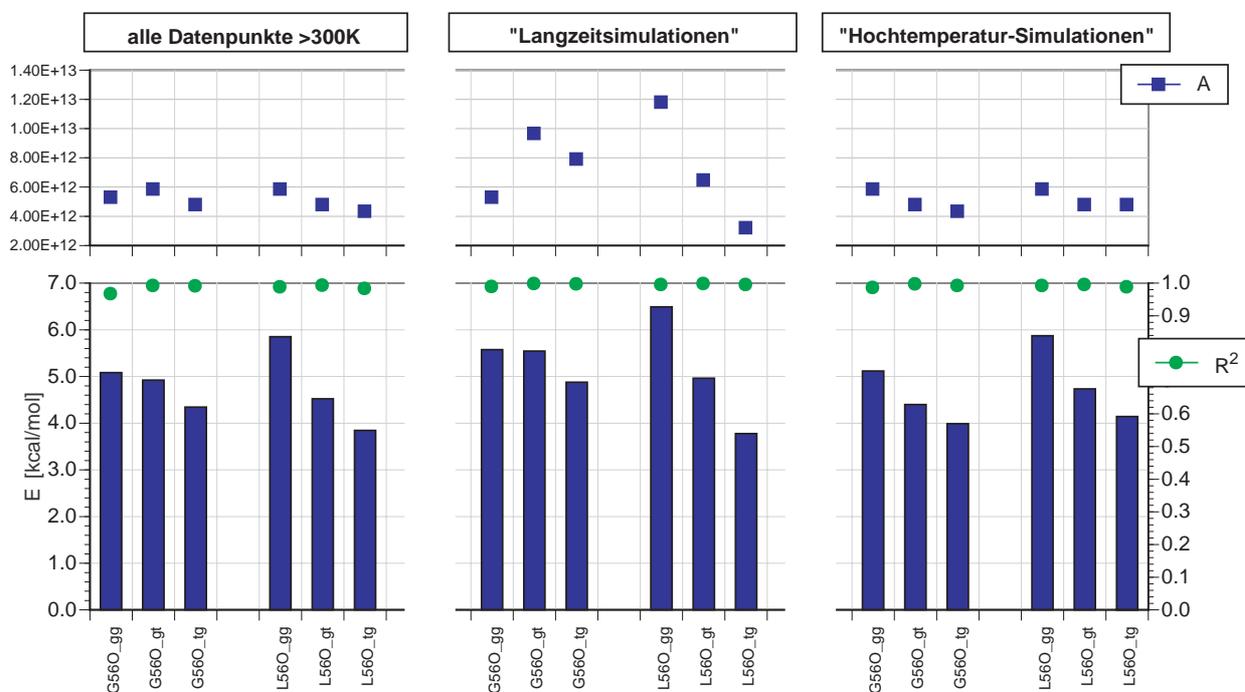
**Abb. X144:** Arrhenius-Analyse für den tg-Zustand von G56O für verschiedene Simulationen A bis G (Maltose, CVFF,  $\varepsilon = 4$ , Gap=10%)



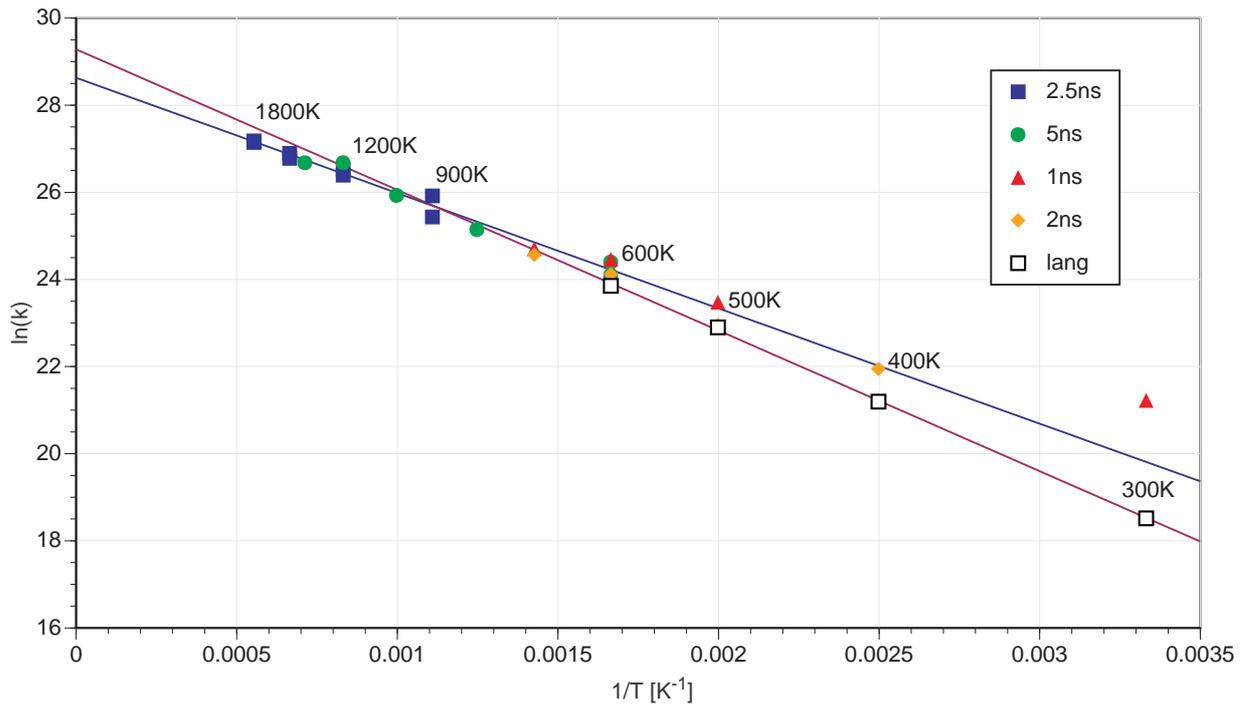
**Abb. X145:** Arrhenius-Analyse für die Rotameren von G56O (Maltose, CVFF,  $\varepsilon = 4$ , Gap = 10%)



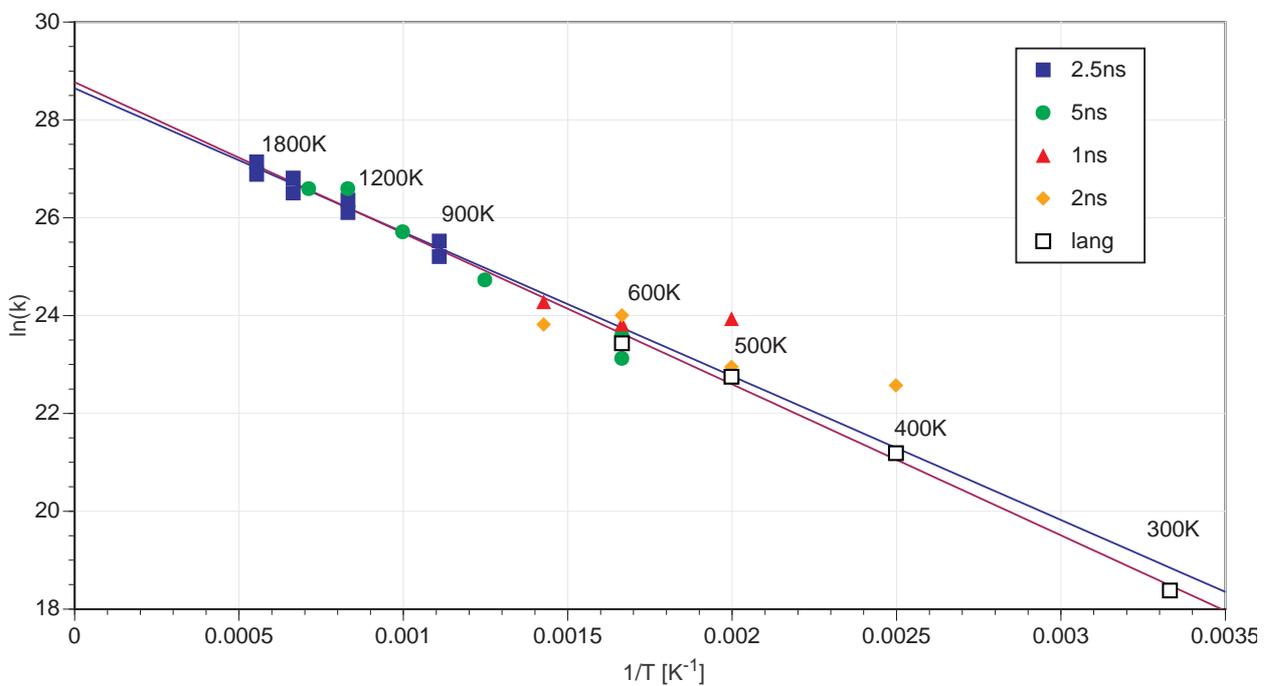
**Abb. X146:** Arrhenius-Analyse für die Rotameren von L56O (Maltose, CVFF,  $\epsilon = 4$ , Gap = 10%)



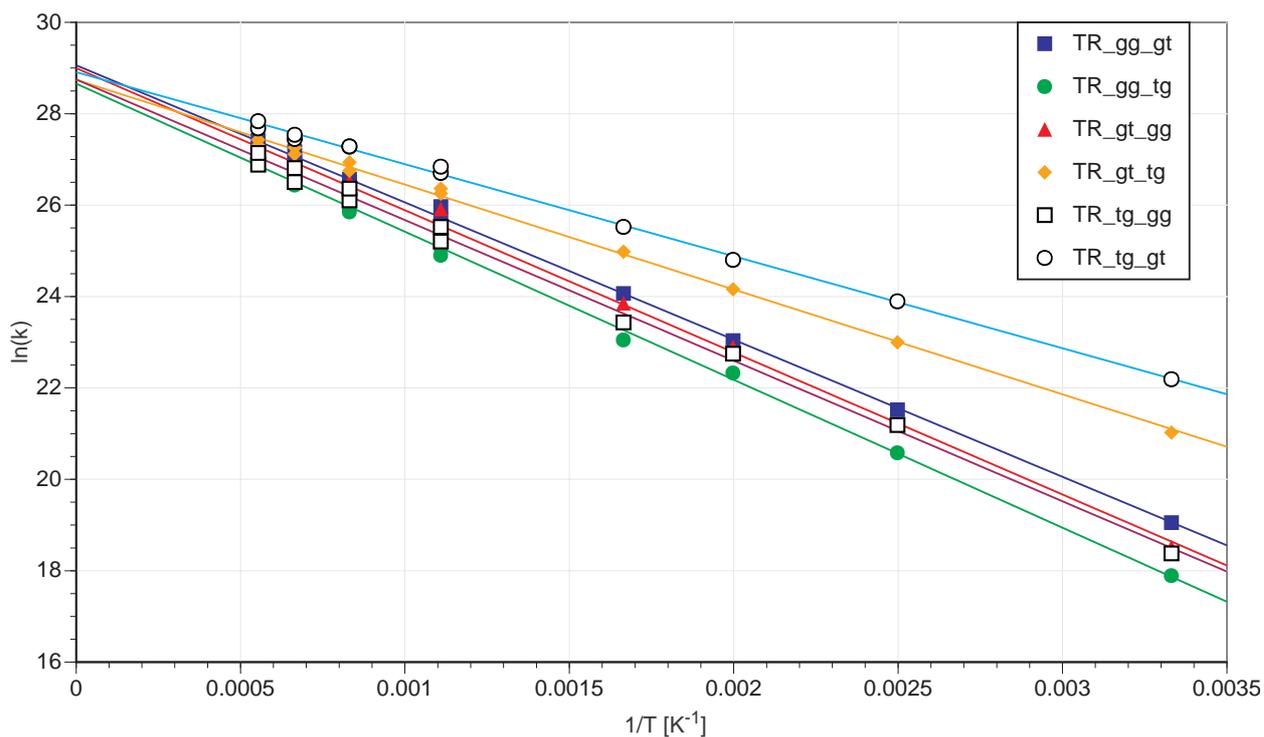
**Abb. X147:** Übergangsennergien aus der Arrhenius-Analyse für die Hydroxymethylgruppen (Maltose, CVFF,  $\epsilon = 4$ , Gap = 10%)



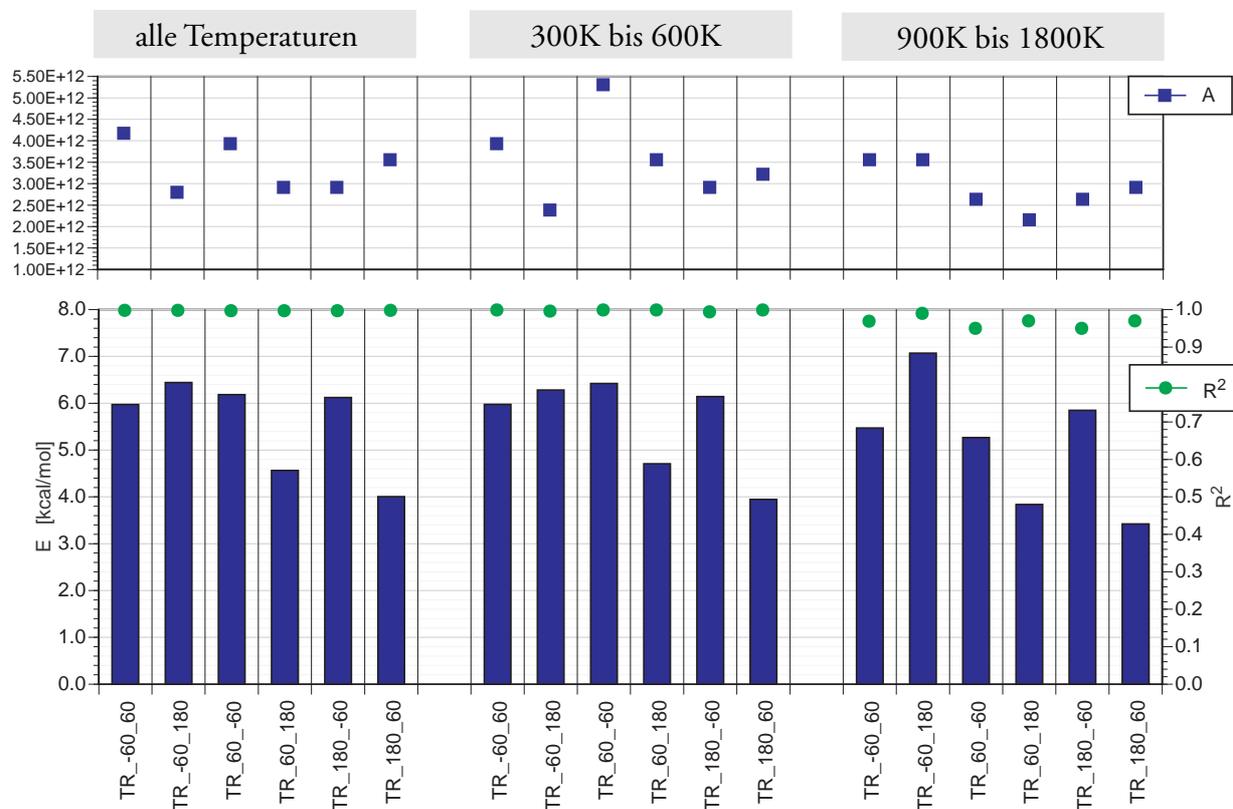
**Abb. X148:** Arrhenius-Analyse für den Übergang  $gt \rightarrow gg$  (Maltose, L56O, CVFF,  $\epsilon = 4$ , Gap=12°)  
 Simulationszeit „lang“:  $t_s(300K) = 2 \mu s$ ;  $t_s(400K) = 300 \text{ ns}$ ;  $t_s(500K) = 100 \text{ ns}$ ;  $t_s(600K) = 30 \text{ ns}$



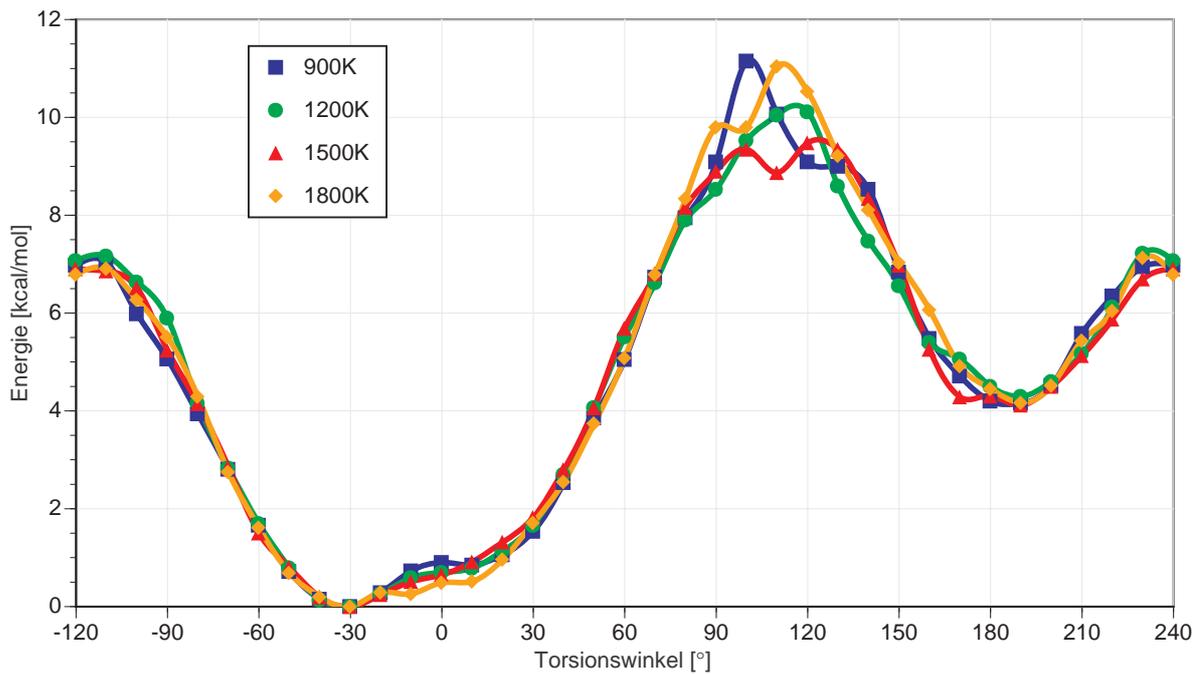
**Abb. X149:** Arrhenius-Analyse für den Übergang  $tg \rightarrow gg$  (Maltose, L56O, CVFF,  $\epsilon = 4$ , Gap=12°)  
 Simulationszeit „lang“:  $t_s(300K) = 2 \mu s$ ;  $t_s(400K) = 300 \text{ ns}$ ;  $t_s(500K) = 100 \text{ ns}$ ;  $t_s(600K) = 30 \text{ ns}$



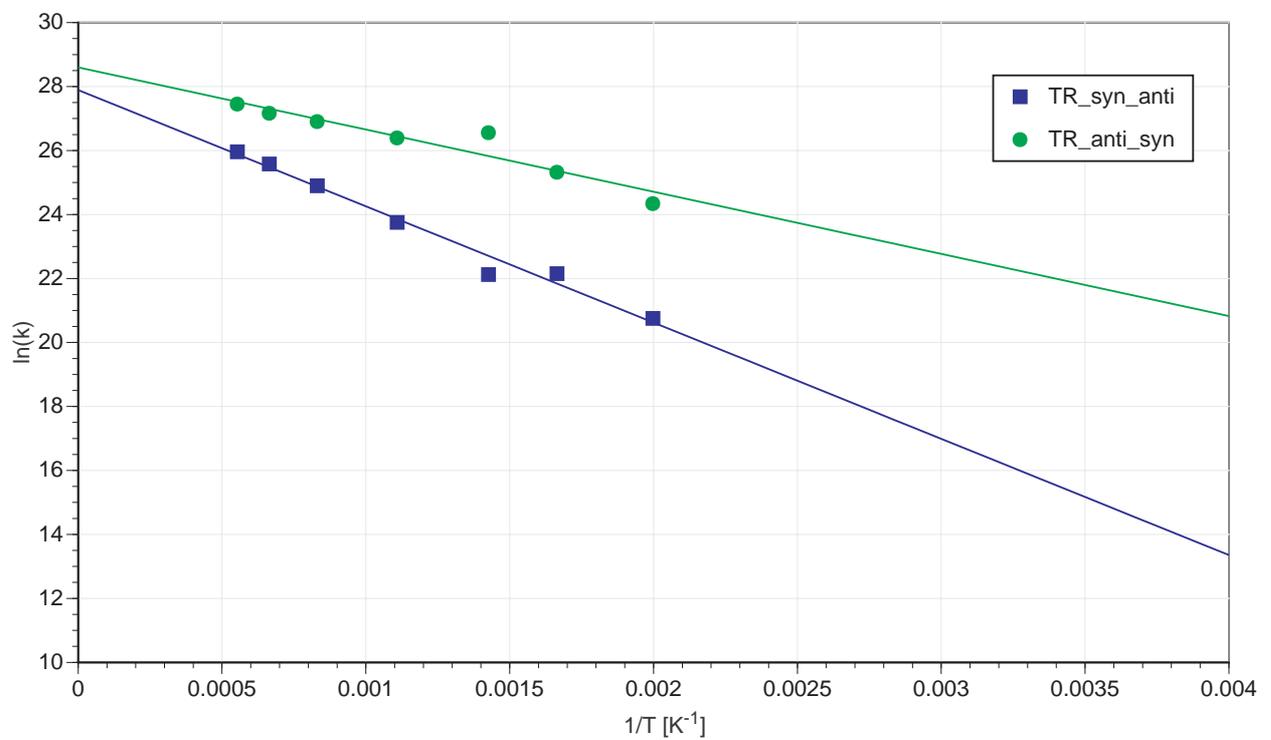
**Abb. X150:** Arrhenius-Analyse von Hydroxymethyl-Rotamerenübergangsbarrieren (Maltose, L56O, CVFF,  $\epsilon = 4$ , Gap =  $12^\circ$ )



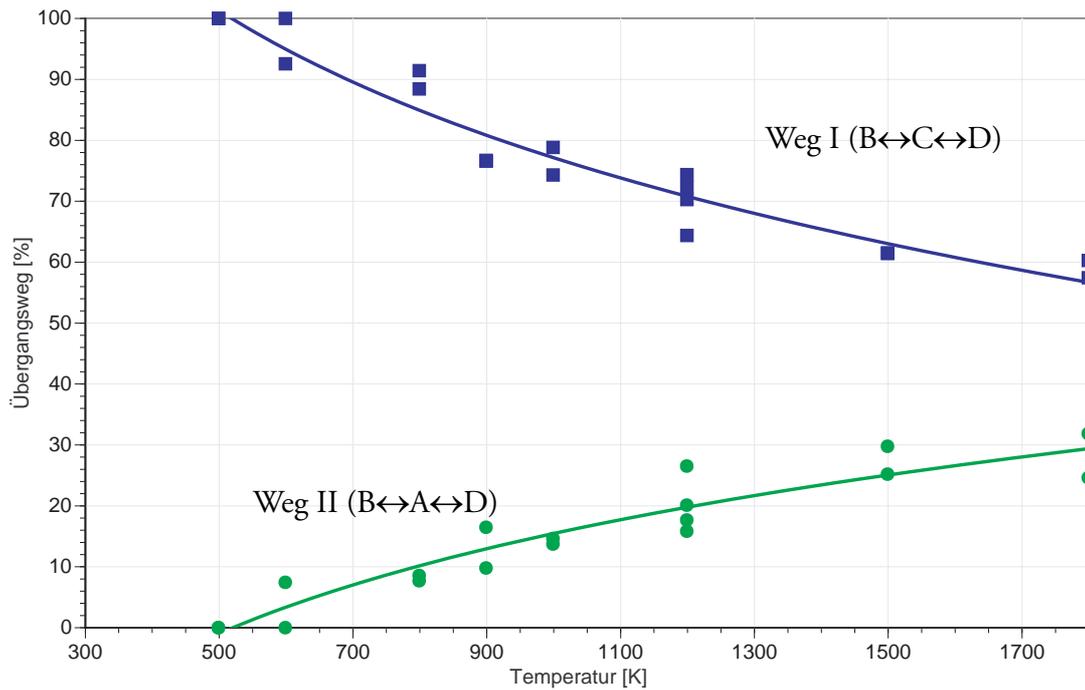
**Abb. X151:** Übergangsenergien aus Arrhenius-Analyse (Maltose, L56O, CVFF,  $\epsilon = 4$ , Gap =  $12^\circ$ )



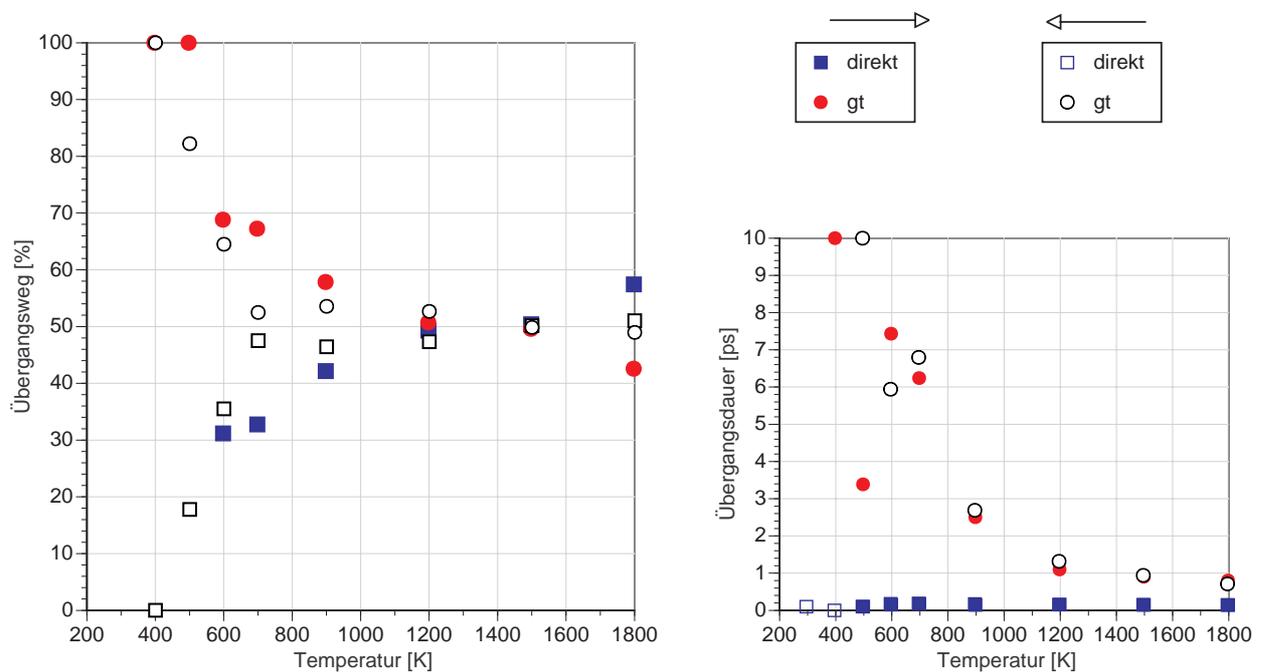
**Abb. X152:**  $\Psi$ -FGE-Profil in Abhängigkeit von der Simulationstemperatur (Maltose: CVFF,  $\epsilon = 4$ )



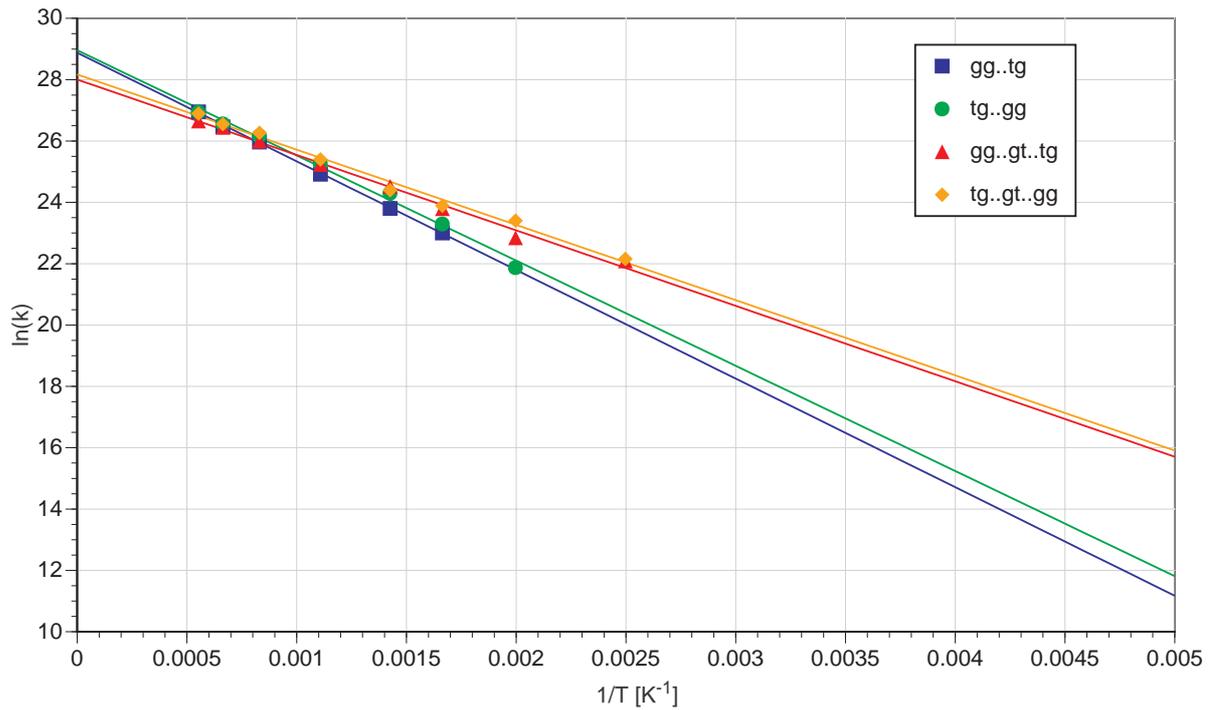
**Abb. X153:** Arrhenius-Analyse von  $\Psi$ -Rotamerenübergangsbarrieren (Maltose, CVFF,  $\epsilon = 4$ )



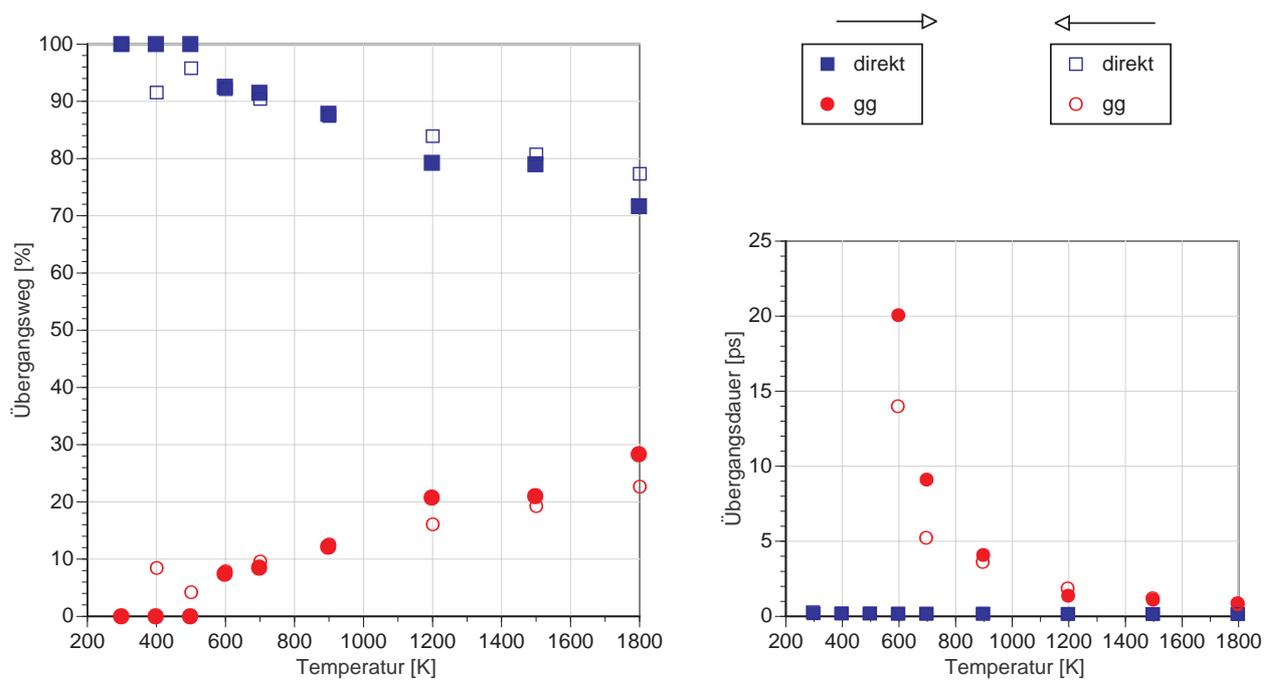
**Abb. X154:** Übergangsweg syn↔anti in Abhängigkeit von der Simulationstemperatur (Maltose, CVFF,  $\varepsilon = 4$ )



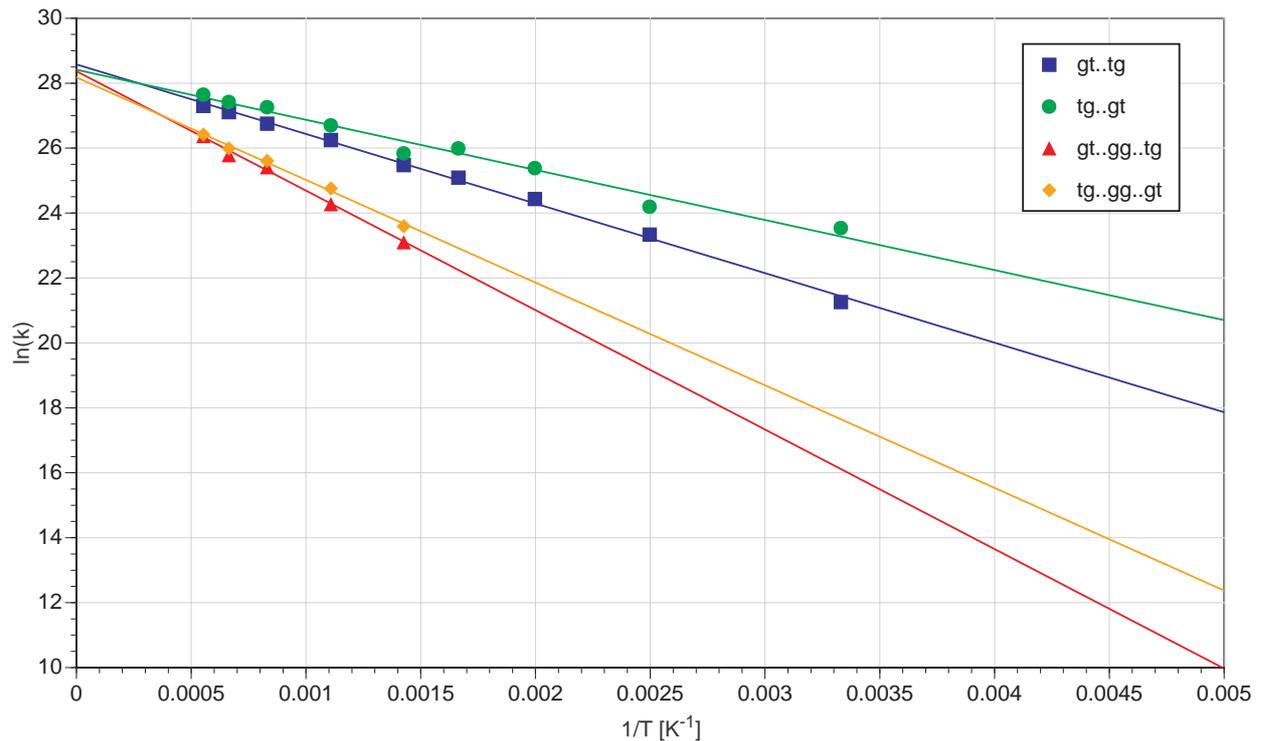
**Abb. X155:** Übergangsweg gg↔tg in Abhängigkeit von der Simulationstemperatur (Maltose, L56O, CVFF,  $\varepsilon = 4$ , Gap=10%)



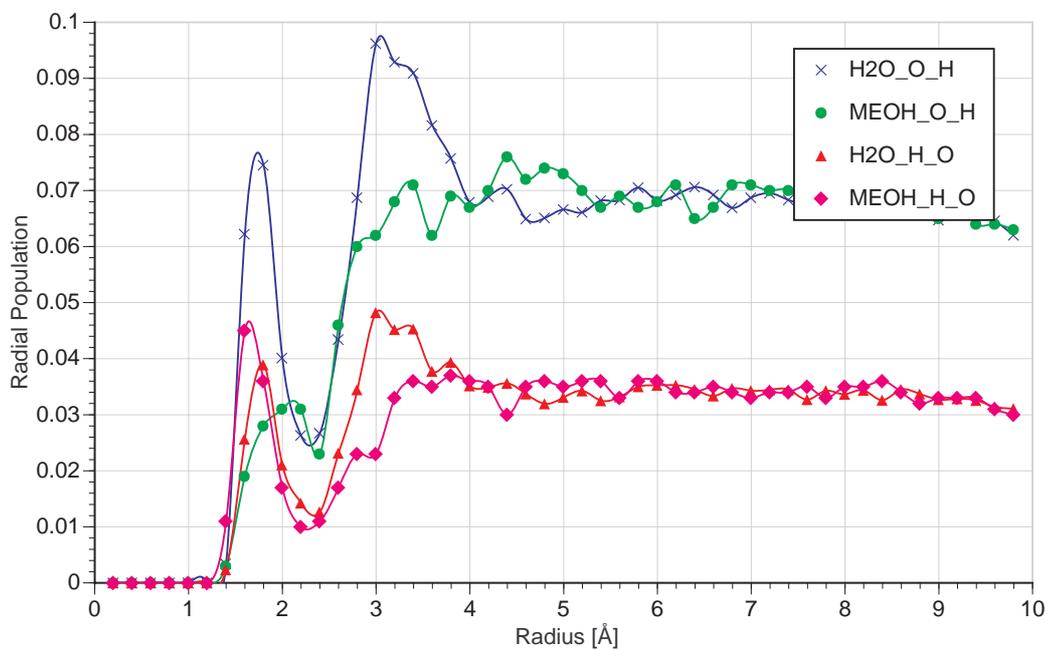
**Abb. X156:** Arrhenius-Analyse für den Übergang  $gg \leftrightarrow tg$  (Maltose, L56O, CVFF,  $\varepsilon = 4$ , Gap = 10%)



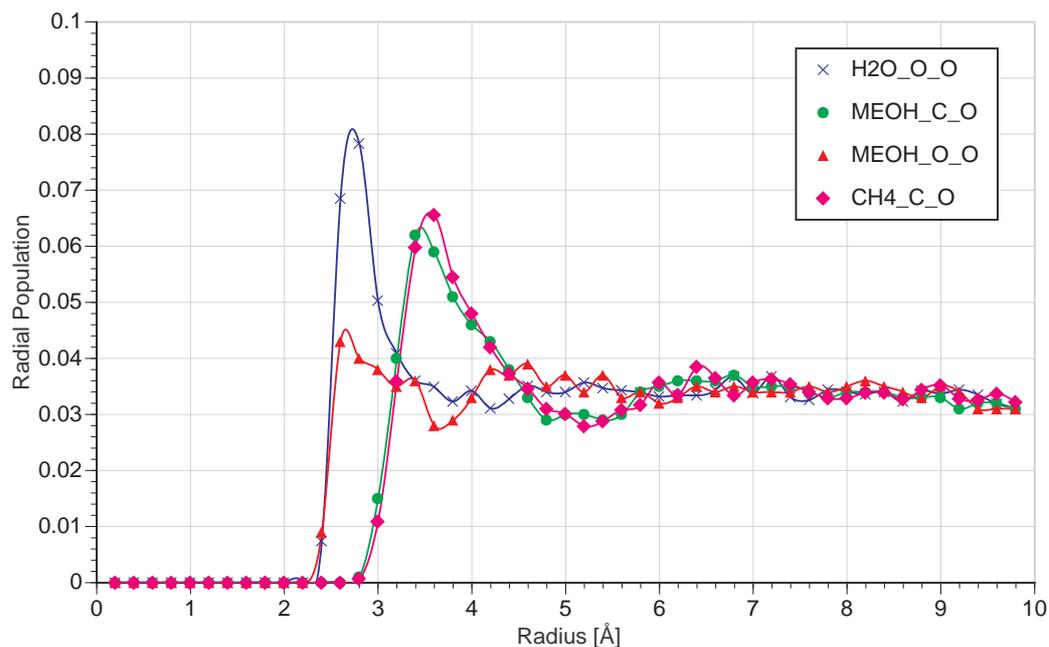
**Abb. X157:** Übergangsweg  $gt \leftrightarrow tg$  in Abhängigkeit von der Simulationstemperatur (Maltose, L56O, CVFF,  $\varepsilon = 4$ , Gap = 10%)



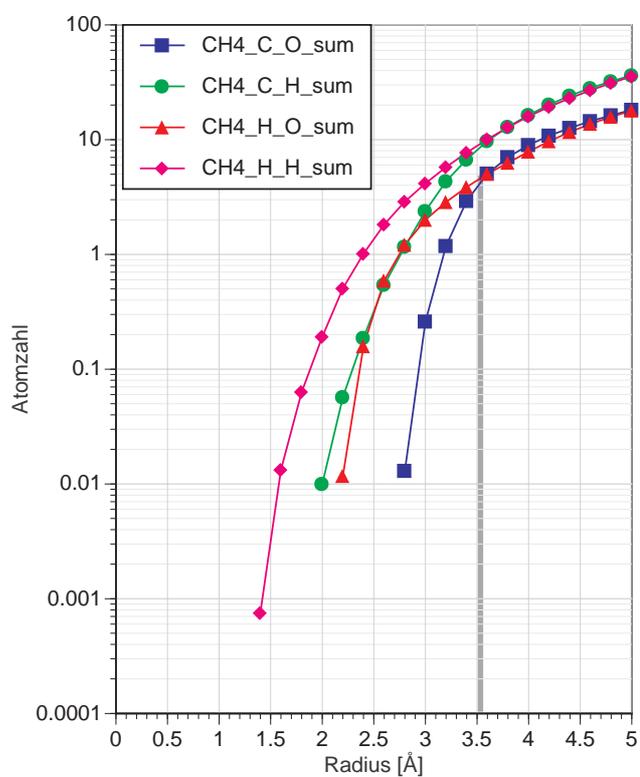
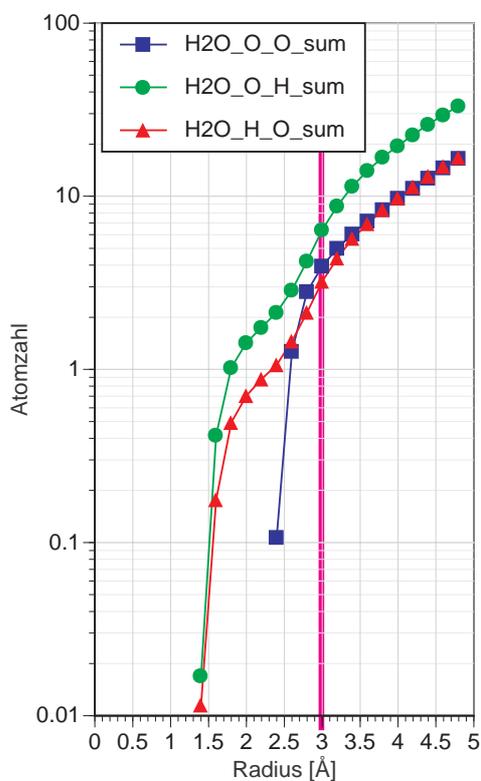
**Abb. X158:** Arrhenius-Analyse für den Übergang  $gt \leftrightarrow tg$  (Maltose, L56O, CVFF,  $\epsilon = 4$ , Gap = 10%)



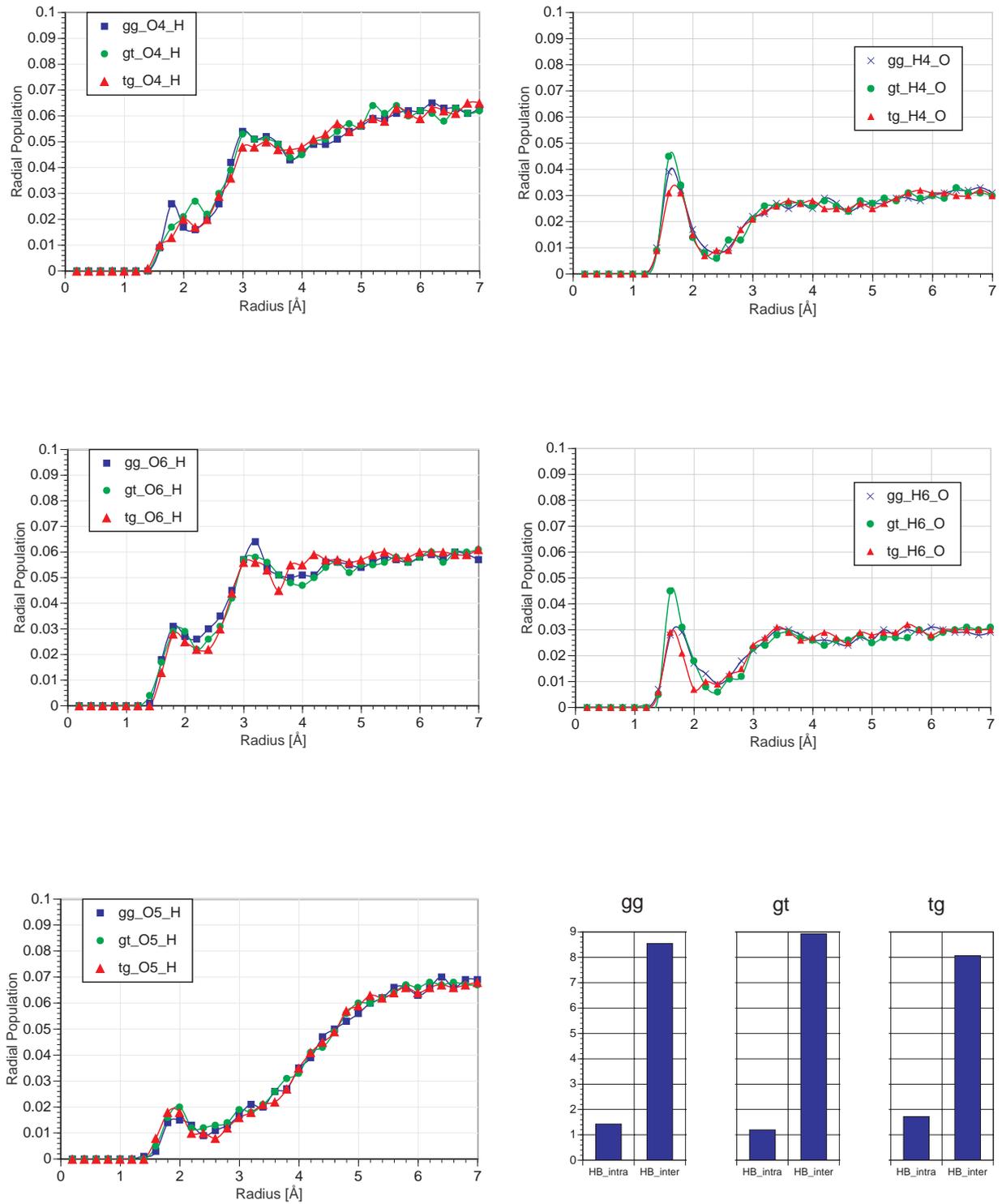
**Abb. X159:** Radial-Histogramme für MD-Simulationen von Wasser, Methan und Methanol in einer Wasserbox der Größe  $20\text{\AA} \times 20\text{\AA} \times 20\text{\AA}$  (CVFF, 300K,  $t_s = 300$  ps,  $\Delta t_s = 1$  ps)



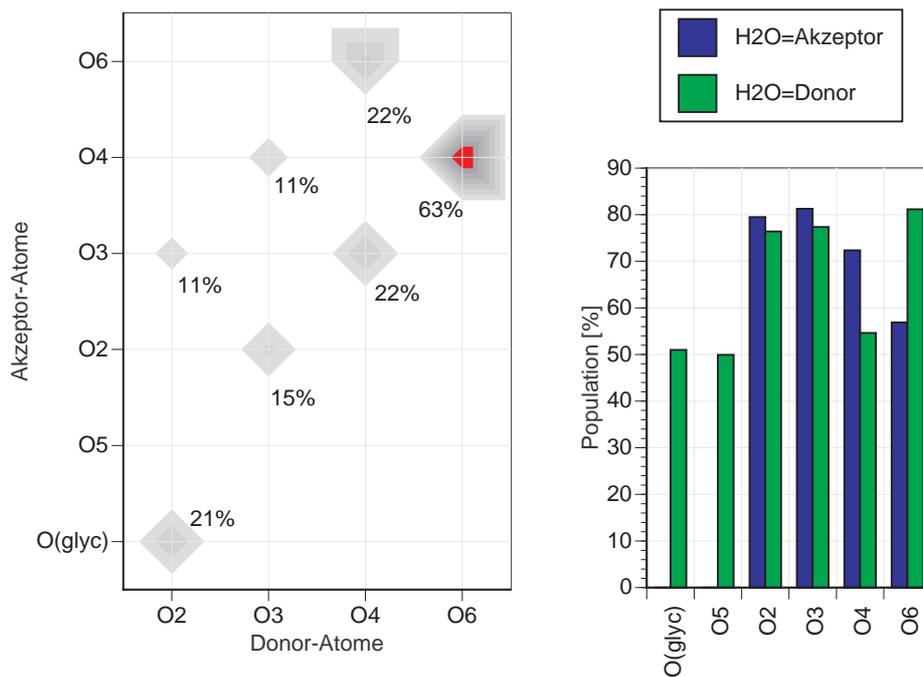
**Abb. X160:** Radial-Histogramme für MD-Simulationen von Wasser, Methan und Methanol in einer Wasserbox der Größe  $20\text{\AA} \times 20\text{\AA} \times 20\text{\AA}$  (CVFF, 300 K,  $t_s = 300$  ps,  $\Delta t_s = 1$  ps)



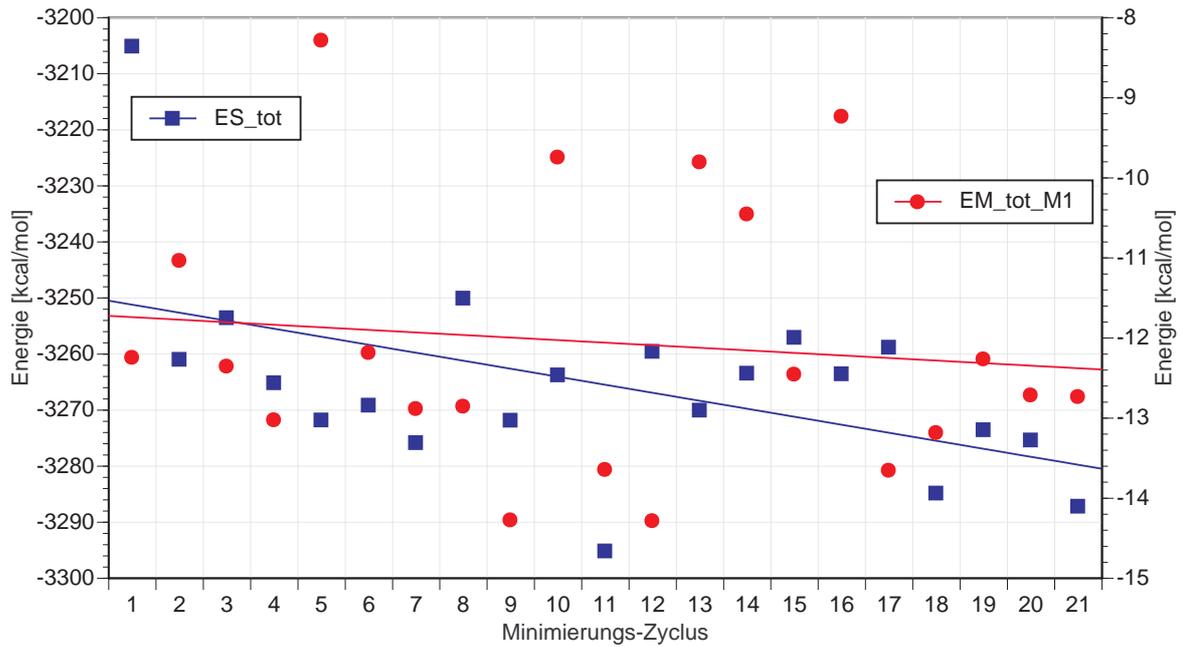
**Abb. X161:** Berechnung der Anzahl der benachbarten Solvensatome in Abhängigkeit vom Suchradius für MD-Simulationen von Wasser und Methan in einer Wasserbox der Größe  $20\text{\AA} \times 20\text{\AA} \times 20\text{\AA}$  (CVFF, 300 K,  $t_s = 300$  ps,  $\Delta t_s = 1$  ps)



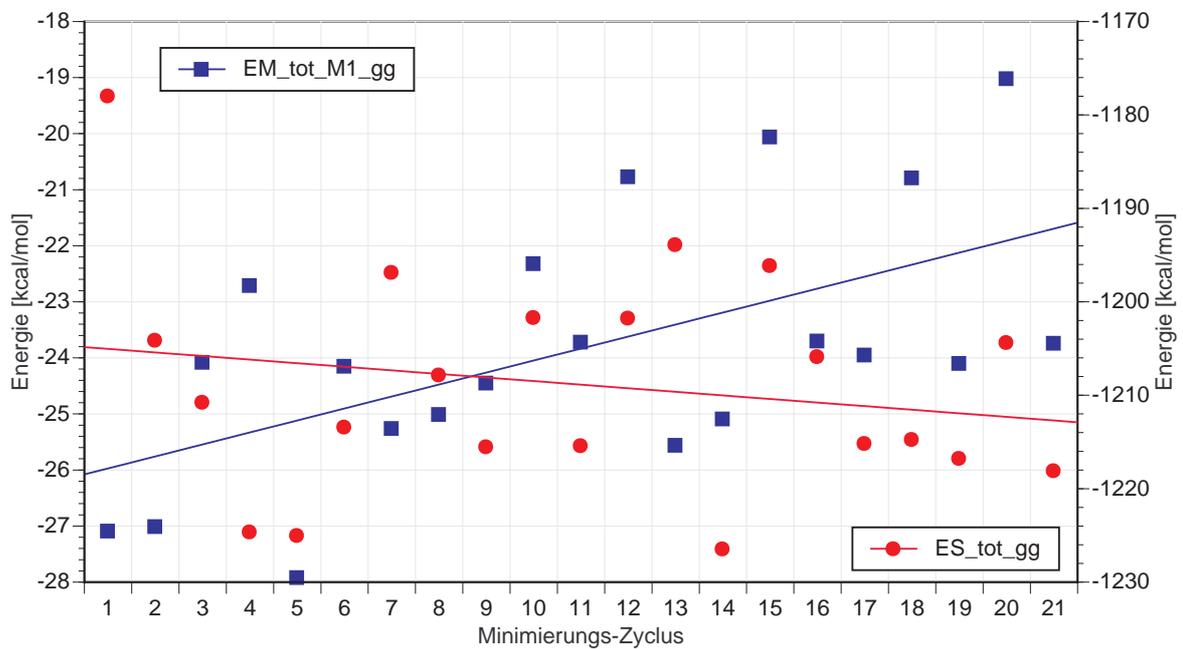
**Abb. X162:** Glc $\alpha$ -O-iProp: Radial-Histogramme und mittlere Anzahl der H-Brücken (CVFF, 300K, PBC (20x20x20),  $\epsilon = 1$ ,  $t_s = 300$  ps,  $\Delta t_s = 1$  ps)



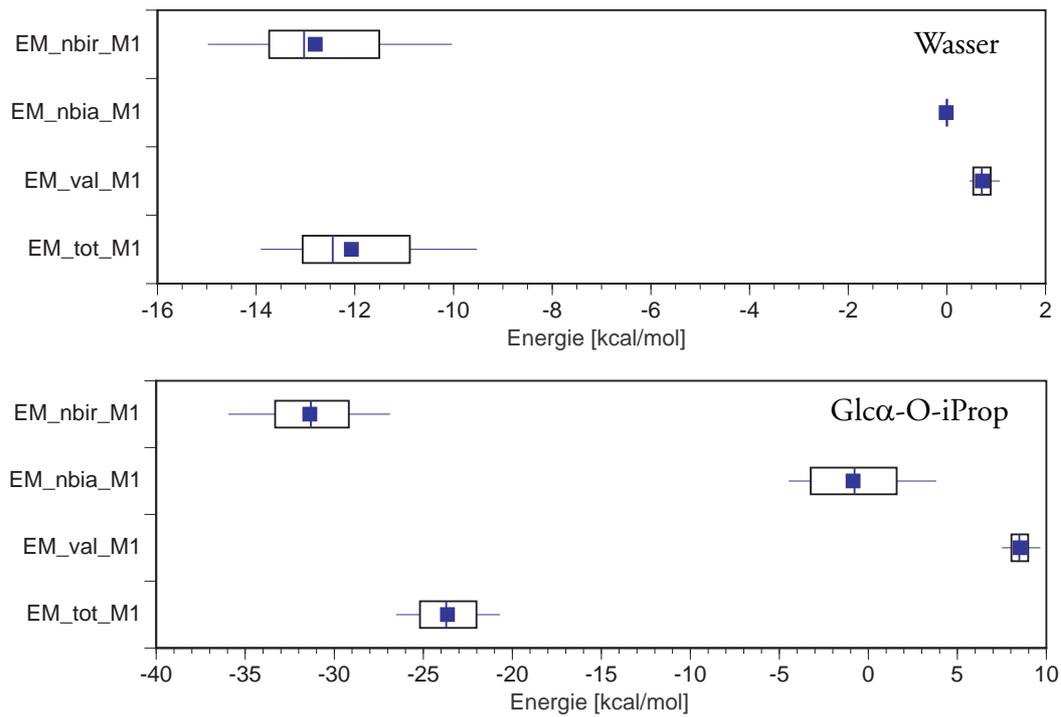
**Abb. X163:** Glc $\alpha$ -O-iProp: Detailanalyse der H-Brücken im Solvens (CVFF, 300K,  $t_s = 300$  ps,  $\Delta t_s = 1$  ps, PBC (20x20x20),  $\epsilon=1$ )



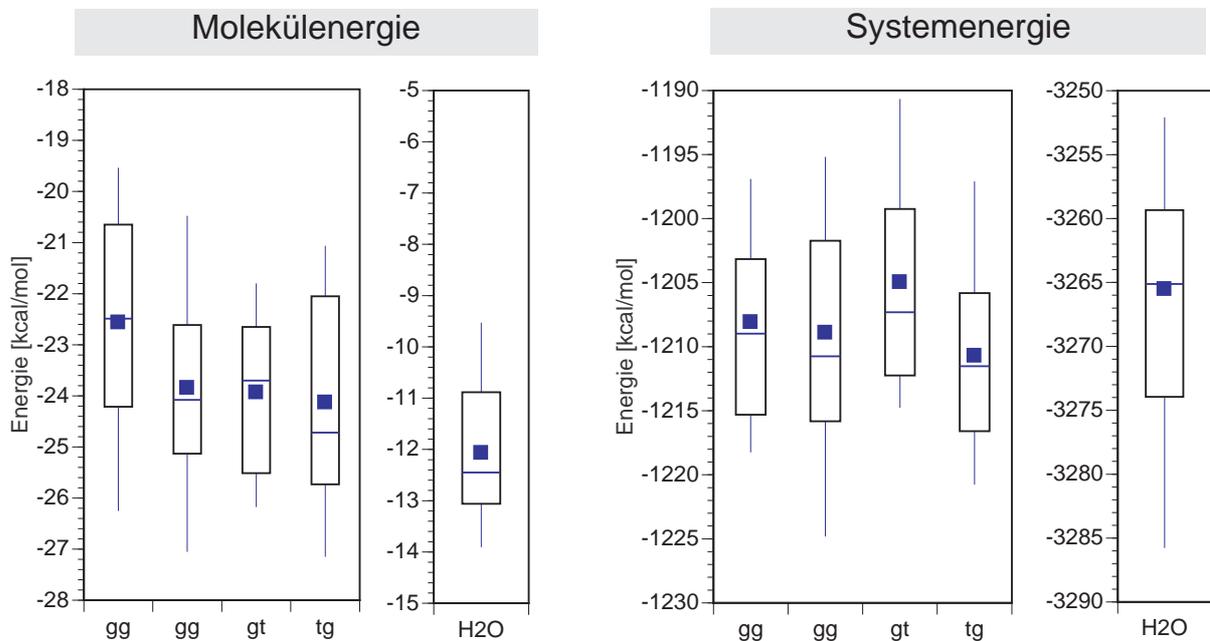
**Abb. X164:** Änderung der Energie bei der MD-Minimierung eines Wassermoleküls in einer Wasserbox (CVFF, 200 K / 10ps Equilibrierung)



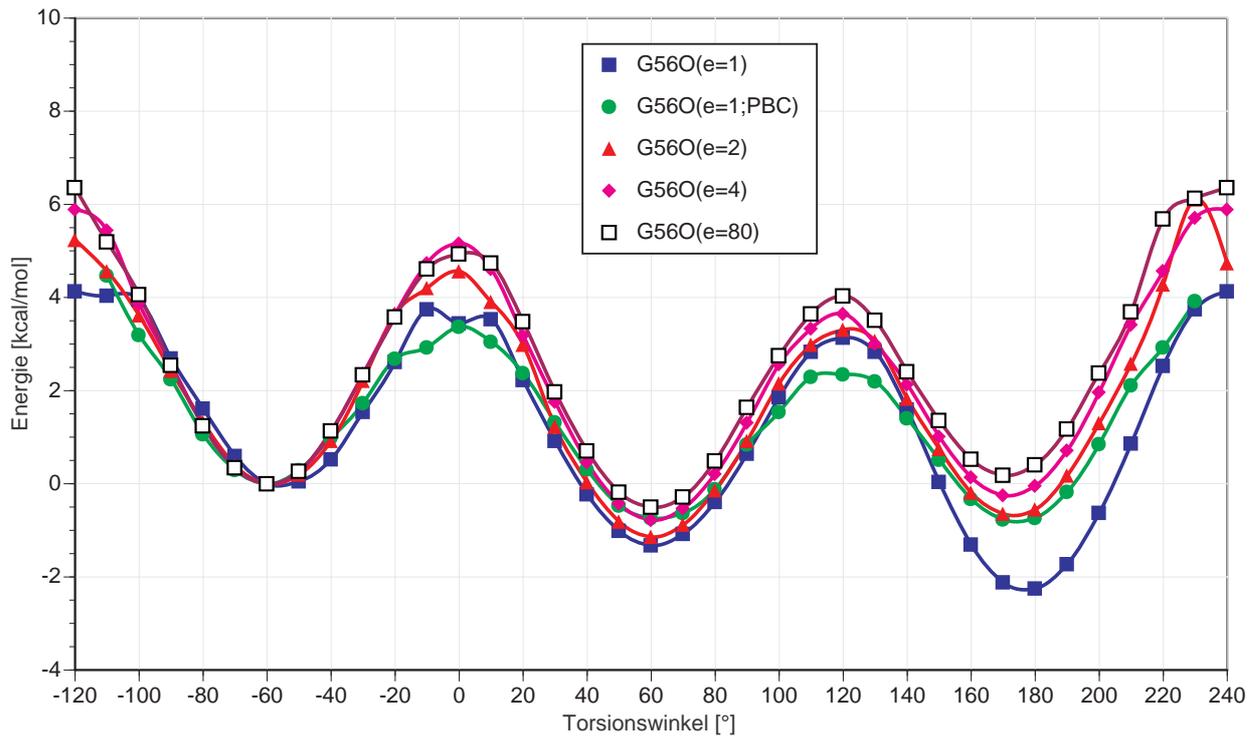
**Abb. X165:** MD-Minimierung von Glc $\alpha$ -O-iProp in einer Wasserbox (CVFF, 200 K / 10ps Equilibrierung)



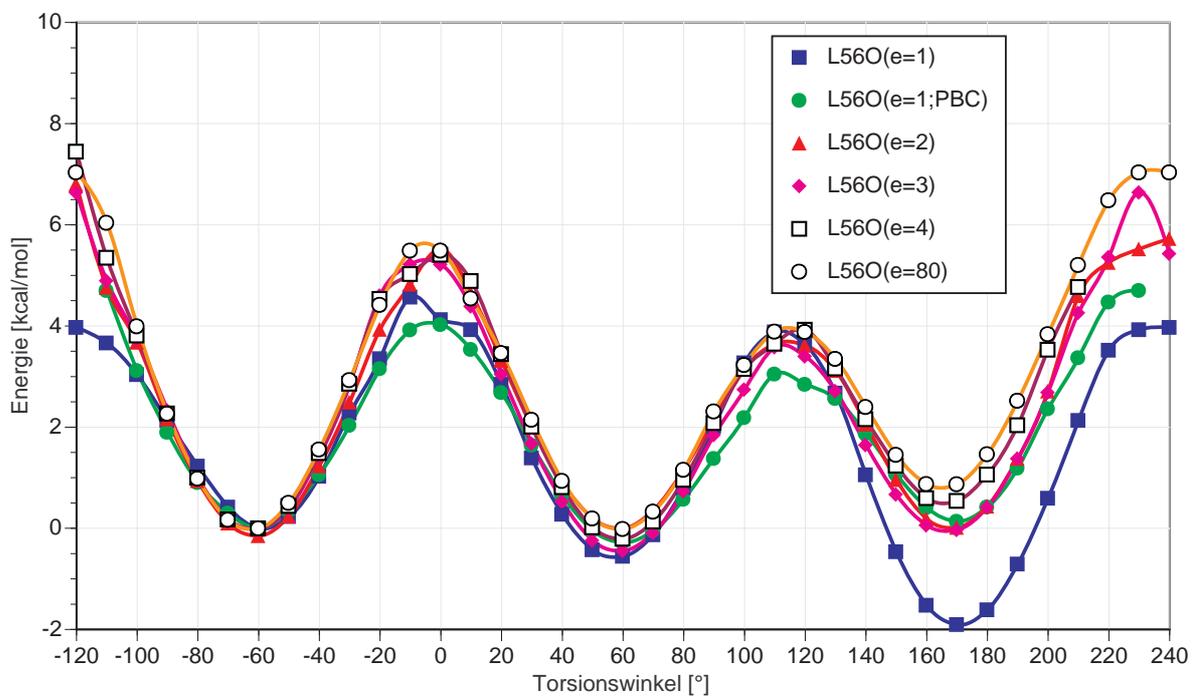
**Abb. X166:** Schwankungen der Energie des solvatisierten Moleküls (MD-Minimierung, CVFF, 200K/10ps Equilibrierung)



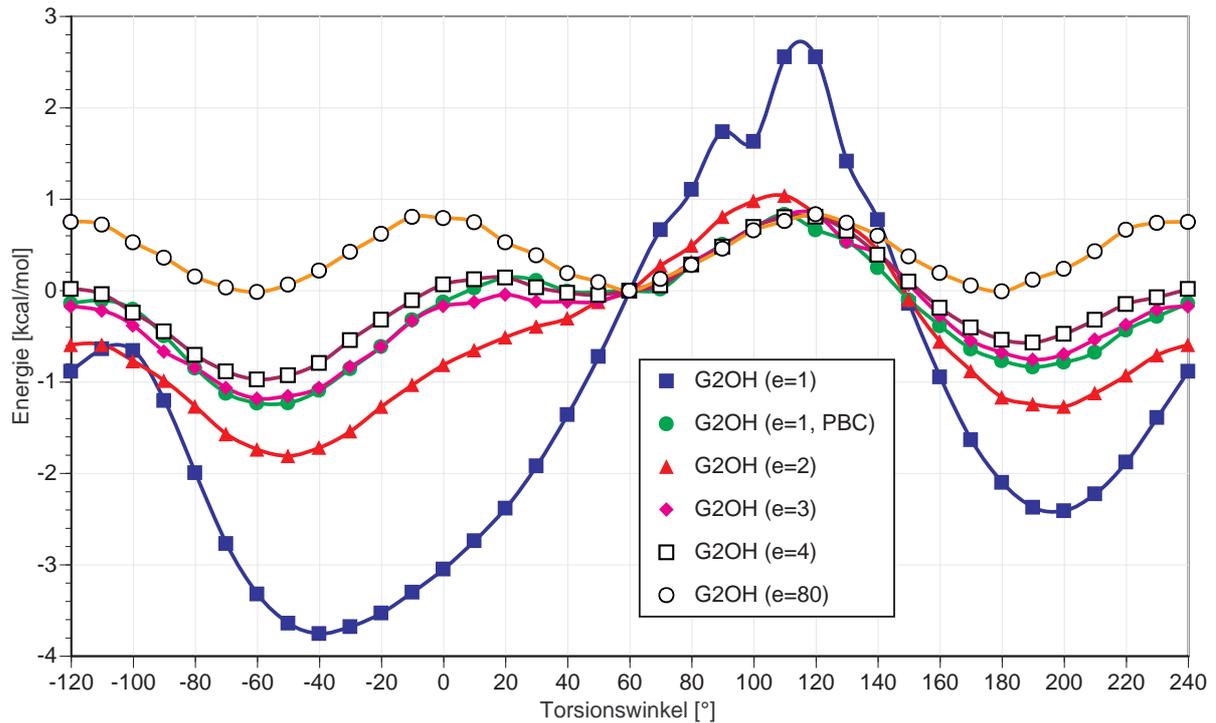
**Abb. X167:** MD-Minimierung eines Wassermoleküls und von Glcα-O-iProp in einer Wasserbox (Boxplots: 50% der Werte liegen innerhalb der Box). Die MD-Minimierung des gg-Zustands wurde zweimal durchgeführt.



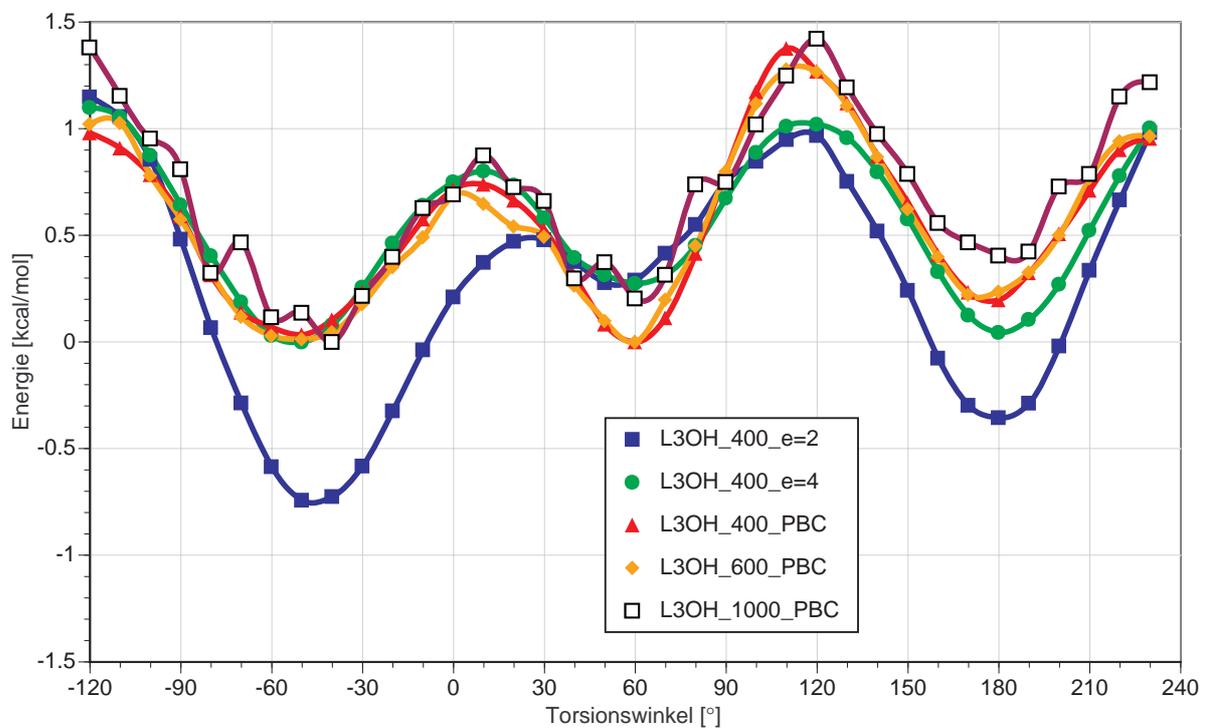
**Abb. X168:** Abhängigkeit der Energieprofils von der Dielektrizitätskonstanten (Maltose, G56O, CVFF, 400K) (PBC = „Periodic Boundary Conditions“, Wasserbox  $20\text{\AA} \times 20\text{\AA} \times 20\text{\AA}$ ) (Normierung der Energie auf  $-60^\circ$ )



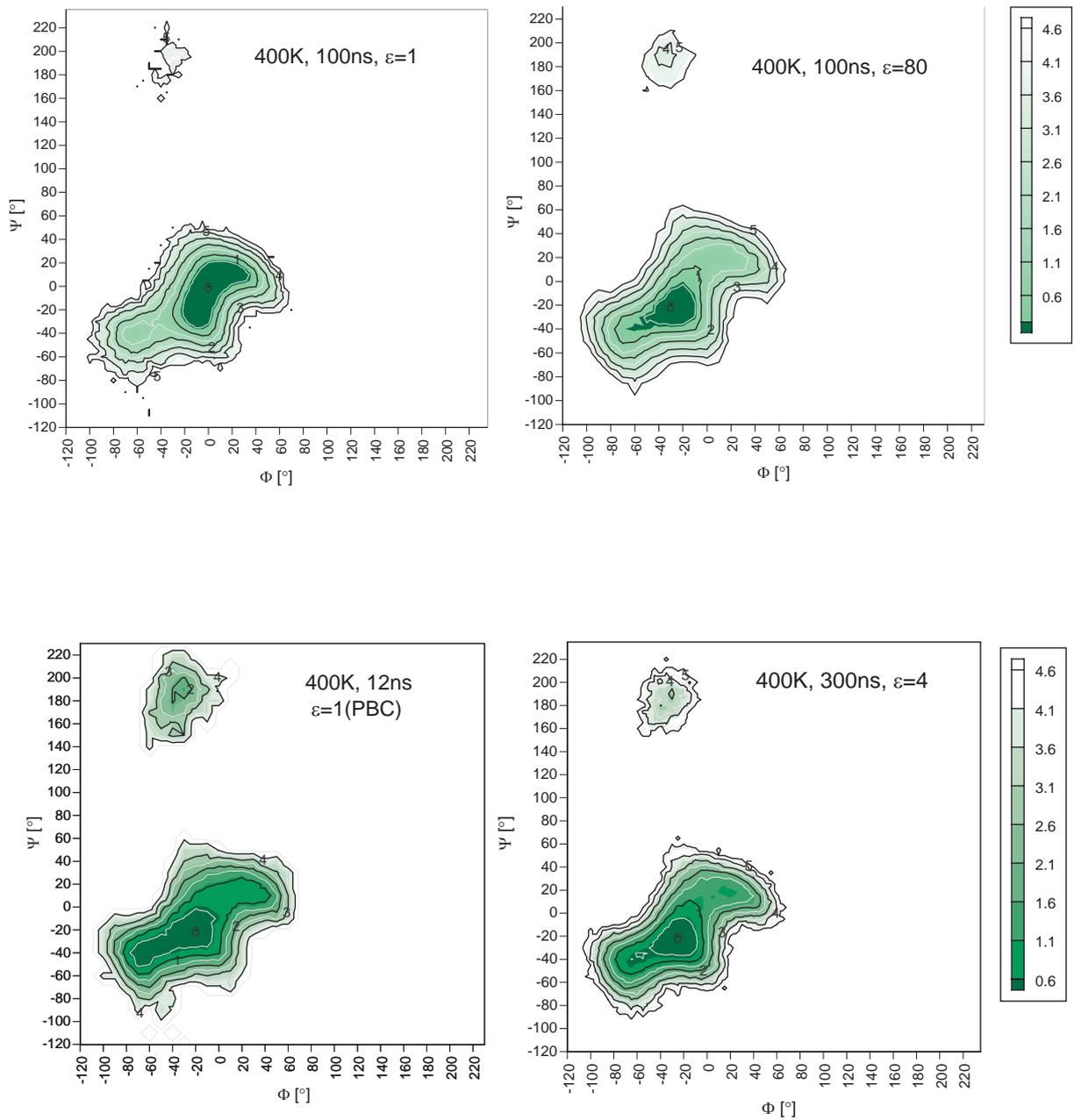
**Abb. X169:** Abhängigkeit der Energieprofils von der Dielektrizitätskonstanten (Maltose, L56O, CVFF, 400K) (PBC = „Periodic Boundary Conditions“, Wasserbox  $20\text{\AA} \times 20\text{\AA} \times 20\text{\AA}$ ) (Normierung der Energie auf  $-60^\circ$ )



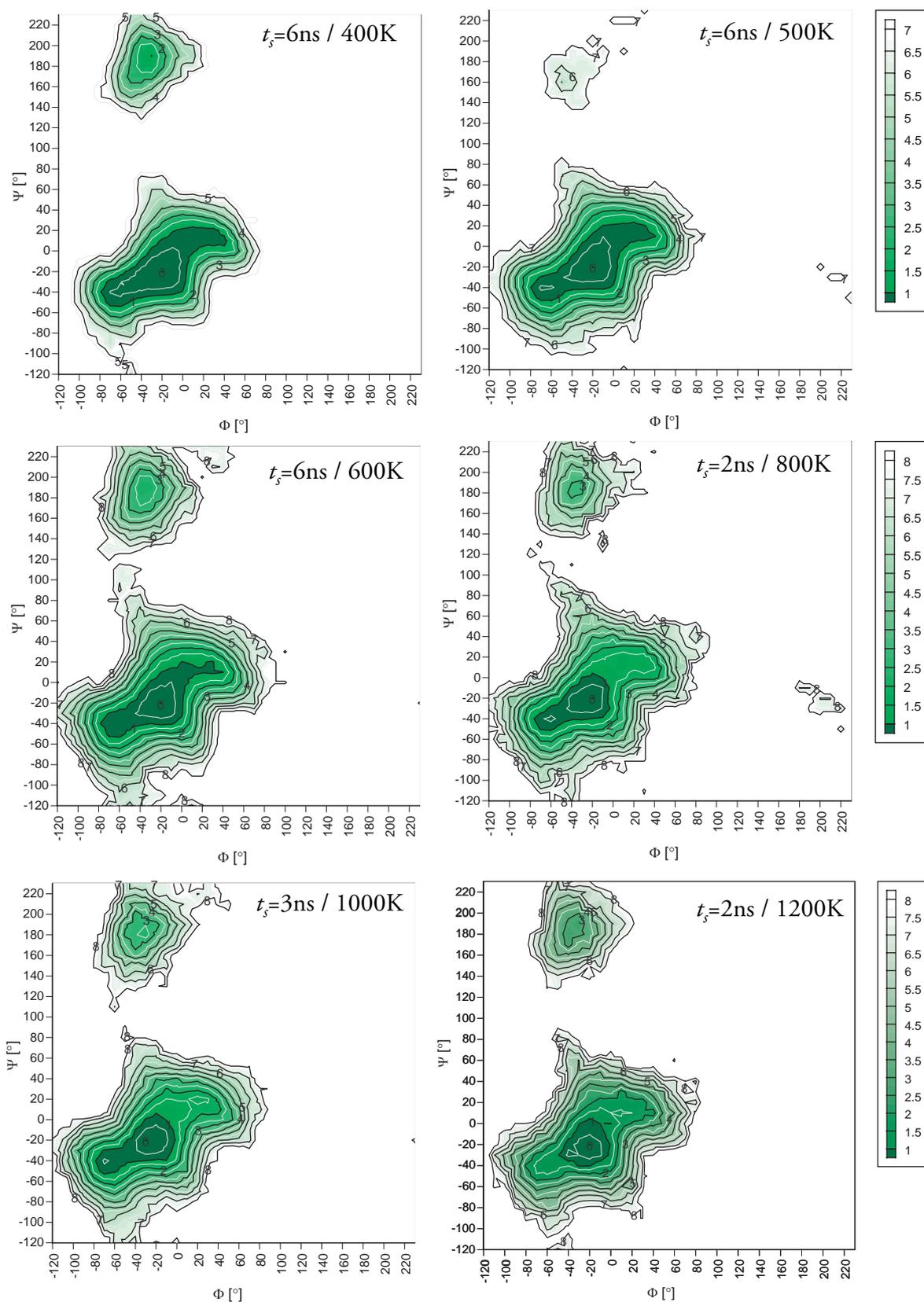
**Abb. X170:** Abhängigkeit der Energieprofils von der Dielektrizitätskonstanten (Maltose, G2OH, CVFF, 400K) (PBC = „Periodic Boundary Conditions“, Wasserbox  $20\text{\AA} \times 20\text{\AA} \times 20\text{\AA}$ ) (Normierung der Energie auf  $+60^\circ$ )



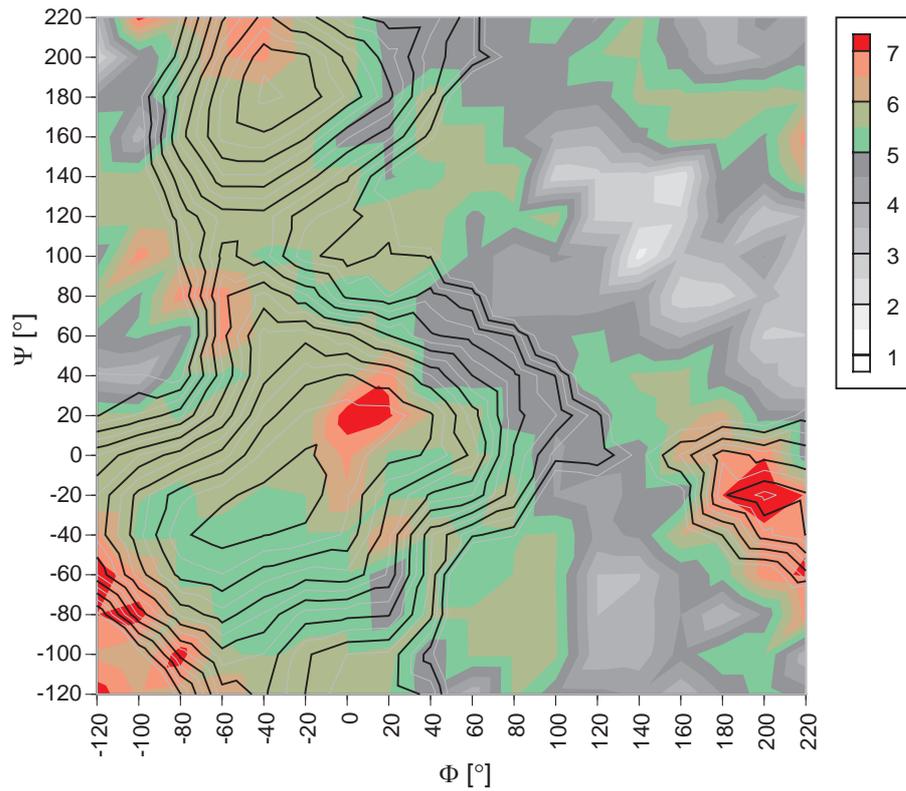
**Abb. X171:** Abhängigkeit der FGE-Profiles von der Temperatur und von der Dielektrizitätskonstanten (Maltose, L3OH, CVFF) (PBC = „Periodic Boundary Conditions“, Wasserbox  $20\text{\AA} \times 20\text{\AA} \times 20\text{\AA}$ )



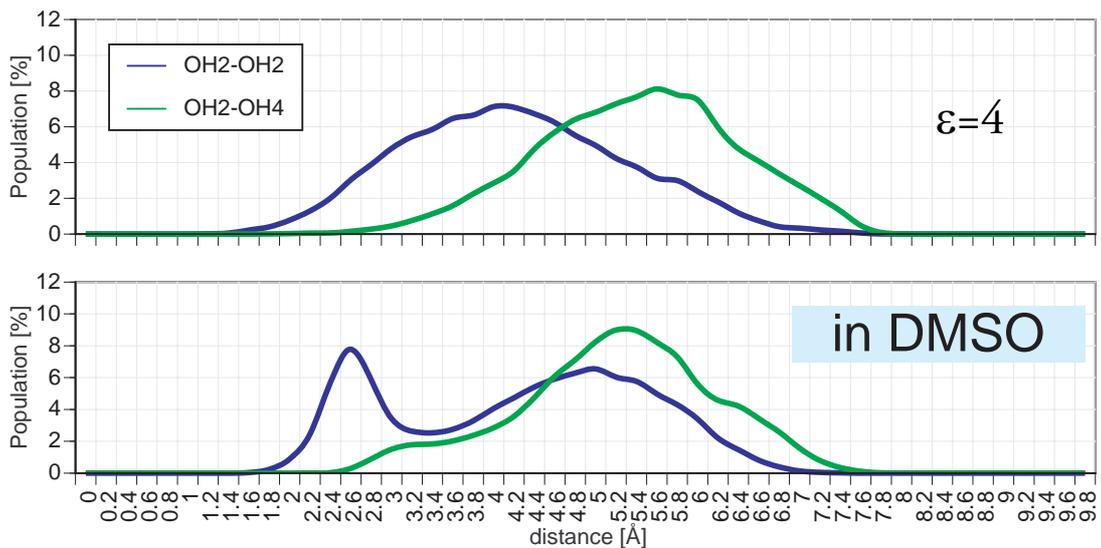
**Abb. X172:**  $\Phi$  /  $\Psi$  -FGE-Map in Abhängigkeit von der Dielektrizitätskonstanten (MD-Simulation von  $\beta$ -Maltose: CVFF, 400 K)



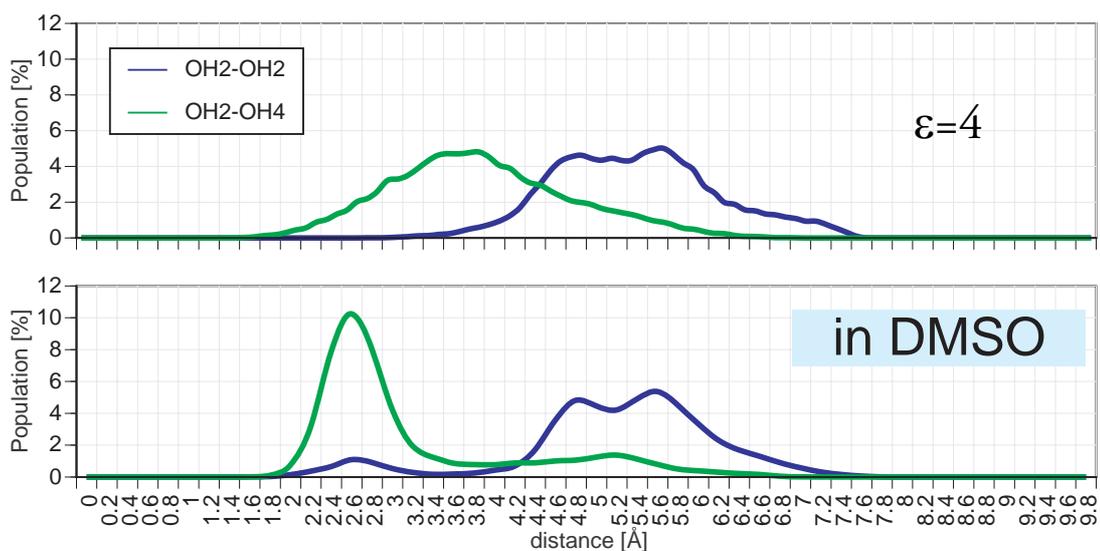
**Abb. X173:**  $\Phi / \Psi$ -FGE-Map in Abhängigkeit von der Temperatur (MD-Simulation von  $\beta$ -Maltose: CVFF, PBC(20x20x20),  $\epsilon=1$ ,  $\Delta t_s=0.1$ ps) (Startstruktur = anti-Konformation)



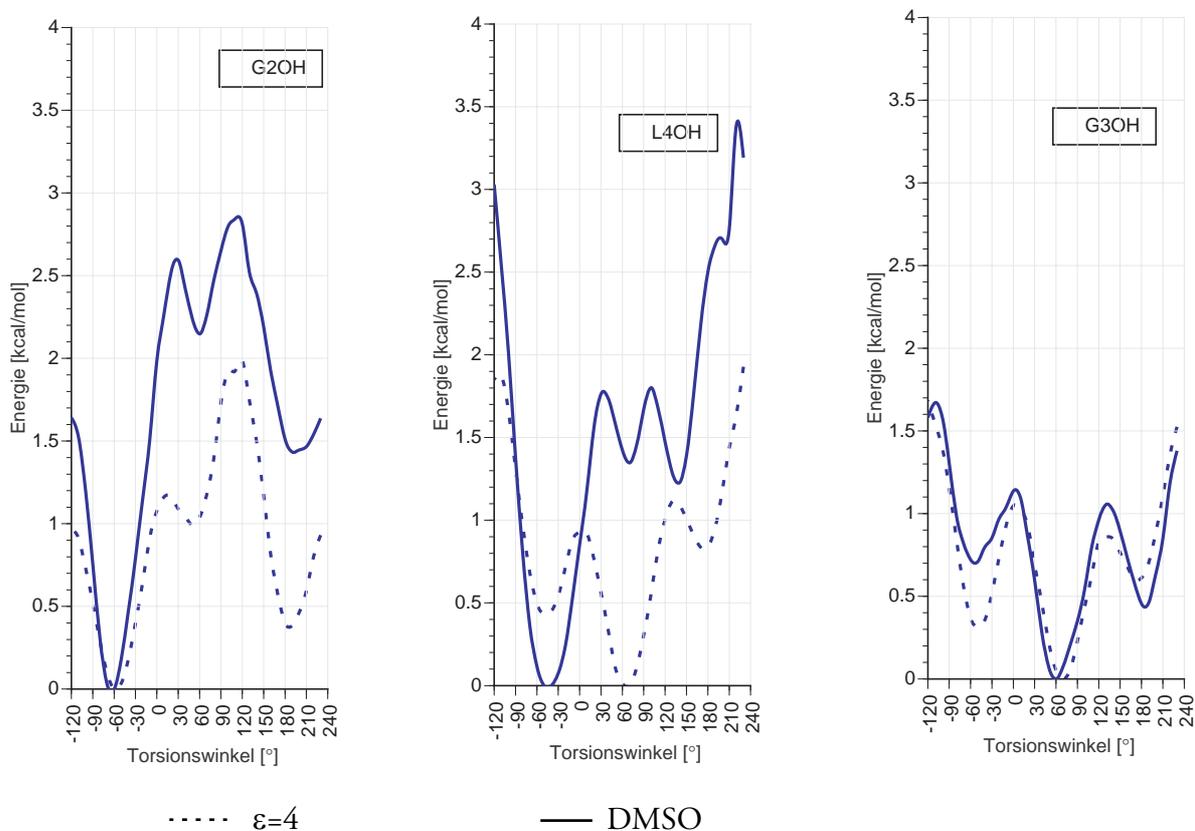
**Abb. X174:** Anzahl der intramolekularen Wasserstoffbrückenbindungen in Abhängigkeit vom  $\Phi/\Psi$  (Adiabatic Map aus  $\Phi/\Psi/\omega_1/\omega_2$ -Grid-Search)



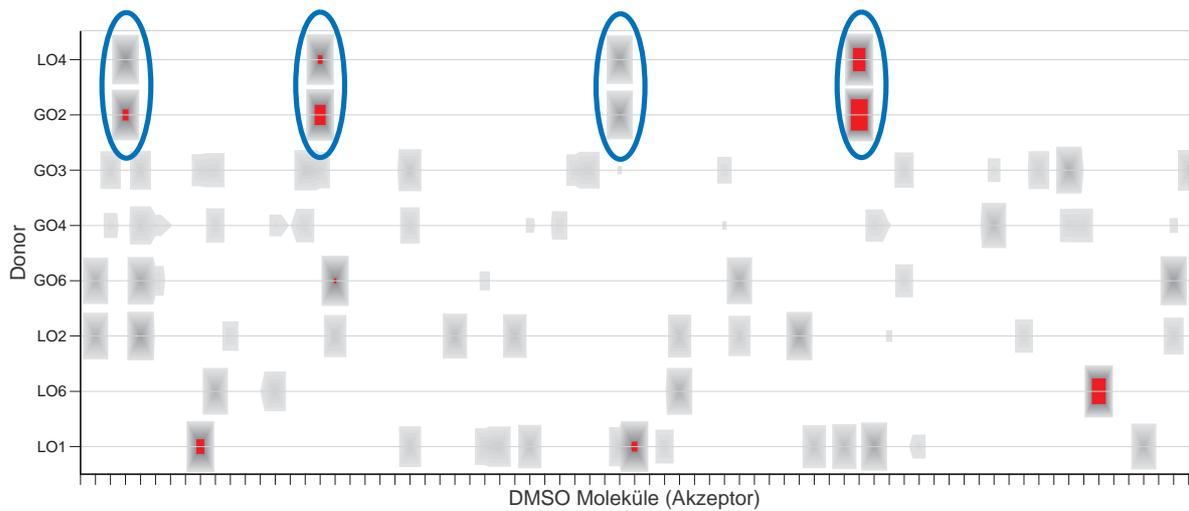
**Abb. X175:** Häufigkeitsverteilung von OH-OH-Abständen für Gal $\beta$ 1-3Gal $\beta$  (MD-Simulation 300 K, CVFE,  $t_s = 1$  ns)



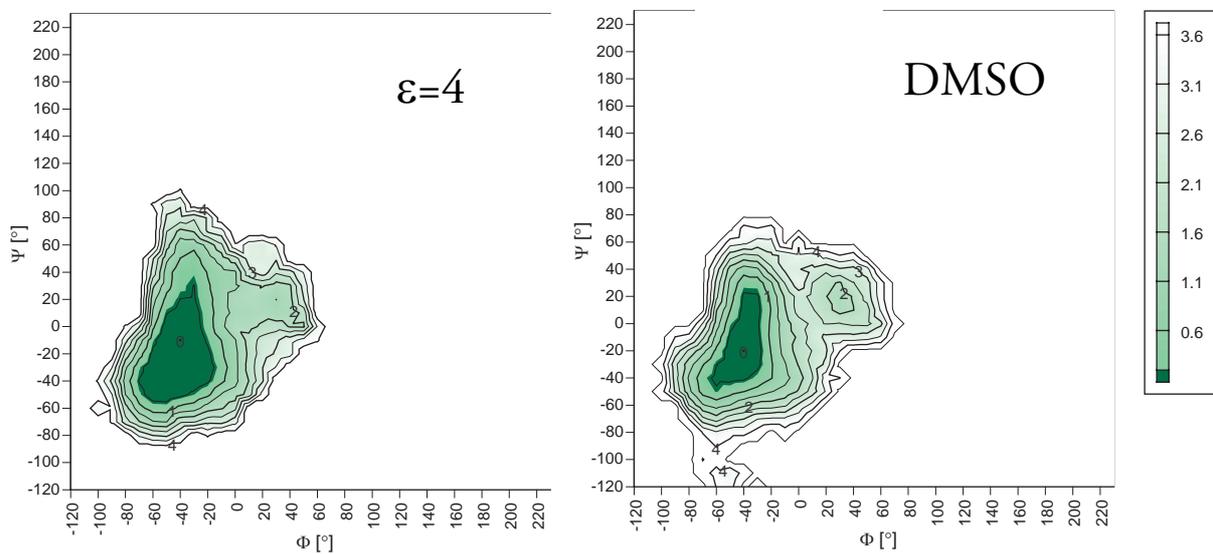
**Abb. X176:** Häufigkeitsverteilung von OH-OH-Abständen für Gal $\alpha$ 1-3Gal $\beta$  (MD-Simulation 300 K, CVFF,  $t_s = 1$  ns)



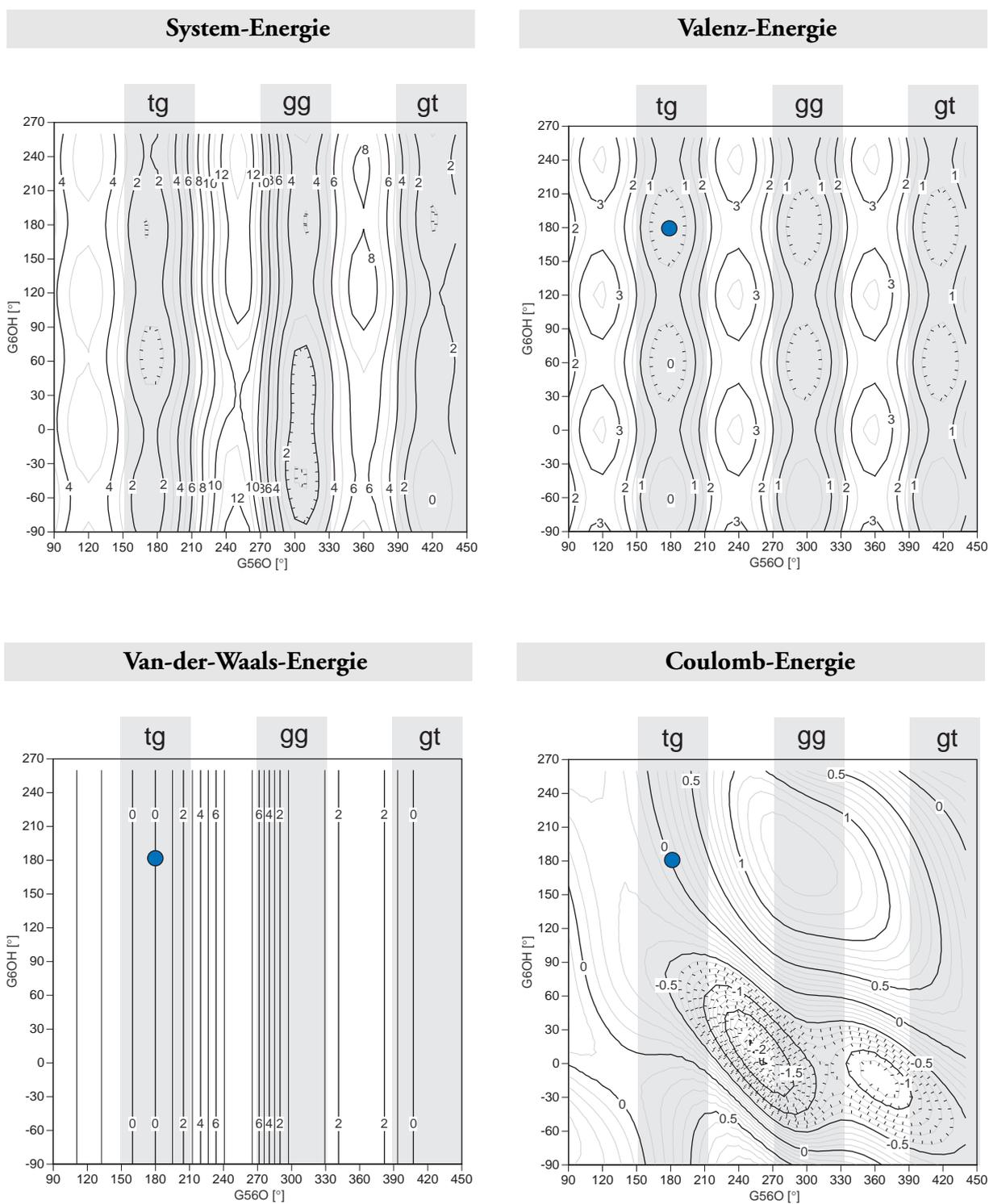
**Abb. X177:** FGE-Profile ausgewählter OH-Gruppen mit und ohne Solvens (Gal $\alpha$ 1-3Gal $\beta$ , 300K, CVFF,  $t_s = 1$  ns)



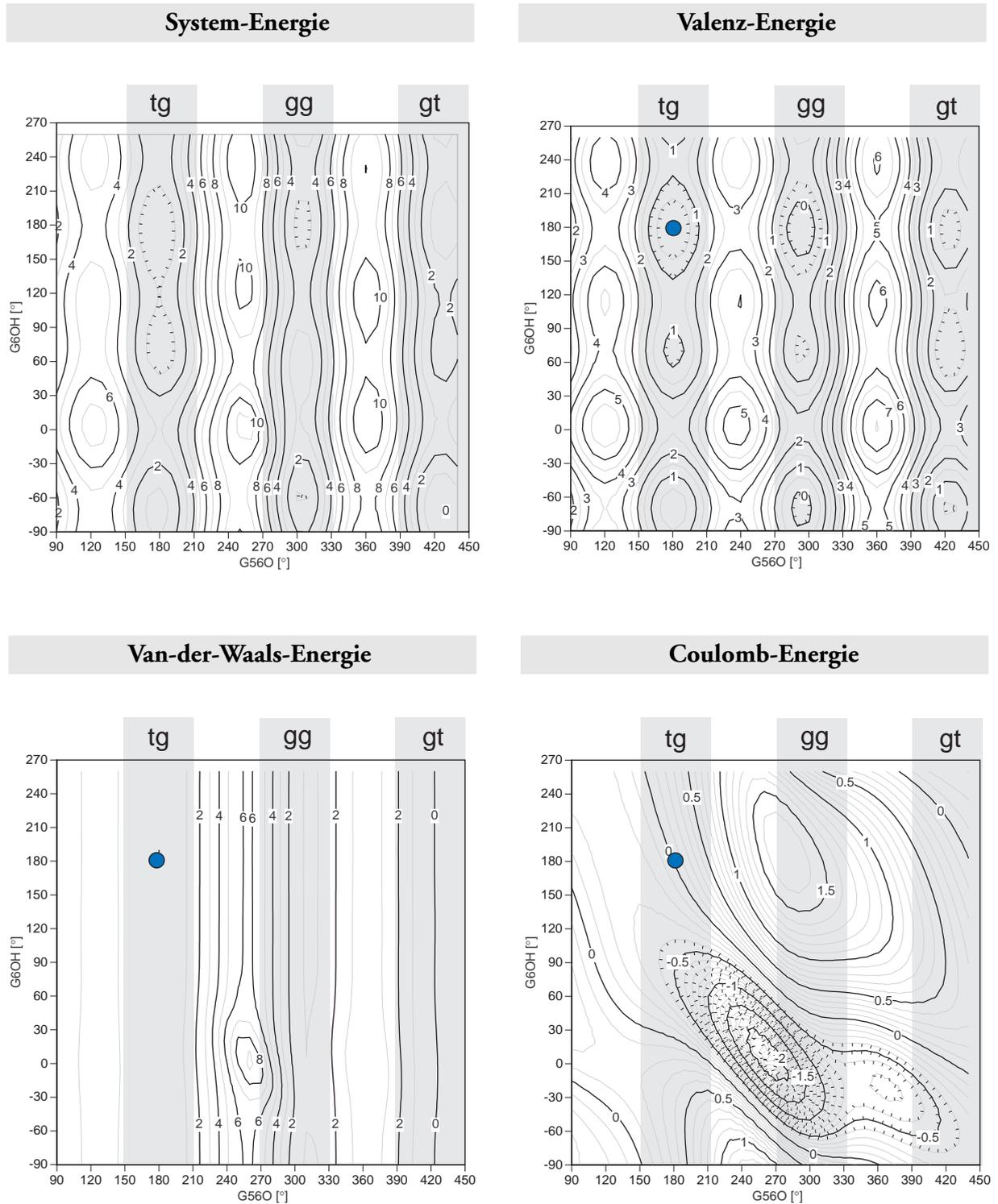
**Abb. X178:** Visualisierung von OH...O...HO-Interaktionen die durch eine Dreizentren-Interaktions-Analyse einer H-Brücken-Trajektorie gefunden wurden (Gal $\alpha$ 1-3Gal $\beta$ , 300K, CVFF, 1 ns, PBC, DMSO). Die „Interaktions-Peaks“ der chelat-artigen Komplexierung einzelner DMSO-Moleküle durch GOH2 und LOH4 sind markiert.



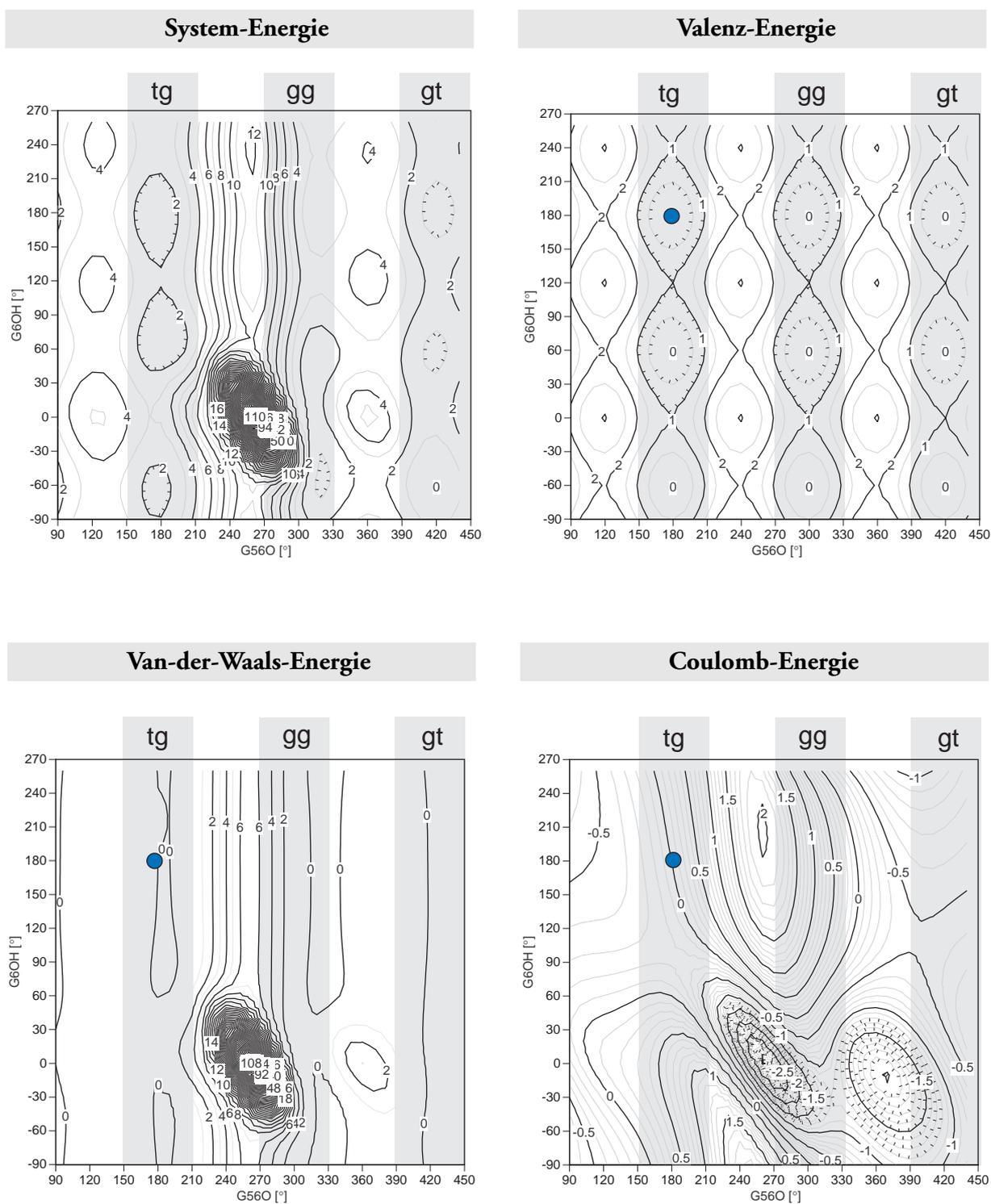
**Abb. X179:** Vergleich der  $\Phi/\Psi$ -FGE-Maps für eine Simulation in der Gasphase ( $\epsilon=4$ ) und in DMSO (Gal $\alpha$ 1-3Gal $\beta$ , 300K, CVFF, 1 ns)



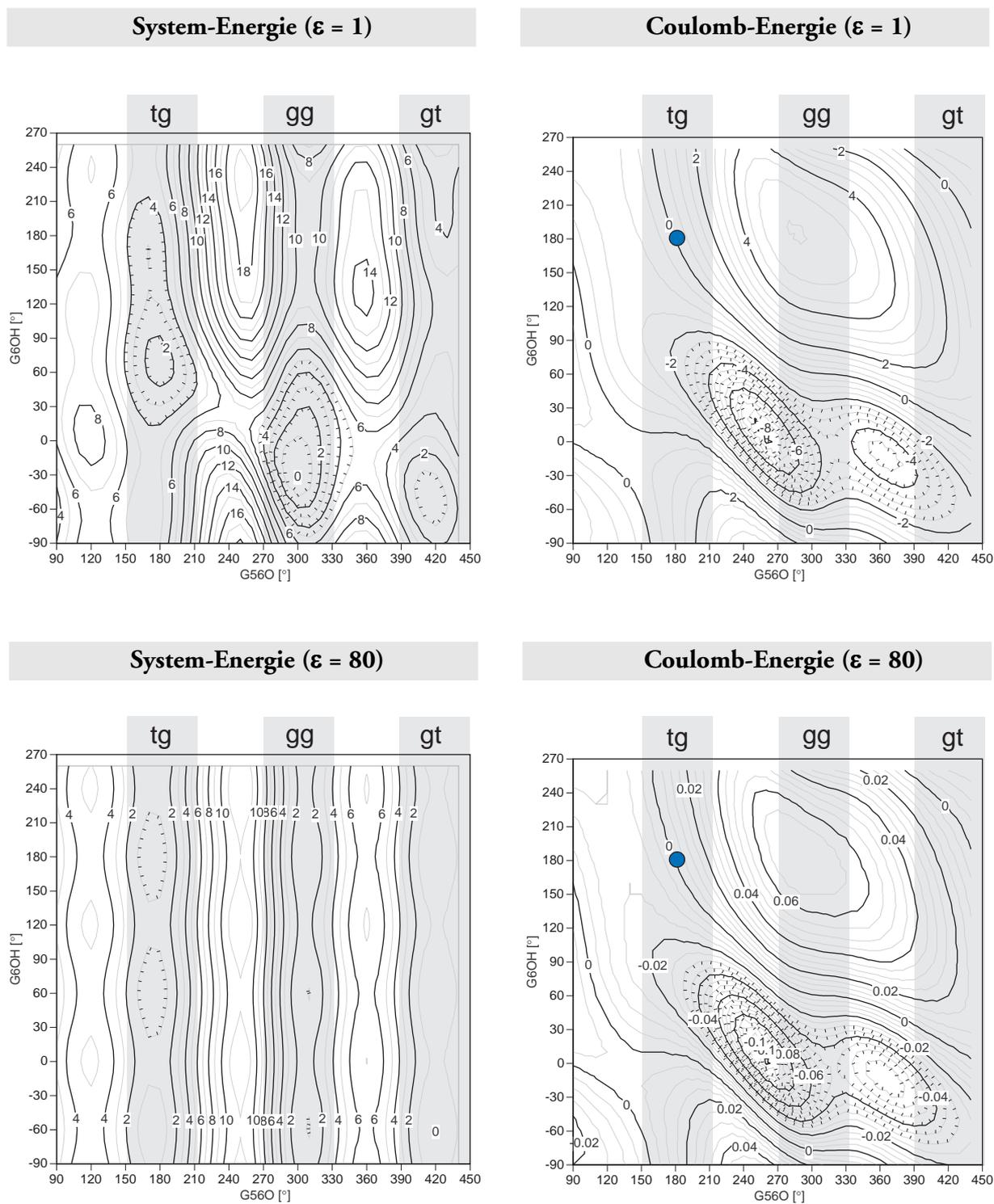
**Abb. X180:** Galβ-O-iProp: Abhängigkeit der Energie von den Torsionswinkeln G560 und G6OH. Rigid Grid; Interval = 10°; Kraftfeld = CVFF;  $\epsilon=4$ . Die System-Energie ist relativ zum globalen Minimum, die Valenz-, VdW- und Coulomb-Energie relativ zur Struktur mit G560 = 180° und G6OH = 180° berechnet.



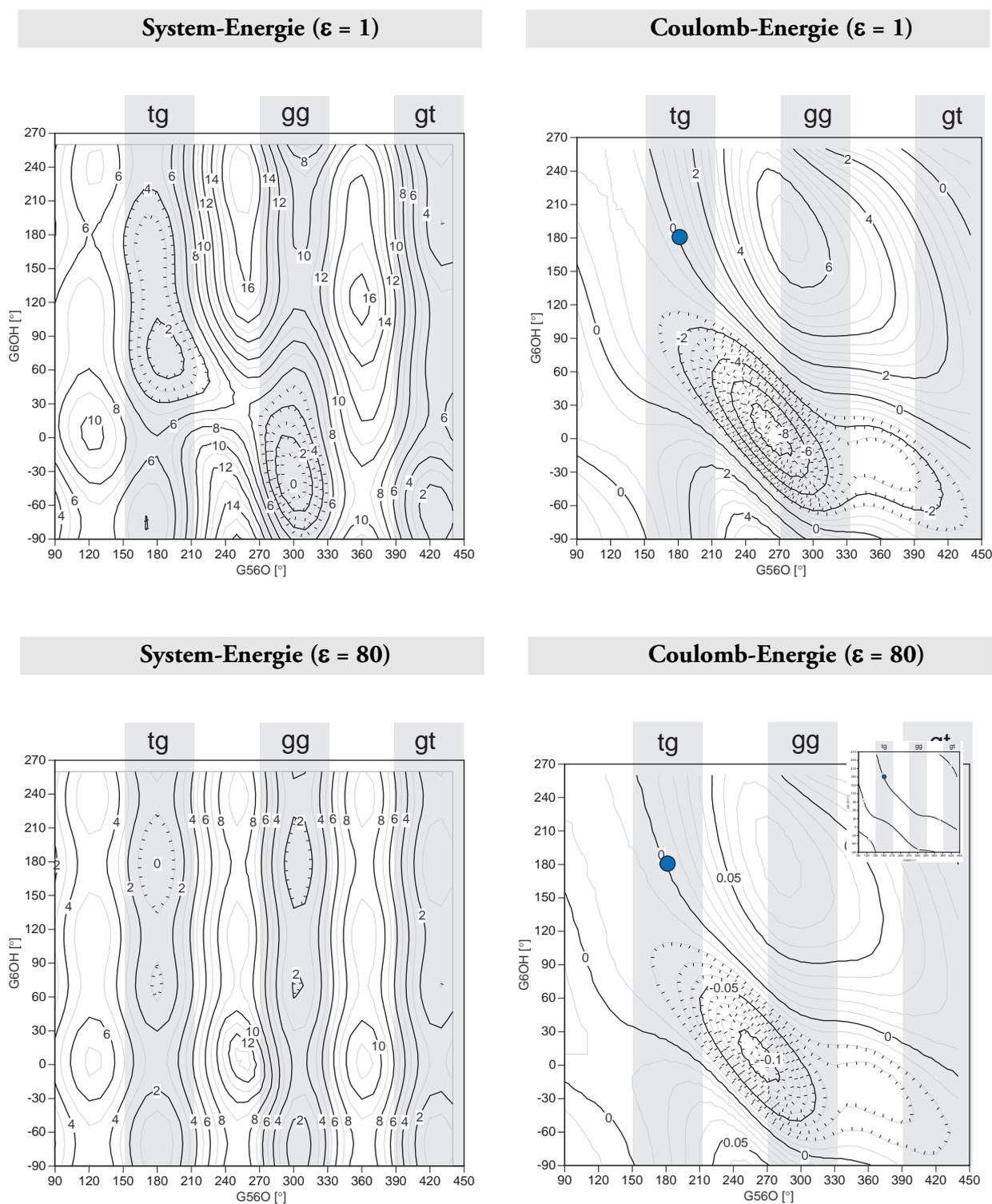
**Abb. X181:** Galβ-O-iProp: Abhängigkeit der Energie von den Torsionswinkeln G56O und G6OH. Rigid Grid; Interval = 10°; Kraftfeld = CFF91;  $\epsilon=4$ . Die System-Energie ist relativ zum globalen Minimum, die Valenz-, VdW- und Coulomb-Energie relativ zur Struktur mit G56O = 180° und G6OH = 180° berechnet.



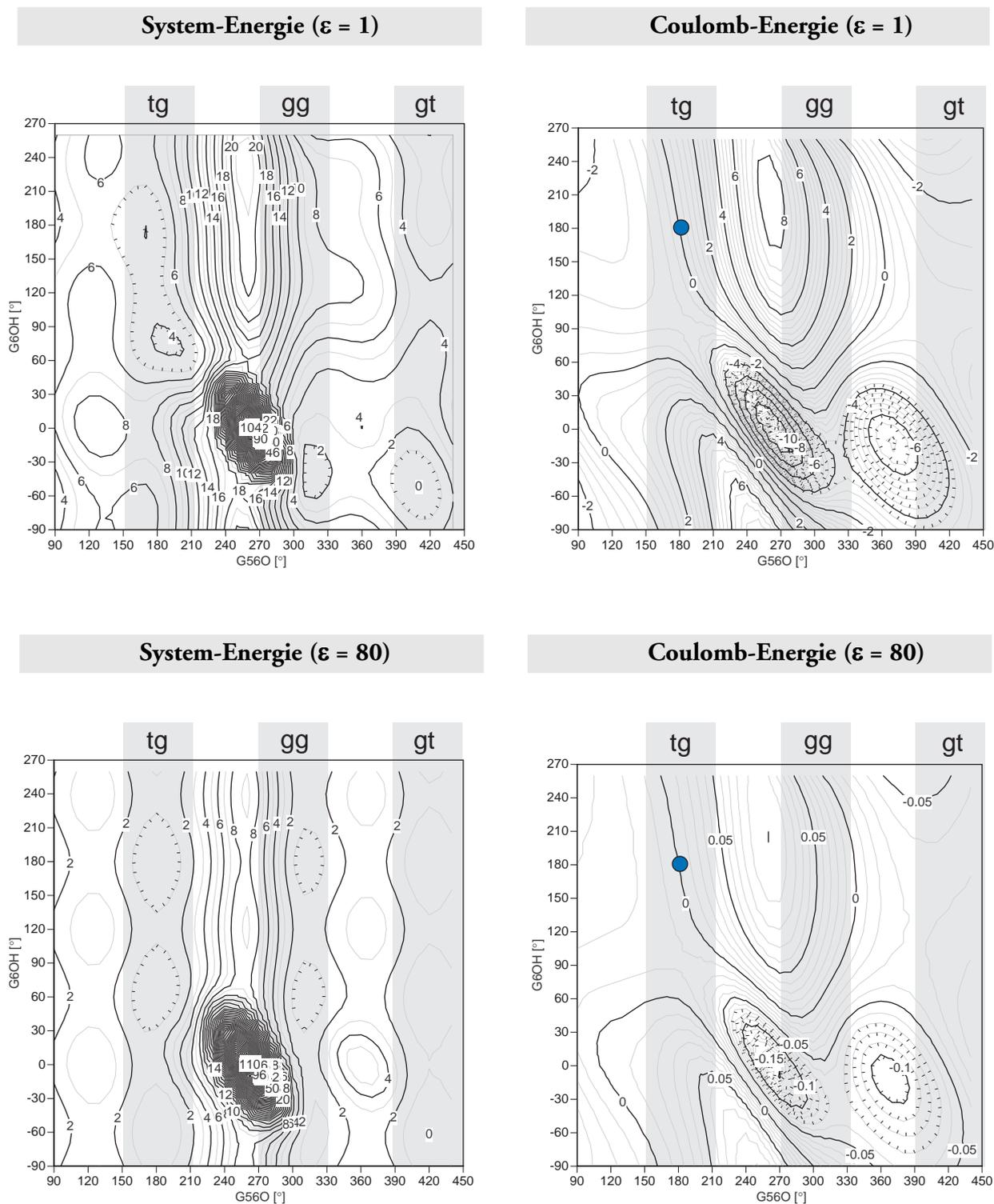
**Abb. X182:** Galß-O-iProp: Abhängigkeit der Energie von den Torsionswinkeln G560 und G6OH. Rigid Grid; Interval = 10°; Kraftfeld = AMBER/Homans;  $\epsilon=4$ . Die System-Energie ist relativ zum globalen Minimum, die Valenz-, VdW- und Coulomb-Energie relativ zur Struktur mit G560 = 180° und G6OH = 180° berechnet.



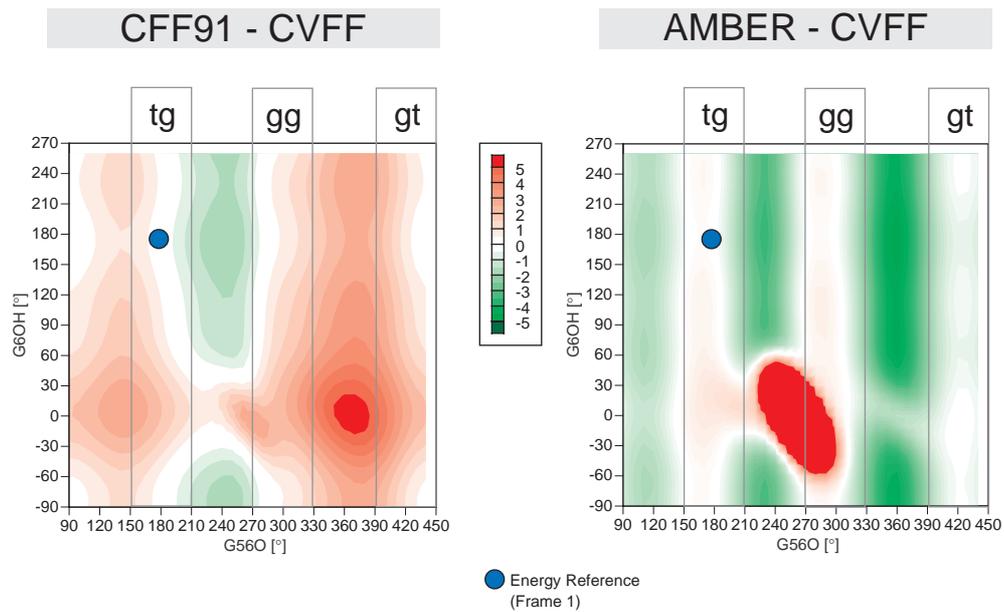
**Abb. X183:** Galß-O-iProp: Abhängigkeit der Energie von den Torsionswinkeln G560 und G6OH und von der Dielektrizitätskonstanten. Rigid Grid; Intervall = 10°; Kraftfeld = CVFF. Die System-Energie ist relativ zum globalen Minimum, die Valenz-, VdW- und Coulomb-Energie relativ zur Struktur mit G560 = 180° und G6OH = 180° berechnet.



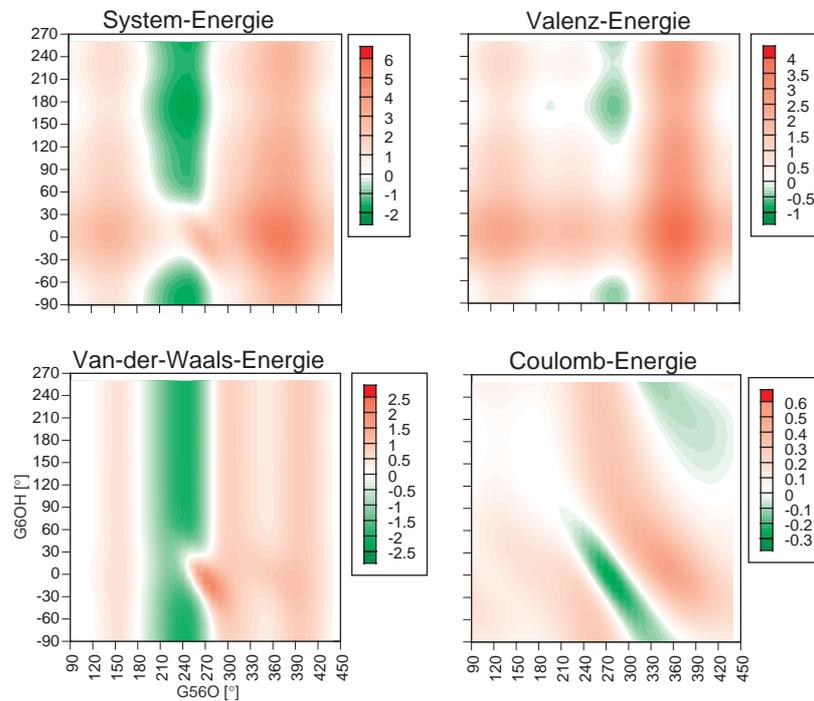
**Abb. X184:** Galß-O-iProp: Abhängigkeit der Energie von den Torsionswinkeln G560 und G6OH und von der Dielektrizitätskonstanten. Rigid Grid; Intervall = 10°; Kraftfeld = CFF91. Die System-Energie ist relativ zum globalen Minimum, die Valenz-, vdW- und Coulomb-Energie relativ zur Struktur mit G560 = 180° und G6OH = 180° berechnet.



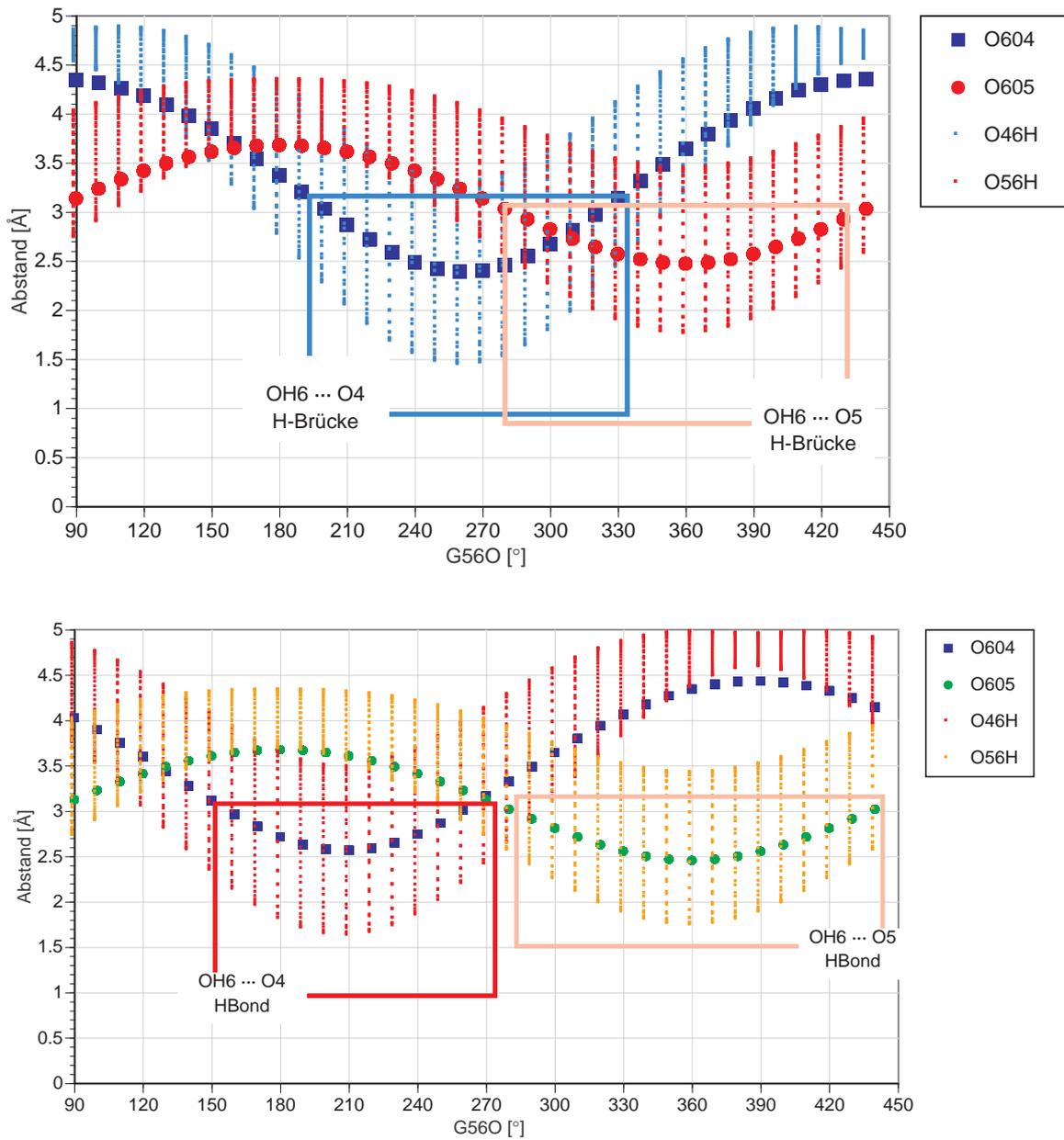
**Abb. X185:** Galß-O-iProp: Abhängigkeit der Energie von den Torsionswinkeln G560 und G6OH und von der Dielektrizitätskonstanten. Rigid Grid; Intervall = 10°; Kraftfeld = AMBER/Homans. Die System-Energie ist relativ zum globalen Minimum, die Valenz-, vdW- und Coulomb-Energie relativ zur Struktur mit G560 = 180° und G6OH = 180° berechnet.



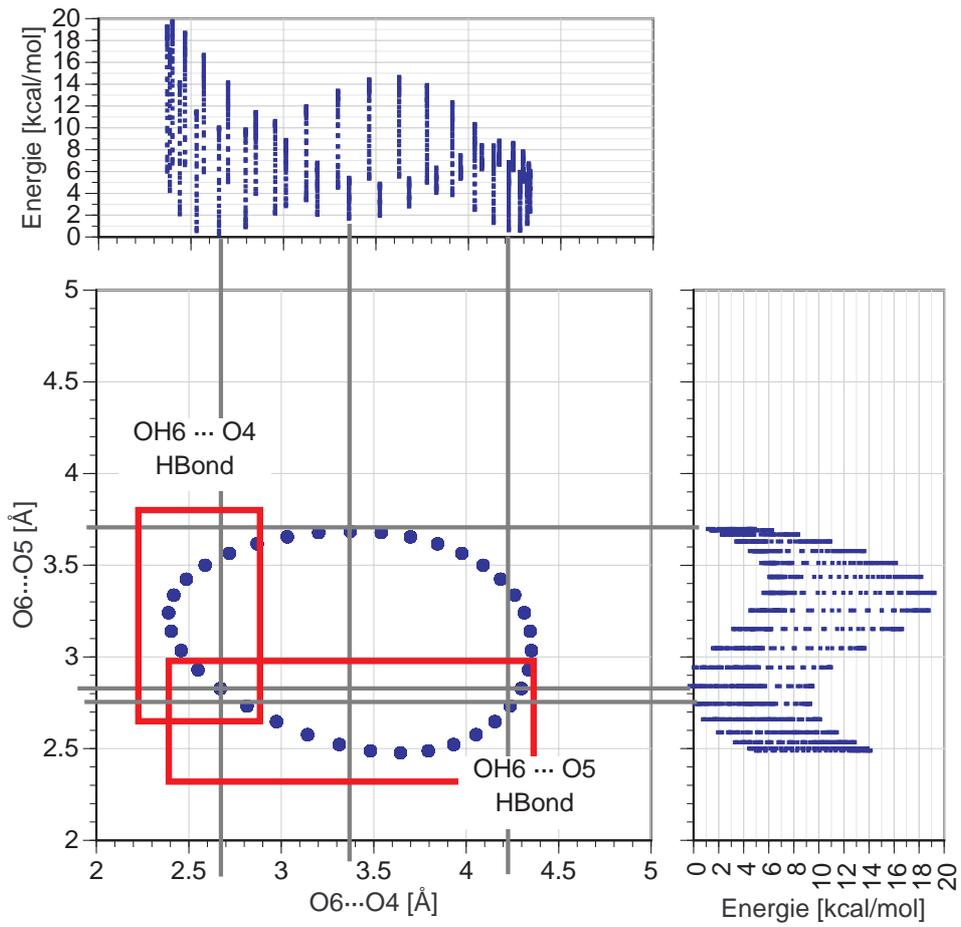
**Abb. X186:** Gal $\beta$ -O-iProp: Differenz-Maps der Systemenergie. Grün: Die Energie von CFF91 bzw. AMBER ist niedriger. Rot: Die Energie von CVFF ist niedriger. CFF91 hat eine deutlich höhere Übergangsbarriere gg  $\leftrightarrow$  gt als CVFF, dafür ist die Barriere gg  $\leftrightarrow$  tg niedriger.



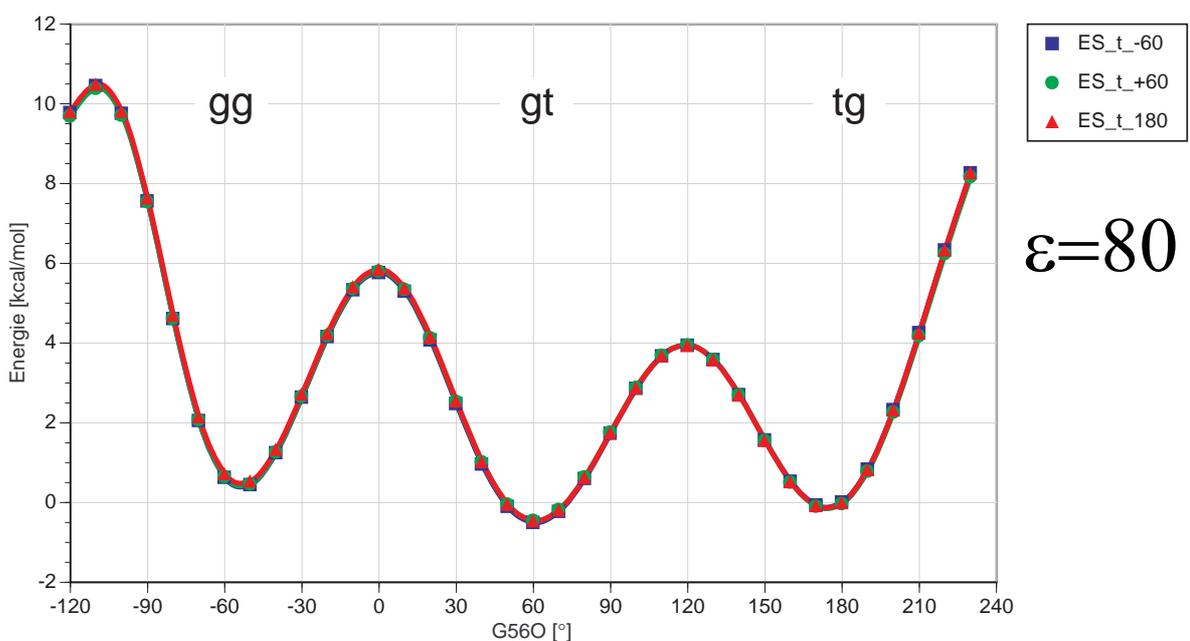
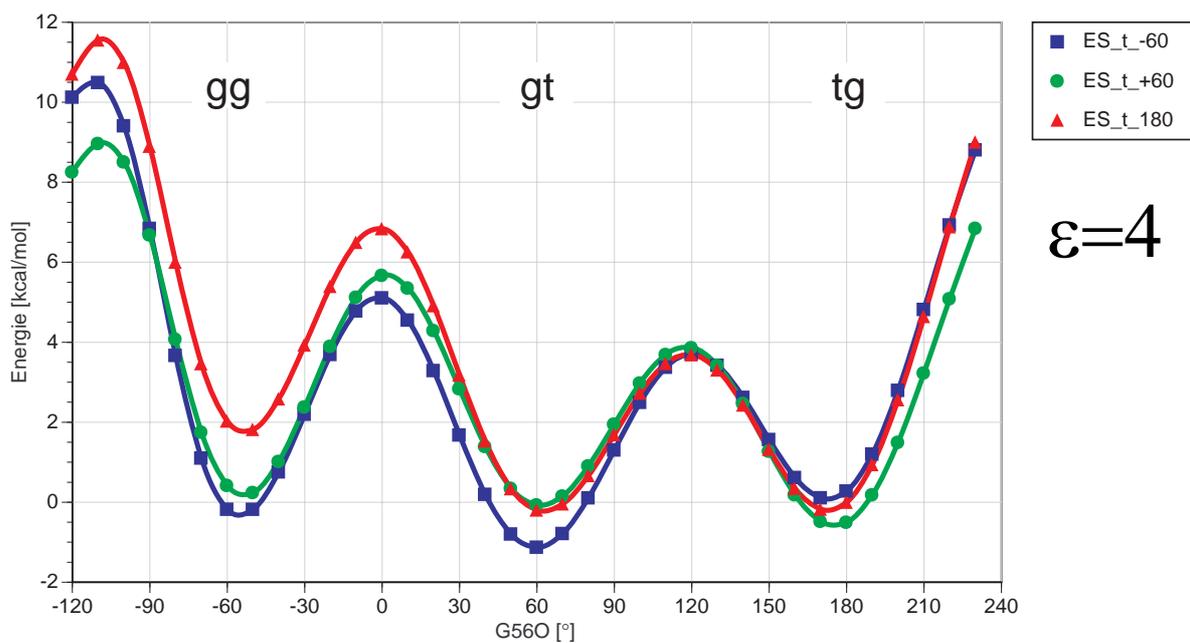
**Abb. X187:** Gal $\beta$ -O-iProp: Differenz-Maps der Energie-Terme (CFF91-CVFF) Grün: Die Energie von CFF91 ist niedriger. Rot: Die Energie von CVFF ist niedriger.



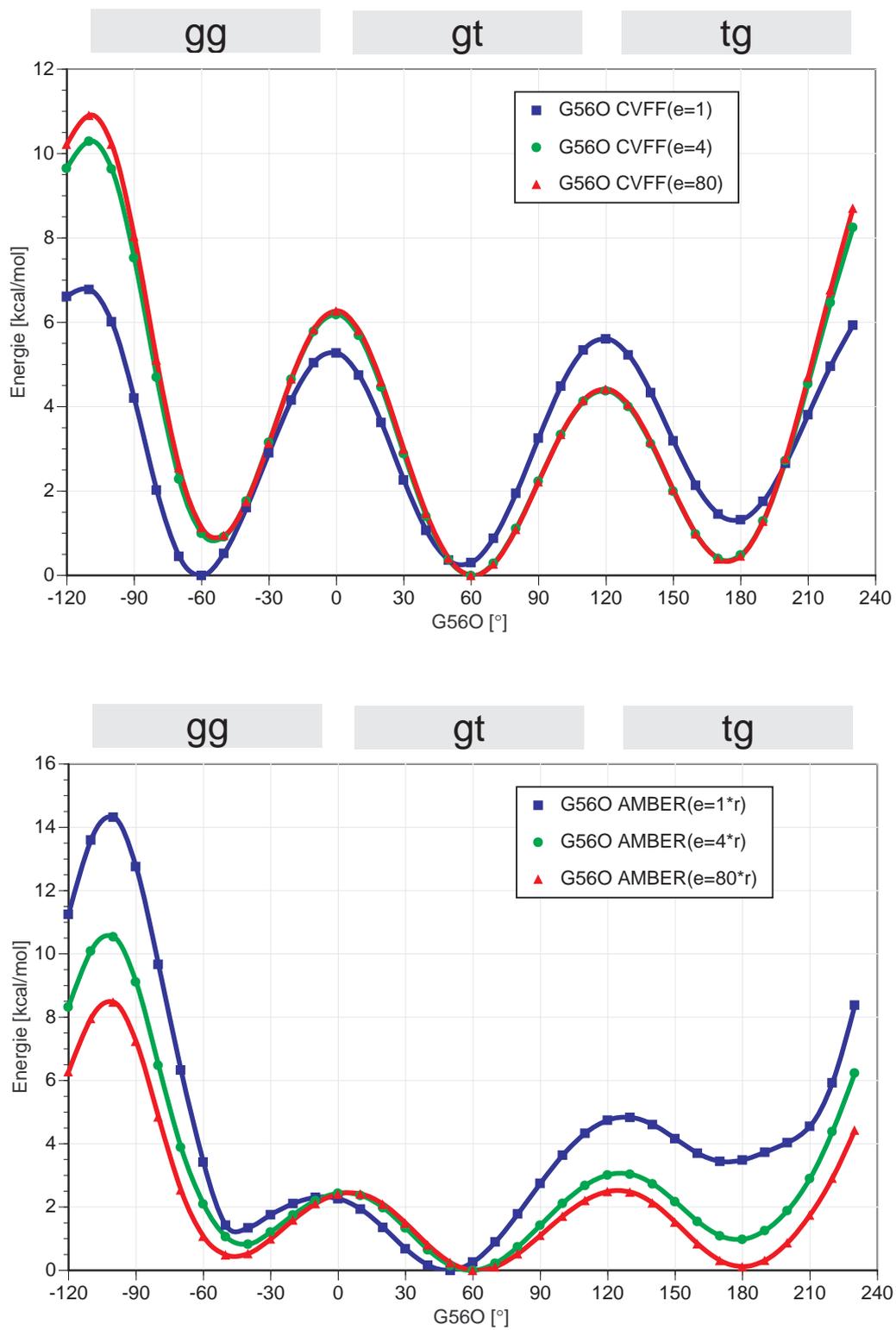
**Abb. X188:** „Rigid Grid“: Abhängigkeit diverser Atomabstände vom Torsionswinkel G560 für Galβ-O-iProp (oben) und Glcα-O-iProp: (unten)



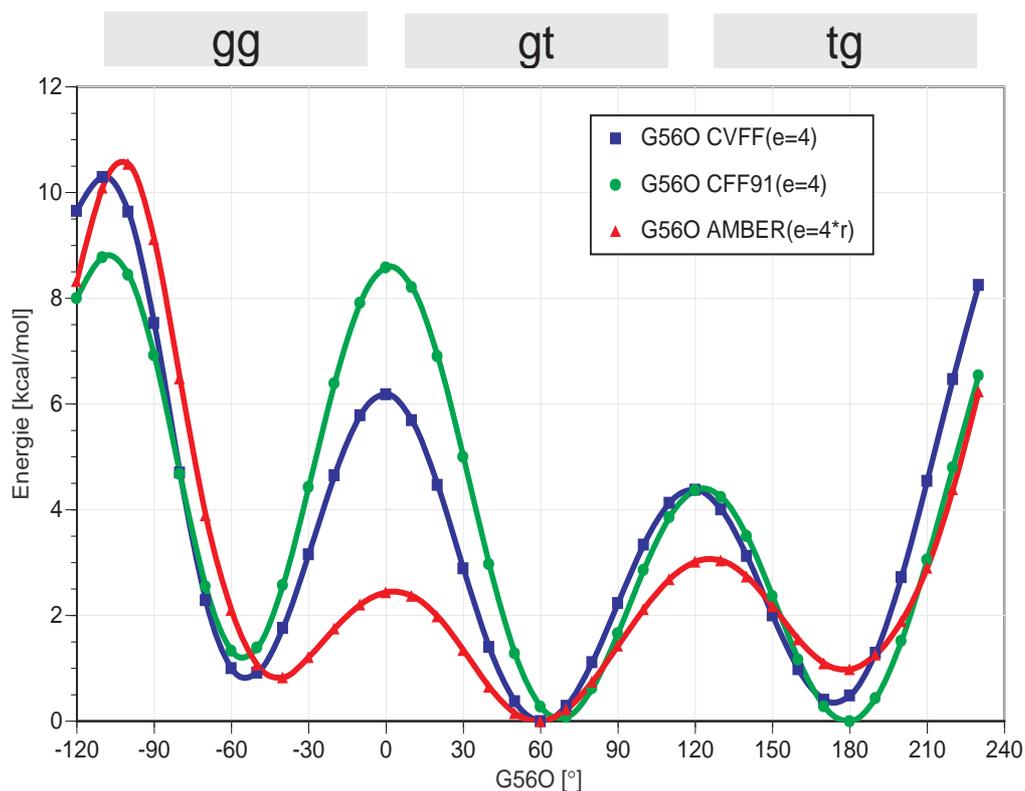
**Abb. X189:** Galβ-O-iProp (Rigid Grid): Abhängigkeit der Systemenergie vom O6··O5- bzw. O6··O4-Abstand. (CVFF,  $\epsilon = 1$ )



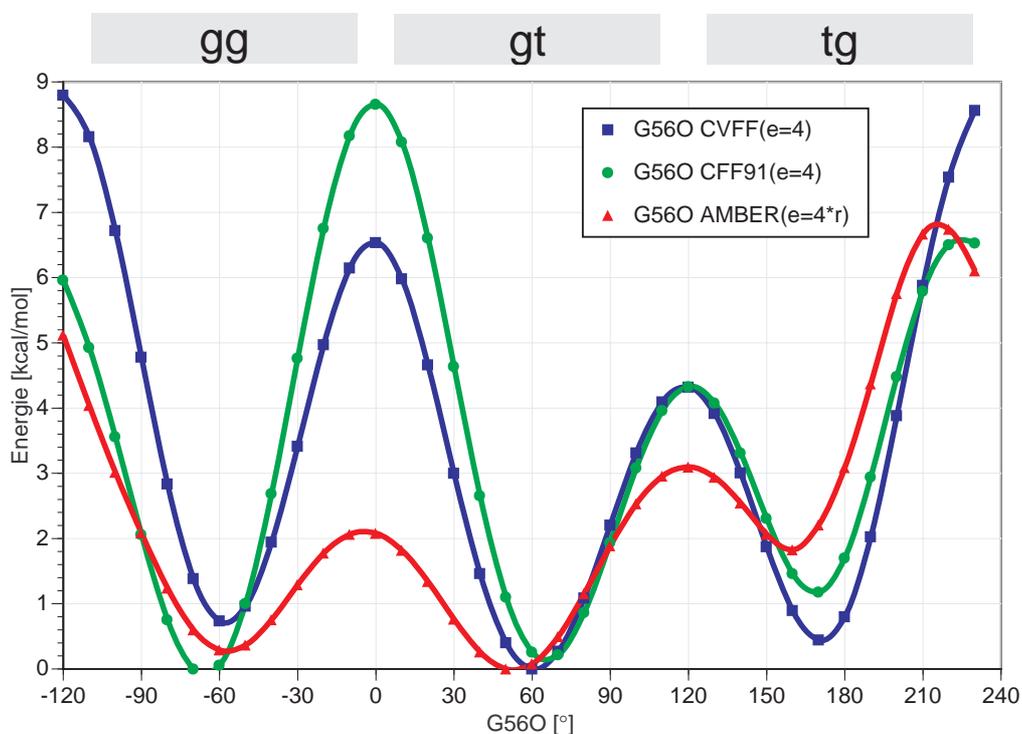
**Abb. X190:** Gal $\beta$ -O-iProp: Rotationsprofil des Torsionswinkels G56O gewonnen aus „Rigid Grid“ Daten gruppiert nach G6OH („-60“ = [-90, -30]; „+60“ = [+30, +90]; „180“ = [150, 210]) „Adiabatic Grid“, d.h. für jeden G56O-Gridpunkt wurde die Struktur mit der geringsten Systemenergie ermittelt. Kraftfeld = CVFF.



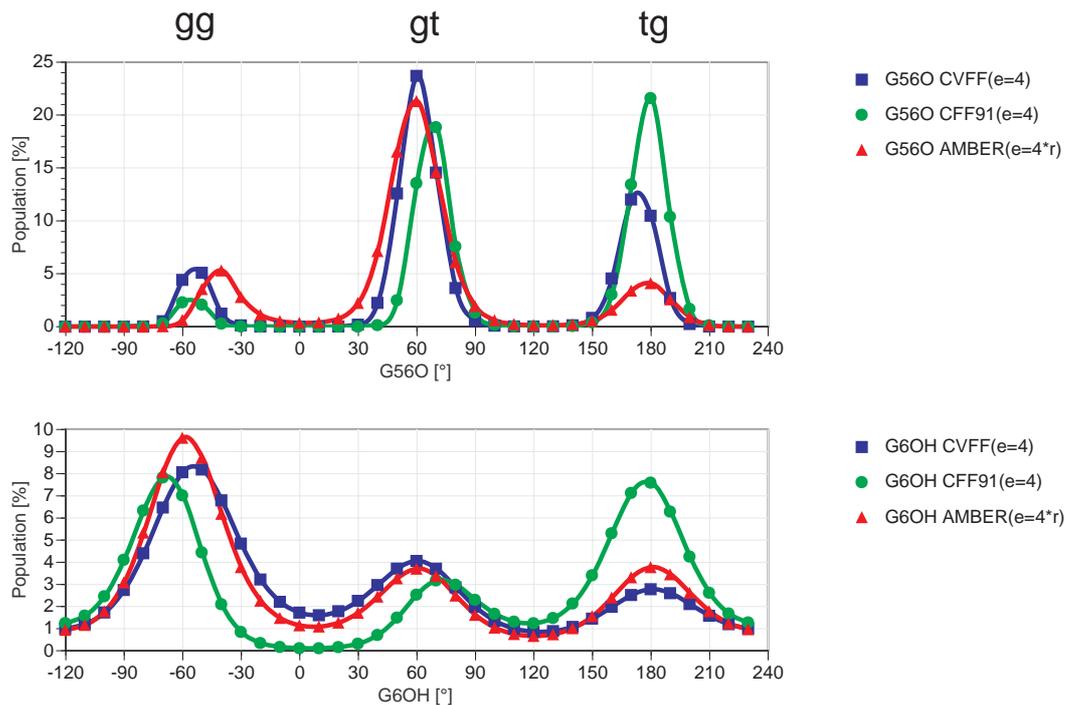
**Abb. X191:** Galβ-O-iProp: Freie Grid-Energie in Abhängigkeit von der Dielektrizitätskonstanten für CVFF (oben) und AMBER/Homans (unten).



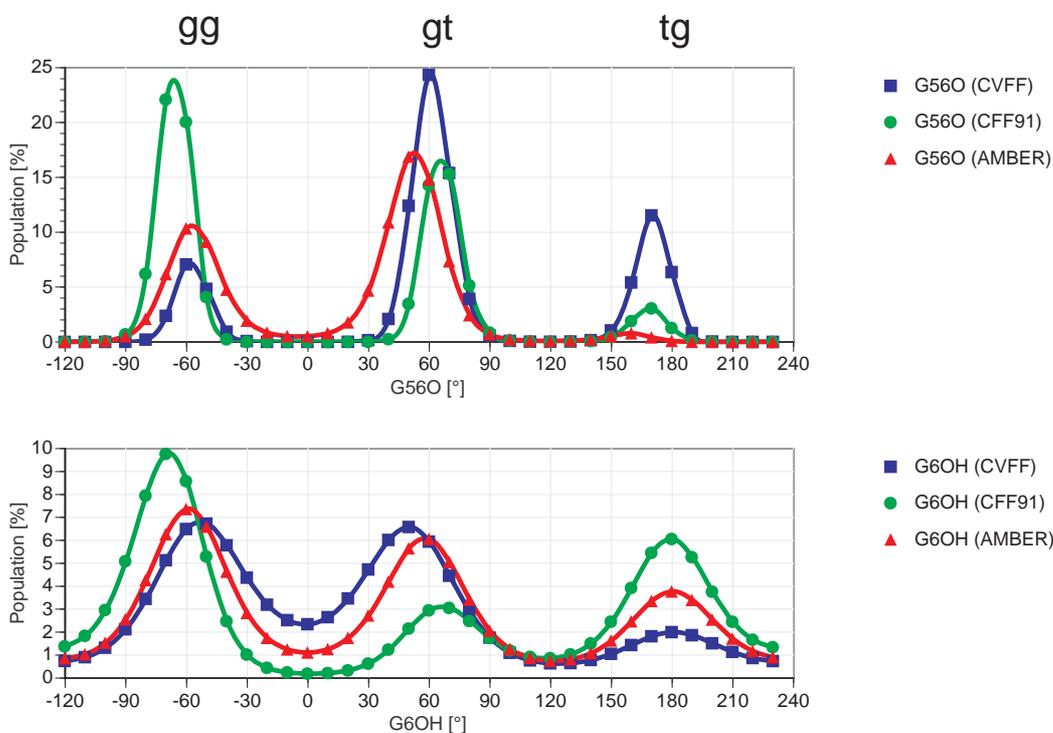
**Abb. X192:** Gal $\beta$ -O-iProp: Energieprofil von G56O berechnet aus Grid-Search Daten. „Freie Gridgruppen-Energie“, Datenbasis: G56O/G6OH Grid (Gridintervall = 10°,  $\epsilon = 4$ )



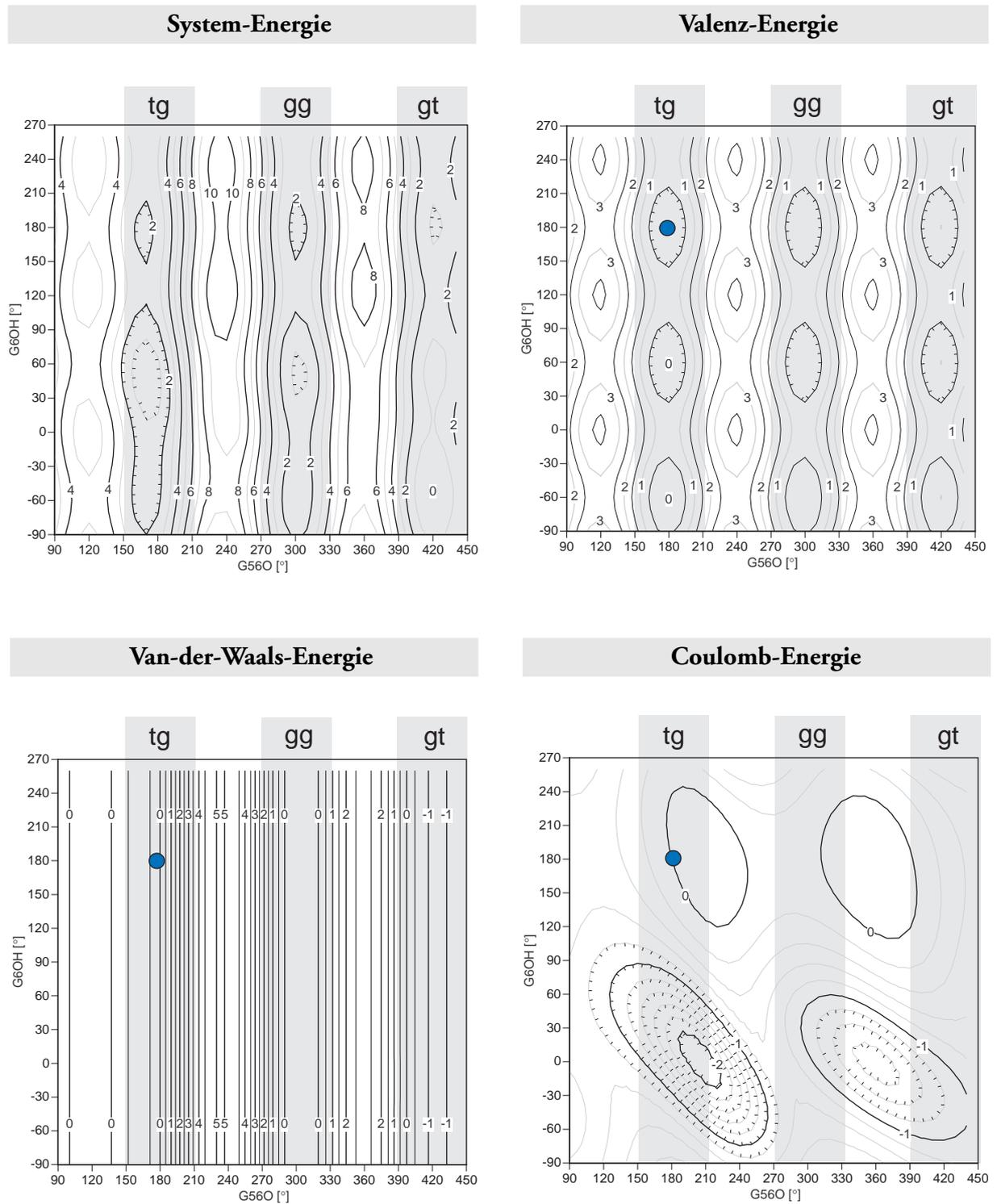
**Abb. X193:** Glc $\alpha$ -O-iProp: Energieprofil von G56O berechnet aus Grid Search Daten. „Freie Gridgruppen-Energie“, Datenbasis: G56O/G6OH Grid (Gridintervall = 10°,  $\epsilon = 4$ )



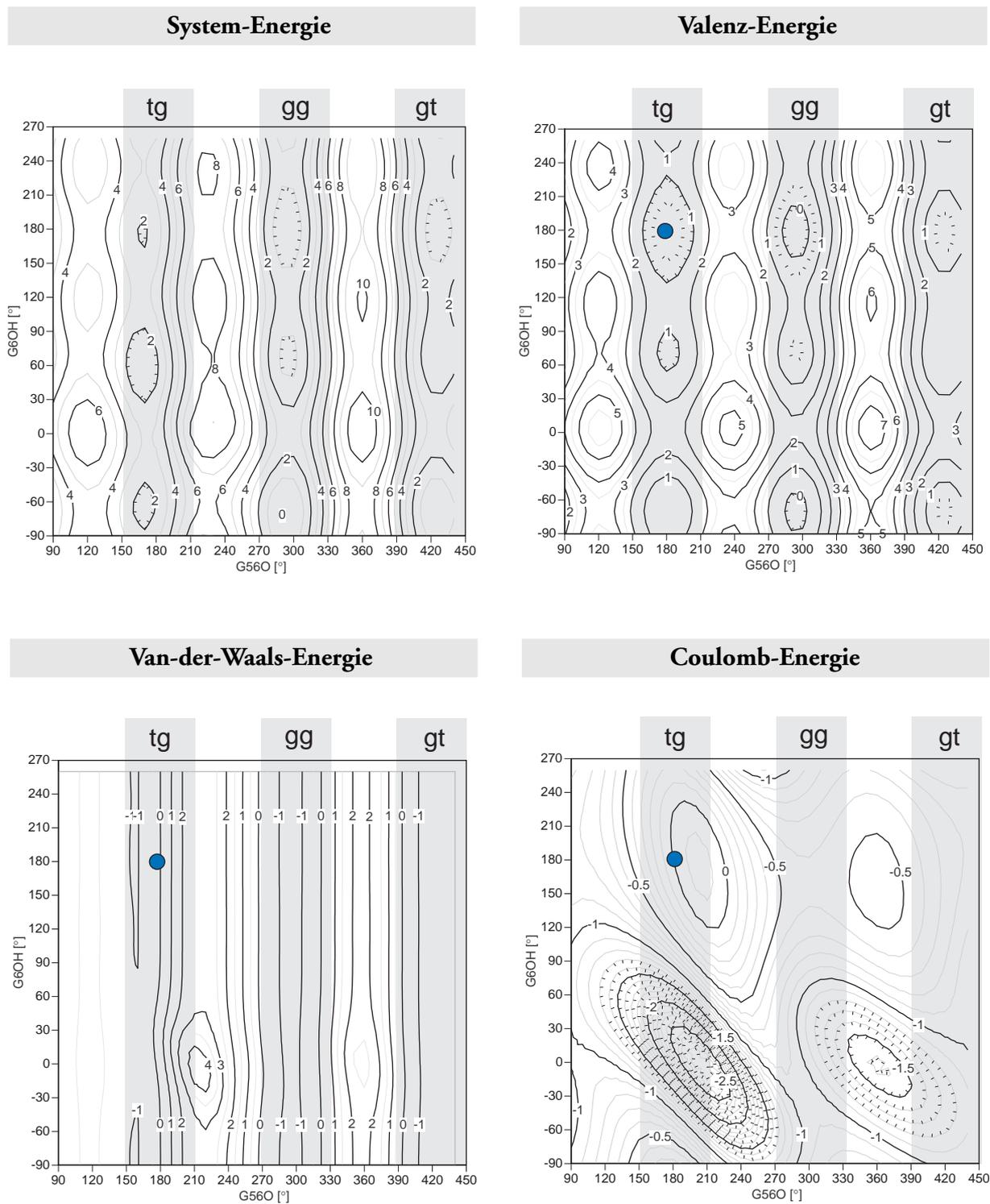
**Abb. X194:** Gal $\beta$ -O-iProp: Rotamerenverteilung der Hydroxymethylgruppe und der OH6-Gruppe Boltzmann-gewichtete Gridgruppen-Population bei 300K, Datenbasis: G56O/G6OH Grid (Gridintervall = 10°,  $\epsilon = 4$ )



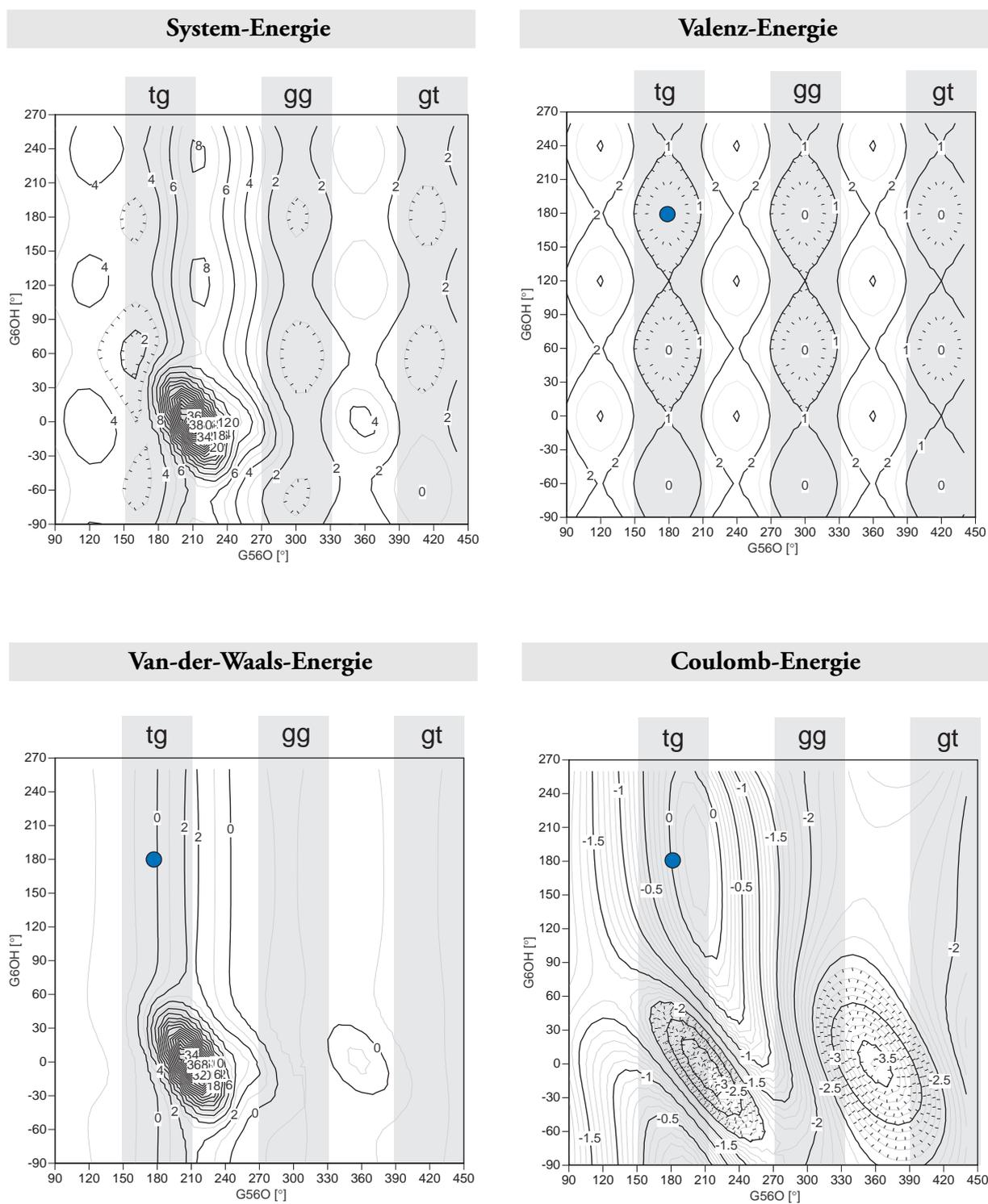
**Abb. X195:** Glc $\alpha$ -O-iProp: Rotamerenverteilung der Hydroxymethylgruppe und der OH6-Gruppe Boltzmann-gewichtete Gridgruppen-Population bei 300K, Datenbasis: G56O/G6OH Grid (Gridintervall = 10°,  $\epsilon = 4$ )



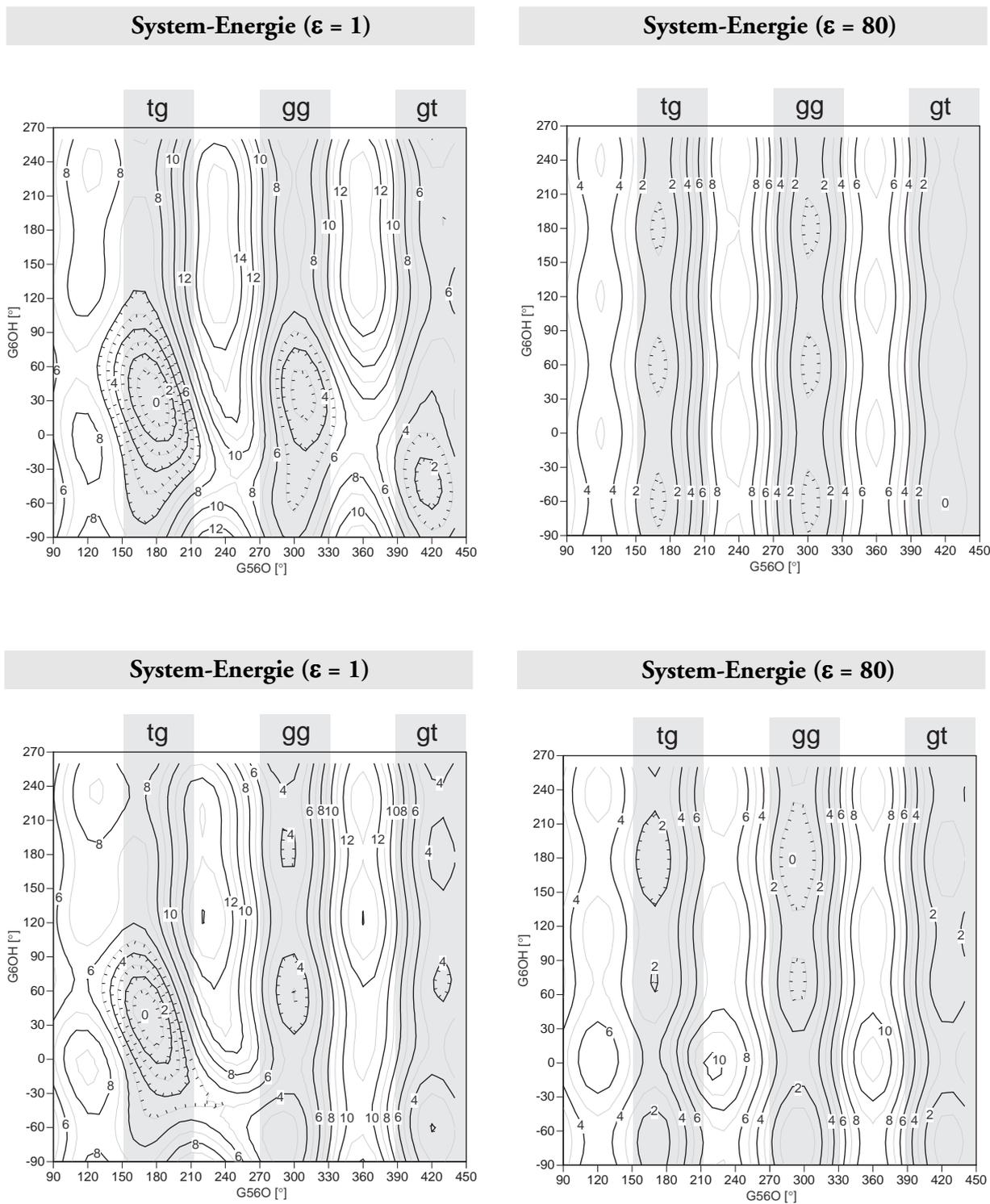
**Abb. X196:** Glc $\alpha$ -O-iProp: Abhängigkeit der Energie von den Torsionswinkeln G56O und G6OH. Rigid Grid; Intervall = 10°; Kraftfeld = CVFF;  $\epsilon=4$ . Die System-Energie ist relativ zum globalen Minimum, die Valenz-, vdW- und Coulomb-Energie relativ zur Struktur mit G56O = 180° und G6OH = 180° berechnet.



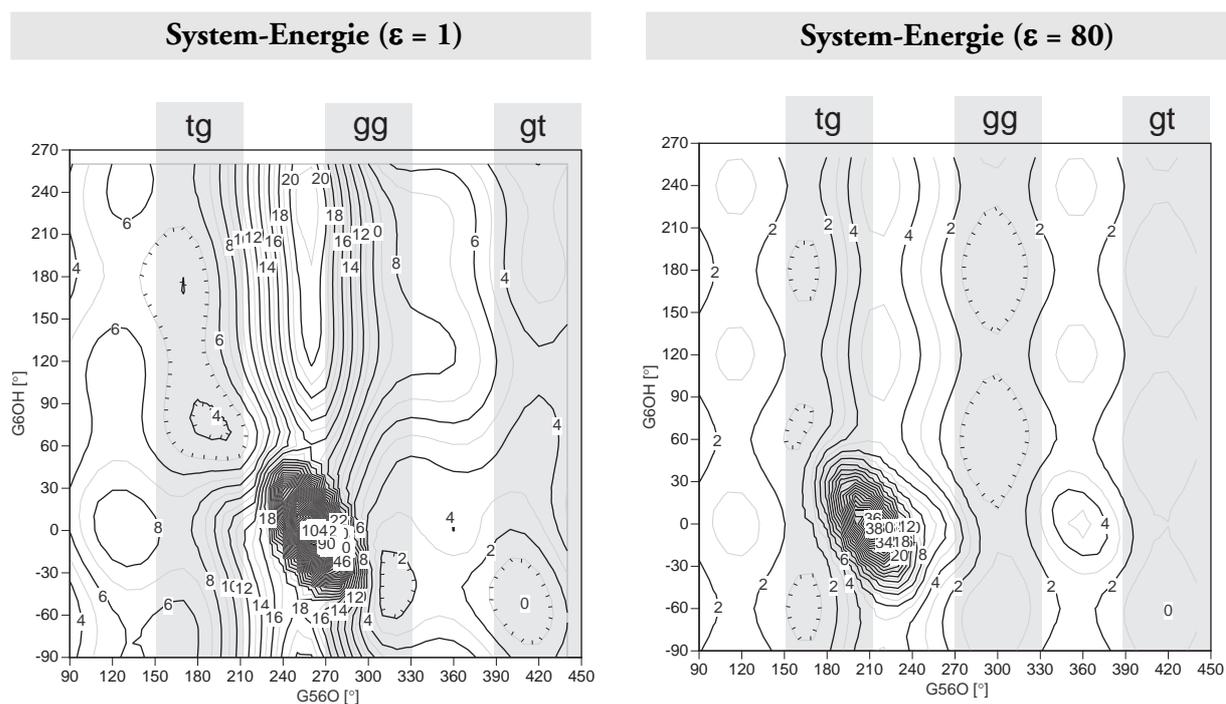
**Abb. X197:** Glc $\alpha$ -O-iProp: Abhängigkeit der Energie von den Torsionswinkeln G56O und G6OH. Rigid Grid; Intervall = 10°; Kraftfeld = CFF91;  $\epsilon=4$ . Die System-Energie ist relativ zum globalen Minimum, die Valenz-, vdW- und Coulomb-Energie relativ zur Struktur mit G56O = 180° und G6OH = 180° berechnet.



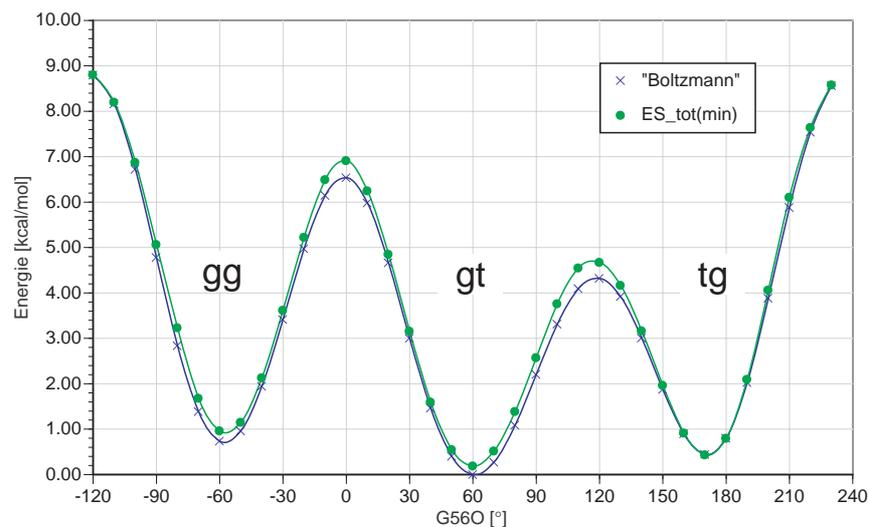
**Abb. X198:** Glc $\alpha$ -O-iProp: Abhängigkeit der Energie von den Torsionswinkeln G560 und G6OH. Rigid Grid; Intervall = 10°; Kraftfeld = AMBER/Homans;  $\epsilon=4$ . Die System-Energie ist relativ zum globalen Minimum, die Valenz-, VdW- und Coulomb-Energie relativ zur Struktur mit G560 = 180° und G6OH = 180° berechnet.



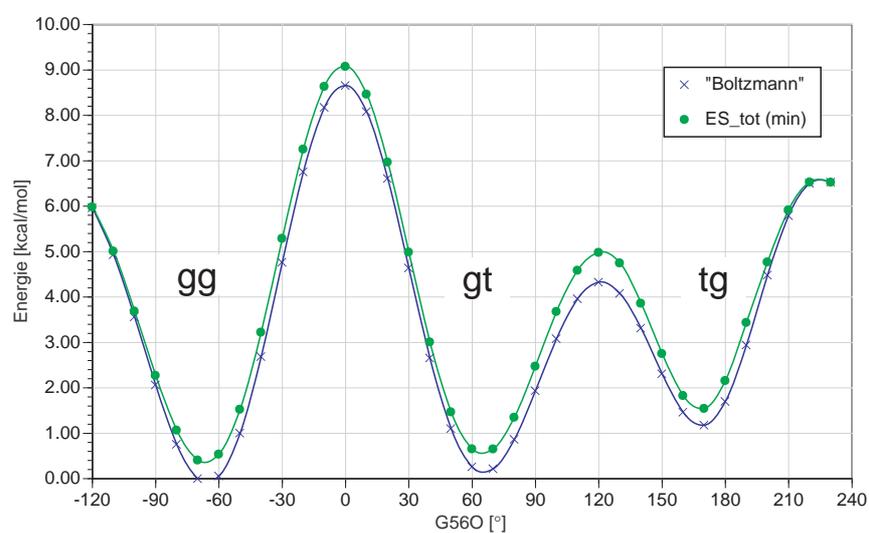
**Abb. X199:** Gl $\alpha$ -O-iProp: Abhängigkeit der Energie von den Torsionswinkeln G56O und G6OH und von der Dielektrizitätskonstanten. Rigid Grid; Intervall = 10°; Kraftfeld = CVFF (oben), CFF91 (unten). Die System-Energie ist relativ zum globalen Minimum, die Valenz-, vdW- und Coulomb-Energie relativ zur Struktur mit G56O = 180° und G6OH = 180° berechnet.



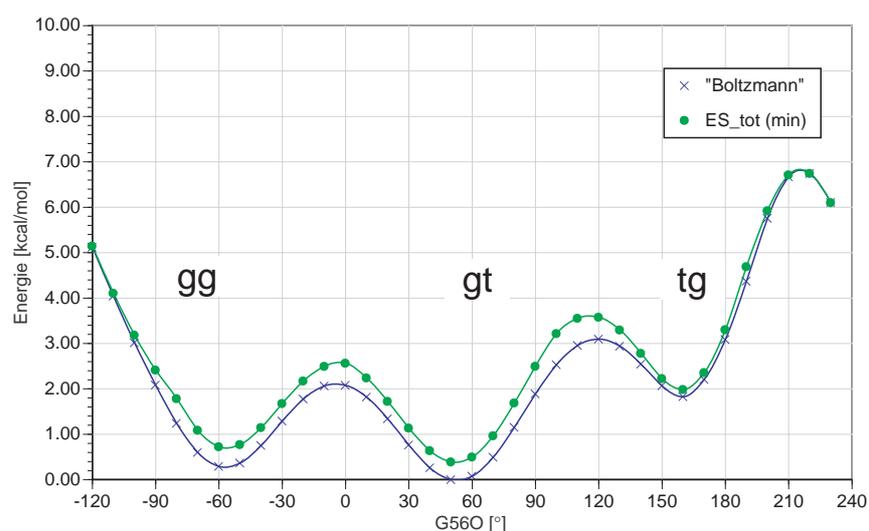
**Abb. X200:** Gl $\alpha$ -O-iProp: Abhängigkeit der Energie von den Torsionswinkeln G56O und G6OH und von der Dielektrizitätskonstanten. Rigid Grid; Intervall = 10°; Kraftfeld = AMBER/Homans. Die System-Energie ist relativ zum globalen Minimum, die Valenz-, vdW- und Coulomb-Energie relativ zur Struktur mit G56O = 180° und G6OH = 180° berechnet.



CVFF

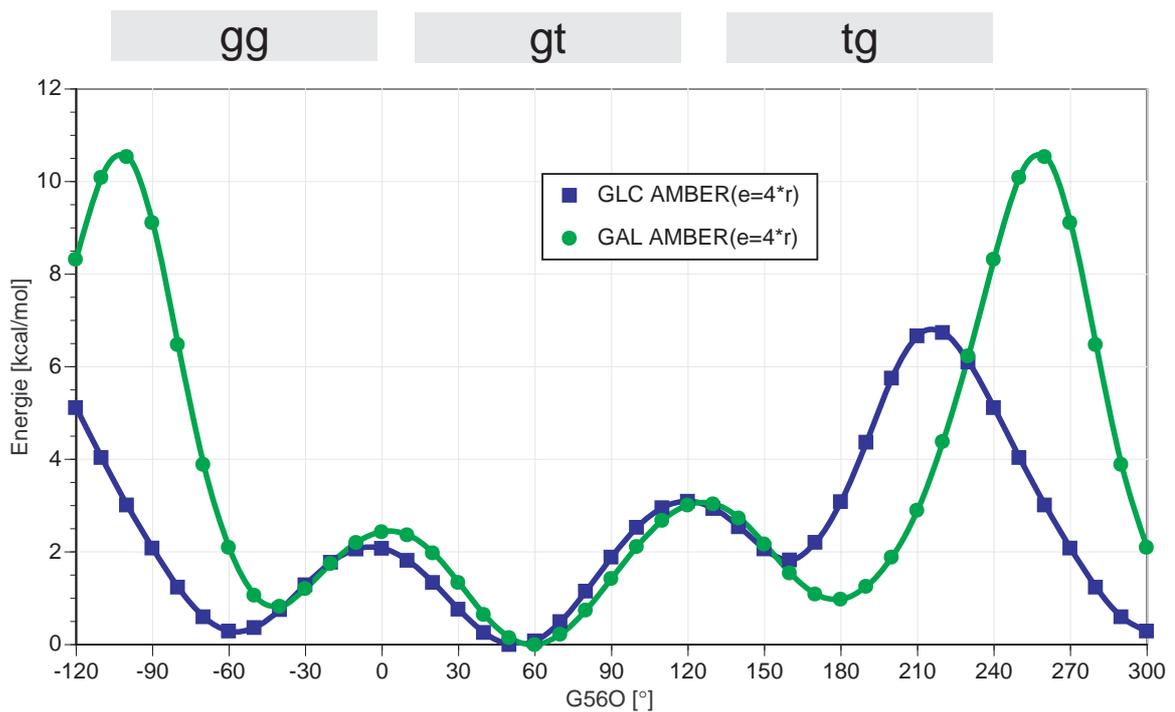
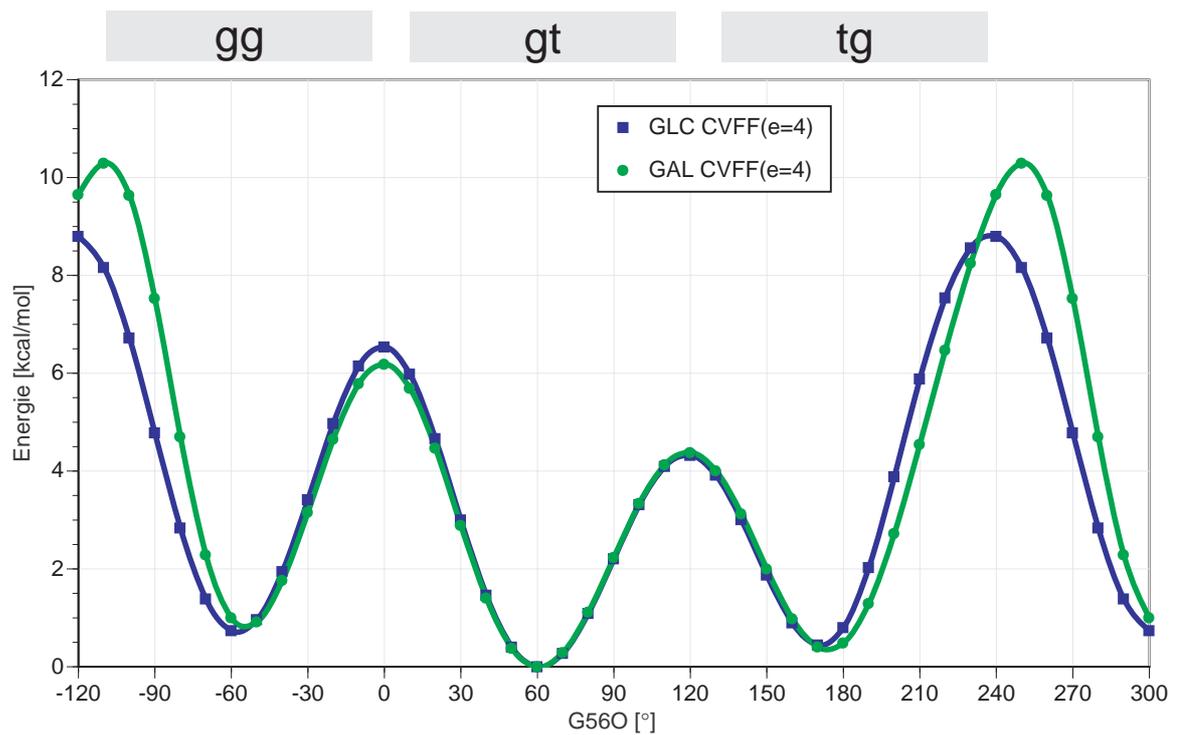


CFF91

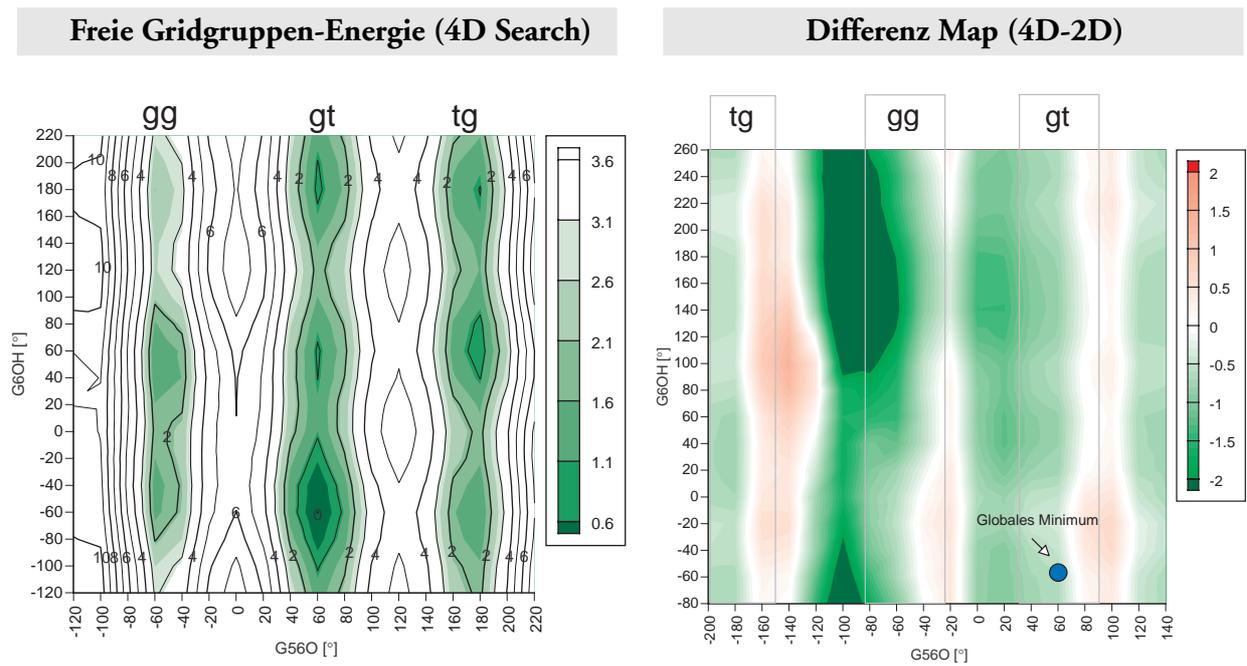


AMBER / Homans

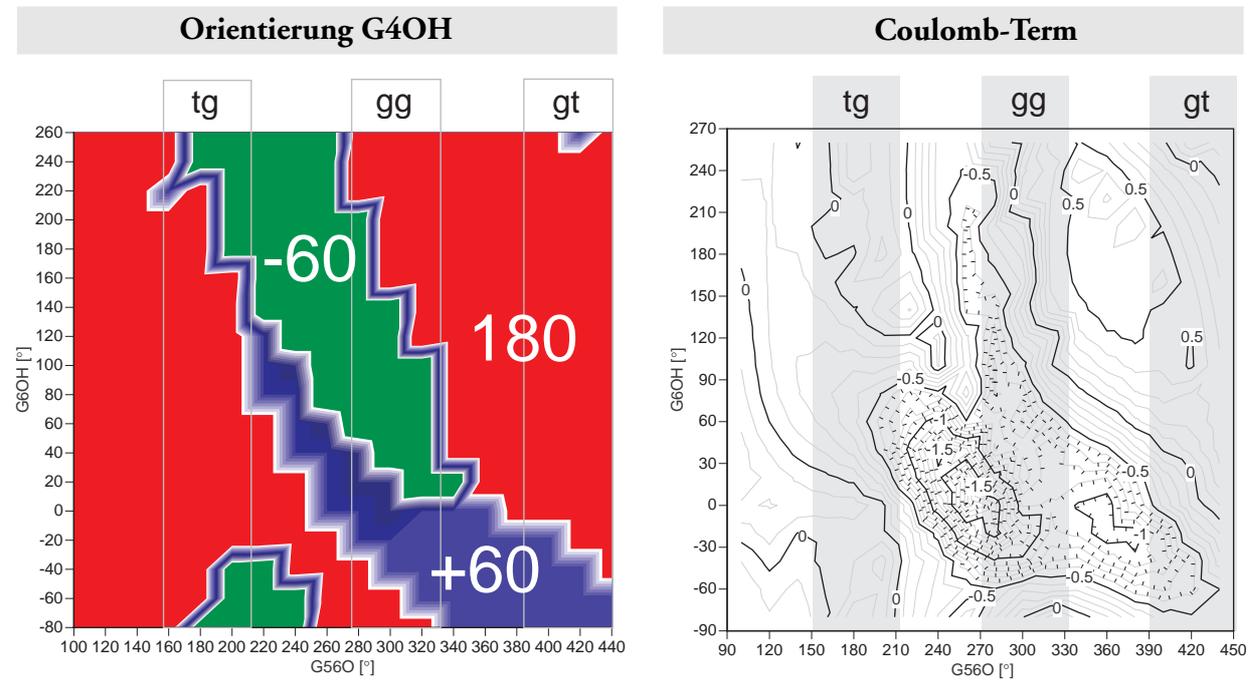
**Abb. X201:** Glc $\alpha$ -O-iProp: Energieprofil von G56O berechnet aus Grid-Search Daten.  
 Boltzmann = „Freie Grid-Energie“, ES-tot(min): Minimumenergie-Methode („Adiabatic Grid“)  
 Datenbasis: G56O/G6OH Grid (Gridintervall = 10°,  $\epsilon = 4$ )



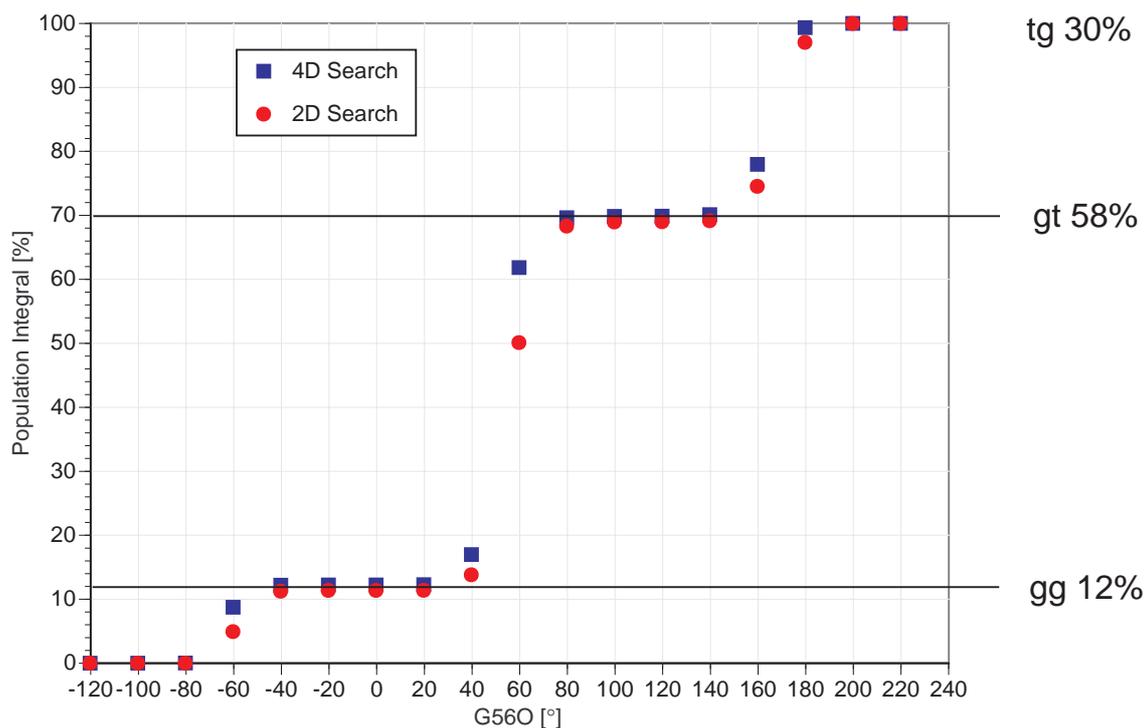
**Abb. X202:** Energieprofile von G56O berechnet aus Grid Search Daten.  
 „Freie Grid-Energie“, Datenbasis: G56O/G6OH Grid (Gridintervall = 10°,  $\epsilon = 4$ , CVFF (oben),  
 AMBER/Homans (unten)), GAL = Gal $\beta$ -O-iProp, GLC = Glc $\alpha$ -O-iProp



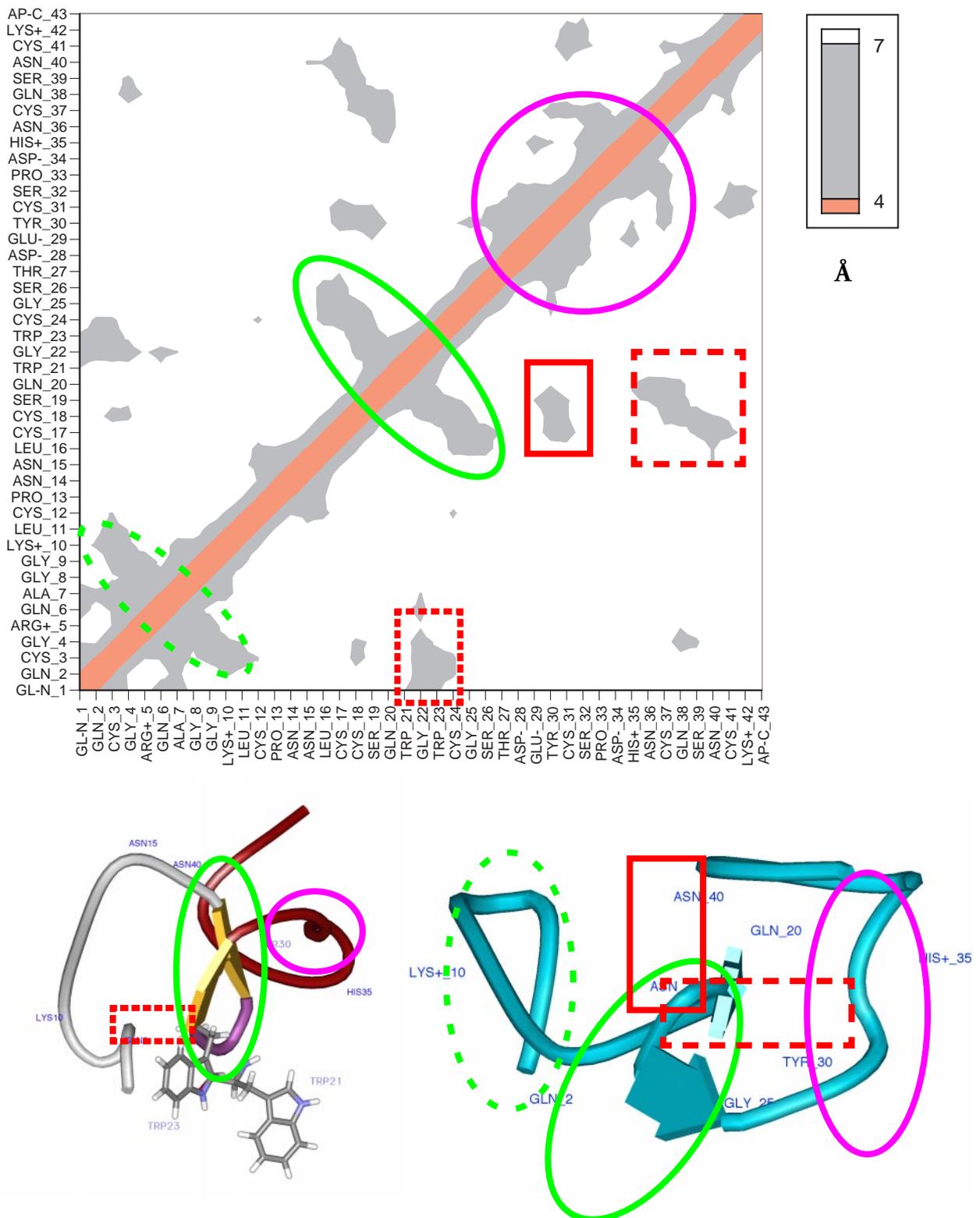
**Abb. X203:** Gal $\beta$ -O-iProp: Relative Freie Gridgruppen-Energie einer G560/G6OH Karte, berechnet aus einem vierdimensionalen G560/G6OH/G4OH/G3OH Grid (Gridintervall = 20°, Kraftfeld = CVFF,  $\varepsilon = 4$ )



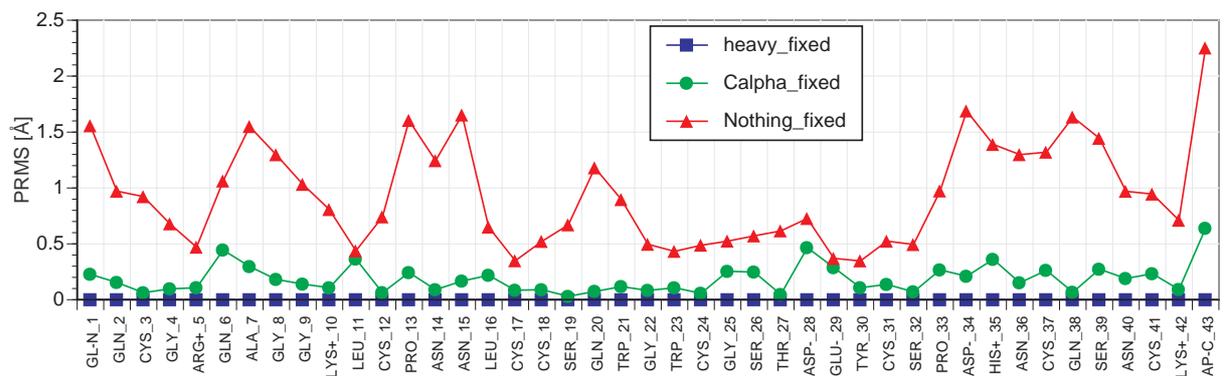
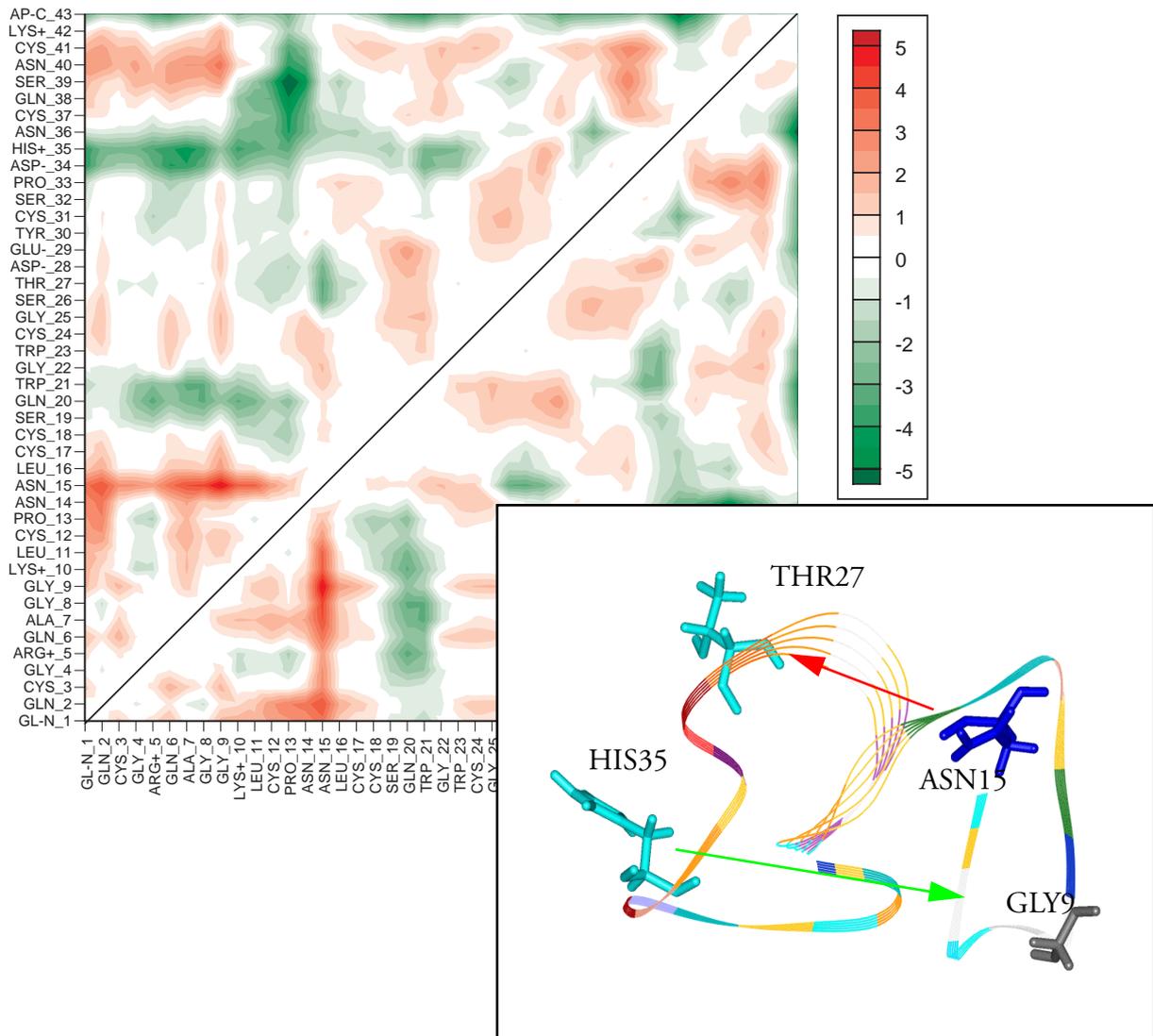
**Abb. X204:** Gal $\beta$ -O-iProp: Einfluß der G4OH Orientierung auf die G560/G6OH Karte. Adiabatic Grid. Datenbasis: vierdimensionales G560/G6OH/G4OH/G3OH Grid (Gridintervall = 20°, Kraftfeld = CVFF,  $\varepsilon = 4$ )



**Abb. X205:** Gal $\beta$ -O-iProp: Rotamerenverteilung der Hydroxymethylgruppe von Gal $\beta$ -O-iProp.  
2D: G56O/G6OH Grid (Gridintervallbreite = 10°, Kraftfeld = CVFF,  $\epsilon = 4$ )  
4D: G56O/G6OH/G4OH/G3OH Grid (Gridintervallbreite = 20°, Kraftfeld = CVFF,  $\epsilon = 4$ )



**Abb. X206:** Beschreibung der Hevein-Struktur durch die Abstandsmatrix der C $\alpha$ -Atome (Abstände zwischen 4 Å und 7 Å sind grau dargestellt). Illustration der zu den „Intensitätsmustern“ gehörenden Sekundärstrukturelemente und „Fernkontakte“



**Abb. X207:** Überprüfung der Stabilität der Hevein-Struktur bei einer Minimierung ohne Fixierung durch Differenzmatrix-Methode der C $\alpha$ -Abstände (oben) und die Berechnung des Rigid-Body-RMS für jede Aminosäure getrennt (unten).

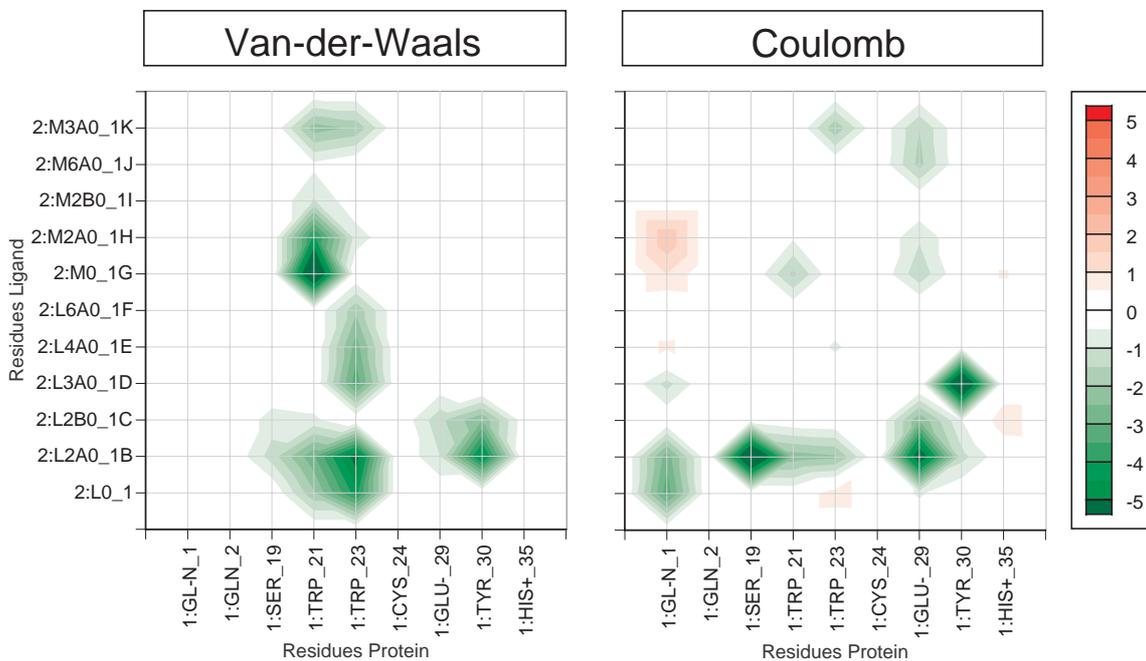
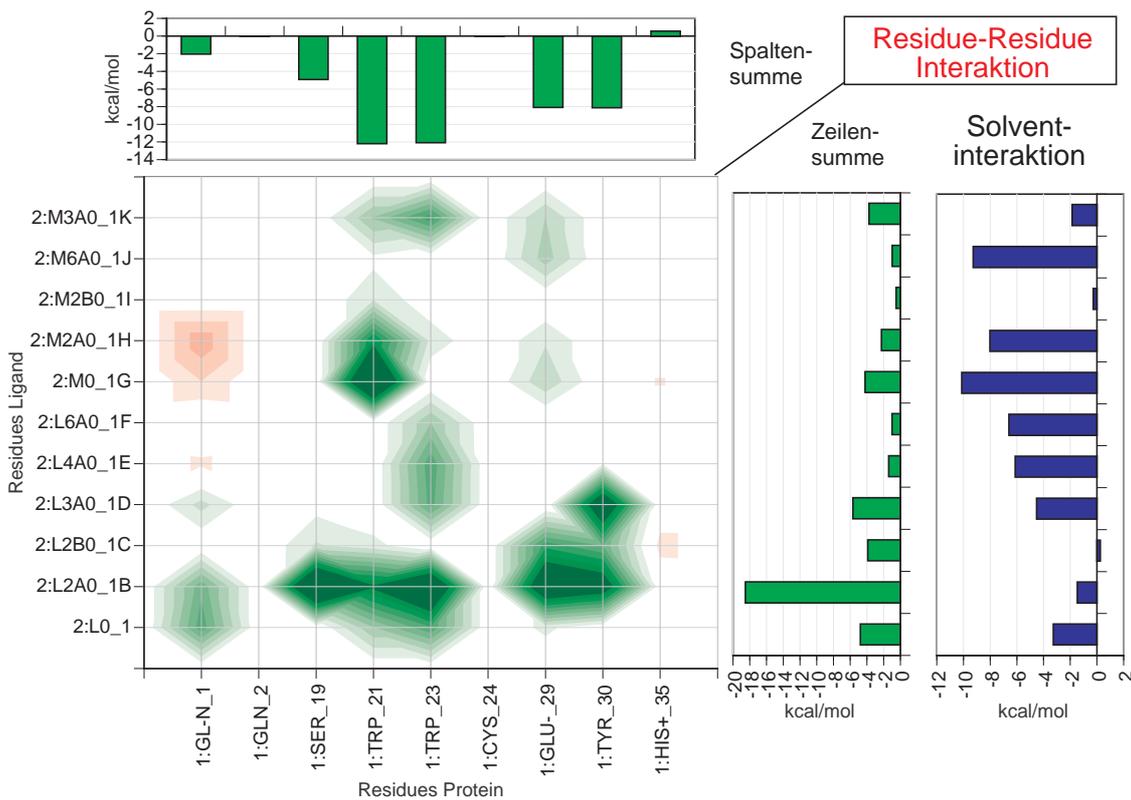
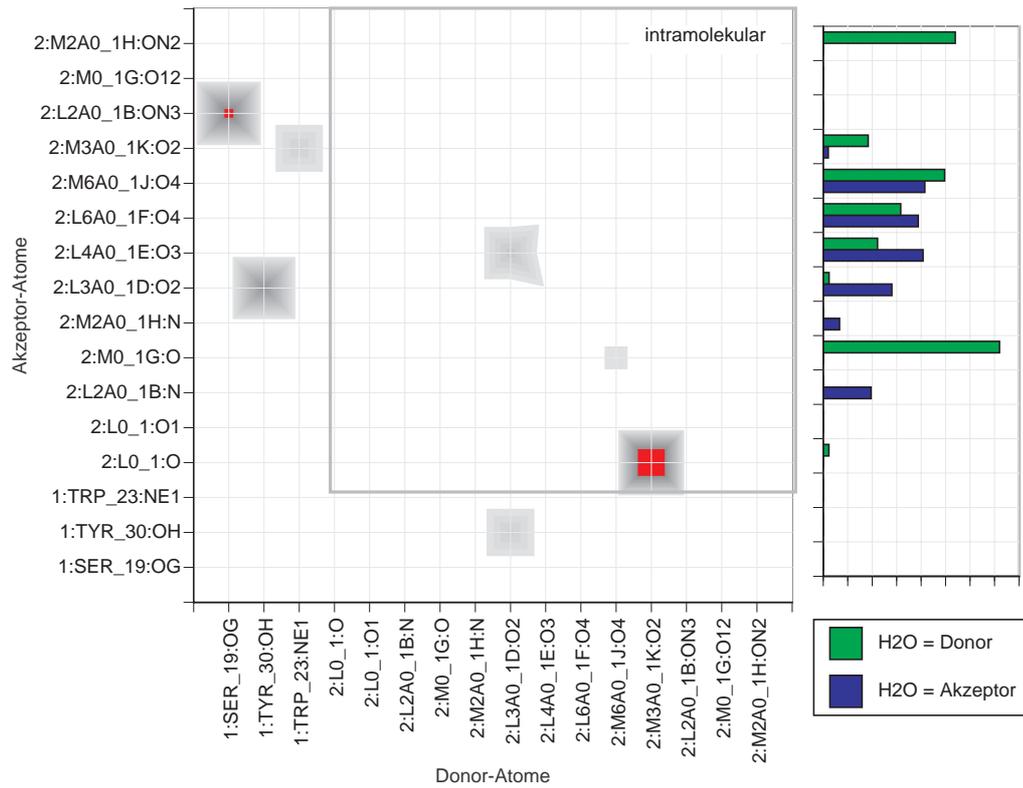
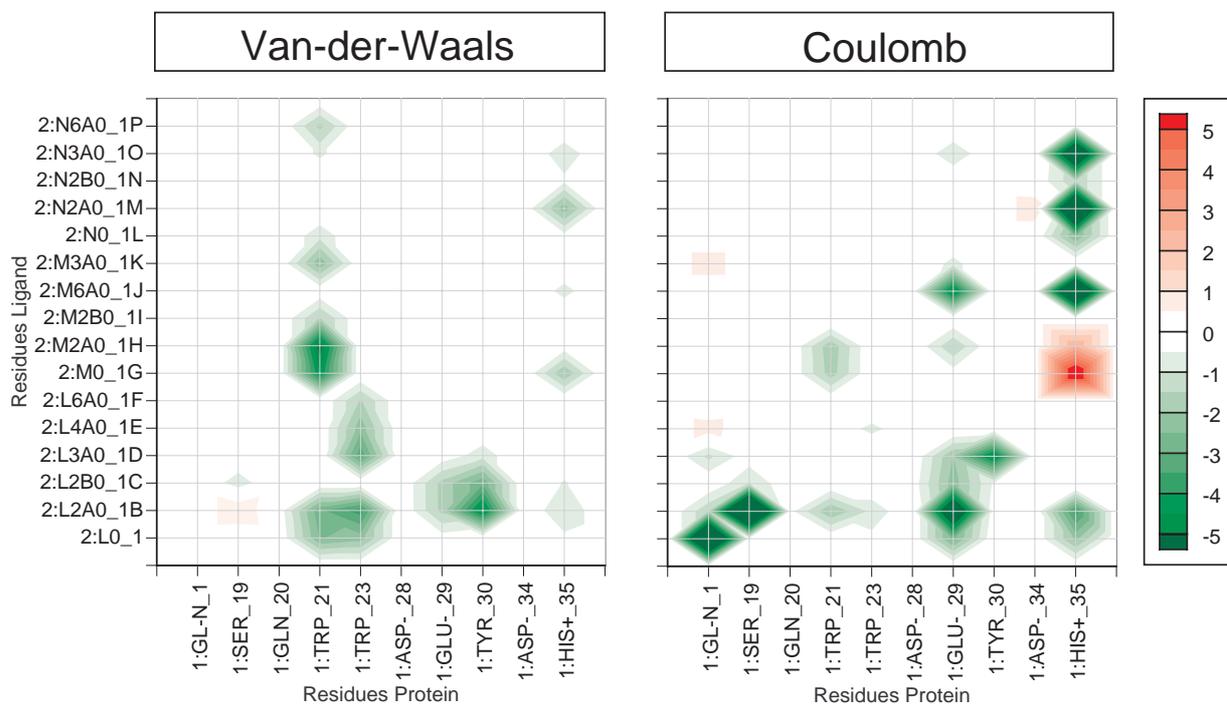
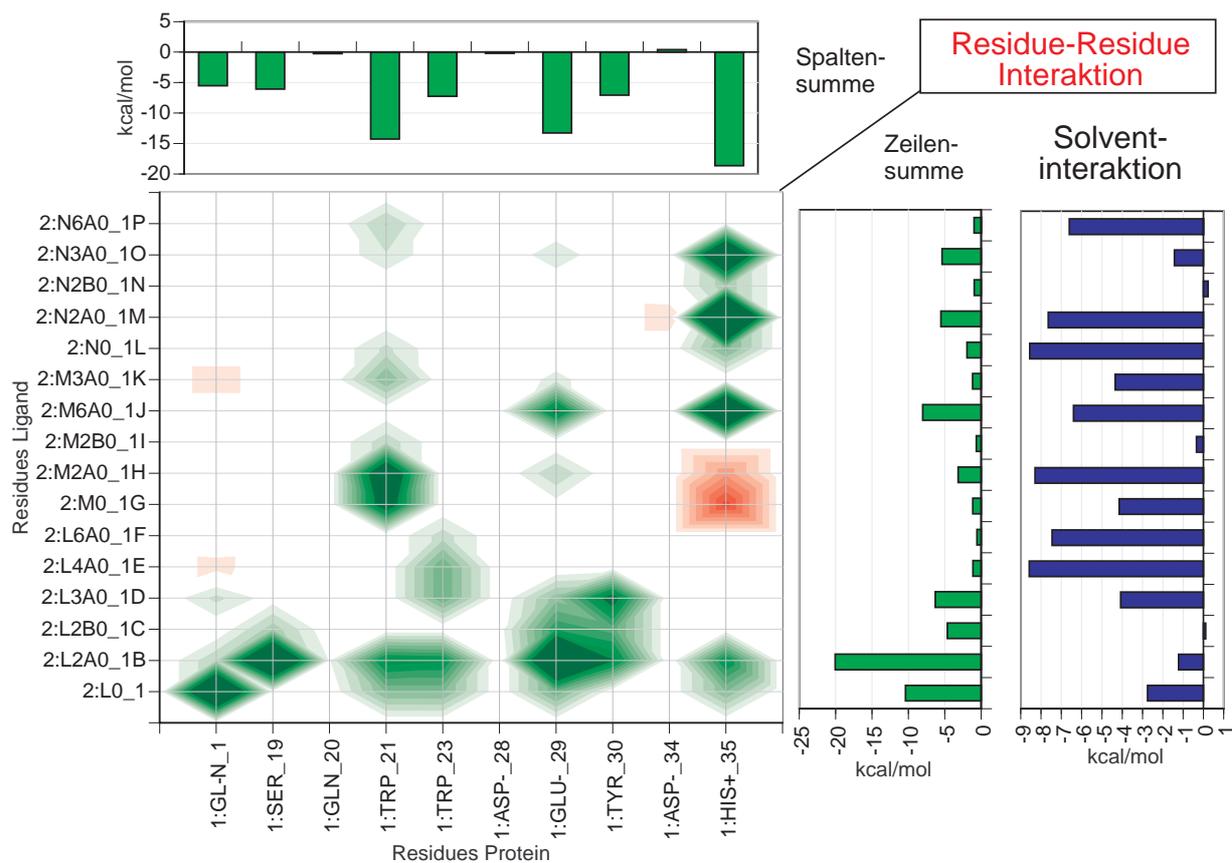


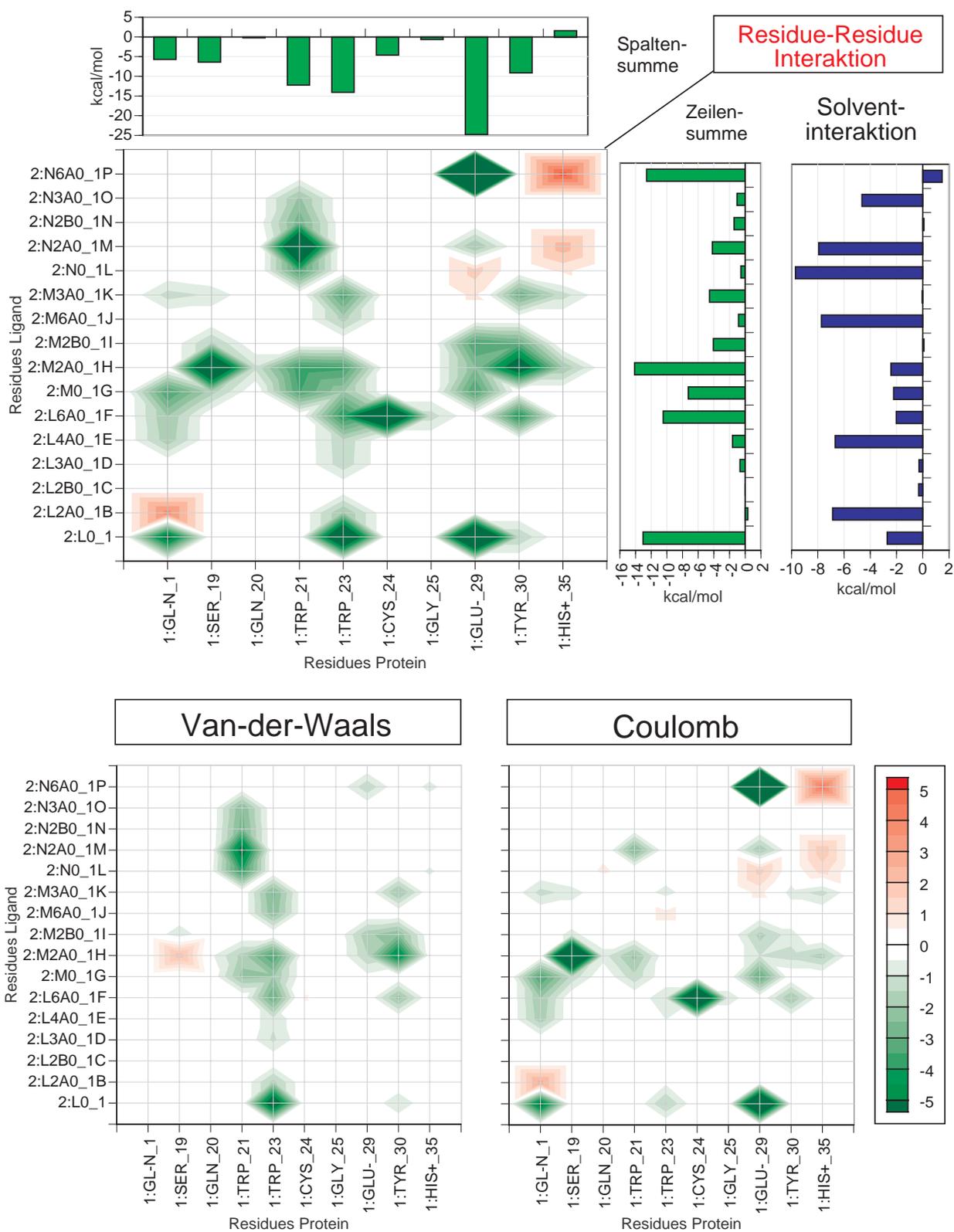
Abb. X208: Wechselwirkungsanalyse für Hevein und Chitobiose



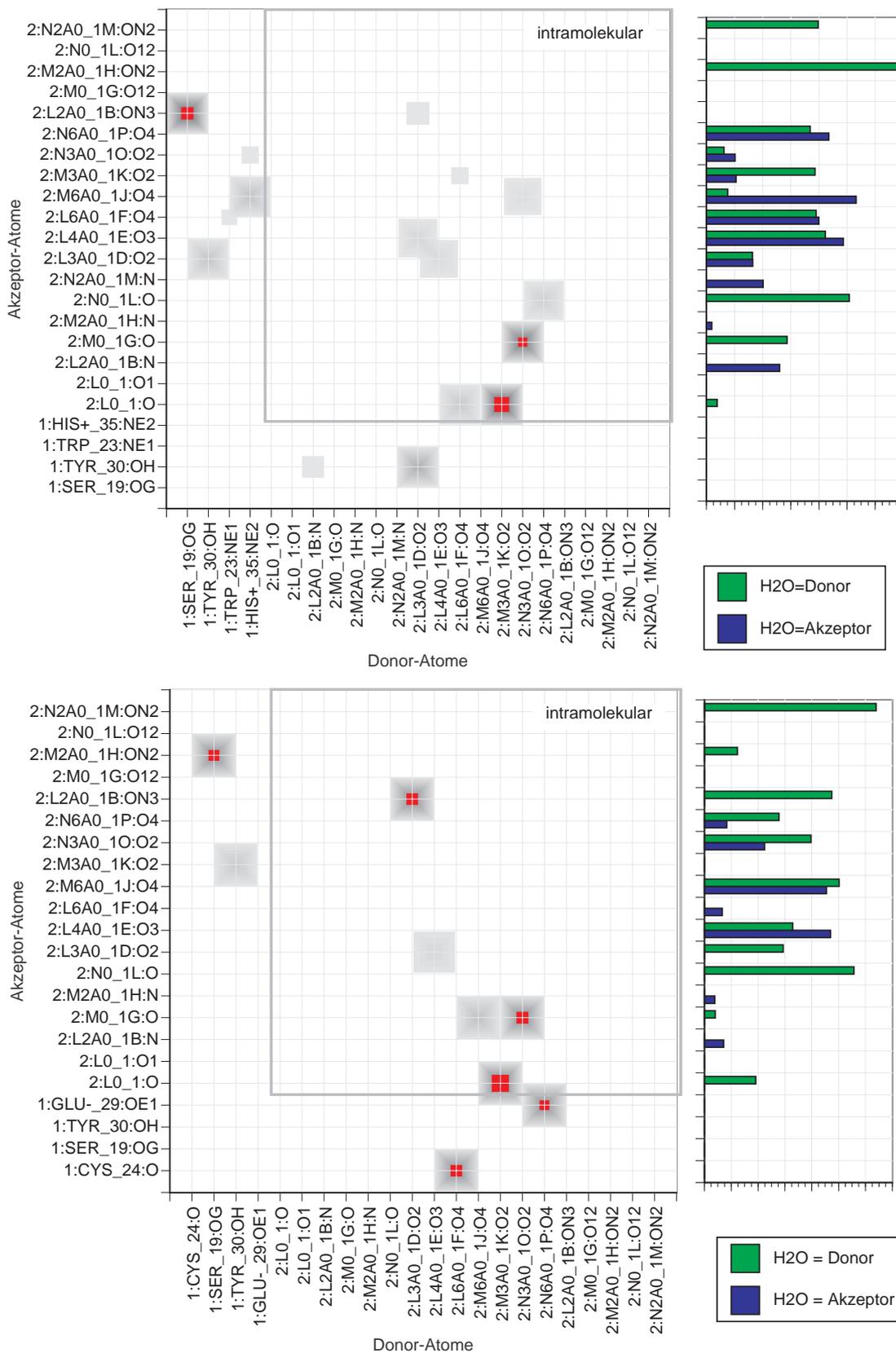
**Abb. X209:** H-Brücken zwischen Hevein und Chitobiose



**Abb. X210:** Wechselwirkungsenergie für Hevein und Chitotriose in der Orientierung A



**Abb. X211:** Wechselwirkungsenergie für Hevein und Chitotriose in der Orientierung B



**Abb. X212:** H-Brücken zwischen Hevein und Chitotriose in der Orientierung A (oben) und B (unten).

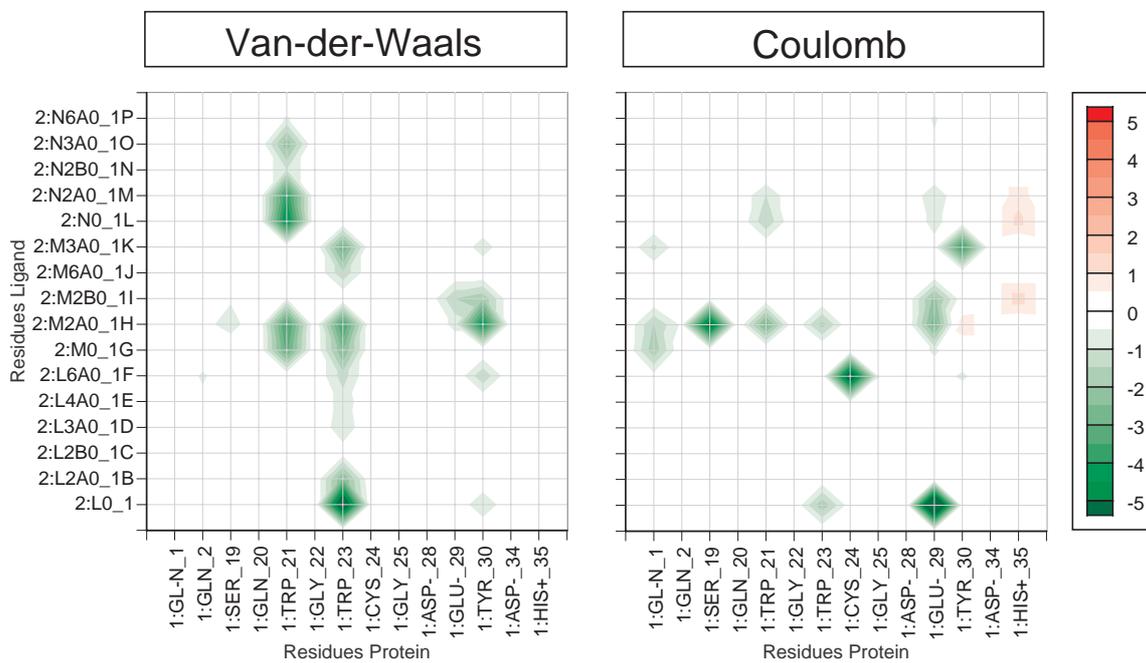
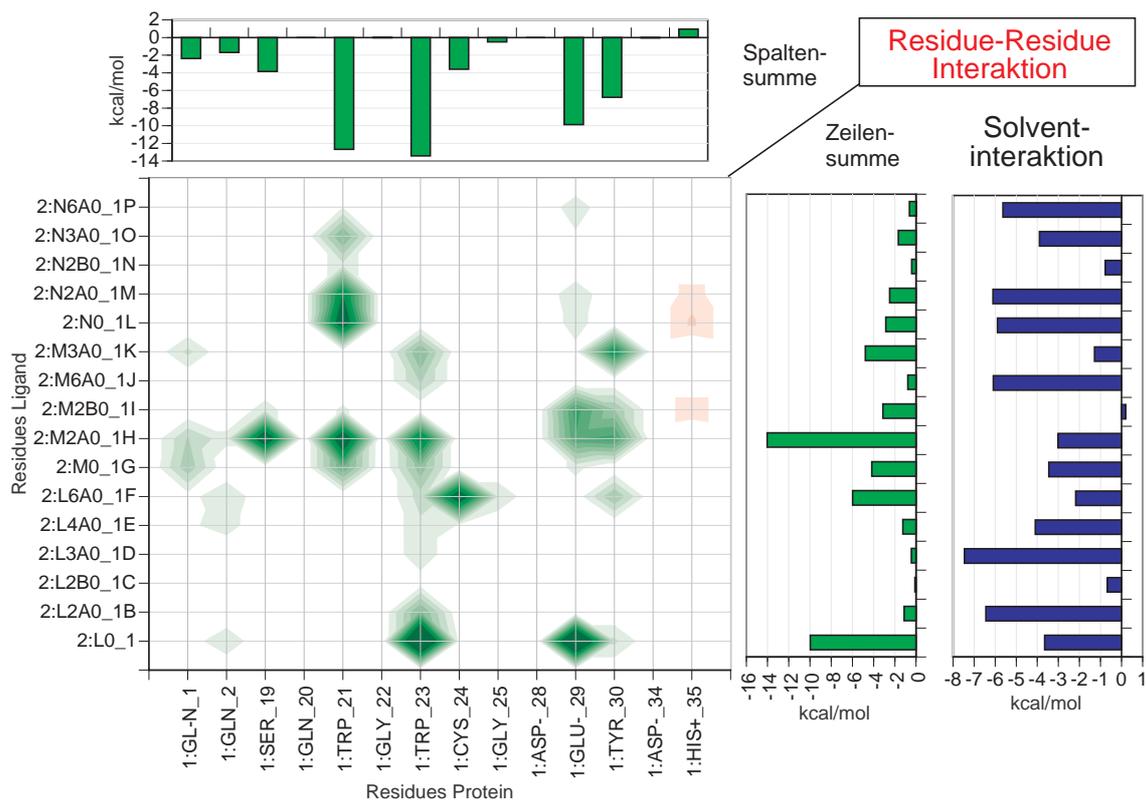
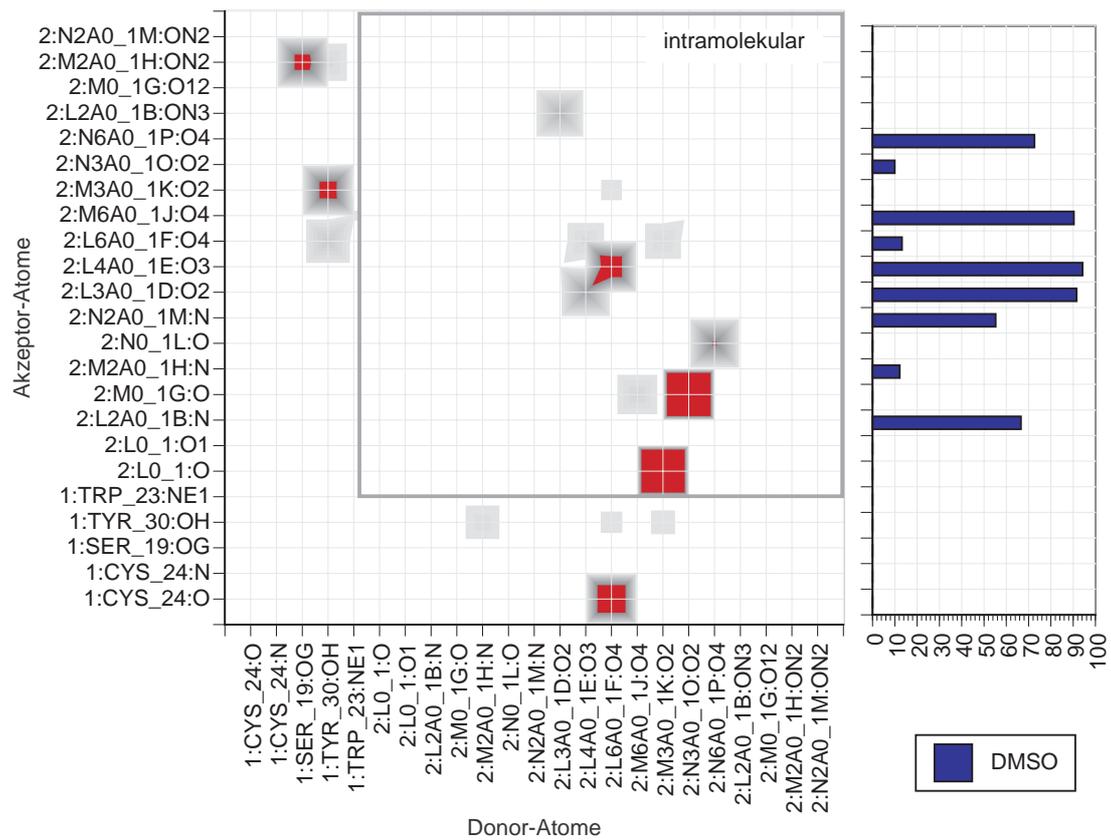
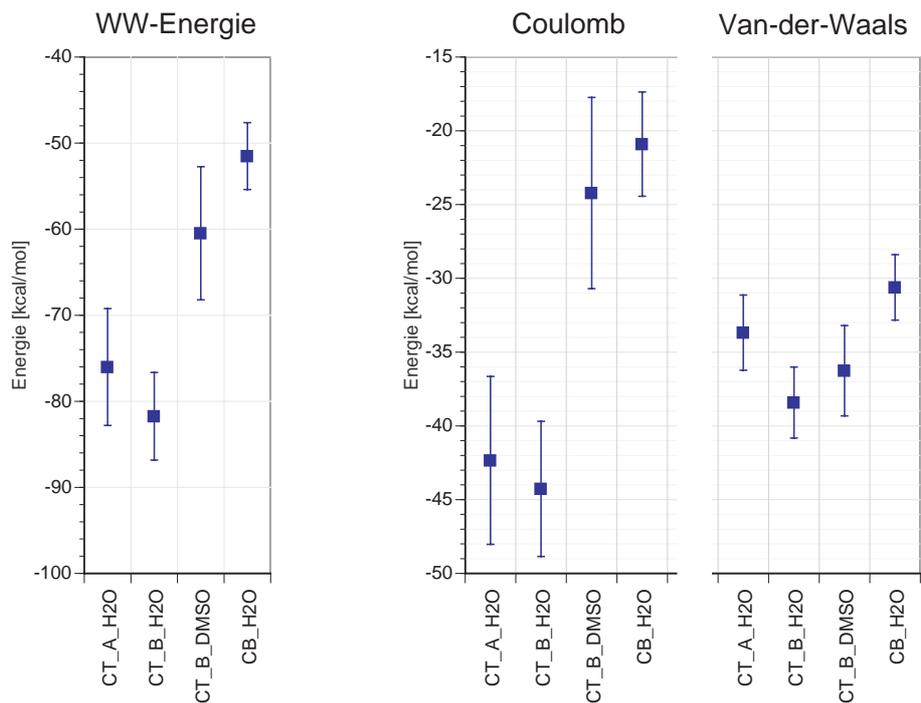


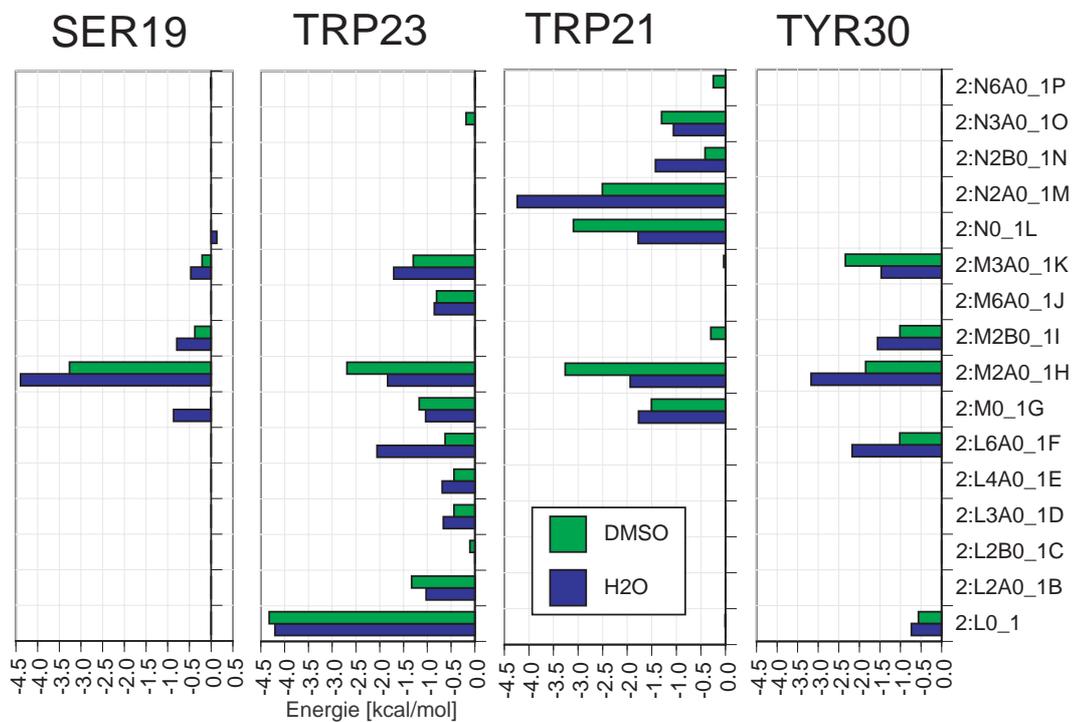
Abb. X213: Interaktion zwischen Hevein und Chitotriose in der Orientierung B in DMSO



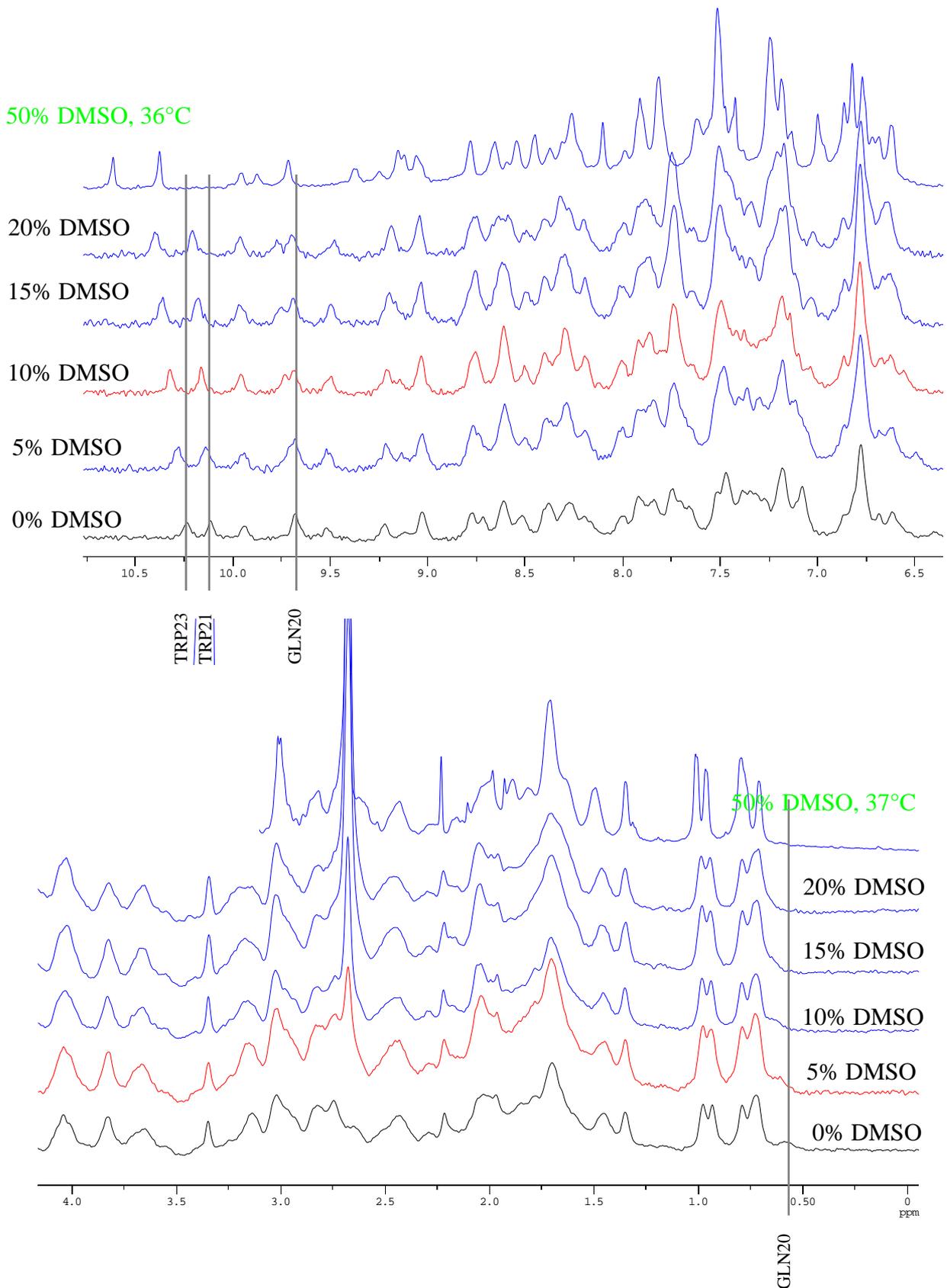
**Abb. X214:** Interaktion zwischen Hevein und Chitotriose in der Orientierung B in DMSO



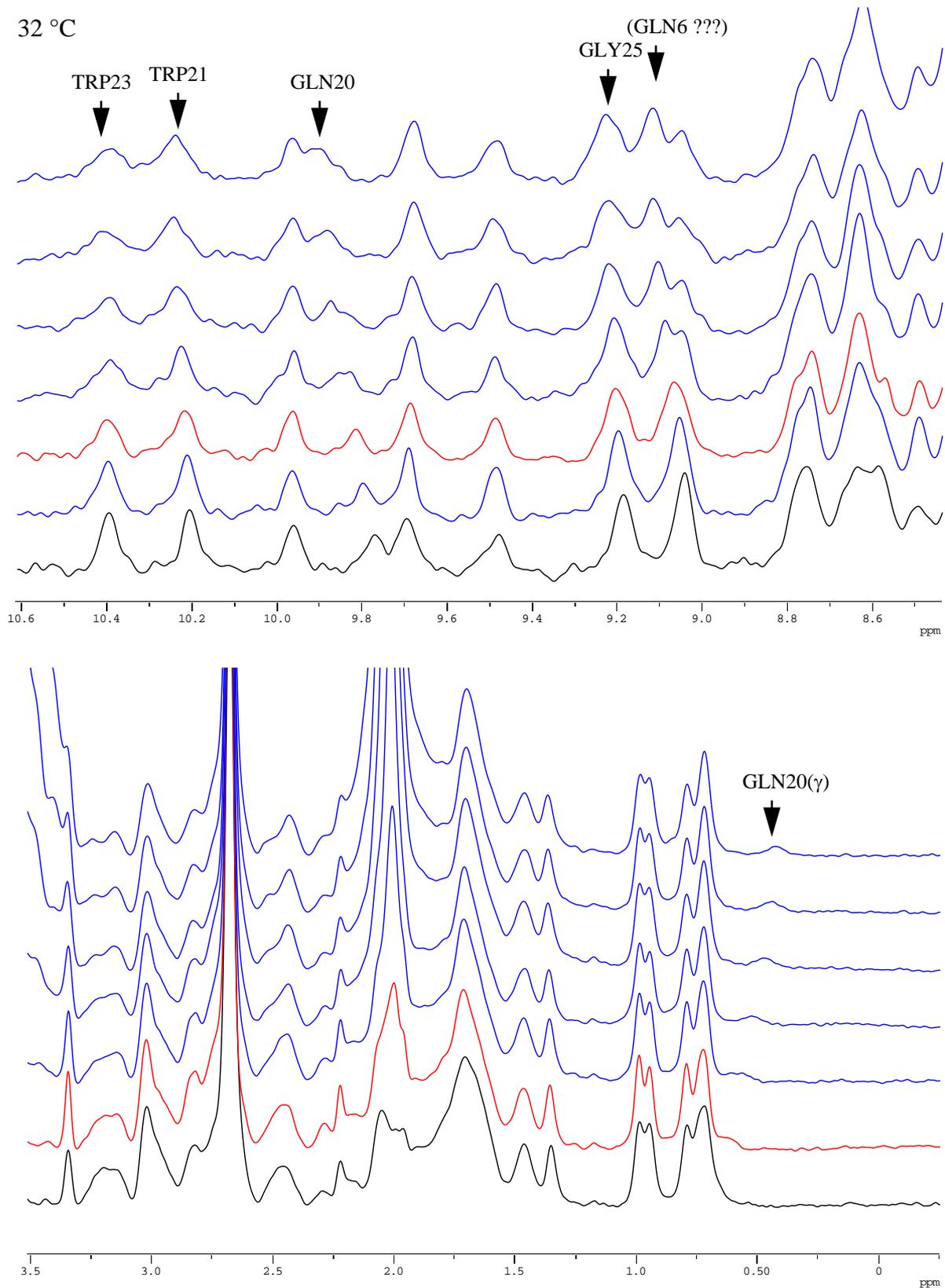
**Abb. X215:** Mittlere Wechselwirkungsenergie zwischen Hevein und Chitotriose (CT) in Wasser (Orientierungen A und B) und in DMSO (Orientierung B), sowie Hevein/Chitobiose (CB) in Wasser. (DISCOVER, „print molecule-molecule interaction“)



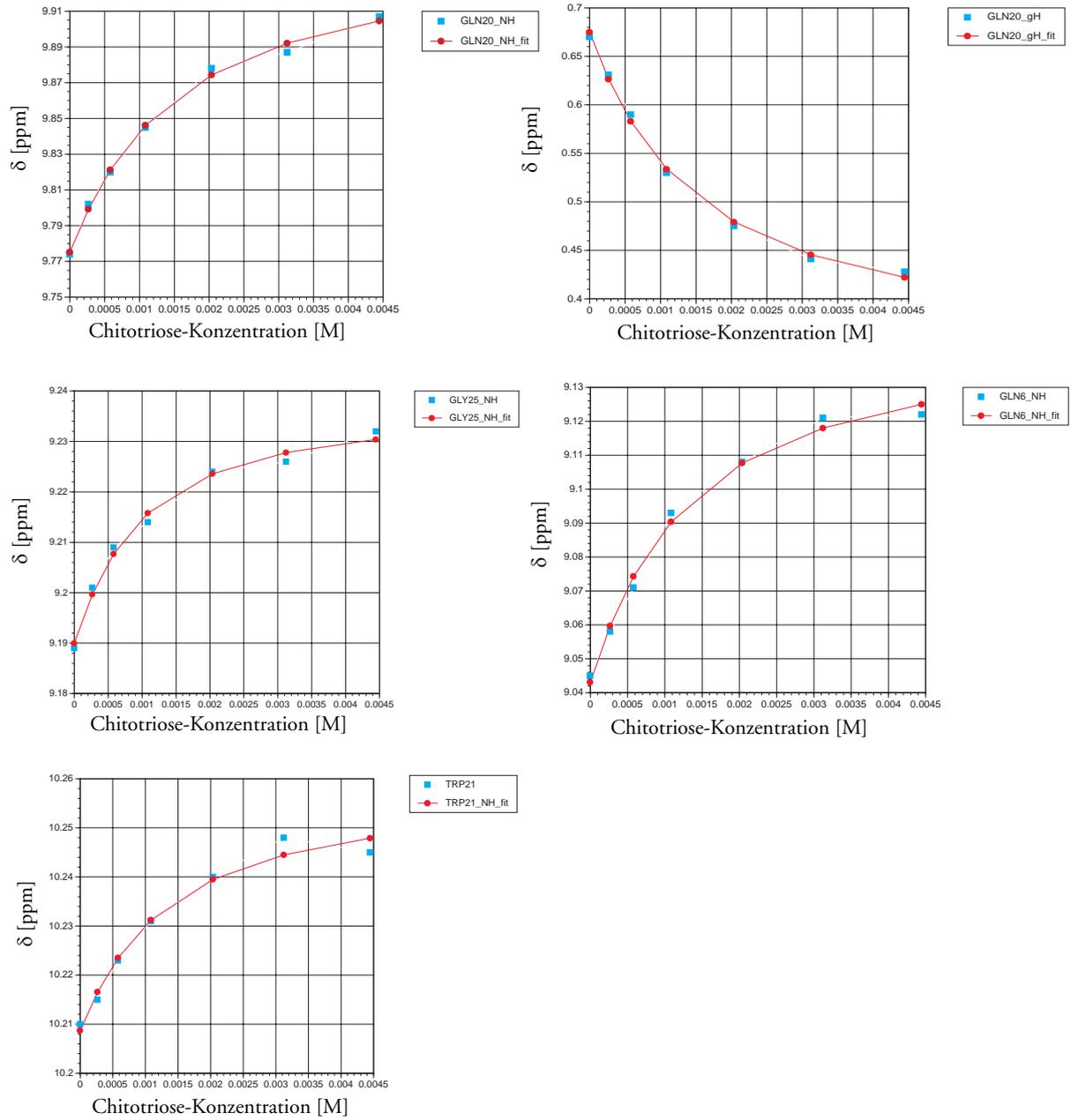
**Abb. X216:** Vergleich der Residue-Residue-Wechselwirkungsenergie mit den für die Bindung wichtigen Aminosäuren für eine MD-Simulation in Wasser und in DMSO



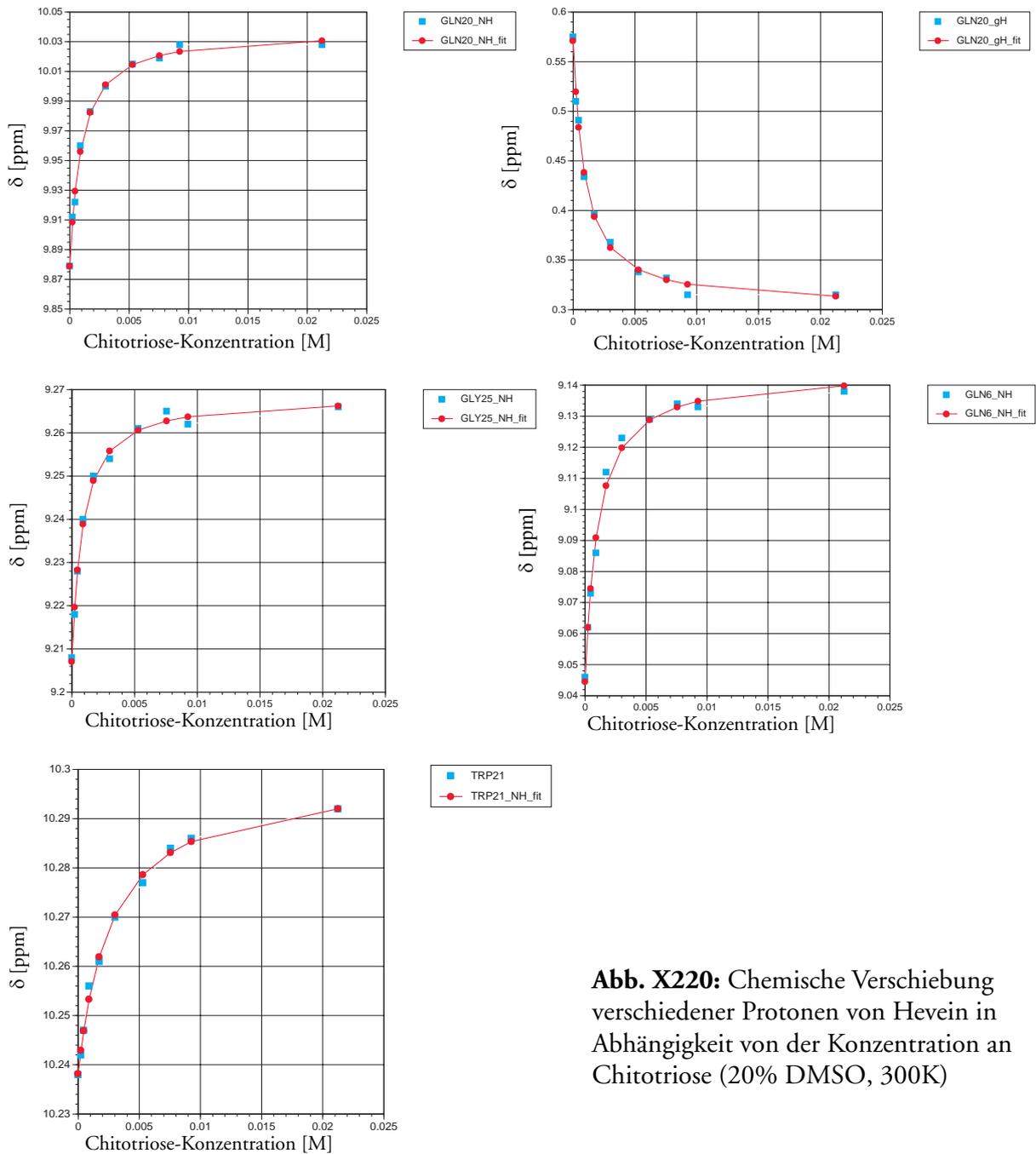
**Abb. X217:**  $^1\text{H-NMR}$ -Spektren einer 0.5 mM Hevein-Lösung mit variabler DMSO-Konzentration ( $T=305$  K, das abgebildete Spektrum mit 50% DMSO stammt aus einer anderen Messreihe und wurde bei  $T=310$  K aufgenommen.)



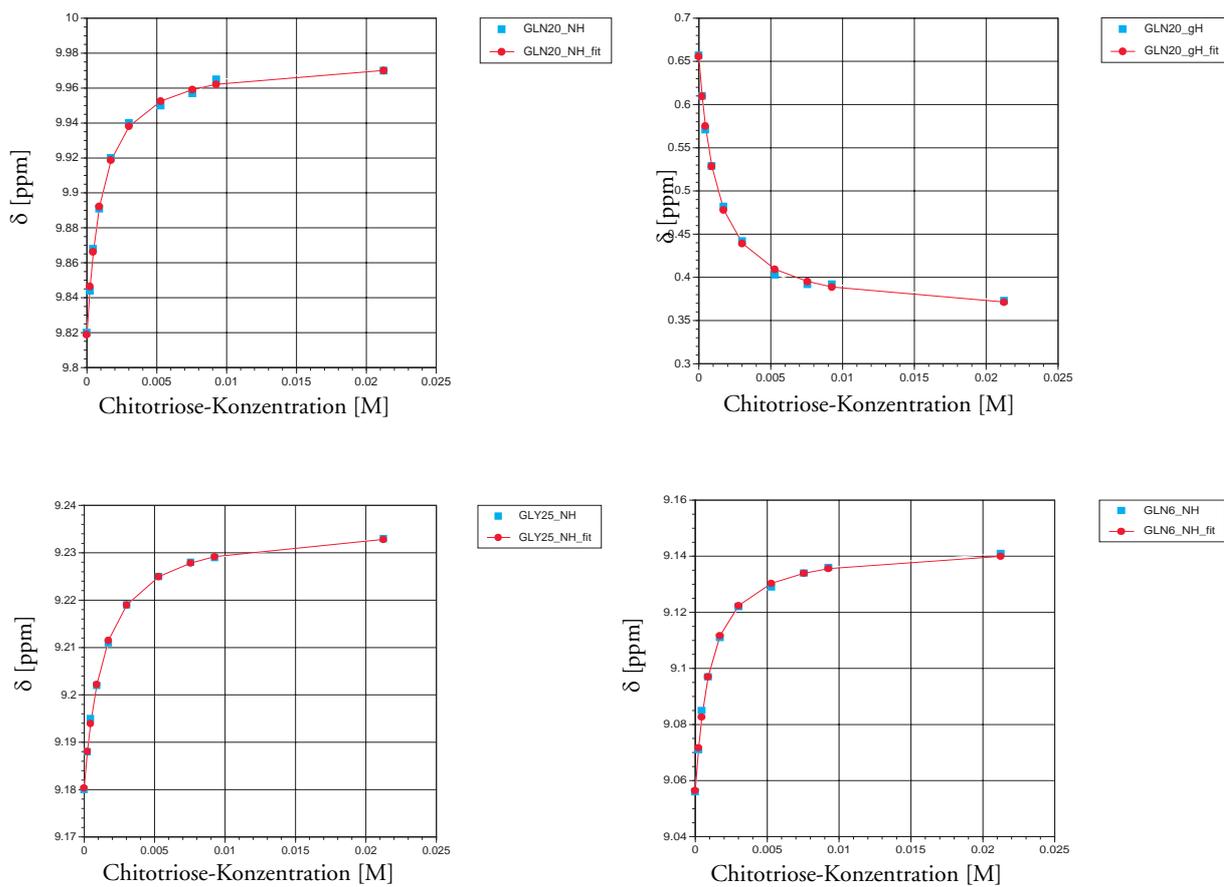
**Abb. X218:**  $^1\text{H}$ -NMR-Spektren einer 0.39 mM Hevein-Lösung mit zunehmender Chitotriose-Konzentration ( $T=305\text{ K}$ )



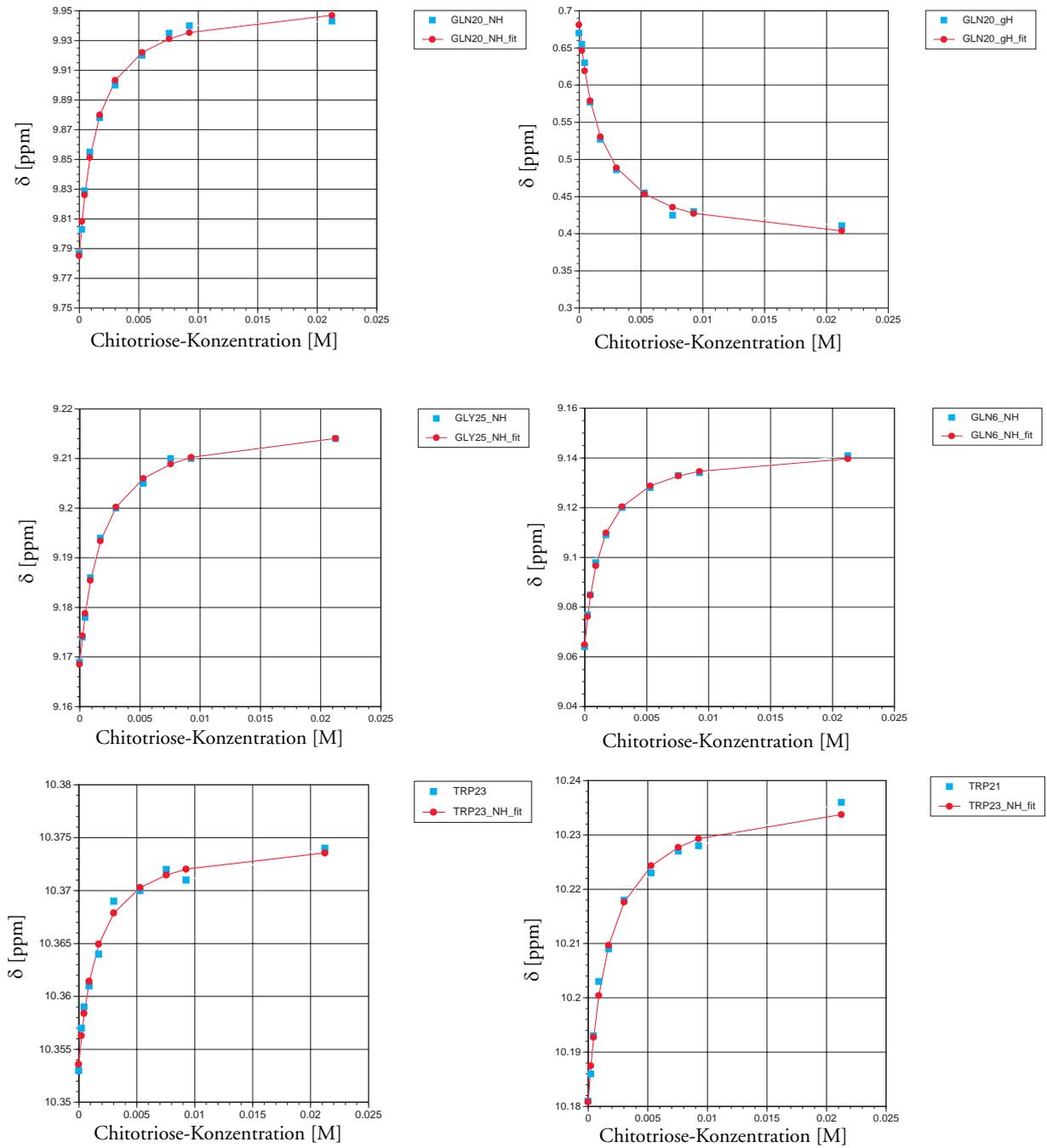
**Abb. X219:** Chemische Verschiebung ausgewählter Protonen von Hevein in Abhängigkeit von der Konzentration an Chitriose (20% DMSO, 305K)



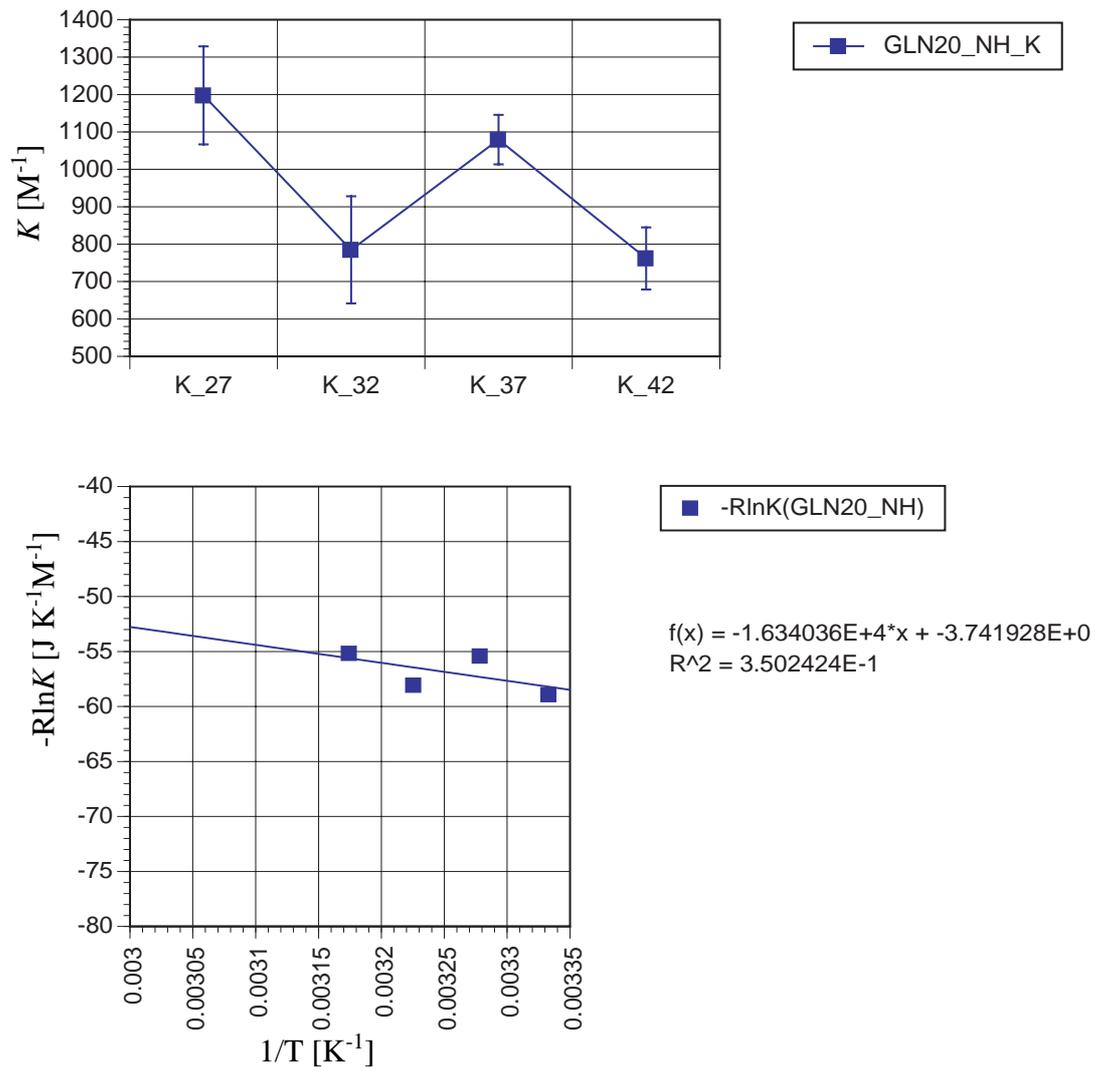
**Abb. X220:** Chemische Verschiebung verschiedener Protonen von Hevein in Abhängigkeit von der Konzentration an Chitriose (20% DMSO, 300K)



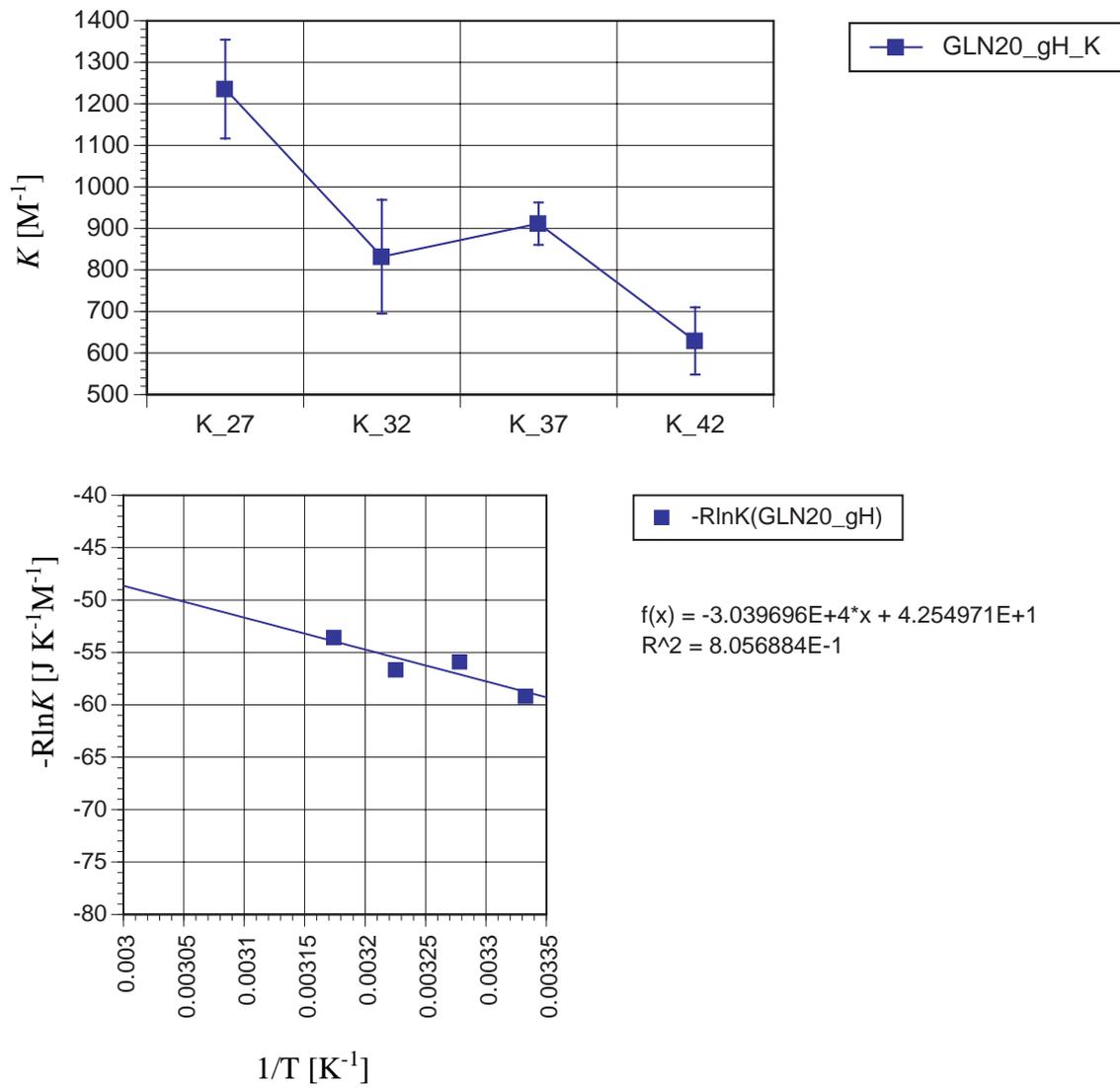
**Abb. X221:** Chemische Verschiebung verschiedener Protonen von Hevein in Abhängigkeit von der Konzentration an Chitotriose (20% DMSO, 310K)



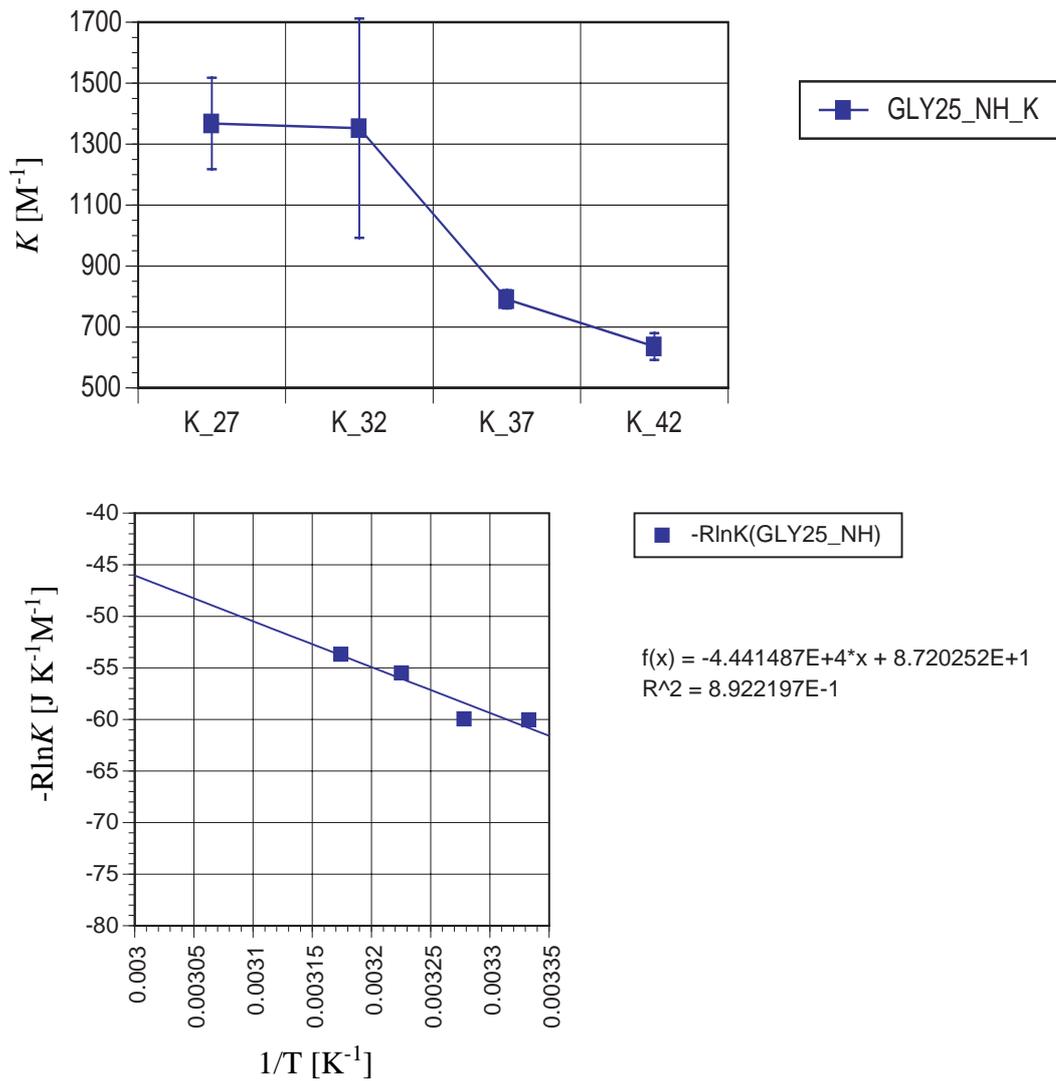
**Abb. X222:** Chemische Verschiebung verschiedener Protonen von Hevein in Abhängigkeit von der Konzentration an Chitriose (20% DMSO, 315K)



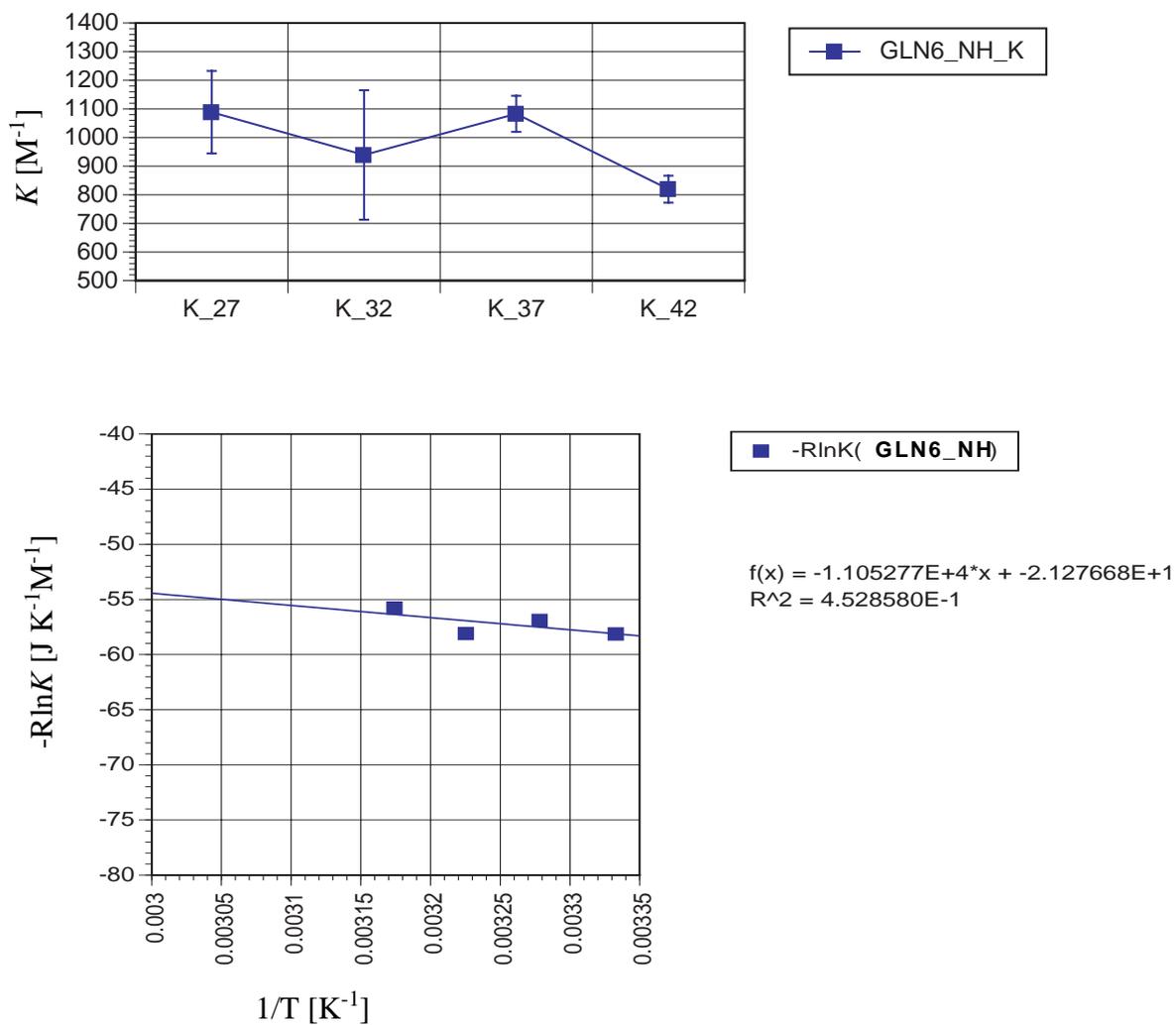
**Abb. X223:** Van't Hoff-Analyse (NH-Proton von Gln20,  $T = 300$  K, 305 K, 310 K, 315 K)



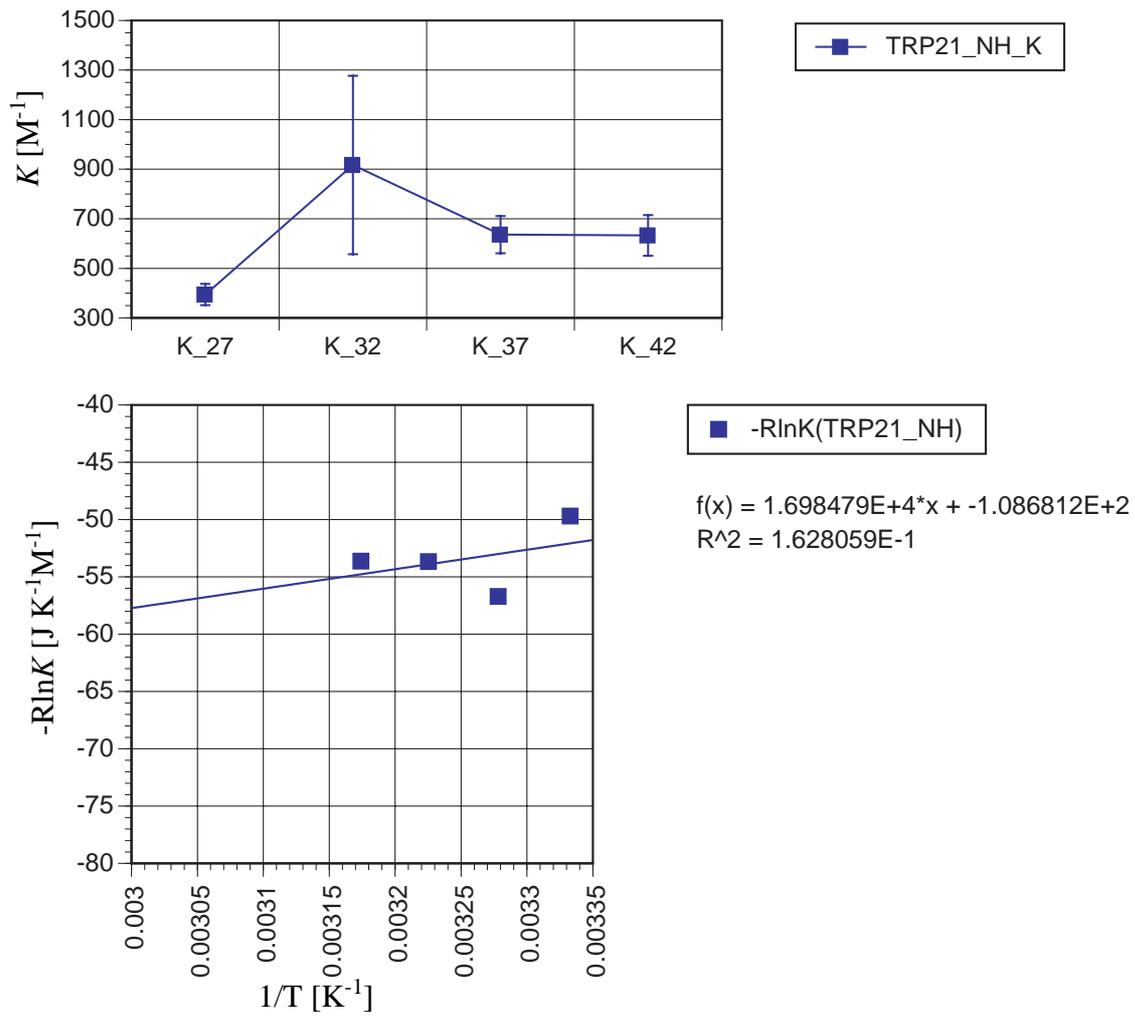
**Abb. X224:** Van't Hoff-Analyse ( $\gamma$ -Proton von Gln20,  $T = 300 \text{ K}, 305 \text{ K}, 310 \text{ K}, 315 \text{ K}$ )



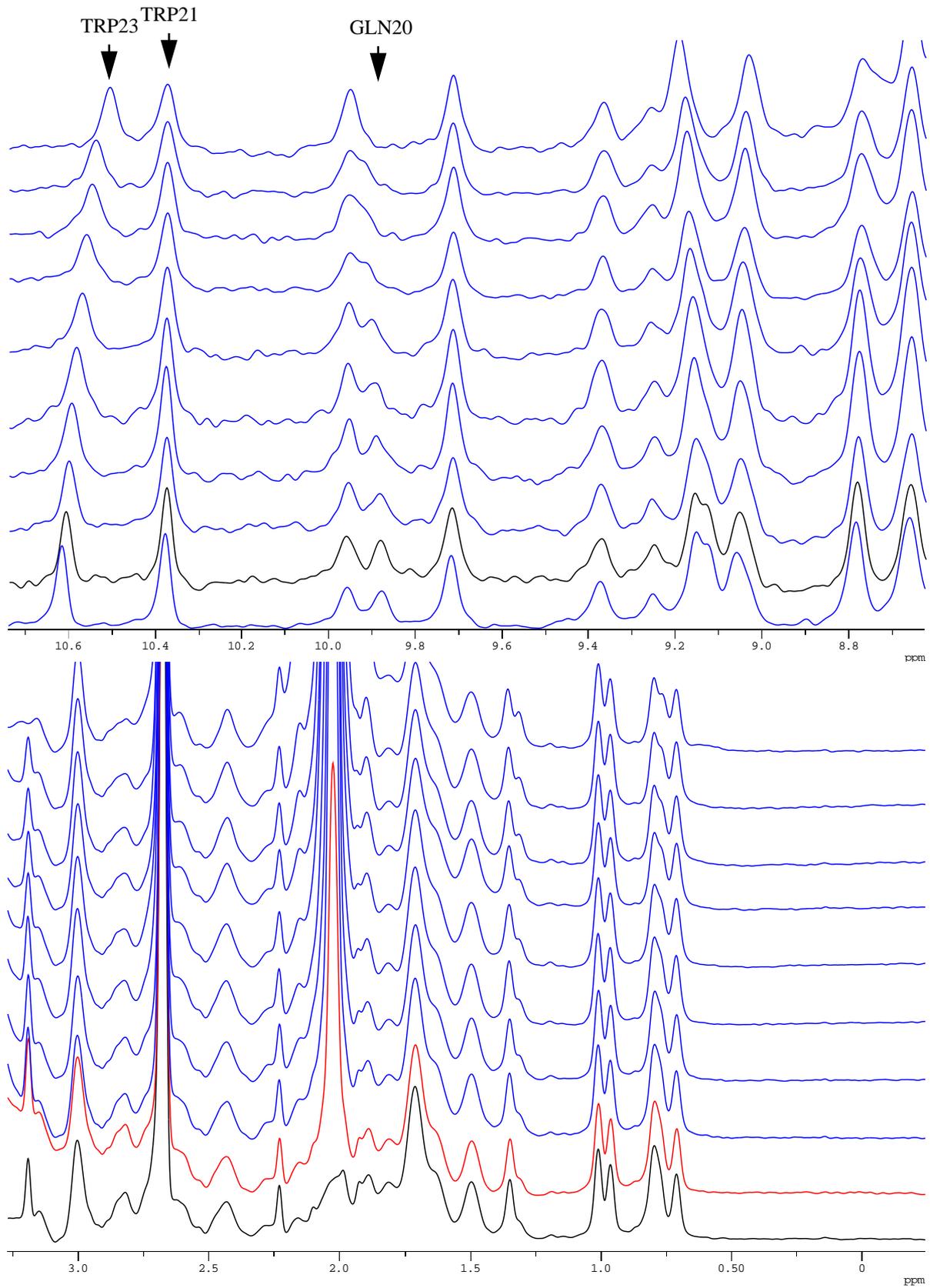
**Abb. X225:** Van't Hoff-Analyse (NH-Proton von Gly25,  $T = 300$  K, 305 K, 310 K, 315 K)



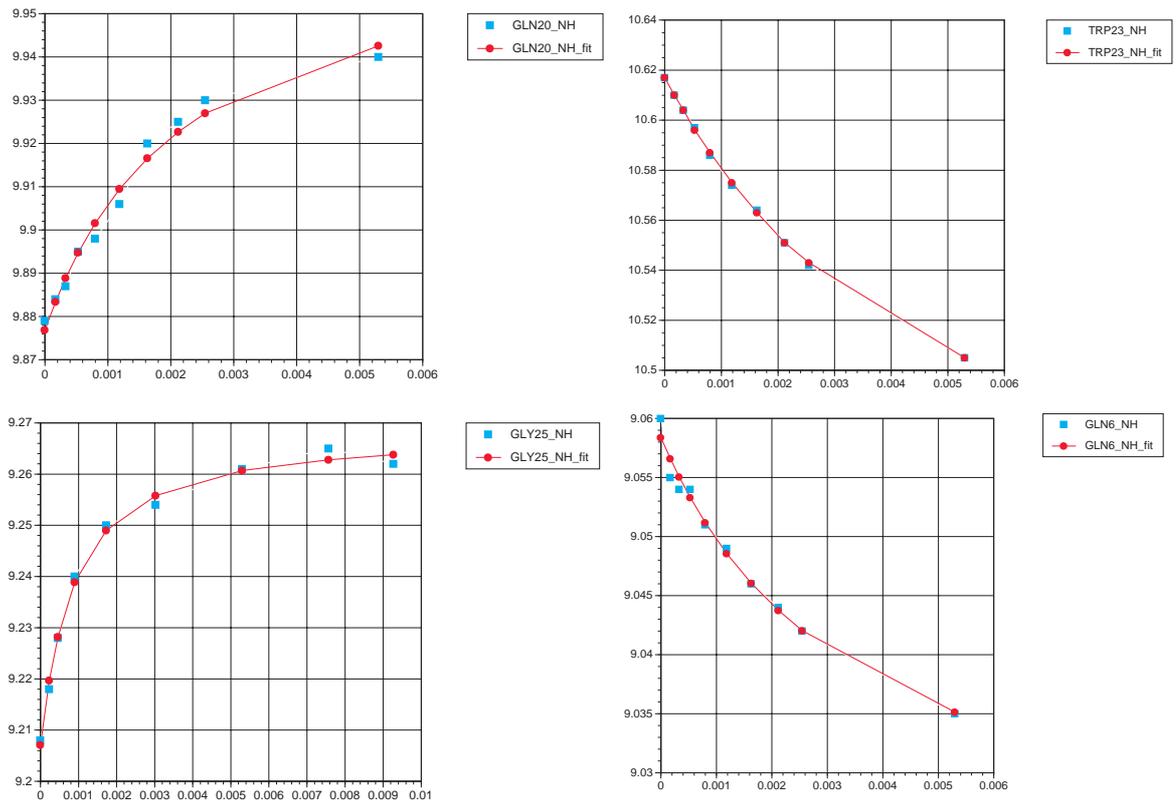
**Abb. X226:** Van't Hoff-Analyse (NH-Proton von „Gln6“,  $T = 300$  K, 305 K, 310 K, 315 K)



**Abb. X227:** Van't Hoff-Analyse (NH-Proton von Trp21,  $T = 300$  K, 305 K, 310 K, 315 K))



**Abb. X228:** NMR-Titration von Chitotriose mit Hevein in 50% DMSO ( $T=309$  K)



**Abb. X229:** Chemische Verschiebung verschiedener Protonen von Hevein in Abhängigkeit von der Konzentration an Chitotriose (50% DMSO, 309K)



## 9 Literatur

- [1] Voet, D.; Voet, J. G. *Biochemistry*; 2nd ed.; John Wiley & Sons, Inc.: New York, **1995**.
- [2] Lehmann, J. *Kohlenhydrate: Chemie und Biologie*; 2., neubearb. und erw. Aufl.; Thieme: Stuttgart, **1996**.
- [3] Kopitz, K. *Glycolipids: Structure and Funktion*; Gabius, H.-J. and Gabius, S., Ed.; Chapman & Hall: London, **1997**, 163-189.
- [4] Sharon, N.; Lis, H. *Glycoproteins: Structure and Funktion*; Gabius, H.-J. and Gabius, S., Ed.; Chapman & Hall: London, **1997**, 133-162.
- [5] Laine, R. A. *The Information-Storing Potential of the Sugar Code*; Gabius, H.-J. and Gabius, S., Ed.; Chapman & Hall: London, **1997**, 1-14.
- [6] Sharon, N.; Lis, H. *Scientific American* **1993**, 268, 82-89.
- [7] Rüdiger, H.; Rougé, P. *Carbohydrates in Europe* **1998**, 23, 18-22.
- [8] Gabius, H.-J.; Romero, A. *Carbohydrates in Europe* **1998**, 23, 24-29.
- [9] Sharon, N. *Protein Science* **1998**, 7, 2042-2048.
- [10] Bundle, D. R. *Antibody-Oligosaccharide Interactions Determined by Crystallography*; Gabius, H.-J. and Gabius, S., Ed.; Chapman & Hall: London, **1997**, 311-331.
- [11] Bundle, D. R.; Young, N. M. *Curr. Opin. Struct. Biol.* **1992**, 2, 666-673.
- [12] von Itzstein, M.; Wu, W. Y.; Kok, G. B.; Pegg, M. S.; Dyason, J. C.; Jin, B.; Phan, T. V.; Smythe, M. L.; White, H. E.; Oliver, S. W.; Colman, P. M.; Varghese, J. N.; Ryan, D. M.; Woods, J. M.; Bethell, R. C.; Hotham, V. J.; Cameron, J. M.; Penn, C. R. *Nature* **1993**, 363, 418-423.
- [13] Gabius, H.-J. *Eur. J. Biochem.* **1997**, 243, 543-576.
- [14] Kaltner, H.; Stierstorfer, B. *Acta Anat.* **1998**, 161, 162-179.
- [15] Lis, H.; Sharon, N. *Chem. Rev.* **1998**, 98, 637-674.
- [16] Villalobo, A. V.; Gabius, H.-J. *Acta Anat.* **1998**, 161, 110-129.
- [17] André, S.; Kojima, S.; Yamazaki, N.; Fink, C.; Kaltner, H.; Kayser, K.; Gabius, H.-J. *J. Cancer Res. Clin. Oncol.* **1999**, 125, 461-474.
- [18] Diaz-Mauriño, T.; Solis, D.; Jiménez-Barbero, J.; Martín-Lomaz, M.; Siebert, H.-C.; Vliegthart, J. F. G. *Carbohydrates in Europe* **1998**, 23, 36-41.
- [19] Wiley, D. C.; Skehel, J. J. *Ann. Rev. Biochem.* **1987**, 56, 365-94.
- [20] Watowich, S. J.; Skehel, J. J.; Wiley, D. C. *Structure* **1994**, 2, 719-731.
- [21] Weis, W.; Brown, J. H.; Cusack, S.; Paulson, J. C.; Skehel, J. J.; Wiley, D. C. *Nature* **1988**, 333, 426-431.
- [22] Böhm, H.-J.; Klebe, G.; Kubinyi, H. *Wirkstoffdesign*; Spektrum, Akad. Verl.: Heidelberg, **1996**.
- [23] Kelm, S.; Paulson, J. C.; Rose, U.; Brossmer, R.; Schmid, W.; Bandgar, B. P.; Schreiner, E.; Hartmann, M.; Zbiral, E. *Eur. J. Biochem.* **1992**, 205, 147-153.
- [24] Machytka, D.; Kharitononkov, I.; Isecke, R.; Hetterich, P.; Brossmer, R.; Klein, R. A.; Klenk, H. D.; Egge, H. *FEBS Letters* **1993**, 334, 117-120.
- [25] Matrosovich, M. N.; Gambaryan, A. S.; Tuzikov, A. B.; Byramova, N. E.; Mochalova, L. V.; Golbraikh, A. A.; Shenderovich, M. D.; Finne, J.; Bovin, N. V. *Virology* **1993**, 196, 111-121.
- [26] Mochalova, L. V.; Tuzikov, A. B.; Marinina, V. P.; Gambaryan, A. S.; Byramova, N. E.; Bovin, N. V.; Matrosovich, M. N. *Antiviral Research* **1994**, 23, 179-190.

- [27] Sauter, N. K.; Hanson, J. E.; Glick, G. D.; Brown, J. H.; Crowther, R. L.; Park, S. J.; Skehel, J. J.; Wiley, D. C. *Biochemistry* **1992**, *31*, 9609-9621.
- [28] von der Lieth, C.-W.; Kozar, T. J. *Mol. Struct. (Theochem)* **1996**, *368*, 213-222.
- [29] Keppler, O.; Herrmann, M.; von der Lieth, C.-W.; Stehling, P.; Reutter, W.; Pawlita, M. *Biochem. Biophys. Res. Commun.* **1998**, *253*, 437-442.
- [30] Taylor, N. R.; von Itzstein, M. *Journal of Computer-Aided Molecular Design* **1996**, *10*, 233-246.
- [31] Laver, W. G.; Bischofberger, N.; Webster, R. G. *Scientific American* **1999**, *1*, 56-65.
- [32] Colman, P. M.; Laver, W. G.; Varghese, J. N.; Baker, A. T.; Tulloch, P. A. *Nature* **1987**, *326*, 358-363.
- [33] Stehle, T.; Harrison, S. C. *Structure* **1996**, *4*, 183-194.
- [34] Gabius, H. J.; Gabius, S. *Glycosciences: Status and Perspectives*; Chapman & Hall: London, **1996**.
- [35] Reuter, G.; Gabius, H. J. *Biol. Chem. Hoppe-Seyler* **1996**, *377*, 325-342.
- [36] Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683-720
- [37] Lasky, L. *Science* **1992**, *258*, 964-969.
- [38] Gabius, H. J. *Angew. Chem. Int. Ed. Engl.* **1988**, *27*, 1267-1276.
- [39] Gabius, H.-J.; André, S.; Kaltner, H.; Siebert, H.-C.; Kayser, G.; von der Lieth, C.-W.; Bovin, N. V.; Brinck, U.; Kayser, K.; Vliegthart, J. F. G. *Bioforum* **1996**, *19*, 503-506.
- [40] Elgavish, S.; Shaanan, B. *Trends in Pharmacological Sciences* **1997**, *22*, 462-467.
- [41] Barondes, S. H.; Cooper, D. N. W.; Gitt, M. A.; Leffler, H. J. *Biol. Chem.* **1994**, *269*, 20807-20810.
- [42] Ohannesian, D. W.; Lotan, R. *Galectins in Tumor Cells*; Gabius, H.-J. and Gabius, S., Ed.; Chapman & Hall: London, **1997**, 459.
- [43] Gabius, H.-J.; Kayser, K.; Gabius, S. *Naturwissenschaften* **1995**, *82*, 533-543.
- [44] Witczak, Z. J. *Curr. Med. Chem.* **1999**, *6*, 92-178.
- [45] Rüdiger, H.; Siebert, H.-C.; Solís, D.; Jiménez-Barbero, J.; Romero, A.; von der Lieth, C.-W.; Diaz-Mauriño, T.; Gabius, H.-J. *Curr. Med. Chem.* **1999**, *in press*.
- [46] Hermann, M.; von der Lieth, C.-W.; Stehling, P.; Reutter, W.; Pawlita, M. *J. Virol.* **1997**, *71*, 5922-5931.
- [47] Karlsson, K. A. *Trends in Pharmacological Sciences* **1991**, *12*, 265-272.
- [48] Neidle, S. *Annals of Oncology* **1994**, *5*, S51-S54.
- [49] Vyas, N. K. *Curr. Opin. Struct. Biol.* **1991**, *1*, 732-740.
- [50] Rice, K. G.; Wu, P.; Brand, L.; Lee, Y. C. *Curr. Opin. Struct. Biol.* **1993**, *3*, 669-674.
- [51] Siebert, H.-C.; von der Lieth, C.-W.; Gilleron, M.; Reuter, G.; Wittmann, J.; Vliegthart, J. F. G.; Gabius, H.-J. *Carbohydrate-Protein Interaction*; Gabius, H.-J. and Gabius, S., Ed.; Chapman & Hall: London, **1997**, 291-310.
- [52] Weis, W. I.; Drickamer, K.; Henrickson, W. A. *Nature* **1992**, *360*, 127-134.
- [53] Poveda, A.; Asensio, J. L.; Espinosa, J. E.; Martinpastor, M.; Canada, F. J.; Jiménez-Barbero, J. *J. Mol. Graphics Mod.* **1997**, *15*, 9-17.
- [54] Poveda, A.; Jiménez-Barbero, J. *Chemical Society Reviews* **1998**, *27*, 133-143.
- [55] Jiménez-Barbero, J.; Asensio, J. L.; Cañada, F. J., and Poveda, A. *Curr. Opin. Struct. Biol.* **1999**, *9*, 549-555.
- [56] Otting, G. *Curr. Opin. Struct. Biol.* **1993**, *3*, 760-768.
- [57] Wagner, G. *Nat. Struct. Biol.* **1997**, *10*, 841-844.

- [58] Peters, T.; Pinto, B. M. *Curr. Opin. Struct. Biol.* **1996**, *6*, 710-720.
- [59] Miller, E. K.; Mukhopadhyay, C.; Perseveranda, C.; Bush, C. A. *Biochemistry* **1992**, *31*, 6703-6709.
- [60] Kessler, H.; Gehrke, M.; Griesinger, C. *Angew. Chem. Int. Ed. Engl.* **1988**, *27*, 490-536.
- [61] Gronenborn, A. M.; Clore, G. M. *Biochem. Pharm.* **1990**, *40*, 115-119.
- [62] Bose, P. K.; Polavarapu, P. L. *J. Am. Chem. Soc.* **1999**, *121*, 6094-6095.
- [63] Leefflang, B. R.; Vliegthart, J. F. G.; Kroon-Batenburg, L. M. J.; van Eijck, B. P.; Kroon, J. *Carbohydr. Res.* **1992**, *230*, 41-61.
- [64] Poppe, L.; van Halbeek, H. *Nature Struct. Biol.* **1994**, *1*, 215-216.
- [65] Casu, B.; Reggiani, M.; Gallo, G. G.; Vigevani, A. *Tetrahedron* **1966**, *22*, 3061-3083.
- [66] Dabrowski, J.; Hanfland, P.; Egge, H.; Dabrowski, U. *Arch. Biochem. Biophys.* **1981**, *210*, 405-411.
- [67] Yan, Z.-Y.; Rao, B. N. N.; Bush, C. A. *J. Am. Chem. Soc.* **1987**, *109*, 7663-7669.
- [68] Gupta, M. N. *Eur. J. Biochem.* **1992**, *203*, 25-32.
- [69] Gómez-Puyou, M. T.; Gómez-Puyou, A. *Crit. Rev. Biochem. Mol. Biol.* **1998**, *33*, 53-89.
- [70] Siebert, H.-C.; von der Lieth, C.-W.; Kaptein, R.; J., B. J.; Dijkstra, K.; van Nuland, N.; Soedjanaamadja, U. M.; Rice, A.; Vliegthart, J. F. G.; Wright, C. S.; Gabius, H. J. *Proteins* **1997**, *28*, 268-284.
- [71] Siebert, H.-C.; Adar, R.; Arango, R.; Burchert, M.; Kaltner, H.; Kayser, G.; Tajkhorshid, E.; von der Lieth, C.-W.; Kaptein, R.; Sharon, N.; Vliegthart, J. F. G.; Gabius, H.-J. *Eur. J. Biochem.* **1997**, *249*, 27-38.
- [72] Tajkhorshid, E.; Siebert, H.-C.; Burchert, M.; Kaltner, H.; Kayser, G.; von der Lieth, C.-W.; Kaptein, R.; Vliegthart, J. F. G.; Gabius, H.-J. *J. Mol. Model.* **1997**, *3*, 325-331.
- [73] Siebert, H.-C.; von der Lieth, C.-W.; Kaptein, R.; Soedjanaamadja, U. M. S.; Vliegthart, J. F. G.; Wright, C. S.; Gabius, H.-J. *J. Mol. Modeling* **1996**, *2*, 351-353.
- [74] Siebert, H.-C.; Kaptein, R.; J., B. J.; Soedjanaamadja, U. M.; Wright, C. S.; Rice, A.; Kleineidam, R. G.; Kruse, S.; Schauer, R.; Pouwels, P. J. W.; Kamerling, J. P.; Gabius, H.-J.; Vliegthart, J. F. G. *Glycoconjugate J.* **1997**, *14*, 531-534.
- [75] Asensio, J. L.; Canada, F. J.; Bruix, M.; Rodríguezromero, A.; Jiménez-Barbero, J. *Eur. J. Biochem.* **1995**, *230*, 621-633.
- [76] Andersen, N. H.; Cao, B.; Rodríguez-Romero, A.; Arreguin, B. *Biochemistry* **1993**, *32*, 1407-1422.
- [77] Williams, B. A.; Chervenak, M. C.; Toone, E. J. *J. Biol. Chem.* **1992**, *267*, 22907-22911.
- [78] Gupta, D.; Brewer, C. F. *Thermodynamic Analysis of Protein-Carbohydrate Interaction*; Gabius, H.-J. and Gabius, S., Ed.; Chapman & Hall: London, **1997**, 333-343.
- [79] Garcia-Hernandez, E.; Zubillaga, R. A.; Rojo-Dominguez, A.; Rodríguez-Romero, A.; Hernandez-Arana, A. *Proteins* **1997**, *29*, 467-477.
- [80] Kronis, K. A.; Carver, J. P. *Biochemistry* **1982**, *21*, 3050-3057.
- [81] Toone, E. J. *Curr. Opin. Struct. Biol.* **1994**, *4*, 719-728.
- [82] Murray-Rust, P.; Glusker, J. P. *J. Am. Chem. Soc.* **1984**, *106*, 1018-1025.
- [83] Lemieux, R. U. *Chem. Soc. Rev.* **1989**, *18*, 347-374.
- [84] Lemieux, R. U. *Acc. Chem. Res.* **1996**, *29*, 373-380.
- [85] Lemieux, R. U.; Delbaere, L. T. J.; Beierbeck, H.; Spohr, U. *Ciba Foundation Symposium* **1991**, *158*, 231-248.
- [86] Carver, J. P.; Michnick, S. W.; Imberty, A.; Cumming, D. A. *Ciba Foundation Symposium* **1989**, *145*, 6-26.

- [87] Carver, J. P. *Pure and Applied Chemistry* **1993**, *65*, 763-770.
- [88] Searle, M. S.; Williams, D. H. *J. Am. Chem. Soc.* **1992**, *114*, 10690-10697.
- [89] Poornima, C. S.; Dean, P. M. *J. Comput.-Aided Mol. Design* **1995**, *9*, 500-512.
- [90] Blokzijl, W.; Engberts, J. B. F. N. *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 1545-1579.
- [91] Mancera, R. L. *J. Comput.-Aided Mol. Design* **1996**, *10*, 321-326.
- [92] Drickamer, K. *Nat. Struct. Biol.* **1995**, *2*, 437-439.
- [93] Mammen, M.; Seok-Ki, C.; Whitesides, M. *Angew. Chem.* **1998**, *110*, 2908-2953.
- [94] Matrosovich, M. N. *FEBS Letters* **1989**, *252*, 1-4.
- [95] Thomas, V. H.; Yang, Y.; Rice, K. G. *J. Biol. Chem.* **1999**, *274*, 19035-19040.
- [96] Mier, W.; von der Lieth, C.-W.; Hull, W. E.; Roth, R.; Kliem, H. C.; Menzler, S.; Wiessler, M. *Carbohydrate Res.* **2000**, to be submitted.
- [97] von der Lieth, C.-W.; Kozár, T.; Hull, W. E. *J. Mol. Struct. (Theochem.)* **1997**, *395-396*, 225-244.
- [98] Imberty, A. *Curr. Opin. Struct. Biol.* **1997**, *7*, 617-623.
- [99] Imberty, A.; Bourne, Y.; Cambillau, C.; Rouge, P.; Perez, S. *Adv. Biophys. Chem.* **1993**, *3*, 71-118.
- [100] Bush, C. A.; Martin-Pastor, M.; Imberty, A. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 269-293.
- [101] Woods, R. J. *Glycoconjugate Journal* **1998**, *15*, 209-216.
- [102] van Gunsteren, W. F.; King, P. M.; Mark, A. E. *Quarterly Reviews of Biophysics* **1994**, *4*, 435-481.
- [103] Imberty, A.; Von der Lieth, C.-W. *Carbohydrates in Europe* **1998**, *23*, 30-35.
- [104] Woods, R. J. *Curr. Opin. Struct. Biol.* **1995**, *5*, 591-598.
- [105] Woods, R. J. *The Application of Molecular Modeling Techniques to the Determination of Oligosaccharide Solution Conformations*; Boyd, K. B. L. a. D. B., Ed.; VCH Publishers, Inc.: New York, **1996**; Vol. 9, 129-165.
- [106] Pettersson, I.; Liljefors, T. *Molecular Mechanics Calculated Conformational Energies of Organic Molecules: A Comparison of Force Fields*; Boyd, K. B. L. a. D. B., Ed.; VCH Publishers, Inc.: New York, **1996**; Vol. 9, 167-189.
- [107] Verlinde, C. L. M. J.; Hol, W. G. J. *Structure* **1994**, *2*, 577-587.
- [108] Böhm, H. J. *J. Comput.-Aided Mol. Design* **1993**, *8*, 243-256.
- [109] Kollman, P. *Chem. Rev.* **1993**, *93*, 2395-2417.
- [110] Kollman, P. A. *Curr. Opin. Struct. Biol.* **1994**, *4*, 240-245.
- [111] Aqvist, J.; Medina, C.; Samuelsson, J. E. *Protein Engineering* **1994**, *7*, 385-391.
- [112] Warshel, A.; Tao, H.; Fothergill, M.; Chu, Z. T. *Israel Journal of Chemistry* **1994**, *34*, 253-256.
- [113] Klebe, G.; Abraham, U. *J. Med. Chem.* **1993**, *36*, 70-80.
- [114] Straatsma, T. P. *Free Energy by Molecular Simulation*; Boyd, K. B. L. a. D. B., Ed.; VCH Publishers, Inc.: New York, **1996**; Vol. 9, 81-127.
- [115] Wang, J.; Szewczuk, Z.; Yue, S. Y.; Tsuda, Y.; Konishi, Y.; Purisima, E. O. *J. Mol. Biol.* **1995**, *253*, 473-492.
- [116] Meirovitch, H. *Calculation of the Free Energy and the Entropy of Macromolecular Systems by Computer Simulation*; Boyd, K. B. L. a. D. B., Ed.; VCH Publishers, Inc.: New York, **1996**; Vol. 12, 1-74.
- [117] Beveridge, D. L.; Dicapua, F. M. *Ann. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431-492.
- [118] Rao, B. G.; Singh, U. C. *J. Am. Chem. Soc.* **1991**, *113*, 6735-6750.

- [119] Merz, K. M.; Kollman, P. A. *J. Am. Chem. Soc.* **1989**, *111*, 5649-5658.
- [120] Williams, D. H.; Cox, J. P. L.; Doig, A. J.; Gardner, M.; Gerhard, U.; Kaye, P. T.; Lal, A. R.; Nicholls, I. A.; Salter, C. J.; Mitchell, R. C. *J. Am. Chem. Soc.* **1991**, *113*, 7020-7030.
- [121] Williams, D. H.; Searle, M. S.; Mackay, J. P.; Gerhard, U.; Maplestone, R. A. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 1172-1178.
- [122] Searle, M. S.; Williams, D. H.; Gerhard, U. *J. Am. Chem. Soc.* **1992**, *114*, 10697-10704.
- [123] Hitzemann, R. *Trends in Pharmaceutical Sciences* **1988**, *9*, 408-411.
- [124] von der Lieth, C.-W.; Adler, M.; Kozár, T. *Detailed analysis of interaction patterns as a basis step towards a better semi-quantitative estimation of binding constants of protein-oligosacchride complexes.*; Trampisch, H. J. and Lange, S., Ed.; MMV Medizin Verlag: München, **1995**, 479-483.
- [125] Siebert, H.-C.; von der Lieth, C.-W.; Dong, X.; Reuter, G.; Gabius, H.-J.; Vliegthart, J. F. G. *Glycobiology* **1996**, *6*, 561-572.
- [126] Siebert, H.-C.; von der Lieth, C.-W.; Dong, X.; Reuter, G.; Schauer, R.; Gabius, H. J.; Vliegthart, J. F. G. *Glycobiology* **1996**, *6*, 561-572.
- [127] Boström, J.; P., N.; Liljefors, T. *J. Comput.-Aided Mol. Design* **1998**, *12*, 383-396.
- [128] Kozar, T.; von der Lieth, C.-W. *Glycoconjugate J.* **1997**, *14*, 925-933.
- [129] Leach, A. R. *Molecular Modelling - Principles and Applications*; Addison Wesley Longman Limited., **1996**.
- [130] Höltje, H.-D.; Folkers, G. *Molecular Modeling: basic principles and applications*; VCH: Weinheim, **1997**; Vol. 5.
- [131] *Discover User Guide (Versions 2.9.7 & 95.0/3.00)*; Oktober; BIOSYM/MSI: San Diego, **1995**.
- [132] Gelin, B.; Karplus, M. *J. Am. Chem. Soc.* **1975**, *97*, 6996-7006.
- [133] Ha, S. N.; Madsen, L. J.; Brady, J. W. *Biopolymers* **1988**, *27*, 1927-1952.
- [134] Tran, V.; Buleon, A.; Imberty, A.; Perez, S. *Biopolymers* **1989**, *28*, 679-690.
- [135] Kozar, T.; Petrak, F.; Galova, Z.; Tvaroska, I. *Carbohydr. Res.* **1990**, *204*, 27-36.
- [136] Perez, S. *Curr. Opin. Struct. Biol.* **1993**, *3*, 675-680.
- [137] Koca, J.; Perez, S.; Imberty, A. *J. Comp. Chem.* **1995**, *16*, 296-310.
- [138] Engelsens, S. B.; Koca, J.; Braccini, I.; Dupenhoat, C. H.; Perez, S. *Carbohydr. Res.* **1995**, *276*, 1-29.
- [139] Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087
- [140] Stuike-Prill, R.; Meyer, B. *Eur. J. Biochem.* **1990**, *194*, 903-913
- [141] Peters, T.; Meyer, B.; Stuike-Prill, R.; Somorjai, R.; J.-R., B. *Carbohydr. Res.* **1993**, *238*, 49-73
- [142] Weimar, T.; Peters, T.; Perez, S.; Imberty, A. *J. Mol. Struct. (Theochem)* **1997**, *395-396*, 297-311.
- [143] Guarnieri, F.; Still, W. C. *J. Comp. Chem.* **1994**, *15*, 1302-1310
- [144] Bernardi, A.; Raimondi, L.; Zanferrari, D. *J. Mol. Struct. (Theochem)* **1997**, *361-373*.
- [145] Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comp. Chem.* **1990**, *11*, 440-467
- [146] van Gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 992-1023.
- [147] Homans, S. W.; Forster, M. *Glycobiology* **1992**, *2*, 143-151.
- [148] Li, F.; Cui, W.; Allinger, N. L. *J. Comp. Chem.* **1994**, *15*, 769-781.
- [149] Rutherford, T. J.; Spackman, D. G.; Simpson, P. J. *Glycobiology* **1994**, *4*, 59-68.

- [150] Yan, Z.-Y.; Bush, C. A. *Biopolymers* **1990**, *29*, 799-811.
- [151] Mukhopadhyay, C.; Bush, C. A. *Biopolymers* **1994**, *34*, 11-20.
- [152] Ott, K.-H.; Meyer, B. *Carbohydrate Research* **1996**, *281*, 11-34.
- [153] Venable, R. M.; Bizik, F.; Henderson, T. J.; Egan, W. J. *Mol. Struct. (Theochem)* **1997**, 375-388.
- [154] Naidoo, K. J.; Brady, J. W. *J. Mol. Struct. (Theochem)* **1997**, 395-396, 469-475.
- [155] Asensio, J. L.; Cañada, F. J.; Bruix, M.; Gonzales, C.; Khair, N.; Rodríguez-Romero, A.; Jiménez-Barbero, J. *Glycobiology* **1998**, *8*, 569-577.
- [156] Qasba, P. K.; Balaji, P. V.; Rao, V. S. R. *J. Mol. Struct. (Theochem)* **1997**, 333-360.
- [157] Klewinghaus, P.; van Eijck, B. P.; Kouwijzer, M. L. C. E.; Kroon, J. J. *Mol. Struct. (Theochem)* **1997**, 395-396, 289-295.
- [158] Kawai, T.; Ichinose, T.; Endo, Y.; Shudo, K.; Itai, A. *J. Med. Chem.* **1992**, *35*, 2248-2253.
- [159] Sun, Y.; Kollman, P. *J. Phys. Chem.* **1996**, *100*, 6760-6763.
- [160] Karaman, R.; Almarsson, O.; Bruice, T. C. *J. Org. Chem.* **1992**, *57*, 1555-1559.
- [161] Frank, M.; Peraus, K.; Staab, H. A. *J. Mol. Modeling* **1996**, *2*, 383-385.
- [162] Karplus, M.; Petsko, G. A. *Nature* **1990**, *347*, 631-639.
- [163] Brooks, B. R. *Chemica Scripta* **1989**, *29A*, 165-169.
- [164] Kollman, P. A.; Grootenhuys, P. D. J.; Lopez, M. A. *Pure and Applied Chemistry* **1989**, *61*, 593-596.
- [165] Levitt, M.; Sharon, R. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 7557-7561.
- [166] Lopez, M. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1989**, *111*, 6212-6222.
- [167] Mierke, D. F.; Kessler, H. *Biopolymers* **1993**, *33*, 1003-1017.
- [168] Siebert, H.-C.; Gilleron, M.; Kaltner, H.; von der Lieth, C.-W.; Kozar, T.; Bovin, N.; Korchagina, E. Y.; Vliegert-hart, J. F. G.; Gabius, H. J. *Biochem. Biophys. Res. Comm.* **1996**, *219*, 205-212.
- [169] Sun, Y. X.; Kollman, P. A. *J. Comp. Chem.* **1992**, *13*, 33-40.
- [170] Nilges, M.; Gronenborn, A. M.; Bruenger, A. T.; Clore, G. M. *Protein Engineering* **1988**, 27-38.
- [171] Homans, S. W.; Forster, M. *Glycobiology* **1992**, *2*, 143-151.
- [172] Scheek, R. M.; van Gunsteren, W. F.; Kaptein, R. *Methods Enzymol.* **1989**, *177*, 204-218.
- [173] Dauber-Osguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M.; Hagler, A. T. *Proteins* **1988**, *4*, 31-47.
- [174] Hagler, A. T.; Lifson, S.; Dauber, P. *J. Am. Chem. Soc.* **1979**, *101*, 5122-5130.
- [175] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
- [176] Engelsens, S. B.; Penhoat, C. H.; Perez, S. *J. Phys. Chem.* **1995**, *99*, 13334-13351.
- [177] Poveda, A.; Asensio, J. L.; Martinpastor, M.; Jiménez-Barbero, J. *Chem. Comm.* **1996**, 421-422.
- [178] Brady, J. W.; Schmidt, R. K. *J. Phys. Chem.* **1993**, *97*, 958-966.
- [179] Takusagawa, F.; Jacobson, G. A. *Acta Crystallogr.* **1978**, *34*, 213-218.
- [180] Mikami, B.; Adachi, M.; Kage, T.; Sarikaya, E.; Nanmori, T.; Shinke, R.; Utsumi, S. *Biochemistry* **1999**, *38*, 7050-7061.
- [181] Stevens, E. S. *Biopolymers* **1992**, *32*, 1571-1579.

- [182] Schmidt, R. K.; Brady, J. W. *J. Phys. Chem.* **1995**, *99*, 11339-11343.
- [183] French, A. D.; Dowd, M. K. *J. Mol. Struct. (Theochem)* **1993**, *286*, 183-201.
- [184] Hagler, A. T.; Huler, E.; Lifson, S. *J. Am. Chem. Soc.* **1974**, *96*, 5319-5327.
- [185] Asensio, J. L.; Martinpastor, M.; Jiménez-Barbero, J. *Int. J. Biol. Macromol.* **1995**, *17*, 137-148.
- [186] Martinpastor, M.; Espinosa, J. E.; Asensio, J. L.; Jiménez-Barbero, J. *Carbohydr. Res.* **1997**, *298*, 15-49.
- [187] Chou, S. S. C.; Jeffrey, G. A. *Acta Crystallogr.* **1967**, *23*, 1038-1040.
- [188] Shashkov, A. S.; Lipkind, G. M.; Kachetkov, N. K. *Carbohydr. Res.* **1986**, *147*, 175-182.
- [189] Rees, D. A.; Thom, D. *J. Chem. Soc. Perkin Trans. II* **1977**, 191-201.
- [190] Reutter, W.; Stäsche, R.; Stehling, P.; Baum, O. *The Biology of Sialic Acids: Insights into their Structure, Metabolism and Function in particular during Viral Infection*; Gabius, H.-J. and Gabius, S., Ed.; Chapman & Hall: London, **1997**, 245-259.
- [191] Schauer, R. *Adv. Carbohydr. Chem. Biochem.* **1982**, *40*, 131-234.
- [192] Gilleron, M.; Siebert, H. C.; Kaltner, H.; von der Lieth, C.-W.; Kozar, T.; Halkes, K. M.; Korchagina, E. Y.; Bovin, N. V.; Gabius, H. J.; Vliegthart, F. G. *Eur. J. Biochem.* **1998**, *252*, 416-427.
- [193] Siebert, H.-C.; Reuter, G.; Schauer, R.; von der Lieth, C.-W.; Dabrowski, J. *Biochemistry* **1992**, *31*, 6962-6971.
- [194] Poppe, L.; Brown, G. S.; Philo, J. S.; Pandurang, V. N.; Shah, B. H. *J. Am. Chem. Soc.* **1997**, *119*, 1727-1736.
- [195] Perez, S.; Mouhous-Riou, N.; Nifantev, N.; Tsvetkov, Y. E.; Bachert, B.; Imberty, A. *Glycobiology* **1996**, *6*, 537-542.
- [196] Yvelin, F.; Zhang, Y.-M.; Mallet, J.-M.; Robert, F.; Jeannin, Y.; Sinay, P. *Carbohydrate Letters* **1996**, *1*, 475-482.
- [197] Scheffler, K.; Ernst, B.; Katopodis, A.; Magnani, J. L.; Wang, W. T.; Weisemann, R.; Peters, T. *Angew. Chem.* **1995**, *107*, 2034-2037.
- [198] Gronenborn, A. M.; Clore, G. M. *Prog. Nucl. Magn. Reson. Spectrosc.* **1985**, *17*, 1-32.
- [199] Acquotti, D.; Poppe, L.; Dabrowski, J.; von der Lieth, C.-W.; Sonnino, S.; Tettamanti, G. *J. Am. Chem. Soc.* **1990**, *112*, 7772-7778.
- [200] Lommerse, J. P. M.; Kroon-Batenburg, J. M. J.; Kroon, J.; Kamerling, J. P.; Vliegthart, J. V. G. *J. Biomol. NMR* **1995**, *5*, 79-94.
- [201] Kroon-Batenburg, L. M. J.; Kroon, J. *Biopolymers* **1990**, *29*, 1243-1248.
- [202] Naidoo, K. J.; Brady, J. W. *J. Am. Chem. Soc.* **1998**, *121*, 2244-2252.
- [203] Zschunke, A. *Molekülstruktur: Form, Dynamik, Funktion*; Spektrum, Akad. Verl.: Heidelberg, Berlin, Oxford, **1993**.
- [204] Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671-680.
- [205] Hassel, O.; Ottar, B. *Acta Chem. Scand.* **1947**, *1*, 929-942.
- [206] Wyn-Jones, E.; Pethrick, R. A. *Top. Stereochem.* **1969**, *5*.
- [207] Lowe, J. P. *Prog. Phys. Org. Chem.* **1968**, *6*.
- [208] Carey, F. A. *Advanced Organic Chemistry: Part A*; 2nd edition; Plenum Press: New York, **1984**.
- [209] Friebolin, H.; Schmid, H. G.; Kabuß, S.; Faißt, W. *Org. Magn. Reson.* **1969**, *1*, 147-162.
- [210] Lui, Q.; Brady, J. W. *J. Phys. Chem.* **1997**, *101*, 1317-1321.
- [211] Poppe, L.; Dabrowski, J.; von der Lieth, C.-W.; Koike, K.; Ogawa, T. *Eur. J. Biochem.* **1990**, *189*, 313-325.

- [212] Nishida, Y.; Ohruai, H.; Meguro, H. *Tetrahedron Letters* **1984**, *25*, 1575-1578.
- [213] Siebert, H.-C.; André, S.; Dong, X.; Espinosa, J.-F.; Frank, M.; Gilleron, M.; Kaltner, H.; Kozár, T.; Beintema, J. J.; Bovin, N. V.; von der Lieth, C.-W.; Vliegthart, J. F. G.; Jiménez-Barbero, J.; Gabius, H.-J. *to be submitted*.
- [214] Ohruai, H.; Nishida, Y.; Higuchi, H.; Hori, H.; Meguro, H. *Can. J. Chem.* **1987**, *65*, 1145
- [215] Nishida, Y.; Hori, H.; Ohruai, H.; Meguro, H. *J. Carbohydr. Chem.* **1988**, *7*, 239
- [216] Marchessault, R. H.; Perez, S. *Biopolymers* **1979**, *18*, 2369-2374
- [217] Brady, J. W. *J. Am. Chem. Soc.* **1986**, *108*, 8153-8160.
- [218] Brady, J. W. *Carbohydr. Res.* **1987**, *165*, 306-312.
- [219] van Eijck, B. P.; Hooft, R. W. W.; Kroon, J. J. *Phys. Chem.* **1993**, *97*, 12093-12099.
- [220] Tvaroska, I.; Kozar, T. *Theor. Chem. Acta* **1986**, *70*, 99-114.
- [221] Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1993**, *115*, 5745-5753.
- [222] Spieser, S. A. H.; van Kuil, J. A.; Kroon-Batenburg, L. M. J.; Kroon, J. *Carbohydr. Res.* **1999**, submitted.
- [223] van Gunsteren, W. F.; Berendsen, H. J. C. *Groningen Molecular Simulation Package GROMOS*; University of Groningen: Netherlands, **1987**.
- [224] Maple, J. R.; Hwang, M.-J.; Stockfisch, T. P.; Dinur, U.; Waldman, M.; Ewig, C. S.; Hagler, A. T. *J. Comput. Chem.* **1994**, *15*, 162-182.
- [225] Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. J. *J. Am. Chem. Soc.* **1984**, *106*, 765-776.
- [226] Homans, S. W. *Biochemistry* **1990**, *29*, 9110-9118.
- [227] Chin, J. T.; Wheeler, S. L.; Klibanov, A. M. *Biotechnol. Bioeng.* **1994**, *44*, 140-145.
- [228] Klibanov, A. M. *Acc. Chem. Res.* **1990**, *23*, 114-120.
- [229] Griebenow, K.; Klibanov, A. M. *J. Am. Chem. Soc.* **1996**, *118*, 11695-11700.
- [230] Rodriguez, A.; Tablero, M.; Barragan, B.; Lara, P.; Rangel, M.; Arreguin, B.; Possani, L.; Soriano-Garcia, M. J. *Crystallogr. Growth* **1986**, *76*, 710-714.
- [231] Gidrol, X.; Chrestin, H.; Tan, H.-L.; Kush, A. J. *Biol. Chem.* **1994**, *209*, 9278-9283.
- [232] Kronis, K. A.; Carver, P. *Biochemistry* **1985**, *24*, 826-833.
- [233] Garcia-Hernandez, E.; Hernandez-Arana, A. *Protein Sci.* **1999**, *8*, 1075-1086.
- [234] Chervenak, M. C.; Toone, E. J. *Biophys. J.* **1994**, *66*, A359-A359.
- [235] Beierbeck, H.; Delbaere, L. T. J.; Vandonselaar, M.; Lemieux, R. U. *Can. J. Chem.* **1994**, *72*, 463-470.
- [236] Vaisman, I. I.; Berkowitz, M. L. *J. Am. Chem. Soc.* **1992**, *114*, 7889-7896.
- [237] Vishnyakov, A.; Widmalm, G.; Laaksonen, A. *Angew. Chem.* **2000**, *112*, 144-146.

