

**Entwicklung eines rechnerischen Verfahrens zur
Simulation der thermischen Entfaltung von Proteinen
und zur Untersuchung ihrer Thermostabilität**

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich Biowissenschaften
der Goethe-Universität
Frankfurt am Main

von
Sebastian Radestock
aus Rottweil

Frankfurt am Main, 2010

(D 30)

Vom Fachbereich Biowissenschaften der Goethe-Universität Frankfurt am Main als
Dissertation angenommen.

Dekanin: Prof. Dr. Starzinski-Powitz

Erster Gutachter: Prof. Dr. Holger Gohlke

Zweiter Gutachter: Prof. Dr. Peter Güntert

Datum der Disputation: 22.11.2010

Die vorliegende Arbeit wurde unter Anleitung von Herrn Prof. Dr. H. Gohlke in der Arbeitsgruppe Molekulare Bioinformatik am Institut für Zellbiologie und Neurowissenschaften der Goethe-Universität, Frankfurt am Main, später in der Arbeitsgruppe Computerunterstützte Pharmazeutische Chemie am Institut für Pharmazeutische Chemie der Christian-Albrechts-Universität zu Kiel und zuletzt in der Arbeitsgruppe Computerunterstützte Pharmazeutische Chemie am Institut für Pharmazeutische Chemie der Heinrich-Heine-Universität, Düsseldorf, in der Zeit von Oktober 2006 bis September 2009 durchgeführt.

Mein Dank gilt allen, die mir die Anfertigung dieser Arbeit ermöglicht und mich bei deren Fertigstellung unterstützt haben. Herrn Prof. Dr. Holger Gohlke sei gedankt für die Überlassung des interessanten Themas, für seine Betreuung und für die Zeit in seiner Arbeitsgruppe. Doris Klein, Christopher Pflieger, Hannes Kopitz und insbesondere Christine Radestock möchte ich herzlich für das Korrekturlesen und für konstruktive Kritik danken. Dr. Domingo Gonzalez-Ruiz sei gedankt für seine ständige Diskussionsbereitschaft und die Impulse, die er mir während der gemeinsamen Zeit in Frankfurt am Main gegeben hat.

Ein Dankeschön geht an Dr. Stefan Häfner, Dr. Wolfgang Höffken und Thomas Mietzner (BASF SE, Ludwigshafen) für das Interesse an meiner Arbeit und die Bereitstellung diverser Daten. Auch möchte ich der Goethe-Universität, Frankfurt am Main, und der Christian-Albrechts-Universität, Kiel, für die Gewährung von Promotionsstipendien danken, ohne die die Fertigstellung der Arbeit unmöglich gewesen wäre. Ich möchte diese Arbeit meiner Frau Ronak, meinen Geschwistern und meinen Eltern widmen, denen ich sehr dankbar bin für ihre Geduld, Unterstützung und Ermutigung.

*Le seul véritable voyage, le seul bain de jouvence,
ce ne serait pas d'aller vers de nouveaux paysages,
mais d'avoir d'autres yeux...*

Die einzige wahre Reise, der einzige Jungbrunnen
wäre für uns, wenn wir nicht neue Landschaften
aufsuchten, sondern andere Augen hätten...

Marcel Proust,

Auf der Suche nach der verlorenen Zeit

Inhaltsverzeichnis

| | |
|--|-----|
| Abkürzungsverzeichnis..... | III |
| 1 Einleitung..... | 1 |
| 1.1 Einführung in die Problemstellung..... | 1 |
| 1.2 Proteinstabilität und thermische Entfaltung..... | 4 |
| 1.2.1 Makroskopische und mikroskopische Proteinstabilität | 4 |
| 1.2.2 Stabilisierende Wechselwirkungen und Kooperativität..... | 6 |
| 1.2.3 Die thermische Entfaltung | 8 |
| 1.2.4 Reversible vs. irreversible Denaturierung..... | 9 |
| 1.3 Thermostabilität und thermophile Anpassung..... | 10 |
| 1.3.1 Industrielle Bedeutung thermostabiler Proteine | 10 |
| 1.3.2 Natürlich vorkommende thermostabile Proteine | 11 |
| 1.3.3 Die Entwicklung künstlicher thermostabiler Proteine | 15 |
| 1.4 Rechnerische Ansätze zur Vorhersage der Proteinstabilität..... | 16 |
| 1.4.1 Vorhersage der Gesamtstabilität..... | 16 |
| 1.4.2 Identifizierung für die Stabilität wichtiger Regionen | 18 |
| 1.4.3 Simulation der thermischen Denaturierung | 19 |
| 1.4.4 Vorhersage der mikroskopischen Stabilität | 21 |
| 2 Rigidität und Flexibilität in statischen Netzwerken..... | 23 |
| 2.1 Rigiditätstheorie..... | 24 |
| 2.1.1 Das Maxwell'sche <i>constraints counting</i> | 24 |
| 2.1.2 Der <i>pebble game</i> -Algorithmus im 2D-Raum..... | 25 |
| 2.1.3 Rigidität in Netzwerken im 3D-Raum | 26 |
| 2.2 Rigiditätsp Perkolation..... | 27 |
| 2.3 Anwendung auf Proteine | 28 |
| 2.3.1 Modellierung der einzelnen Interaktionen..... | 30 |
| 2.3.2 Simulierte thermische Entfaltung von Proteinen | 33 |
| 3 Aufgabenstellung | 34 |
| 4 Theorie und Methoden..... | 36 |
| 4.1 Konstruktion der Netzwerke..... | 36 |
| 4.2 Flexibilitätsanalyse | 38 |
| 4.3 Simulation der thermischen Entfaltung | 38 |
| 4.4 Analyse der Entfaltungssimulationen | 40 |

| | | |
|-------|--|-----|
| 4.4.1 | Rigiditätsordnungsparameter | 41 |
| 4.4.2 | Perkolationsindex..... | 42 |
| 4.4.3 | Clusterkonfigurationsentropie | 42 |
| 4.4.4 | Entfaltungsregionen | 45 |
| 4.4.5 | Anpassung der Temperaturskala..... | 45 |
| 4.5 | Ensemblebasierte Analyse statischer Netzwerke..... | 46 |
| 4.6 | Erzeugung von Strukturensamples | 47 |
| 4.7 | Datensätze | 49 |
| 4.7.1 | Homologe meso- und thermophile Proteine | 49 |
| 4.7.2 | Phytasen | 52 |
| 4.8 | Computerprogramme und programmiertechnische Hilfsmittel..... | 54 |
| 5 | Ergebnisse und Diskussion | 56 |
| 5.1 | Analyse statischer Netzwerke thermophiler Proteine..... | 56 |
| 5.1.1 | Datensatz homologer meso- und thermophiler Proteine..... | 58 |
| 5.1.2 | Thermostabilitäten homologer meso- und thermophiler Proteine | 62 |
| 5.1.3 | Molekulare Ursachen erhöhter Thermostabilität | 71 |
| 5.1.4 | Konserviertheit der mikroskopischen Stabilität..... | 83 |
| 5.1.5 | Mikroskopische Stabilität und Bioaktivität | 89 |
| 5.2 | Analyse statischer Netzwerke von Phytasen | 96 |
| 5.2.1 | Datensatz..... | 98 |
| 5.2.2 | Erzeugung konformationeller Ensembles..... | 101 |
| 5.2.3 | Thermostabilitäten von Phytasen..... | 103 |
| 5.2.4 | Molekulare Ursachen der Thermostabilität | 106 |
| 5.2.5 | Vorhersage und Testung thermostabilisierender Mutationen..... | 112 |
| 5.3 | Abschließende Diskussion und Ausblick | 116 |
| 6 | Zusammenfassung | 120 |
| 7 | Anhang..... | 123 |
| 7.1 | Die Gibbs-Helmholtz-Gleichung..... | 123 |
| 7.2 | Weitere Parameter zur Analyse der Entfaltungssimulationen | 123 |
| 8 | Literatur | 124 |

Abkürzungsverzeichnis

Aminosäuren

| | | |
|--------------|--------------------|---|
| Alanin | Ala | A |
| Arginin | Arg | R |
| Asparagin | Asn | N |
| Aspartat | Asp | D |
| Cystein | Cys, Cyx | C |
| Glutamat | Glu | E |
| Glutamin | Gln | Q |
| Glycin | Gly | G |
| Histidin | His, Hie, Hid, Hip | H |
| Isoleucin | Ile | I |
| Leucin | Leu | L |
| Lysin | Lys | K |
| Methionin | Met | M |
| Phenylalanin | Phe | F |
| Prolin | Pro | P |
| Serin | Ser | S |
| Threonin | Thr | T |
| Tryptophan | Trp | W |
| Tyrosin | Tyr | Y |

Organismen

| | |
|-------------------------------|-------------------------------------|
| <i>A. ficum</i> | <i>Aspergillus ficum</i> |
| <i>A. fumigatus</i> | <i>Aspergillus fumigatus</i> |
| <i>B. cereus</i> | <i>Bacillus cereus</i> |
| <i>B. stearothermophilus</i> | <i>Bacillus stearothermophilus</i> |
| <i>B. subtilis</i> | <i>Bacillus subtilis</i> |
| <i>B. thermoproteolyticus</i> | <i>Bacillus thermoproteolyticus</i> |
| <i>E. coli, E.c.</i> | <i>Escherichia coli</i> |
| <i>T. brucei</i> | <i>Trypanosoma brucei</i> |
| <i>T. fusca</i> | <i>Thermomonospora fusca</i> |

| | |
|------------------------------------|-----------------------------------|
| <i>T. thermophilus</i> | <i>Thermus thermophilus</i> |
| <i>T. vulgaris</i> | <i>Thermoactinomyces vulgaris</i> |
| <i>Y. mollaretii</i> , <i>Y.m.</i> | <i>Yersinia mollaretii</i> |

Sonstige Abkürzungen

| | |
|------------------------|--|
| $\langle r \rangle$ | mittlere Koordinationszahl |
| $\langle r \rangle_c$ | kritische mittlere Koordinationszahl |
| 2D | zweidimensional |
| 3D | dreidimensional |
| $\Delta\Delta G$ | Stabilitätsdifferenz |
| ΔC_p | Wärmekapazitätsdifferenz |
| ΔG | freie Enthalpiedifferenz, Stabilität |
| ΔH | Enthalpiedifferenz |
| ΔH_m | Enthalpiedifferenz bei der Schmelztemperatur |
| ΔS | Entropiedifferenz |
| ΔT_{50} | Halbwertszeittemperaturdifferenz |
| ΔT_m | Schmelztemperaturdifferenz |
| ΔT_{og} | Differenz der optimalen Wachstumstemperaturen der Organismen, aus denen zwei Proteine stammen |
| ΔT_p | Phasenübergangstemperaturdifferenz |
| γ, θ, ϕ | Winkel zwischen Bindungen |
| a, c, d, r | Abstände zwischen Atomen oder Atomradien |
| A | Protonenakzeptor |
| ADP | Adenosindiphosphat |
| b | Anzahl von Beschränkungen |
| $C\alpha$ | Kohlenstoff im Rückgrad einer Aminosäure |
| $C\beta$ | Erster Kohlenstoff in der Seitenkette einer Aminosäure |
| c_{cut} | Distanzschwellenwert |
| CNA | <i>constraint network analysis</i> , Analyse statischer Netzwerke |
| CP | Consensus-Phytase |
| C_p | Wärmekapazität |
| C-Terminus | Carboxyterminus |
| D | denaturierter Zustand; Protonendonor |
| DHFR | Dihydrofolatreduktase |

| | |
|--------------------------|---|
| E | Energie |
| E_{CUT} | Energieschwellenwert |
| E_{HB} | H-Brückenenergie |
| E_{SB} | Salzbrückenenergie |
| EU | Europäische Union |
| F | finaler Zustand |
| F | Anzahl Nullenergiemoden (<i>floppy modes</i>) |
| FEP | <i>free energy perturbation</i> , freie Enthalpiestörungsrechnung |
| FIRST | <i>floppy inclusion of rigid substructure topology</i> |
| GAPDH | Glyceraldehyd-3-Phosphatdehydrogenase |
| GB | generalisiertes Born Modell |
| GB^{OBC} | GB nach Onufried, Bashford und Case |
| G_{D} | freie Enthalpie des denaturierten Zustands |
| GHz | Gigahertz |
| G_{N} | freie Enthalpie des nativen Zustands |
| GNM | elastisches Gauß'sches Netzwerkmodell |
| IPM | 3-Isopropylmalat |
| IPMDH | 3-Isopropylmalatdehydrogenase |
| k_{cat} | spezifische Aktivität |
| K_{m} | Michaelis-Menten-Konstante |
| M1, M2 | Mutanten |
| MC | Monte Carlo |
| MD | Molekulardynamik |
| MM | Molekularmechnik |
| N | nativer Zustand |
| N | Anzahl von Atomen |
| n | Anzahl von Atomen |
| NAD^+ | Nikotinamid-Adenindinucleotid |
| N_{c} | Anzahl von Beschränkungen |
| NMA | Normalmodenanalyse |
| NMR | Kernmagnetresonanz |
| N_{r} | Anzahl redundanter Beschränkungen |
| Nr. | Nummer |
| n_{s} | normalisierte Anzahl von Clustern der Größe s |

| | |
|--------------|---|
| N-Terminus | Aminoterminus |
| P_{∞} | Rigiditätsordnungsparameter |
| PA | Pseudoatom |
| PCA | <i>principal component analysis</i> , Hauptkomponentenanalyse |
| PDB | Proteindatenbank |
| PGTdb | <i>prokaryotic growth temperature database</i> |
| PME | Particle-Mesh-Ewald |
| RCD | <i>rigid cluster decomposition</i> , Zerlegung eines Netzwerks in rigide Cluster und flexible Regionen |
| R_{HB} | H-Brücken-Referenzabstand |
| r_i | kartesische Koordination von Atom i |
| $RMSD$ | <i>root mean square deviation</i> , mittlere quadratische Abweichung |
| R_{SB} | Salzbrücken-Referenzabstand |
| s | Größe eines rigiden Clusters |
| S | Clusterkonfigurationsentropie |
| SRP | Signalerkennungspartikel |
| T | Temperatur |
| T_{50} | Habwertszeittemperatur |
| TI | thermodynamische Integrationsrechnung |
| TLP | thermolysinähnliche Protease |
| T_m | Schmelztemperatur |
| T_{og} | optimale Wachstumstemperatur des Organismus, aus dem ein Protein stammt |
| T_{opt} | Temperatur der optimalen Aktivität eines Enzyms |
| T_p | Phasenübergangstemperatur |
| T_r | Referenztemperatur |
| UE | Untereinheit |
| V_{HB} | H-Brücken-Referenzenergie |
| V_{SB} | Salzbrücken-Referenzenergie |
| w | Wahrscheinlichkeit |
| w_{ER} | Entfaltungsregionwahrscheinlichkeit |
| WT | Wildtyp |

1 Einleitung

1.1 Einführung in die Problemstellung

Proteine sind die Bausteine des Lebens mit einer Vielzahl biochemischer Eigenschaften. Das macht sie für die Biotechnologie interessant, wo sie (teilweise unter Extrembedingungen) als Biokatalysatoren in industriellen Prozessen, als Futtermittelzusatzstoffe oder als pharmazeutische Wirkstoffe eingesetzt werden (Braun *et al.* 2006). Proteine sind lineare Heteropolymere, deren Monomereinheiten 20 verschiedene Aminosäuren sind. Ihre biochemischen Eigenschaften hängen stark von ihrem Zustand, insbesondere ihren lokalen und globalen Faltungseigenschaften ab, die empfindlich auf Änderungen der Temperatur, des Drucks oder der Konzentration bestimmter Chemikalien reagieren. Im nativen Zustand werden Proteine durch eine Vielzahl nichtkovalenter Wechselwirkungen gegenüber ihrer Umgebung stabilisiert und behalten ihre enzymatische Aktivität. Ändert sich die Umgebung des Proteins, beispielsweise durch Temperaturerhöhung, dann kann das Protein denaturieren und global zerfallen, d.h. seine natürliche Struktur verlieren. Im denaturierten Zustand ist die Aminosäurekette des Proteins entfaltet, was in der Regel zum Verlust der Bioaktivität eines Enzyms führt. In der Praxis ist es wichtig, das Temperaturintervall zu kennen, in dem ein biotechnologisch eingesetztes Enzym stabil und aktiv ist. Den Anforderungen industrieller Prozesse genügen oft nur Thermozyeme, d.h. Proteine mit einer hohen thermischen Stabilität.

Zur Aufklärung der molekularen Grundlagen der Thermostabilität wurden Proteine aus mesophilen Organismen (optimale Wachstumstemperatur von 15–45 °C) mit natürlich vorkommenden thermostabilen Proteinen aus thermophilen oder hyperthermophilen Organismen (optimale Wachstumstemperatur von 50–80 bzw. mehr als 80 °C) verglichen. Diese Studien haben entgegen früherer Erwartungen keine deutlichen systematischen Unterschiede zwischen meso- und thermostabilen Proteinen erkennen lassen. Vielmehr erfolgt die Anpassung thermophiler Proteine durch gezielte Veränderungen der nichtkovalenten Wechselwirkungen, die sowohl die elektrostatischen Kräfte (H- und Salzbrücken), als auch die Packung der Aminosäureseitenketten und die Stärke des hydrophoben Effekts betreffen. Experimentelle Untersuchungen haben gezeigt, dass diese Veränderungen im Netzwerk nichtkovalenter Wechselwirkungen zu einer Verringerung der Flexibilität und somit zu einer Rigidisierung thermophiler Proteine führt (Vielle und Zeikus 2001; Sterner und Brunner 2008). Wenn die Untersuchung homologer meso- und thermophiler Proteine aber bei ihrer jeweiligen Arbeitstemperatur erfolgt, dann

befinden sie sich in korrespondierenden Zuständen ähnlicher Bioaktivität, was trotz globaler Rigidisierung auf eine Konservierung wichtiger flexibler Bereiche schließen lässt (Somero 1978; Jaenicke und Böhm 1998; Sterner und Brunner 2008).

Um den Bedarf maßgeschneiderter Thermozyyme für die Biotechnologie zu decken, müssen natürliche Proteine mit Hilfe molekularbiologischer Methoden verändert und an die gewünschten Reaktionen und Reaktionsbedingungen angepasst werden (Atomi 2005). Dies kann mit Hilfe des „Protein Engineering“ erfolgen, bei dem die Strukturen durch Aminosäureaustausche (Mutationen) verändert werden. Aufgrund der großen Anzahl theoretisch möglicher Mutationen und der damit verbundenen Menge an Experimenten, die durchgeführt werden müssen, spielen rechnerische Methoden zur Unterstützung des Protein Engineering eine immer größer werdende Rolle. Computeransätze müssen einerseits die thermische Stabilität einer Struktur sowie den Effekt einer Mutation abschätzen können, andererseits Hinweise darauf geben, in welchen Bereichen der Struktur thermostabilisierende Mutationen eingeführt werden können. Die Methoden sollten außerdem in der Lage sein, die Auswirkung einer thermostabilisierenden Mutation auf die Aktivität vorherzusagen. Dies kann gelingen, indem untersucht wird, ob die Mutation die Struktur des aktiven Zentrums oder dessen Flexibilität stört (Korkegian *et al.* 2005). Als rechnerische Methoden kommen Ansätze in Frage, mit denen die freie Stabilisierungsenthalpie von Proteinen vorhergesagt oder die Denaturierung des Proteins simuliert werden kann. Von Interesse sind außerdem Ansätze, mit denen die Flexibilität der Struktur untersucht werden kann.

So kann mit Hilfe von Molekulardynamik(MD)-Simulationen die Bewegung einzelner Atome oder bestimmter Bereiche einer Proteinstruktur untersucht und daraus auf ihre Flexibilität geschlossen werden. Diese Simulationen können bei verschiedenen Temperaturen oder im thermodynamischen Ungleichgewicht durchgeführt werden, um strukturelle Veränderungen aufzuzeichnen, die während der thermischen Denaturierung des Proteins auftreten und letztendlich zum globalen Zerfall der Struktur führen (van Gunsteren *et al.* 1995; Finkelstein 1997). Mit Hilfe solcher Methoden wurde beispielsweise die Entfaltung einer Cutinase untersucht (Creveld *et al.* 1998). Über den Vergleich von Simulationen unter verschiedenen Bedingungen konnte zwischen flexiblen Regionen, die für die Aktivität des Enzyms wichtig waren, und flexiblen Regionen, von denen die Entfaltung ausging, unterschieden werden. So konnten Schwachstellen (*weak spots*) vorhergesagt werden, deren Mutation die thermische Stabilität, nicht aber die Aktivität beeinflussen dürfte. Der Nachteil solcher Simulationen besteht darin, dass die-

se Rechnungen sehr viel Zeit kosten. Zudem ist unklar, ob die verwendeten Kraftfelder und insbesondere Lösungsmittelmodelle geeignet sind, das Verhalten des Systems bei hohen Temperaturen korrekt wiederzugeben.

Eine Alternative zu Dynamiksimulationen stellen Methoden dar, wie beispielsweise die Normalmodenanalyse (NMA), mit deren Hilfe relevante Bewegungsmoden in der Proteinstruktur ausgehend von einer einzelnen Konformation vorhergesagt werden können. Su *et al.* haben einen solchen Ansatz dazu verwendet, Veränderungen in der Fluktuation einzelner Reste während der thermischen Entfaltung eines Proteins zu untersuchen (Su *et al.* 2008). Die mit Hilfe von MD-Simulationen oder NMA-ähnlichen Verfahren gemessene Dynamik (Bewegung) ist immer nur ein *indirektes* Maß für die tatsächlich vorhandene Flexibilität (Beweglichkeit). Eine *direkte* Bestimmung rigider und flexibler Regionen einer Proteinstruktur ist von großem Interesse, da mit Hilfe dieser Eigenschaften eine Aussage über die Stabilität der jeweiligen Region oder gesamten Struktur getroffen werden kann, da sie beschreiben, ob ein bestimmter Strukturbereich deformiert werden kann und dann zur Destabilisierung des Proteins beiträgt.

In dieser Arbeit wird ein rechnerisches Verfahren vorgestellt, mit dem die Simulation der thermischen Entfaltung von Proteinen und die Untersuchung ihrer Thermostabilität möglich ist. Das Verfahren beruht auf der Modellierung der Strukturen als eine bestimmte Art von *bar-and-joint*-Netzwerke, in denen die Atome durch Knoten und die Wechselwirkungen zwischen den Atomen durch Beschränkungen beschrieben werden. Neben kovalenten und nichtkovalenten Bindungsbeschränkungen werden auch Winkelbeschränkungen in das Netzwerk eingefügt. Diese *bond-bending*- bzw. äquivalente *bond-and-bar*-Netzwerke können hinsichtlich ihrer Flexibilität analysiert werden, d.h. die direkte Identifizierung rigider Cluster und flexibler Bereiche in atomarer Auflösung ist möglich. Dabei kommen Methoden aus der Rigiditätstheorie zum Einsatz. Die thermische Entfaltung des Proteins wird durch sukzessives Aufbrechen nichtkovalenter Wechselwirkungen und anschließender erneuter Flexibilitätsanalyse simuliert. Weitergehende Analysen ermöglichen die Beschreibung des globalen Zerfalls, die Quantifizierung der Thermostabilität des Proteins, die Identifizierung von Bereichen, von denen der globale Zerfall ausgeht, und somit die Vorhersage von Entfaltungsregionen, in die thermostabilisierende Mutationen eingeführt werden können. Darüber hinaus ist auch die Untersuchung der für die Bioaktivität wichtigen lokalen Flexibilität möglich.

Diese unter dem Begriff der Analyse statischer Netzwerke (*constraint network analysis*, CNA) zusammengefassten Methoden werden in der vorliegenden Arbeit auf einen

Datensatz homologer meso- und thermophiler Proteine sowie auf eine Serie von Phyta-
sen angewendet. Dabei wird gezeigt, dass die thermophile Anpassung tatsächlich über
eine Rigidisierung der Struktur erfolgt. Zudem wird gezeigt, dass trotz globaler Rigidi-
sierung für die Bioaktivität wichtige flexible Bereiche der Struktur konserviert sind. Es
wird gezeigt, dass die CNA hilfreich zur Unterstützung des Protein Engineering ist,
indem sie einerseits die thermische Stabilität einer Struktur sowie den Effekt einer Mu-
tation abschätzen kann, andererseits Hinweise darauf gibt, in welchen Bereichen der
Struktur thermostabilisierende Mutationen eingeführt werden können.

1.2 Proteinstabilität und thermische Entfaltung

1.2.1 Makroskopische und mikroskopische Proteinstabilität

Proteine sind in ihrem nativen (natürlichen) Zustand (N) in der Regel dicht gepackt.
Ihre Aminosäurenkette ist stark gefaltet. In diesem Zustand sind sie stabil gegenüber
ihrer Umgebung und behalten deswegen ihre enzymatische Aktivität. Ändert sich die
Umgebung des Proteins, beispielsweise durch Temperaturerhöhung, dann kann das Pro-
tein von seinem nativen in einen denaturierten Zustand (D) übergehen und global zerfal-
len, d.h. seine natürliche Struktur verlieren. Dieser Prozess wird Entfaltung oder Den-
aturierung genannt (Tanford 1968).

Im denaturierten Zustand ist die Aminosäurenkette des Proteins entfaltet (Baldwin
und Zimm 2000), was in der Regel zum Verlust der Bioaktivität eines Enzyms führt.
Der native und denaturierte Zustand liegen immer im Gleichgewicht ($N \rightleftharpoons D$) vor
(Anfinsen 1973; Pace 1990). Daher wird die Proteinstabilität als die freie Enthalpiedif-
ferenz (ΔG) zwischen dem nativen und denaturierten Zustand definiert. Sie setzt sich
aus einem enthalpischen (ΔH) und temperaturabhängigen entropischen Term ($T \cdot \Delta S$)
zusammen:

$$\Delta G = G_N - G_D = \Delta H - T \cdot \Delta S \quad (1)$$

Wegen der unterschiedlichen Wärmekapazitäten (C_p) des nativen und denaturierten Zu-
stands (Prabhu und Sharp 2005) sind der enthalpische und entropische Term (und somit
auch die Proteinstabilität) temperaturabhängig:

$$\Delta G(T) = \Delta H(T) - T \cdot \Delta S(T) \quad (2)$$

Unter der vereinfachenden Annahme eines temperaturunabhängigen ΔC_p (Prabhu und Sharp 2005) kann die Stabilität eines Proteins mit der modifizierten Gibbs-Helmholtz-Gleichung beschrieben werden:

$$\Delta G(T) = \Delta H_m (1 - T/T_m) + \Delta C_p \cdot [(T - T_m) - T \cdot \ln(T/T_m)] \quad (3)$$

Hier bedeuten T_m die Schmelztemperatur, bei der $\Delta G = 0$ ist, und ΔH_m der Wert von ΔH am Schmelzpunkt. Die Schmelztemperatur T_m definiert die thermische Stabilität des Proteins. Unterhalb dieser Temperatur (unter nativen Bedingungen) befindet sich das Gleichgewicht auf der Seite des nativen Zustands. Oberhalb dieser Temperatur (unter denaturierenden Bedingungen) befindet sich das Gleichgewicht auf der Seite des denaturierten Zustands. Die Gesamtstabilität von Proteinen kann in Form einer „Stabilitätskurve“ (Becktel und Schellman 1987) aufgetragen werden. Eine typische Stabilitätskurve ist in Abbildung 1 gezeigt.

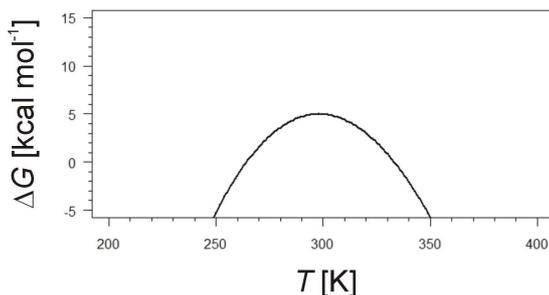


Abbildung 1. Typische Stabilitätskurve eines Proteins. Die Kurve wurde mit Hilfe der im Anhang beschriebenen Gibbs-Helmholtz-Gleichung simuliert, wobei die Werte der Parameter $\Delta H(T_r)$, $\Delta S(T_r)$, ΔC_p und T_r 5 kcal mol⁻¹, 0 kcal mol⁻¹ K⁻¹, 2,5 kcal mol⁻¹ K⁻¹ bzw. 298 K betragen. Nach (Motono *et al.* 2001).

Im nativen Zustand werden Proteine durch eine Vielzahl nichtkovalenter Wechselwirkungen gegenüber ihrer Umgebung stabilisiert. Insbesondere der Entropieverlust und die Desolvatisierung polarer Gruppen während der Faltung wirken dieser Stabilisierung aber entgegen, so dass die Proteinstabilität mit 5–15 kcal mol⁻¹ nur marginal ist (Richards 1997). Die marginale Stabilität erlaubt der Struktur lokal flexibel („beweglich“) zu sein, ohne dass dies eine Auswirkung auf den globalen Faltungszustand hat. Der native Zustand stellt makroskopisch also einen Durchschnittswert mikroskopisch „ähnlicher“ gefalteter Konformationen dar. Die Menge mikroskopisch ähnlicher gefalteter Konformationen wird auch konformationelles Ensemble bezeichnet. Die Flexibilität bestimmter Bereiche der Struktur wird durch lokale Entfaltungsereignisse ermöglicht.

Diese können mit Hilfe der mikroskopischen Stabilität der jeweiligen Region beschrieben werden (Závodszky *et al.* 1998), die als die freie Enthalpiedifferenz (ΔG^{mic}) zwischen dem nativen (N) und einem lokal entfalteten Zustand (N') definiert wird. Die durch die Mikro stabilität bedingte Flexibilität ist die Voraussetzung für die Bioaktivität von Enzymen (Henzler-Wildman und Kern 2007). Durch Anpassung der Flexibilität im Bereich des aktiven Zentrums kann beispielsweise die spezifische Aktivität (k_{cat}) und/oder die Michaelis-Menten-Konstante (K_{m}) moduliert werden (Fields 2001).

1.2.2 Stabilisierende Wechselwirkungen und Kooperativität

Die nichtkovalenten Wechselwirkungen, die zur Stabilisierung des nativen Zustands von Proteinen beitragen, sind seit vielen Jahren bekannt (Kauzmann 1959), allerdings wird über die Bedeutung der einzelnen Kräfte auch heute noch kontrovers diskutiert. Der wichtigste Beitrag zur Proteinstabilität ist ohne Frage der hydrophobe Effekt (Kauzmann 1959; Dill 1990), der zur spontanen Aggregation der hydrophoben Seitenketten im Innern von Proteinen führt. Unter nativen Bedingungen, d.h. bei niedrigen Temperaturen, ist der hydrophobe Effekt entropiegetrieben, wohingegen er bei hohen Temperaturen enthalpiegetrieben ist. Die Ursache dieses Effekts ist die schlechte Wasserlöslichkeit hydrophober Gruppen: Nach dem Mischen von Wasser und Öl werden die unpolaren Moleküle wieder aus dem Wasser hinausgedrängt, was zur Bildung großer Öltröpfchen und schließlich zur Abtrennung einer unpolaren Phase führt. Entsprechend diesem „Öltröpfchenmodell“ aggregieren die hydrophoben Seitenketten von Proteinen und bilden einen hydrophoben Kern im Innern der Struktur. Studien an Modellkomponenten lassen vermuten, dass für jeden Rest, der aus der wässrigen Lösung in den hydrophoben Kern gebracht wird, etwa $1-2 \text{ kcal mol}^{-1}$ gewonnen werden (Wolfenden 2007).

Neben dem hydrophoben Kern können auch an der Proteinoberfläche kleine hydrophobe Kerne gefunden werden (Tisi und Evans 1995). Diese „hydrophoben Oberflächencluster“ bestehen meist nur aus wenigen Seitenketten und spielen insbesondere bei der Stabilisierung von Sekundärstrukturelementen, insbesondere von β -Faltblättern, eine Rolle (Ceruso *et al.* 1999; Espinosa *et al.* 2001; Colombo *et al.* 2003). Eine erst kürzlich durchgeführte Untersuchung des Auftretens hydrophober Cluster in Proteinstrukturen ergab (Arunachalam und Gautham 2008), dass fast alle Proteine neben dem hydrophoben Kern im Inneren der Struktur, der möglicherweise die Triebkraft für die Faltung ist, mehrere kleinere hydrophobe Cluster an der Peripherie besitzen, welche die Struktur stabilisieren. Die hydrophoben Reste in hydrophoben Kernen von Proteinen sind dicht gepackt (Eriksson *et al.* 1992; Chen und Stites 2001), was darauf hindeutet, dass auch

van der Waals-Wechselwirkungen einen wichtigen Beitrag zur Stabilität leisten. Dabei handelt es sich um schwache Kräfte, die aber aufgrund ihrer Vielzahl eine große Rolle spielen.

Kontakte zwischen hydrophoben Atomen, die aufgrund des hydrophoben Effekts und der van der Waals-Wechselwirkungen zustande gekommen sind, werden in der Literatur auch „hydrophobe Bindungen“ oder „hydrophobe Kontakte“ bezeichnet (Scheraga *et al.* 1962). Streng genommen ist diese Bezeichnung nicht sinnvoll, denn sie impliziert eine gewisse Spezifität des hydrophoben Effekts und der van der Waals-Wechselwirkungen, die in der Realität nicht immer vorhanden ist. Die fehlende Spezifität hydrophober Bindungen bedeutet aber, dass diese Wechselwirkungen (entgegen früherer Annahmen) alleine nicht für die Stabilisierung und Rigidisierung einer einzigen, definierten Struktur ausreichen können (Rose und Wolfenden 1993). Die Hauptursache struktureller Spezifität in Proteinen müssen spezifische Wechselwirkungen sein (Richards 1974), wofür hauptsächlich Wasserstoffbrücken (H-Brücken) in Frage kommen, ohne die das Innere von Proteinen einem Tropfen einer unpolaren Flüssigkeit ähneln würde, der zwar kompakt ist, aber keinerlei definierte Struktur hat. Tatsächlich haben Untersuchungen von Proteinkristallstrukturen gezeigt, dass es polare Gruppen im Innern von Proteinen gibt, und dass diese Gruppen fast immer H-Brücken zu anderen polaren Gruppen ausbilden (Baker und Hubbard 1984; Stickle *et al.* 1992; Savage *et al.* 1993; McDonald und Thornton 1994).

H-Brücken verbinden Atome X und Y, deren Elektronegativität größer ist als die von Wasserstoff, in Proteinen vornehmlich N und O, ferner C und S. Die X–H-Gruppe wird allgemein als Protonendonator- (D) und das Y-Atom als Protonenakzeptor-Gruppe (A) bezeichnet. Die stärksten H-Brücken werden zwischen den Atomen N und O als X und Y gebildet, obgleich auch C–H als Donor und ein π -Elektronensystem als Akzeptor fungieren kann (Taylor und Kennard 1982). H-Brücken zwischen geladenen Gruppen werden Salzbrücken genannt. Die ersten theoretischen Modelle legten nahe, dass H-Brücken ausschließlich als elektrostatische Wechselwirkung zwischen dem partiell positiven Wasserstoffatom des Donors und dem freien Elektronenpaar des Akzeptors zu betrachten sind (Pauling 1928). Heute wird allgemein akzeptiert, dass H-Brücken sowohl den Charakter einer elektrostatischen als auch einer schwachen kovalenten Wechselwirkung haben und entsprechend distanz- und winkelabhängig sind (Desiraju 2002). Der Beitrag von H-Brücken zur Proteinstabilität darf nicht auf Kosten der hydrophoben Wechselwirkungen überschätzt werden. Tatsächlich beträgt die durchschnittliche Ener-

gie einer neutralen H-Brücke in Lösung nicht mehr als 1–2 kcal mol⁻¹ (Myers und Pace 1996) und im Proteininnern 3,5 kcal mol⁻¹ (Shirley *et al.* 1992), was bedeutet, dass die Bildung von H-Brücken unmöglich die Haupttriebkraft der Faltung sein kann. H-Brücken erhöhen allerdings die Packung der Struktur (Pace 2001; Pace *et al.* 2004; Schell *et al.* 2006), was zu einer weiteren Verstärkung der van der Waals-Wechselwirkungen im Innern des Proteins führt.

Zusammenfassend kann gesagt werden, dass die nativen nichtkovalenten Wechselwirkungen auf eine komplexe Weise zusammenwirken. Dieser Überlegung zufolge kann das Innere eines Proteins eher mit einer (rigiden) kristallinen Feststoffphase als mit einem Tropfen einer unpolaren Flüssigkeit verglichen werden, was mit Hilfe kristallographischer und densimetrischer (Hvidt 1983; Kundrot und Richards 1987; Zhou *et al.* 1999) sowie theoretischer Studien (Cooper 2000) auch belegt wurde. Das komplexe Zusammenwirken nichtkovalenter Wechselwirkungen führt zu einer gewissen Kooperativität, d.h. die einzelnen Kräfte summieren sich im nativen Zustand nicht nur, sondern sie verstärken sich. Zusätzlich treten deswegen langreichweitige Effekte auf, so genannte allosterische Effekte, bei denen die Stärke bzw. die An- oder Abwesenheit von Wechselwirkungen in einer Region des Proteins einen Einfluss auf die Mikrostabilität einer anderen Region hat (Cui und Karplus 2008). Die Bereiche, in denen die Wechselwirkungen zusammenwirken, werden auch kooperative Faltungseinheiten genannt. Kleine Proteine besitzen meist nur eine einzige kooperative Einheit, wohingegen große Proteine, die aus mehreren Modulen (Untereinheiten, Domänen oder Subdomänen) bestehen, mehrere kooperative Einheiten besitzen, die unabhängig voneinander falten.

1.2.3 Die thermische Entfaltung

Eine durch die Temperatur hervorgerufene „Veränderung der Quartär-, Tertiär- und/oder Sekundärstruktur durch Zerstörung der nativen nichtkovalenten Wechselwirkungen ohne Veränderungen der Primärstruktur oder kovalenter Bindungen“ (Tanford 1968) wird thermische Entfaltung oder Denaturierung genannt. Die während dieses Prozesses zugeführte Energie in Form von Wärme beeinflusst die intramolekularen nichtkovalenten Wechselwirkungen auf eine sehr komplexe Weise und verursacht letztendlich den globalen Zerfall der Struktur. Der hydrophobe Effekt ist bei niedrigen Temperatur endotherm und bei höheren Temperaturen exotherm (Scheraga *et al.* 1962). Der hydrophobe Effekt nimmt deswegen mit steigender Temperatur erst einmal zu, bis er ab einer Temperatur von ungefähr 75 °C wieder abnimmt (Makhatadze und Privalov 1995).

Van der Waals-Wechselwirkungen und H-Brücken sind exotherm und werden deswegen mit zunehmender Temperatur schwächer. Für isolierte H-Brücken konnte mit Hilfe quantenchemischer Methoden gezeigt werden, dass ihre Stärke linear mit zunehmender Temperatur abnimmt (Dougherty 1998). Die Abschwächung gilt auch für Salzbrücken, deren Bildung ebenfalls exotherm sind. Hier muss allerdings berücksichtigt werden, dass die Desolvatation geladener Protonendonator- und Protonenakzeptor-Gruppen bei zunehmender Temperatur günstiger wird, so dass Salzbrücken in Proteinstrukturen mit zunehmender Temperatur möglicherweise sogar leicht verstärkt werden (Elcock 1998).

In erster Näherung kann somit davon ausgegangen werden, dass Salzbrücken mit zunehmender Temperatur gleich stark bleiben oder sogar stärker werden, H-Brücken schwächer werden und hydrophobe Bindungen bis etwa 75 °C gleich stark bleiben und danach schwächer werden. Das Aufbrechen der H-Brücken genügt allerdings schon, um die Entfaltung einzuleiten, da mit dem Verlust dieser Bindungen die strukturelle Spezifität verloren geht. Dies führt dazu, dass die Struktur global flexibel wird und sich deswegen entfaltet.

1.2.4 Reversible vs. irreversible Denaturierung

Theoretisch ist der Prozess der Denaturierung immer reversibel, da es sich bei der Entfaltung (und Faltung) um eine Gleichgewichtsreaktion handelt (Anfinsen 1973). In der Regel treten neben dem reversiblen Vorgang aber auch irreversible Denaturierungsprozesse auf. In der Praxis spielt die irreversible Denaturierung sogar eine weitaus größere Rolle als der reversible Prozess (Vieille und Zeikus 2001). Die thermische Stabilität irreversibel denaturierender Proteine wird mit der Halbwertszeittemperatur (T_{50}) beschrieben, die angibt, bei welcher Temperatur die Halbwertszeit des Proteins eine bestimmte Zeit (beispielsweise 10 Minuten) unterschreitet. Der Zusammenhang reversibler und irreversibler Denaturierungsprozesse kann mit Hilfe des Lumry-Eyring-Modells verstanden werden (Lumry und Eyring 1954).

In diesem Modell entspricht der erste Schritt der Denaturierung einem reversiblen Prozess, über den der native mit dem denaturierten oder einem partiell denaturierten Zustand im Gleichgewicht steht ($N \rightleftharpoons D$). Der zweite Schritt entspricht einer irreversiblen Veränderung des denaturierten oder eines partiell denaturierten Zustands zu einem finalen Zustand ($D \rightarrow F$). Der irreversible Schritt ist das Ergebnis tiefgreifender Modifikationen des Proteins, die entgegen der obigen Definition der Denaturierung auch die

Primärstruktur (d.h. die kovalenten Bindungen) betreffen können. Dabei handelt es sich um Aggregation, Koagulation, Autolyse oder die chemische Veränderung von Resten (Ahern und Klibanov 1988; Vieille und Zeikus 2001). Wenn der irreversible Schritt sehr schnell abläuft, handelt es sich um irreversible Denaturierung. Ansonsten bleibt der Vorgang reversibel.

1.3 Thermostabilität und thermophile Anpassung

1.3.1 Industrielle Bedeutung thermostabiler Proteine

Aufgrund ihrer Vielseitigkeit werden Proteine in der Biotechnologie als Biokatalysatoren in industriellen Prozessen, als Futtermittelzusatzstoffe oder als pharmazeutische Wirkstoffe eingesetzt. Das Potential biotechnologisch eingesetzter Enzyme ist hoch (Braun *et al.* 2006), denn für jede bekannte chemische Reaktion wird eine biokatalytische Entsprechung vermutet (Rozzell 1999). In einer Vielzahl industrieller und biotechnologischer Prozesse werden insbesondere thermostabile Enzyme (so genannte Thermozyyme) eingesetzt (Braun *et al.* 2006). Neben der Stabilität gegenüber hohen Temperaturen zeichnen sich Thermozyyme meist auch durch eine geringere Empfindlichkeit gegenüber Schwankungen des pH-Werts und/oder erhöhten Konzentrationen denaturierender Reagenzien in der Lösung aus (Bruins *et al.* 2001).

Der Einsatz von Thermozyymen hat gegenüber dem Einsatz mesostabiler Biokatalysatoren mehrere entscheidende Vorteile (Bruins *et al.* 2001): So wird durch das Ansetzen hoher Temperaturen im Bioreaktor das Risiko bakterieller Kontamination herabgesetzt. Besonders in der Nahrungsmittelindustrie werden deswegen Prozesstemperaturen von 60–70 °C angestrebt. Die Erhöhung der Temperatur bringt darüber hinaus auch eine Reihe verfahrenstechnischer Vorteile mit sich, wie beispielsweise die Erniedrigung der Viskosität der Reaktionslösung, die Erhöhung von Diffusionsraten sowie die Erhöhung der Löslichkeit von Substraten. Aus diesen Gründen erfolgt beispielsweise die Verarbeitung von Stärke bei Temperaturen von 100–110 °C. Ein weiterer Vorteil von Thermozyymen besteht darin, dass sie oft über einen größeren Temperaturbereich hinweg aktiv sind und generell eine höhere Aktivität als homologe mesophile Enzyme aufweisen.

Neben der Bedeutung als Biokatalysatoren spielen Thermozyyme auch als Futtermittelzusatzstoffe oder pharmazeutische Wirkstoffe in große Rolle (Braun *et al.* 2006). So wird beispielsweise das Phytinsäure zersetzende Enzym Phytase in der Schweine- und Geflügelzucht als Zusatzstoff in pflanzlichem Futter eingesetzt (Haefner *et al.* 2005). Im Gegensatz zu Wiederkäuern können Schweine und Geflügel den zu 80 % in Form von

Phytinsäure gebundenen Phosphor in pflanzlicher Nahrung nicht nutzen. Dies macht die Zufütterung anorganischen Phosphats erforderlich, die bei falscher Dosierung zu erheblichen Umweltbelastungen führen kann. Durch den alternativen Zusatz von Phytase wird der in den Futterpflanzen vorhandene Phosphor als Nährstoff erschlossen. Auf die sonst übliche Zufütterung von Phosphat kann dann mit positivem Effekt für die Umwelt verzichtet werden: Allein in der Schweinezucht konnte der Phosphataustrag über die Gülle nach Einsatz von Phytasen um 30 % verringert werden (Braun *et al.* 2006). Die bei der Produktion des Futters (Pellettierung) auftretenden hohen Temperaturen machen es erforderlich, dass die eingesetzte Phytase thermostabil ist. Gleichzeitig muss sie aber bei niedrigeren Temperaturen im Magen der Tiere aktiv sein.

1.3.2 Natürlich vorkommende thermostabile Proteine

Der Vergleich von Proteinen aus mesophilen Organismen (optimale Wachstumstemperatur von 15–45 °C) mit natürlich vorkommenden thermostabilen Proteinen aus thermophilen oder hyperthermophilen Organismen (optimale Wachstumstemperatur von 50–80 bzw. mehr als 80 °C) erlaubt die Aufklärung der molekularen Grundlagen der Thermostabilität (Jaenicke und Böhm 1998). Dieses Wissen kann einerseits zum besseren Verständnis der Proteinstabilität im Allgemeinen beitragen (Razvi und Scholtz 2006) und andererseits bei der Entwicklung künstlicher thermostabiler Proteine eingesetzt werden (Russell und Taylor 1995). Es ist zu beachten, dass neben intrinsischen Faktoren für die thermophile Anpassung auch extrinsische Faktoren (Umgebungsfaktoren) eine Rolle spielen können (Bruins *et al.* 2001), wie die Anwesenheit von Salzen, Koenzymen oder Substraten sowie die hohe Konzentration von Proteinen in der Zelle. Auch eine Glykosylierung kann die Thermostabilität von Proteinen extrinsisch verändern (Jenkins und Curling 1994).

Thermodynamische Untersuchungen der Stabilität thermophiler Proteine haben gezeigt, dass die Anpassung dieser Proteine über drei verschiedene Mechanismen (und deren Kombinationen) erfolgen kann (Nojima *et al.* 1977; Kumar *et al.* 2001; Razvi und Scholtz 2006), die letztendlich alle zum gleichen Ergebnis führen, nämlich zur Verschiebung der Denaturierung zu höheren Temperaturwerten. Abbildung 2 zeigt, wie sich die drei Mechanismen auf die Temperaturabhängigkeit der Stabilität auswirken: Der erste Mechanismus (Kurve I) führt zu einer Stabilisierung des thermophilen Proteins über den gesamten Temperaturbereich, d.h. zu einer vertikalen Verschiebung der Stabilitätskurve. Dies ist der am häufigsten verwendete Mechanismus der thermophilen Anpassung (Rees und Robertson 2001; Razvi und Scholtz 2006) und wird durch eine

Erhöhung von ΔH in Abwesenheit kompensatorischer Veränderungen von ΔS erreicht. Der zweite Mechanismus (Kurve II) führt zu einer horizontalen Verschiebung der Stabilitätskurve. Dabei wird die maximale Stabilität des thermophilen Proteins gegenüber dem mesophilen Protein nicht verändert, aber entlang der Temperaturskala verschoben. Er wird durch eine Verringerung von ΔS erreicht. Der dritte Mechanismus (Kurve III) führt zu einer Abflachung der Stabilitätskurve, was durch eine Verringerung der Wärmekapazitätsdifferenz (ΔC_p) erreicht wird.

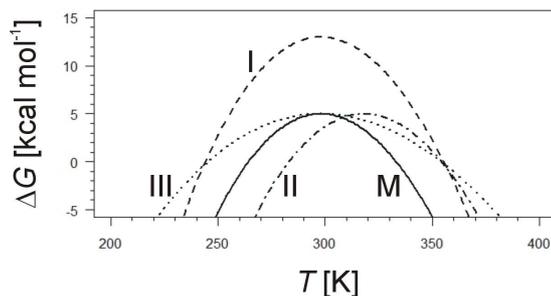


Abbildung 2. Drei Mechanismen zur thermophilen Anpassung. Die Stabilitätskurve M zeigt die Abhängigkeit der Stabilität von der Temperatur für ein mesophiles Protein. Kurve I, II und III sind im Text erläutert. Die Kurven wurden mit Hilfe der im Anhang beschriebenen Gibbs-Helmholtz-Gleichung simuliert, wobei die Werte der Parameter $\Delta H(T_r)$, $\Delta S(T_r)$, ΔC_p und T_r 5 kcal mol⁻¹, 0 kcal mol⁻¹ K⁻¹, 2,5 kcal mol⁻¹ K⁻¹ bzw. 298 K für Kurve M, 13 kcal mol⁻¹, 0 kcal mol⁻¹ K⁻¹, 2,5 kcal mol⁻¹ K⁻¹ bzw. 298 K für Kurve I (gestrichelte Linie), 5 kcal mol⁻¹, 0 kcal mol⁻¹ K⁻¹, 2,5 kcal mol⁻¹ K⁻¹ bzw. 318 K für Kurve II (gestrichelpunktete Linie) und 5 kcal mol⁻¹, 0 kcal mol⁻¹ K⁻¹, 1 kcal mol⁻¹ K⁻¹ bzw. 298 K für Kurve III (gepunktete Linie) betragen. Nach (Motono *et al.* 2001).

Vergleichsstudien der Sequenzen und Strukturen meso- und thermophiler Proteine haben entgegen früherer Erwartungen (Somero 1978; Argos *et al.* 1979) keine deutlichen systematischen Unterschiede zwischen meso- und thermostabilen Proteinen erkennen lassen (Jaenicke und Böhm 1998; Vieille und Zeikus 2001). In Abbildung 3 sind die Strukturen meso- und thermophiler thermolysinähnlicher Protease (TLP) aus *Bacillus cereus* bzw. *thermoproteolyticus* gegenübergestellt. Die Sequenzen der beiden Proteine sind zu 73 % identisch. Auch ihre Strukturen unterscheiden sich kaum voneinander: Sie haben eine mittlere quadratische Abweichung (*root mean square deviation, RMSD*) von 0,9 Å für die Ca-Atome. Offenbar erfolgt die Anpassung thermophiler Proteine durch gezielte Veränderungen der nichtkovalenten Wechselwirkungen, die sowohl die elektrostatischen Kräfte (H- und Salzbrücken), als auch die Packung der Aminosäureseitenketten und die Stärke des hydrophoben Effekts betreffen (Querol *et al.* 1996; Vogt und Argos 1997; Vogt *et al.* 1997; Szilágyi und Závodszy 2000; Gianese *et al.* 2001; Gro-

miha 2001; Kumar und Nussinov 2001; Gianese *et al.* 2002; Robinson-Rechavi *et al.* 2006; Sadeghi *et al.* 2006; Folch *et al.* 2008; Saelensminde *et al.* 2009). So zeigte der Vergleich ganzer Genome thermophiler und mesophiler Organismen, dass thermophile Proteine weniger ungeladene polare Reste (Gln, Asn, Thr und Ser) und mehr geladene Reste (Glu, Arg und Lys) enthalten (Haney *et al.* 1999; Cambillau und Claverie 2000; Das und Gerstein 2000; Sadeghi *et al.* 2006). Thermophile Proteine enthalten außerdem mehr hydrophobe Reste (Sadeghi *et al.* 2006), insbesondere solche, die β - und γ -verzweigt sind (Gromiha *et al.* 1999; Chakravarty und Varadarajan 2000).

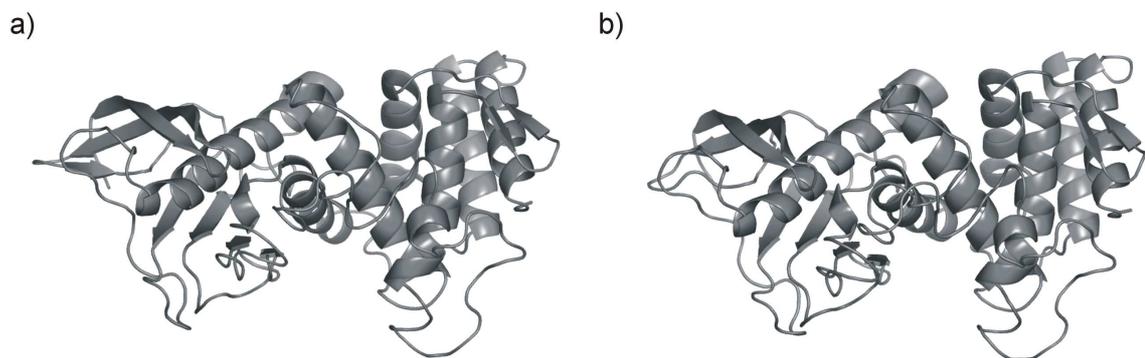


Abbildung 3. Gegenüberstellung der Strukturen meso- (a) und thermophiler thermolysinähnlicher Protease (TLP) aus *Bacillus cereus* bzw. *thermoproteolyticus*. Die Sequenzen der beiden Proteine sind zu 73 % identisch, ihre Strukturen haben eine *RMS*-Abweichung von 0,9 Å für die *Ca*-Atome.

Beim Vergleich der Strukturen meso- und thermophiler Proteine wurden vereinzelt eine erhöhte Anzahl von Disulfidbrücken (Hopfner *et al.* 1999; Singleton *et al.* 1999), das vermehrte Auftreten aromatischer Cluster (Kannan und Vishveshwara 2000), stabilere Sekundärstrukturelemente (Facchiano *et al.* 1998), stabilere α -Helixdipole, stärkere Wechselwirkungen zwischen Untereinheiten oder Domänen (Kirino *et al.* 1994), eine verringerte Spannung im Rückgrad der Polypeptidkette, eine höhere Affinität zu Metallionen, kürzere Loop-Regionen (Russell *et al.* 1997) oder eine erhöhte Anzahl schwacher Wechselwirkungen (Ibrahim und Pattabhi 2004) festgestellt. Die meisten dieser Stabilisierungsmechanismen können auf Veränderungen der elektrostatischen Wechselwirkungen, der Packung der Aminosäureseitenketten und der Stärke des hydrophoben Effekts zurückgeführt werden (Vogt und Argos 1997).

Die Bedeutung elektrostatischer Wechselwirkungen mag bei der Anpassung hyperthermophiler Proteine größer sein als bei der Anpassung thermophiler Proteine (Karshikoff und Ladenstein 2001), da Salzbrücken in Proteinstrukturen mit zunehmender Temperatur möglicherweise leicht verstärkt werden (Elcock 1998). In vielen Fällen

liegen die für die Thermostabilisierung verantwortlichen elektrostatischen Wechselwirkungen an der Proteinoberfläche (Alsop *et al.* 2003; Strickler *et al.* 2006; Saelensminde *et al.* 2009), was für die Tatsache spricht, dass Oberflächeneigenschaften eine wichtige Rolle bei der thermophilen Anpassung spielen und deswegen als Unterscheidungsmerkmal zwischen meso- und thermophilen Proteinen verwendet werden können (Braiuca *et al.* 2007).

Die Bedeutung von Veränderungen der Packung der Aminosäureseitenketten und der Stärke des hydrophoben Effekts für die thermophile Anpassung wird kontrovers diskutiert (Karshikoff und Ladenstein 1998; Pack und Yoo 2005; Robinson-Rechavi *et al.* 2006; Paiardini *et al.* 2008). Generell scheint nicht nur eine erhöhte Anzahl (Vogt und Argos 1997; Vogt *et al.* 1997), sondern in erster Linie auch eine bessere Verteilung (d.h. ein besseres Netzwerk) nichtkovalenter Wechselwirkungen zur thermophilen Anpassung beizutragen (Xiao und Honig 1999; Brinda und Vishveshwara 2005; Robinson-Rechavi *et al.* 2006). Eine Untersuchung von Proteinen aus dem thermophilen Bakterium *Thermatoga maritima* ergab, dass diese Strukturen eine erhöhte Anzahl hochgradig vernetzter Reste (so genannter „Hubs“) besitzen (Robinson-Rechavi *et al.* 2006). Zu diesem Ergebnis kamen auch Brinda *et al.* beim Vergleich homologer meso- und thermophiler Enzyme (Brinda und Vishveshwara 2005).

Die strukturellen Veränderungen thermophiler Proteine tragen alle zur Verbesserung des Netzwerkes nichtkovalenter Wechselwirkungen bei und erhöhen letztendlich die mechanische Rigidität der Struktur (Vieille und Zeikus 2001; Sterner und Brunner 2008). Tatsächlich zeigen Flexibilitätsindizes, die von kristallographischen Temperaturfaktoren abgeleitet wurden, dass die Flexibilität verringert wird, wenn die Thermostabilität erhöht wird (Vihinen 1987). Die Dynamikeigenschaften homologer Proteine verschiedener Thermostabilität sind mit einer ganzen Reihe experimenteller Methoden untersucht worden, wie beispielsweise dem H/D-Austausch, der Neutronenstreuung und der NMR-Spektroskopie (Sterner und Brunner 2008): Wenn die Strukturen bei ihrer jeweiligen optimalen Temperatur (entsprechend der optimalen Wachstumstemperatur des Organismus, aus dem sie stammen) verglichen werden, dann befinden sie sich in „korrespondierenden Zuständen“ ähnlicher Bioaktivität, was trotz globaler Rigidisierung auf eine Konservierung wichtiger flexibler Bereiche schließen lässt (Somero 1978; Jaenicke und Böhm 1998; Sterner und Brunner 2008).

In einigen Fällen stehen experimentelle Ergebnisse zur Theorie der korrespondierenden Zustände im Widerspruch (Jaenicke 2000), was daran liegen mag, dass die expe-

rimentelle Beschreibung und Quantifizierung der Proteinflexibilität insbesondere hinsichtlich der verschiedenen Arten und Zeitskalen der interessierenden Bewegungen schwierig ist (Kamerzell und Middaugh 2008; Sterner und Brunner 2008). Die gemessene Dynamik (Bewegung) ist nur ein indirektes Maß für die tatsächlich vorhandene Flexibilität (Beweglichkeit). Zudem bleibt die mit der Aktivität des Enzyms verknüpften Flexibilität ein schwer messbarer Parameter, da Flexibilitätsänderungen auf limitierte, aber dennoch wichtige Bereiche der Struktur beschränkt sein können (Georlette *et al.* 2004).

1.3.3 Die Entwicklung künstlicher thermostabiler Proteine

Um den steigenden Bedarf (Braun *et al.* 2006) an neuen Biokatalysatoren zu decken, wird zunächst die Natur systematisch nach thermophilen oder hyperthermophilen Organismen durchsucht (Demirjian *et al.* 2001; Atomi 2005; Ferrer *et al.* 2007). Die Enzyme dieser Mikroorganismen liegen jedoch in der Regel nur in sehr geringen Mengen vor. Zudem wachsen ihre natürlichen Wirtsorganismen häufig sehr schlecht unter üblichen Kulturbedingungen (Antranikian 2008). Die Gene für diese Proteine werden daher isoliert und in geeignete Produktionsstämme transformiert. Natürlich vorkommende Thermozyyme werden trotzdem nur selten für industrielle Prozesse vermarktet, da sie hinsichtlich ihrer Stabilität und Aktivität noch weiter verbessert werden können (Antranikian 2008). Die isolierten Enzyme werden daher mit Hilfe molekularbiologischer Methoden (Gentechnologie) verändert und an gewünschte Reaktionen und Reaktionsbedingungen angepasst (Atomi 2005).

Grundsätzlich existieren zwei Ansätze, bekannte Enzyme zu verändern. Das so genannte „rationale Design“ erfordert eine genaue Kenntnis der Beziehungen zwischen der Struktur, der Stabilität und der Funktion eines Enzyms (Eijsink *et al.* 2004). Durch Experimente oder mit Hilfe rechnerischer Verfahren werden dann Regionen vorhergesagt, wo das Einführen einer ganz bestimmten Mutation vielversprechend ist. Strukturmodelle der Mutanten werden dann mit Computerprogrammen bewertet, gegebenenfalls hergestellt und experimentell untersucht. Die „gerichtete Evolution“ verfolgt eine andere Strategie (Eijsink *et al.* 2005). Hier sind keinerlei Kenntnisse der Struktur notwendig, sondern es wird das natürliche Evolutionsprinzip kopiert: Sequenzen des ursprünglichen Proteins werden zufällig mutiert und exprimiert. Die Bibliothek der so erzeugten Sequenzen kann bis zu 10^6 Mutanten enthalten. Anschließend wird nach gewünschten Kriterien (Stabilität, Aktivität, Selektivität) *gescreent* oder selektiert.

In der Praxis stellt die Größe der Bibliothek einen limitierenden Faktor bei der gerichteten Evolution dar. Selbst für ein kleines Protein müssten bei rein zufälliger Mutation der gesamten Sequenz theoretisch weit über 10^{10} Mutanten erzeugt werden. Daher sind Ansätze von Interesse, die bei dem Aufbau „zielgerichteter“ Bibliotheken helfen. Beim so genannten „datengetriebenen“ Protein Engineering wird deswegen jede verfügbare Information über ein Protein genutzt, um die Anzahl der zu erzeugenden Mutanten einzuschränken, aber dennoch (bis zu einem gewissen Umfang) die Identifizierung unvorhersagbarer Mutationen zuzulassen (Chaparro-Riggers *et al.* 2007). Als Datengrundlage kann jede Art struktureller Information dienen, wie beispielsweise Kristallstrukturen, NMR-Strukturen oder Homologiemodelle (Chaparro-Riggers *et al.* 2007). So wurden beispielsweise Informationen über die Beweglichkeit einzelner Reste im nativen Zustand (aus B-Faktoren) verwendet (Reetz *et al.* 2006). Im so genannten Consensus-Ansatz werden Sequenz- und Phylogenieinformationen verwendet (Steipe *et al.* 1994; Lehmann *et al.* 2000). Der Ansatz beruht auf der Hypothese, dass Consensus-Aminosäuren aus Sequenzalignments homologer Proteine mehr zur Stabilität beitragen als Nichtconsensus-Aminosäuren. Der Consensus-Ansatz wurde erfolgreich von Lehmann *et al.* zur Entwicklung von Phytasen mit verbesserten thermischen Eigenschaften verwendet (Lehmann *et al.* 2000; Lehmann und Wyss 2001; Lehmann *et al.* 2002).

Immer häufiger werden auch rechnerische Methoden zur Unterstützung des Protein Engineering herangezogen (Chaparro-Riggers *et al.* 2007). Computeransätze müssen einerseits die thermische Stabilität einer Struktur sowie den Effekt einer Mutation abschätzen können, andererseits Hinweise darauf geben, in welchen Bereichen der Struktur thermostabilisierende Mutationen eingeführt werden können. Bei der Entwicklung von Thermostzymen sollten die Methoden außerdem in der Lage sein, die Auswirkung einer thermostabilisierenden Mutation auf die Aktivität vorherzusagen, indem beispielsweise bestimmt wird, ob die Mutation die Struktur des aktiven Zentrums oder dessen Flexibilität stört (Korkegian *et al.* 2005). Eine Auswahl rechnerischer Ansätze, die diesen Ansprüchen (weitestgehend) genügen, wird im nächsten Kapitel vorgestellt.

1.4 Rechnerische Ansätze zur Vorhersage der Proteinstabilität

1.4.1 Vorhersage der Gesamtstabilität

Aus Molekulardynamik(MD)-Simulationen kann die freie Stabilisierungsenthalpie von Proteinen bestimmt werden, wenn alle relevanten Bereiche des Konformationsraums ausreichend durchmustert werden. Die freie Enthalpiedifferenz zwischen dem

nativen und denaturierten Zustand von Proteinen ergibt sich dann direkt aus dem Verhältnis der Häufigkeiten, mit der native oder denaturierte Konformationen auftreten. Die Unterschiede in den freien Enthalpien sowie die zu überwindenden Aktivierungsbarrieren sind allerdings so groß, dass wichtige Bereiche des Konfigurationsraums nicht oder nur unzureichend durchmustert werden. Zudem stellt die enorme Größe des Konfigurationsraums ein Problem dar: Sali *et al.* schätzen die Zahl der Konfigurationen im nativen und denaturierten Zustand eines kleinen Proteins auf etwa 10^{10} bzw. 10^{16} (Sali *et al.* 1994). Selbst mit Hilfe modernster Computer und spezieller Verfahren der freien Enthalpierechnung (van Gunsteren 1988; van Gunsteren *et al.* 1995) muss dafür mehrere Wochen lang simuliert werden. In der Praxis ist die Berechnung der Gesamtstabilität von Proteinen (ΔG) deswegen nicht möglich.

Mit Hilfe exakter Verfahren wie der thermodynamischen Integrationsrechnung (TI) oder der freien Enthalpiestörungsrechnung (FEP) kann zumindest die Stabilität von Proteinmutanten bestimmt werden (Tidor und Karplus 1991; Shi *et al.* 1993; Funahashi *et al.* 2003). Diese ergibt sich aus dem Stabilitätsunterschied ($\Delta\Delta G$) zwischen einem Wildtyp und seiner Mutante. Bei dieser Anwendung kann die explizite Simulation des denaturierten Zustands durch die Simulation eines Peptidmodells ersetzt werden (Shi *et al.* 1993; Funahashi *et al.* 2003). Die Ergebnisse aus freien Energierechnungen sind allerdings sehr sensibel gegenüber der Qualität des verwendeten Kraftfelds, der Dauer und dem Umfang der Durchmusterung der Zustände sowie der Modellierung des denaturierten Zustands (Shi *et al.* 1993; Funahashi *et al.* 2003), was die Praxistauglichkeit dieser Methoden in Frage stellt (van Gunsteren *et al.* 1995). Eine gangbare, aber immer noch rechenaufwändige Alternative zu freien Energierechnungen sind „semiempirische“ Ansätze, bei denen Wechselwirkungen des Proteins mit dem Lösungsmittel und entropische Effekte implizit berücksichtigt werden können (Lee *et al.* 2000).

Für praktische Anwendungen wurden sehr schnelle Methoden entwickelt, die von einer einfacheren Darstellung des Proteins und einer approximativen Energiefunktion ausgehen (Alvizo *et al.* 2007). Mit diesen „nichtphysikalischen“ Methoden können relative Stabilitäten (so genannte „Scores“) berechnet werden. Nur die Scores ähnlicher Proteine oder Proteine ähnlicher Faltung können verglichen werden. Scoringmethoden werden daher hauptsächlich zur Abschätzung der Stabilität von Proteinmutanten (im Protein Engineering) oder zur Vorhersage von Sequenzen für bestimmte Faltungen (beim so genannten „Protein Design“) verwendet, wo eine sehr schnelle Bewertung ähnlicher Strukturen nötig ist. Generell basieren diese Methoden auf vereinfachten

Kraftfeldern, wissensbasierten Potentialen oder empirischen Modellen. Auf die Simulation des nativen oder denaturierten Zustands wird verzichtet, und die Rechnung wird anhand einer Einzelstruktur (Kristallstruktur oder Homologiemodell) durchgeführt.

Grundsätzlich ist nicht immer davon auszugehen, dass mit diesen Verfahren zur Vorhersage der Gesamtstabilität auch Aussagen über die Thermostabilität eines Proteins gemacht werden können. Wie in Kap. 1.3.2 beschrieben, muss eine Erhöhung der Thermostabilität nicht zwangsläufig durch eine Erhöhung der Gesamtstabilität des Proteins erfolgen. Dennoch ist Malakauskas und Mayo mit Hilfe energiebasierter Methoden der Entwurf eines hyperthermostabilen Proteins gelungen (Malakauskas und Mayo 1998), wobei sie (ausgehend von der Sequenz eines mesostabilen Proteins) über 10^6 Mutanten erzeugten und diese bewerteten. Die verwendete Scoringfunktion berücksichtigte mehrere empirische Komponenten, die unter anderem die Packung, die Stärke des hydrophoben Effekts und das Vermögen eines Rests zur Sekundärstrukturausbildung betrafen (Dahiyat und Mayo 1997). Alle diese Komponenten wurden als wichtig für die Thermostabilisierung angesehen.

Korkegian *et al.* (Korkegian *et al.* 2005) verwendeten einen vergleichbaren Ansatz in Verbindung mit einer kraftfeldbasierten Scoringfunktion (Kuhlman und Baker 2000). Ihnen gelang die Erhöhung der thermischen Stabilität eines Enzyms ohne negative Beeinflussung der Aktivität, indem alle Reste, die weniger als 4 Å von der aktiven Tasche entfernt lagen, nicht „berührt“ wurden. Auch Gribenko *et al.* haben einen Ansatz vorgestellt, mit dem die Thermostabilisierung von Enzymen ohne die negative Beeinflussung der Aktivität möglich ist (Gribenko *et al.* 2009). Ihre Methode beruht auf der rationalen Verbesserung des Salzbrückennetzwerks an der Proteinoberfläche, wobei die für die Aktivität wichtigen Reste ebenfalls wieder nicht berührt werden.

1.4.2 Identifizierung für die Stabilität wichtiger Regionen

Hilfreich zur Unterstützung des Protein Engineering und insbesondere des rationalen oder datengetriebenen Designs sind rechnerische Methoden, die Hinweise darauf geben, in welchen Bereichen der Struktur thermostabilisierende Mutationen eingeführt werden können. Basu und Sen entwickelten ein einfaches Verfahren, das auf der Bestimmung von Resten beruht, deren Wechselwirkungsenergien zu den Nachbarresten schwach ist (Basu und Sen 2009). Zur Bestimmung relevanter Regionen können neben energiebasierten Verfahren auch Methoden eingesetzt werden, bei denen die Proteinstruktur als ein Netzwerk wechselwirkender Atome oder Reste beschrieben wird.

Vendruscolo *et al.* und andere haben gezeigt, dass Strukturnetzwerke von Proteinen die Eigenschaften einer „kleinen Welt“ (*small world*) haben (Dokholyan *et al.* 2002; Vendruscolo *et al.* 2002). *Small-world*-Netzwerke zeichnen sich dadurch aus, dass bestimmte Atome oder Reste identifiziert werden können, die für die Topologie des Netzwerks bestimmend sind (Böde *et al.* 2007). Heringa und Argos verwendeten Strukturnetzwerke von Proteinen, um Cluster dicht gepackter Seitenketten zu identifizieren (Heringa und Argos 1991). Später konnten sie zeigen, dass solche Cluster für die thermische Stabilität der Struktur wichtig sind (Heringa *et al.* 1995). Brinda *et al.* verwendeten einen ähnlichen Ansatz (Kannan und Vishveshwara 1999), um Unterschiede zwischen meso- und thermophilen Proteinen zu finden (Brinda und Vishveshwara 2005). Auch Robinson-Rechavi *et al.* verwendeten ein solches Verfahren (Greene und Higman 2003) zur Untersuchung struktureller Eigenschaften thermophiler Proteine (Robinson-Rechavi *et al.* 2006).

1.4.3 Simulation der thermischen Denaturierung

Mit Hilfe von Simulationen der Denaturierung von Proteinen können strukturelle Veränderungen untersucht werden, die während der Entfaltung auftreten und letztendlich zum globalen Zerfall der Struktur führen. Diese Ansätze liefern keine Informationen über die Gesamtstabilität von Proteinen (ΔG) oder den Stabilitätsunterschied ($\Delta\Delta G$) zwischen einem Wildtyp und seiner Mutante, können aber direkte Hinweise auf die thermische Stabilität von Proteinen (T_m oder ΔT_m) geben. Zudem können neben dem globalen Zerfall auch lokale Entfaltungsereignisse untersucht werden, welche die Flexibilität bestimmter Bereiche der Struktur ermöglichen und mit Hilfe der mikroskopischen Stabilität der jeweiligen Region beschrieben werden können. Entfaltungssimulationen werden zur Unterstützung des Protein Engineering eingesetzt, wenn der Einfluss einer oder mehrerer Mutationen auf die Denaturierung untersucht werden soll. Klassische Simulationen der thermischen Denaturierung sind MD-Simulationen bei hohen Temperaturen. Hier wird vorausgesetzt, dass die Entfaltung weitestgehend temperaturabhängig ist und die gleichen Eigenschaften aufweist wie die Faltung, was für mehrere Beispiele bestätigt werden konnte (Cavalli *et al.* 2002; Day *et al.* 2002), aber nicht allgemein gilt (Finkelstein 1997).

MD-Entfaltungssimulationen können grundsätzlich auf zwei Arten durchgeführt werden (van Gunsteren *et al.* 1995). Bei der ersten Methode werden einzelne „Schritte“ der Entfaltung im Gleichgewicht simuliert. Dazu werden Simulationen bei verschiedenen Temperaturen gestartet. Anschließend werden die Eigenschaften der verschiedenen

Gleichgewichtssimulationen bestimmt und mit experimentellen Daten verglichen. Mit Hilfe dieser Methode wurden erfolgreich Xylanasen verschiedener Thermostabilität untersucht (Purmonen *et al.* 2007). Auch der Vergleich einer mesophilen und thermophilen Dihydrofolatreduktase (DHFR) gelang mit diesem Verfahren (Sham *et al.* 2002; Pang und Allemann 2007). Bei der zweiten Methode wird eine einzelne Ungleichgewichtssimulation durchgeführt. Dazu wird eine Gleichgewichtssimulation des gefalteten Proteins durch plötzliche Erhöhung der Temperatur gestört. Anschließend wird die Anpassung des Systems an das neue Gleichgewicht beobachtet. Mit Hilfe dieser Methode wurde die Entfaltung einer Cutinase untersucht (Creveld *et al.* 1998). Über den Vergleich mit einer Gleichgewichtssimulation des gefalteten Zustands konnte zwischen flexiblen Regionen, die für die Aktivität des Enzyms wichtig waren, und flexiblen Regionen, von denen die Entfaltung ausging, unterschieden werden. So konnten Schwachstellen (*weak spots*) vorhergesagt werden, deren Mutation die thermische Stabilität, nicht aber die Aktivität beeinflussen dürfte.

MD-Entfaltungssimulationen sind erfolgreich eingesetzt worden, um Bereiche zu identifizieren, von denen der globale Zerfall der Struktur ausgeht (Finkelstein 1997). Der große Nachteil dieser Verfahren besteht aber darin, dass diese Rechnungen (besonders im Fall von Gleichgewichtssimulationen) sehr viel Zeit kosten. Zudem ist unklar, ob die verwendeten Kraftfelder und insbesondere Lösungsmittelmodelle geeignet sind, das Verhalten des Systems bei hohen Temperaturen korrekt wiederzugeben. Eine Alternative zu Dynamiksimulationen stellen Methoden dar, bei denen die Proteinstruktur als ein Netzwerk wechselwirkender Atome oder Reste beschrieben wird. Brinda *et al.* haben untersucht, wie sich die Topologie des Netzwerks verändert, wenn eine Temperaturerhöhung durch das Brechen von Wechselwirkungen simuliert wird (Brinda und Vishveshwara 2005). Mit Hilfe dieser Methoden können zudem die relevante Bewegungsmoden in der Proteinstruktur ausgehend von einer einzelnen Konformation vorhergesagt werden. So haben Su *et al.* das elastische Gauß'sche Netzwerkmodell (GNM) verwendet, um den Einfluss brechender Kontakte auf die Fluktuationen der Reste zu untersuchen (Su *et al.* 2008). Die GNM-Analyse ähnelt einer Normalmodenanalyse (NMA), wobei aber nur die C α -Atome für die Berechnung der Normalmoden berücksichtigt werden. Die Entfaltung wurde simuliert, indem sukzessive native Kontakte gebrochen wurden, deren Distanzfluktuationen am größten waren. Die mit Hilfe von MD-Simulationen oder NMA-ähnlichen Verfahren gemessene Dynamik (Bewegung) ist immer nur ein indirektes Maß für die tatsächlich vorhandene Flexibilität (Beweglich-

keit). Eine direkte Bestimmung rigider und flexibler Regionen einer Proteinstruktur ist von großem Interesse, da mit Hilfe dieser Eigenschaften eine Aussage über die makroskopische und mikroskopische Stabilität der jeweiligen Region oder gesamten Struktur getroffen werden kann, da sie beschreiben, ob ein bestimmter Strukturbereich deformiert werden kann und dann zur Destabilisierung des Proteins beiträgt.

Rader *et al.* haben die Entfaltung von Proteinen mit Hilfe einer bestimmten Art von *bar-and-joint*-Netzwerken simuliert, den *bond-bending*-Netzwerken (Rader *et al.* 2002). Diese Netzwerke bestehen aus Knoten und kovalenten sowie nichtkovalenten Bindungs- und Winkelbeschränkungen und werden daher auch statische Netzwerke (*constraint networks*) genannt. Mit Hilfe der Software FIRST (*floppy inclusion of rigid substructure topology*) (Jacobs *et al.* 2001), die Methoden aus der Rigiditätstheorie implementiert, können rigide Cluster und flexible Regionen in diesen Netzwerken bestimmt werden (Flexibilitätsanalyse). Es wurde untersucht, wie sich die Flexibilität im System mit Abnehmender mittlerer Koordinationszahl (durch eine Reduzierung der Anzahl nichtkovalenter Wechselwirkungen) veränderte. So beobachteten Rader *et al.* während der Entfaltung das Auftreten von rigiden Clustern, die „Faltungskernen“ entsprachen. Diese vorhergesagten Faltungskerne stimmten gut mit experimentell bestimmten Faltungskernen überein (Hespenheide *et al.* 2002; Rader und Bahar 2004).

Auch das in dieser Arbeit vorgestellte rechnerische Verfahren basiert auf der Anwendung von Methoden aus der Rigiditätstheorie auf *bond-bending*- bzw. äquivalente *body-and-bar*-Netzwerke von Proteinen und damit auf einer direkten Bestimmung der Rigidität und Flexibilität in der Struktur. Die Verdünnung des Netzwerks wird hier allerdings so durchgeführt, dass ein Vergleich verschiedener Strukturen auf einer relativen Temperaturskala möglich wird. Zudem werden Analysemethoden aus der Perkulations- und Netzwerktheorie herangezogen, mit denen die Identifizierung eines Phasenübergangs zwischen „gefalteten“ und „ungefalteten“ Netzwerken möglich wird. So können Schmelztemperaturen bestimmt und verglichen werden, d.h. die Thermostabilität von Proteinen kann untersucht werden.

1.4.4 Vorhersage der mikroskopischen Stabilität

Die durch die Mikro stabilität bedingte Flexibilität bestimmter Bereiche der Struktur ist eine Voraussetzung für die Bioaktivität von Enzymen (Henzler-Wildman und Kern 2007). Indem untersucht wird, ob eine thermostabilisierende Mutation die Struktur des aktiven Zentrums oder dessen Flexibilität stört (Korkegian *et al.* 2005), kann die Auswirkung einer thermostabilisierenden Mutation auf die Aktivität vorhergesagt werden.

Mit Hilfe von Methoden zur Beschreibung dynamischer Eigenschaften kann ein Ensemble von Konformationen erzeugt werden, das den nativen Zustand des Proteins darstellt. Anhand dieses konformationellen Ensembles können dann Bereiche identifiziert werden, in denen große Fluktuationen auftreten. Hierzu eignen sich insbesondere MD- oder so genannte Monte-Carlo(MC)-Simulationen sowie geometriebasierte Ansätze, wie beispielweise CONCOORD (de Groot *et al.* 1997). Aus den erzeugten Ensembles kann die „essenzielle Dynamik“ (*essential dynamics*) mit einer Art Hauptkomponentenanalyse (PCA) bestimmt werden (Amadei *et al.* 1993). Es konnte gezeigt werden, dass die Moden der essenziellen Dynamik für die Bioaktivität von Proteinen wichtige Bewegungen beschreiben (Amadei *et al.* 1993; van Aalten *et al.* 1995). Zur Flexibilitätsbestimmung können auch Veränderungen in der Packungsdichte untersucht werden, von denen man weiß, dass sie mit Fluktuationen in Proteinen korreliert sind (Halle 2002).

Relevante Bewegungsmoden können auch mit Hilfe elastischer Netzwerke bestimmt werden (Case 1994; Bahar *et al.* 1997; Ming *et al.* 2002; Tama *et al.* 2002; Ming *et al.* 2003). Yang *et al.* zeigten, dass Reste des aktiven Zentrums mit Hilfe globaler Moden aus dem Gauß'schen Netzwerkmodell identifiziert werden können (Yang und Bahar 2005). Katalytische Reste wiesen eine geringe Beweglichkeit auf, wohingegen an der Substratbindung beteiligte Reste moderat flexibel waren. Auch Luque *et al.* beschrieben diesen „dualen Charakter“ aktiver Zentren von Enzymen (Luque und Freire 2000). Mit Hilfe der Software COREX (Wooll *et al.* 2000), die Methoden aus der statistischen Thermodynamik verwendet, um Konformationen des Ensemble zu erzeugen und gewichten, berechneten sie die mikroskopische Stabilität entlang der Aminosäuresequenz von Enzymen und zeigten, dass das aktive Zentrum sowohl von stabilen, als auch von labilen Resten ausgemacht wird. Der Nachteil der Berechnungen mit COREX besteht hauptsächlich darin, dass die Methode relativ zeitaufwändig ist.

Eine sehr schnelle Identifizierung flexibler Bereiche kann gelingen, wenn Proteine als *bond-bending-* bzw. äquivalente *body-and-bar-*Netzwerke modelliert werden. Die Analyse der Flexibilität in diesen statischen Netzwerken ist mit Hilfe von Methoden aus der Rigiditätstheorie möglich, wie sie in der Software FIRST implementiert sind (Jacobs *et al.* 2001). Gohlke *et al.* haben diesen Ansatz zur Untersuchung von Veränderungen in der Flexibilität von Proteinen bei Komplexbildung verwendet (Gohlke *et al.* 2004). Erst kürzlich wurde der Ansatz auch für die Untersuchung der Flexibilität von Nukleinsäuren adaptiert (Fulle und Gohlke 2008). Auch in dieser Arbeit werden flexible Bereiche von Proteinen mit Hilfe von FIRST bestimmt.

2 Rigidität und Flexibilität in statischen Netzwerken

Makromoleküle können als eine bestimmte Art von *bar-and-joint*-Netzwerken modelliert werden, den *bond-bending*-Netzwerken, wenn die wichtigsten Kräfte in der Struktur Bindungsdehnungskräfte zwischen direkt benachbarten Atomen (so genannte Zentral- oder Bindungskräfte) und Bindungswinkelkräfte (so genannte Winkelkräfte) sind (Hendrickson 1992). Die Netzwerke werden auch statische Netzwerke (*constraint networks*) genannt. Unter anderem können damit glasbildende Materialien (Gläser) und Proteine repräsentiert werden. In diesen Netzwerken werden Atome durch Knoten (*joints*) und Wechselwirkungen zwischen Atomen durch Beschränkungen (*bars*) modelliert. Neben kovalenten und nichtkovalenten Bindungsbeschränkungen werden auch Winkelbeschränkungen in das Netzwerk eingefügt.

Die *bond-bending*-Netzwerke können dann mit Hilfe schneller Verfahren hinsichtlich ihrer Flexibilität analysiert werden, d.h. die Identifizierung rigider Cluster und flexibler Bereiche ist möglich. Dabei kommen Methoden aus der Rigiditätstheorie zum Einsatz. In einem „rigiden“ Bereich eines Netzwerks können einzelne Atome nicht mehr relativ zueinander bewegt werden. Die Atome in einem rigiden Bereich können nur als Ganzes (d.h. als ein Cluster) bewegt werden. In den „flexiblen“ Bereichen des Netzwerks ist die relative Bewegung der Atome zueinander dagegen nicht eingeschränkt.

Für kleine Moleküle können rigide und flexible Bereiche wie folgt identifiziert werden: Ein unbeschränktes Atom im dreidimensionalen (3D) Raum hat drei Freiheitsgrade. N Atome eines N -Rings haben $3N$ Freiheitsgrade abzüglich der N Bindungsbeschränkungen und N Winkelbeschränkungen. Es bleiben N Freiheitsgrade übrig, wovon sechs den makroskopischen, externen Freiheitsgraden (Rotation und Translation) entsprechen. Somit bleiben $N - 6$ interne Freiheitsgrade (so genannte *floppy modes* oder Nullenergiemoden). Die Deformation des Rings über diese (nullfrequenten) Moden ist mit keinem Energieaufwand verbunden. Wenn $N = 6$, dann gibt es keine Nullenergiemoden, und der Ring ist isostatisch rigide. Das Modell sagt somit richtig vorher, dass die Konformationsänderung zwischen der Sessel- und der Wannenkongformation des Sechsrings Energie kostet. Wenn $N > 6$, dann ist der Ring flexibel („unterbeschränkt“). Ein Siebenring hat eine *floppy mode*, und das Modell sagt somit richtig vorher, dass die Konformationsänderung zwischen den beiden möglichen Konformeren keine Energie kostet. Ein Fünfring hat eine Beschränkung mehr als zur Rigidisierung notwendig und

ist somit „überbeschränkt“. In einem überbeschränkten System treten Spannungen (*stress*) auf.

Dieses von Maxwell vorgeschlagene „Zählen von Beschränkungen“ (*constraints counting*) liefert für einfache Systeme die korrekte Anzahl an *floppy modes*, wenn keine redundanten (abhängigen) Beschränkungen im Netzwerk vorhanden sind (Maxwell 1864). Für komplexere Systeme ist es sehr schwierig zu überprüfen, welche Beschränkungen unabhängig voneinander sind. Der Grund dafür besteht darin, dass entfernte Beschränkungen die Rigidität einer lokalen Region beeinflussen können, wenn die Beschränkungen über einen Ring oder mehrere miteinander verbundene Ringe gekoppelt sind. Diese Abhängigkeiten müssen beim *constraints counting* berücksichtigt werden, was bei komplexen Systemen sehr aufwändig ist. Hier können aber Konzepte aus der Graphentheorie angewendet werden. Diese Konzepte werden unter dem Begriff „Rigiditätstheorie“ zusammengefasst.

2.1 Rigiditätstheorie

Obwohl Rigidität bereits von Lagrange (1788) (Lagrange 1788) und Maxwell (1864) (Maxwell 1864) untersucht wurde, kommt sie erst seit rund 40 Jahren in praktischen Fragestellungen vor. Als Begründer der modernen Rigiditätstheorie gilt Laman (Laman 1970), der die Grundlagen geschaffen hat, um die Rigiditätstheorie auf Systeme im zwei- (2D) und später auch im dreidimensionalen (3D) Raum anzuwenden (Jacobs und Thorpe 1995; Jacobs und Hendrickson 1997). Seinen Arbeiten folgte der Entwurf von Algorithmen, mit denen Systeme mit mehreren Millionen Knoten innerhalb weniger Sekunden auf ihre Rigiditätseigenschaften untersucht werden können (Jacobs 1998). Ursprünglich wurde die Rigiditätstheorie dazu benutzt, die Stabilität von Bauwerken vorherzusagen, ohne aufwändige Simulationen durchführen zu müssen. Heute dient die Theorie vor allem der Vorhersage physikalischer Eigenschaften von Gläsern, Proteinen und anderen Materialien.

2.1.1 Das Maxwell'sche *constraints counting*

Wenn man davon ausgeht, dass es in einem *bar-and-joint*-Netzwerk keinerlei abhängige (redundante) Beschränkungen gibt, dann genügt es die Anzahl der Beschränkungen zu zählen und von der Anzahl der internen Freiheitsgrade abzuziehen, um die Anzahl der *floppy modes* zu erhalten (siehe oben). Dieses von Maxwell vorgeschlagene „Zählen von Beschränkungen“ (*constraints counting*) liefert schnelle Information über den globalen Zustand des Systems (Maxwell 1864). Ursprünglich wurde der Ansatz

dazu benutzt, um die Stabilität von Brückenbauwerken zu untersuchen. Eine Brücke (oder ein Gerüst) ist nur dann stabil, wenn keine *floppy modes* im System vorhanden sind. Phillips (Phillips 1979; Phillips 1981) und Thorpe (Thorpe 1983) verwendeten den Ansatz, um die ideale Koordination von Atomen zur Glasbildung vorherzusagen. Für Systeme mit redundanten Beschränkungen ist die Anwendung des *constraint counting* allerdings problematisch. Redundante Beschränkungen ändern nicht die Anzahl der *floppy modes* oder die Konfiguration der rigiden Cluster.

In einem völlig unbeschränkten Netzwerk im 3D-Raum mit N Knoten beträgt die Anzahl der Nullenergiemoden $F = 3N - 6$. Fügt man diesem Netzwerk N_c Beschränkungen hinzu, dann beträgt die Anzahl der Nullenergiemoden $F = 3N - 6 - N_c$, wenn keine der hinzugefügten Beschränkungen redundant ist. Pro internem Freiheitsgrad beträgt die Anzahl der Nullenergiemoden in diesem Fall $f = (3N - 6 - N_c) / 3N$. Geht man aber davon aus, dass N_r der hinzugefügten Beschränkungen redundant sind, dann erhöht sich die Anzahl der Nullenergiemoden auf $F = 3N - 6 - N_c + N_r$. In der Praxis bringt diese Formel aber keinen Nutzen, da die Anzahl der redundanten Beschränkungen N_r nicht ermittelt werden kann ohne das Eigenwert-Problem zu lösen. Da N_r stets größer oder gleich Null ist, liefert das Zählen von Beschränkungen aber eine untere Schranke für die Anzahl der *floppy modes*.

2.1.2 Der *pebble game*-Algorithmus im 2D-Raum

Für generische *bar-and-joint*-Netzwerke im 2D-Raum kann mit Hilfe eines Theorems von Laman (Laman 1970) festgestellt werden, ob redundante Beschränkungen im Netzwerk vorhanden sind. In generischen Netzwerken gibt es keine geometrischen Singularitäten, die auftreten können, wenn Beschränkungen parallel oder kollinear gesetzt werden. Das Theorem von Laman besagt, dass ein generisches Netzwerk im 2D-Raum genau dann keine redundanten Beschränkungen hat, wenn kein Teilnetzwerk von n Knoten (mit $n \geq 2$) und b Beschränkungen die Beziehung $b \leq 2n - 3$ verletzt (die so genannte Laman-Bedingung). Von diesem Theorem ausgehend hat Hendrickson einen Algorithmus vorgeschlagen, mit dem für Netzwerke im 2D-Raum die exakte Anzahl der *floppy modes* bestimmt werden kann und mit dem überbeschränkte (flexible) Bereiche sowie rigide Cluster identifiziert werden können (Hendrickson 1992).

Der Grundgedanke des letztendlich von Jacobs und Thorpe entwickelten *pebble game*-Algorithmus besteht darin (Jacobs und Thorpe 1995; Jacobs und Hendrickson 1997), das Theorem von Laman rekursiv anzuwenden, während das Netzwerk Beschränkung für Beschränkung konstruiert wird. Wegen der Rekursion müssen immer

nur die Teilnetzwerke untersucht werden, welche die neu hinzugefügte Beschränkung enthalten. Wenn jedes der untersuchten Teilnetzwerke die Laman-Bedingung erfüllt, dann ist die hinzugefügte Beschränkung unabhängig. Indem auf diese Weise die Anzahl der redundanten Beschränkungen gezählt wird, kann die exakte Anzahl der Nullenergiemoden bestimmt werden. Überbeschränkte Bereiche des Netzwerks werden identifiziert, sobald eine redundante Beschränkung gefunden wird. Die rigiden Cluster werden bestimmt, nachdem das Netzwerk vollständig rekonstruiert wurde. Zuerst werden dazu isolierte Knoten identifiziert. Anschließend wird die Rigidität der übrigen Knoten bezüglich einer Referenzbeschränkung zwischen zwei Referenzknoten bestimmt. Falls eine Beschränkung zwischen einem der beiden Referenzknoten und einem zu testenden Knoten abhängig ist, dann befindet sich diese Beschränkung im selben rigiden Cluster wie die Referenzbeschränkung.

2.1.3 Rigidität in Netzwerken im 3D-Raum

Das *pebble game* kann nicht auf allgemeine *bar-and-joint*-Netzwerke im 3D-Raum erweitert werden, da der Algorithmus auf dem Theorem von Laman beruht, das nur im 2D-Raum gilt (Jacobs 1998). Der Grund dafür, dass das Theorem nicht für den 3D-

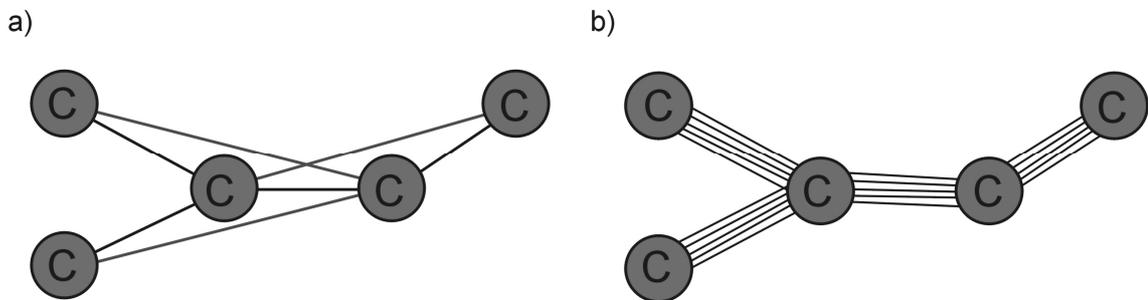


Abbildung 4. Beispiel für die Modellierung eines einfachen Moleküls im 2D-Raum als *bond-bending*-Netzwerk (a) und als äquivalentes *body-and-bar*-Netzwerk (b). Im *bond-bending*-Netzwerk (a) müssen Bindungs- (schwarze Linien) und Winkelbeschränkungen (graue Linien) eingefügt werden. Im *body-and-bar*-Netzwerk müssen nur Bindungsbeschränkungen eingefügt werden (b), die in Form mehrerer Verstrebungen (*bars*) modelliert werden.

Raum generalisiert werden kann, ist die Existenz so genannter bananenförmiger Netzwerke (Jacobs 1998). Auf der Grundlage von Arbeiten von Tay und Whiteley (Tay und Whiteley 1985) konnte Jacobs aber zeigen, dass das *pebble game* auf eine bestimmte Art der *bar-and-joint*-Netzwerke im 3D-Raum angewendet werden kann, die so genannten *bond-bending*-Netzwerke (Jacobs 1998). Lange Zeit gab es keinen mathematischen Beweis, der die Anwendbarkeit eines Theorems ähnlich dem von Laman für *bond-*

bending-Netzwerke im 3D-Raum begründete. Trotzdem ist diese Vermutung (die so genannte *molecular framework conjecture*) trotz zahlreicher Arbeiten nie widerlegt worden (Jacobs 1998). Erst kürzlich konnte die Vermutung bewiesen werden (Naoki und Shin-Ichi 2009).

Die Anwendung des *pebble game*-Algorithmus auf *bond-bending*-Netzwerke im 3D-Raum ist analog zur Anwendung auf allgemeine *bar-and-joint*-Netzwerke im 2D-Raum, aber wesentlich komplizierter, weswegen an dieser Stelle nicht weiter darauf eingegangen werden soll. Interessant ist aber die Tatsache, dass es einen zu den *bond-bending*-Netzwerken äquivalenten Typ von Netzwerken gibt (Jacobs 1998), auf die eine deutlich einfachere Version des *pebble game* angewendet werden kann. Es handelt sich dabei um so genannte *body-and-bar*-Netzwerke (Whiteley 1979; Moukarzel 1996). In Abbildung 4 wird das *bond-bending*-Netzwerk eines einfachen Moleküls mit dem *body-and-bar*-Netzwerk des selben Moleküls verglichen. Ein formaler Vorteil der *body-and-bar*-Netzwerken besteht offensichtlich darin, dass nur noch Bindungsbeschränkungen verwendet werden müssen. Im *body-and-bar*-Netzwerk werden Atome als Körper (*bodies*) modelliert, die jeweils sechs Freiheitsgrade besitzen. Bindungen zwischen den Atomen werden in Form mehrerer Verstrebungen (*bars*) repräsentiert, die jeweils eine Beschränkung darstellen. Für eine drehbare Bindung werden fünf Verstrebungen eingefügt, für eine nicht drehbare Bindung sechs.

2.2 Rigiditätsperturbation

Eine klassische Anwendung der Rigiditätstheorie ist die Suche nach guten glasbildenden Materialien (Gläsern). Man untersucht Materialien des Typs $\text{Ge}_x\text{As}_y\text{Se}_{1-x-y}$, in denen Ge für jedes vierfach koordinierte Atom stehen kann, As für jedes dreifach koordinierte Atom und Se für jedes zweifach koordinierte Atom. Mit Hilfe der Rigiditätstheorie soll nun herausgefunden werden, welche Kombinationen von x und y ein gutes Glas ergeben. Um das herauszufinden, wird die Rigidität des Materials als Funktion der mittleren Koordinationszahl $\langle r \rangle$ untersucht. Verschiedene Kombinationen von x und y ergeben verschiedene mittlere Koordinationszahlen. Die wichtigsten Kräfte in solchen Materialien sind Bindungs- und Winkelkräfte, weshalb sie als *bond-bending*- bzw. äquivalente *body-and-bar*-Netzwerke modelliert werden können. Diese wiederum können mit dem *pebble game*-Algorithmus im 3D-Raum untersucht werden. Man geht davon aus, dass die Bindungswinkelnetzwerke für diese Materialien kontinuierlich zufällig (*continuous random*) sind, was die Analyse vorerst vereinfacht. Darin unterscheiden

sich Gläser beispielsweise von Quarz, einem Kristall, für das sowohl eine Nah- als auch eine Fernordnung gewährleistet ist, und das deswegen als ein regelmäßiges Netzwerk modelliert werden muss. Kontinuierlich bedeutet in diesem Fall, dass es innerhalb des Netzwerks keine identifizierbaren Grenzen gibt, die Regionen verschiedenen Typs von Struktur unterscheiden. Solche Grenzen würden das Material spröde machen.

Geht man von einem Material mit sehr hoher mittlerer Koordinationszahl aus, so ist das Netzwerk wegen der großen Zahl an Beschränkungen vollständig rigide, d.h. es befindet sich in der „rigiden Phase“. Für ein Material mit sehr niedriger Koordinationszahl ist das Netzwerk dagegen vollständig flexibel, d.h. es befindet sich in der „flexiblen Phase“. Der kritische Wert für die mittlere Koordinationszahl, bei der das Netzwerk erstmalig in die flexible Phase übergeht, kann nun bestimmt werden. Die kritische mittlere Koordinationszahl $\langle r \rangle_c$ nennt man auch Rigiditätspertikulationsschwellenwert, da bei diesem Wert die Rigidität nicht mehr das gesamte Netzwerk durchzieht („perkoliert“). Untersuchungen der Rigiditätspertikulation von Netzwerken glasbildender Materialien haben gezeigt, dass das System den Phasenübergang von rigide nach flexibel bei einem $\langle r \rangle_c$ von ungefähr 2,4 hat. Materialien mit einer mittleren Koordinationszahl nur unwesentlich größer als $\langle r \rangle_c$ bilden die besten Gläser (Thorpe 1983). Bei diesen Materialien ist die Anzahl der Beschränkungen gerade groß genug, um alle internen Freiheitsgrade auszugleichen. Wäre die Anzahl der Beschränkungen größer, dann wäre das Material zwar immer noch rigide, aber die hohe Energie durch die Überbeschränkung des Netzwerks würde dazu führen, dass sich die Atome neu anordnen würden, was die Kristallisation des Materials zur Folge hätte.

2.3 Anwendung auf Proteine

Auch die Flexibilität von Proteinstrukturen kann mit Hilfe der Rigiditätstheorie analysiert werden (Whiteley 2005). Zur Modellierung von Proteinen als *bond-bending*- bzw. äquivalente *body-and-bar*-Netzwerke müssen neben kovalenten Bindungen und Bindungswinkeln allerdings auch wichtige nichtkovalente Wechselwirkungen als Beschränkungen modelliert werden. Letztere umfassen H- und Salzbrücken sowie hydrophobe Kontakte (siehe oben). Die ursprüngliche Version des *pebble game*-Algorithmus erlaubt aber nur zwei Typen von Bindungen zu modellieren, drehbare und nicht drehbare Bindungen. Diese werden im *body-and-bar*-Netzwerk durch das Einsetzen von fünf bzw. sechs Verstrebungen zwischen zwei Körpern repräsentiert. Diese Repräsentation ist ausreichend für die Modellierung kovalent verknüpfter Materialien, ist aber unzurei-

chend, wenn auch schwächere (H- und Salzbrücken) und weniger spezifische Wechselwirkungen (hydrophobe Bindungen) modelliert werden sollen.

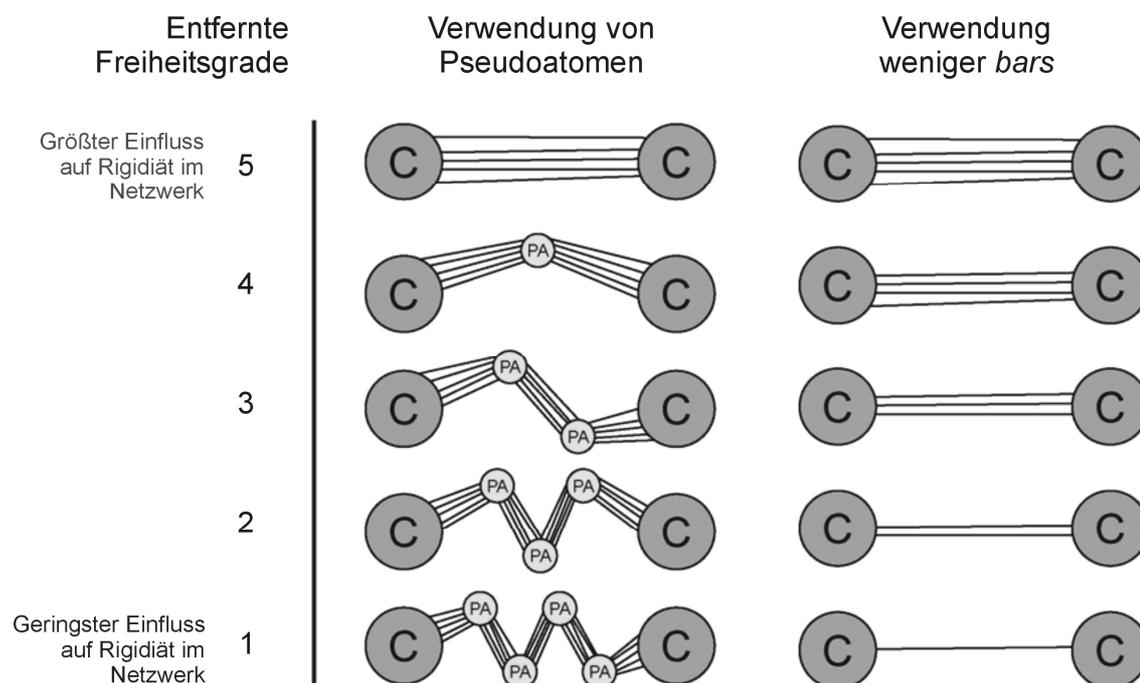


Abbildung 5. Schwächere, weniger spezifische Wechselwirkungen können im *body-and-bar*-Netzwerk durch die Verwendung von Pseudoatomen (PA) oder einer geringeren Anzahl an Verstrebungen (*bars*) modelliert werden. Nach (Chubynsky *et al.* 2004).

Der *pebble game*-Algorithmus für Proteine ist in der Software FIRST (*floppy inclusion of rigid substructure topology*) implementiert (Jacobs *et al.* 2001). In der ersten Version wurden die Strukturen als *bond-bending*-Netzwerke repräsentiert (Jacobs *et al.* 2001), später als *body-and-bar*-Netzwerke (Rader *et al.* 2002). Im *body-and-bar*-Netzwerk wurden die schwächeren Wechselwirkungen ursprünglich modelliert, indem eine einzelne (direkte) Beschränkung aus fünf Verstrebungen durch eine indirekte Beschränkung ersetzt wurde, die über mehrere so genannte Pseudoatome geführt wurde (Rader *et al.* 2002). Der Sinn dieses Ansatzes lag darin, dass jedes Pseudoatom einen internen Freiheitsgrad mehr in das System bringt als von den Beschränkungen ausgeglichen werden kann. Der Einfluss einer Kette aus Pseudoatomen auf ein System ist in Abbildung 5 gezeigt. Eine einzelne Beschränkung aus fünf Verstrebungen zwischen einem Paar von Atomen (wie sie zur Modellierung einer drehbaren kovalenten Bindung verwendet wird) verringert die Anzahl der internen Freiheitsgrade um fünf. Eine indirekte Beschränkung, die über vier Pseudoatome geführt wird, verringert die Anzahl der internen Freiheitsgrade nur um eins, da jede der fünf Beschränkungen die Anzahl der

internen Freiheitsgrade zwar um fünf verringert, jedes Pseudoatom aber sechs zusätzliche interne Freiheitsgrade ins System einführt ($4 \cdot 6 - 5 \cdot 5 = -1$).

Pseudoatome erhöhen die Größe des Systems und damit die Laufzeit des *pebble-game*-Algorithmus. Die *worst-case*-Laufzeit des Algorithmus ist in $O(n^2)$, wobei n die Anzahl der Atome beschreibt (Jacobs *et al.* 2001). Durchschnittlich liegt die Laufzeit in $O(n)$ und nimmt somit linear mit der Größe des Systems zu (Jacobs *et al.* 2001). In späteren Versionen von FIRST wurde aus diesem Grund auf die Verwendung von Pseudoatomen verzichtet. Diese Implementierung erlaubt dagegen das Einfügen von Beschränkungen, die aus weniger als fünf Verstrebungen bestehen (Abbildung 5). Die Anzahl der Verstrebungen pro Beschränkung entsprechen direkt der Anzahl der internen Freiheitsgrade, die aus dem System entfernt werden. Diese Implementierung ist äquivalent zur Modellierung des Systems mit Pseudoatomen (Abbildung 5). Ihr Vorteil besteht darin, dass sie schneller, aber auch intuitiver ist. Diese Implementierung ist in der Version 6.2 der Software FIRST verfügbar.

2.3.1 Modellierung der einzelnen Interaktionen

Netzwerke von Proteinen unterscheiden sich von denen von Gläsern darin, dass neben kovalenten Bindungen auch H- und Salzbrücken sowie hydrophobe Kontakte modelliert werden müssen. H- und Salzbrücken werden aufgrund ihrer Stärke und Direktionalität genau so wie drehbare kovalente Bindungen modelliert. Hydrophobe Kontakte dagegen führen weniger Beschränkungen in das System ein.

2.3.1.1 H- und Salzbrücken

H- und Salzbrücken spielen aufgrund ihrer Stärke und Direktionalität eine bedeutende Rolle bei der Ausbildung von Sekundärstrukturelementen und bei der Stabilisierung der Tertiärstruktur eines Proteins (Kap. 1.2.2). Die Stärke dieser Wechselwirkungen kann stark variieren. So gibt es in einer einzigen Proteinstruktur ein ganzes Spektrum von H- und Salzbrücken unterschiedlicher Stärke (Fersht 1987; Jeffrey 1997). Je nachdem, bei welcher Temperatur die Rigiditätsanalyse durchgeführt wird, werden nur H- und Salzbrücken einer bestimmten Mindeststärke ($E \text{ kcal mol}^{-1}$) bei der Konstruktion des Netzwerks berücksichtigt. Da die Stärke der einzelnen Wechselwirkungen experimentell kaum bestimmt werden kann, muss die Energie einer jeden H- und Salzbrücken mit Hilfe rechnerischer Methoden vorhergesagt werden.

Zur Identifizierung und Bewertung der H- und Salzbrücken wird in der Software FIRST (Version 6.2) ein empirisches Potential verwendet, bei dem die Energie einer H-

Brücke als van der Waals-Wechselwirkung approximiert wird, die durch Distanz- und Winkelterme gewichtet wird. Dazu muss zuerst der Hybridisierungszustand und die Funktion (Donor oder Akzeptor) für jedes Stickstoff-, Sauerstoff- und Schwefelatom gemäß einer Arbeit von Stickle *et al.* bestimmt werden (Stickle *et al.* 1992). Zur späteren Identifikation von Salzbrücken werden das Stickstoffatom in der Hauptkette des aminoterminalen Rests sowie protonierte Stickstoffe der Seitenketten von Arginin, Lysin und Histidin als geladene Donoren berücksichtigt, ebenso wie das Sauerstoffatom in der Hauptkette des carboxyterminalen Rests sowie deprotonierte Sauerstoffe der Seitenketten von Glutamat und Aspartat als geladene Akzeptoren.

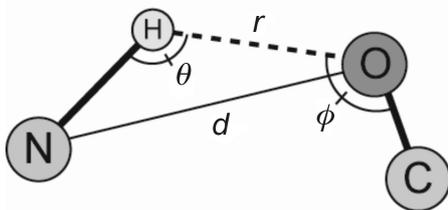


Abbildung 6. H- und Salzbrücken werden bestimmt, indem zuerst alle Donor-Akzeptor-Paare ermittelt werden, deren geometrische Eigenschaften bestimmte Distanz- und Winkelschwellenwerte unter- oder überschreiten. Die N-H-Gruppe stellt einen typischen Donor, das O-Atom einen Akzeptor dar. Die geometrischen Eigenschaften sind der Abstand zwischen dem Wasserstoffatom und dem Akzeptor (r), der Abstand zwischen dem Donor und dem Akzeptor (d), der Winkel zwischen dem Donor, dem Wasserstoffatom und dem Akzeptor (θ) und der Winkel zwischen dem Wasserstoffatom, dem Akzeptor und dem an den Akzeptor gebundenen Atom (ϕ). Die selben geometrischen Eigenschaften werden anschließend dazu verwendet, die Energie der H- oder Salzbrücke quantitativ zu bestimmen.

In einem ersten Schritt (*prescreening*) werden alle Donor-Akzeptor-Paare ermittelt, deren geometrische Eigenschaften bestimmte Distanz- und Winkelschwellenwerte unter- oder überschreiten, die sie als H- oder Salzbrücke qualifizieren. Diese geometrischen Eigenschaften sind der Abstand zwischen dem Wasserstoffatom und dem Akzeptor (r), der Abstand zwischen dem Donor und dem Akzeptor (d), der Winkel zwischen dem Donor, dem Wasserstoffatom und dem Akzeptor (θ) und der Winkel zwischen dem Wasserstoffatom, dem Akzeptor und dem an den Akzeptor gebundenen Atom (ϕ) (Abbildung 6). Generell muss für alle H- und Salzbrücken $d \leq 5$, $r \leq 4$ und $\theta \geq 100^\circ$ gelten. Außerdem muss für Salzbrücken $\phi > 80^\circ$, für H-Brücken mit einem sp^2 -Akzeptor $\phi > 90^\circ$ und für H-Brücken mit einem sp^3 -Donor und einem sp^3 -Akzeptor $\phi - 109,5^\circ \leq 90^\circ$ gelten. Für H-Brücken mit einem sp^2 -Donor und einem sp^3 -Akzeptor muss kein bestimmter Wert für ϕ erfüllt werden.

In einem zweiten Schritt wird die Energie einer H- (E_{HB}) oder Salzbrücke (E_{SB}) mit Hilfe eines empirischen Potentials nach Dahiyat *et al.* (Dahiyat *et al.* 1997) quantitativ bestimmt. Dabei ist

$$E_{HB} = V_{HB} \left\{ 5 \cdot \left(\frac{R_{HB}}{d} \right)^{12} - 6 \cdot \left(\frac{R_{HB}}{d} \right)^{10} \right\} \cdot F(\theta, \phi, \gamma) \quad (4)$$

und

$$E_{SB} = V_{SB} \left\{ 5 \cdot \left(\frac{R_{SB}}{d+a} \right)^{12} - 6 \cdot \left(\frac{R_{SB}}{d+a} \right)^{10} \right\} \quad (5)$$

mit

$$F = \cos^2 \theta \cdot e^{-(\pi-\theta)^6} \cdot \cos^2(\phi - 109,5^\circ) \quad \text{für } sp^3\text{-Donor und } sp^3\text{-Akzeptor} \quad (6)$$

$$F = \cos^2 \theta \cdot e^{-(\pi-\theta)^6} \cdot \cos^2(\phi) \quad \text{für } sp^3\text{-Donor und } sp^2\text{-Akzeptor} \quad (7)$$

$$F = \cos^4 \theta \cdot e^{-(\pi-\theta)^6} \quad \text{für } sp^2\text{-Donor und } sp^3\text{-Akzeptor} \quad (8)$$

$$F = \cos^2 \theta \cdot e^{-(\pi-\theta)^6} \cdot \cos^2(\phi - 109,5^\circ) \quad \text{für } sp^2\text{-Donor und } sp^2\text{-Akzeptor} \quad (9)$$

und $V_{HB} = 8 \text{ kcal mol}^{-1}$, $R_{HB} = 2,8 \text{ \AA}$, $V_{SB} = 10 \text{ kcal mol}^{-1}$, $R_{SB} = 3,2 \text{ \AA}$ und $a = 0,375 \text{ \AA}$, wobei γ dem Winkel zwischen den Normalen der Ebenen entspricht, die durch zwei sp^2 -Zentren beschrieben werden. Die Formeln zur Bestimmung der Energie einer H- oder Salzbrücke wurden aus dem Quellcode der Software FIRST (Version 6.2) abgeleitet. Sie unterscheiden sich geringfügig von den im Handbuch zur Software FIRST angegebenen Formeln.

Eine H- oder Salzbrücke wird nur berücksichtigt, wenn ihre Energie kleiner als ein bestimmter Energieschwellenwert (E_{CUT}) ist. Somit werden nur H- und Salzbrücken eines bestimmten Mindeststärke zur Konstruktion des Netzwerks herangezogen. Die resultierenden Beschränkungen werden jeweils indirekt über ein Pseudoatom modelliert, oder, wie in der neuesten Version der Software FIRST (Version 6.2) implemen-

tiert, direkt durch vier Verstreben. Somit verringert jede berücksichtigte H- oder Salzbrücke die Anzahl der internen Freiheitsgrade des Systems um vier. Diese Parametrisierung wurde festgelegt, da sie für verschiedene Proteine die beste Korrelation der Flexibilität im nativen Zustand und der Lokalisierung von Faltungskernen mit dem Experiment ergab (Hespenheide *et al.* 2002), was in späteren Arbeiten bestätigt wurde (Rader und Bahar 2004).

2.3.1.2 Hydrophobe Kontakte

Hydrophobe Wechselwirkungen machen einen großen Beitrag zur Stabilität der Proteinstruktur und sind ein kritischer Faktor bei der Proteinfaltung (Kap. 1.2.2). In der allerersten Version der Software FIRST wurden hydrophobe Wechselwirkungen nicht in die Netzwerke mit einbezogen (Jacobs *et al.* 2001). Später wurden sie dann mit einbezogen, wobei aber ihre Schwäche und Unspezifität berücksichtigt wurde. Hydrophobe Kontakte zwischen Kohlenstoff- oder Schwefelatomen werden berücksichtigt, wenn der Abstand zwischen diesen Atomen kleiner als die Summe ihrer van der Waals-Radien (C: 1,7 Å, S: 1,8 Å) plus einem bestimmten Distanzschwellenwert (c_{CUT}) ist (Hespenheide *et al.* 2002; Rader *et al.* 2002). Die resultierenden Beschränkungen werden jeweils indirekt über drei Pseudoatome modelliert, oder direkt durch zwei Verstreben. Somit verringert jeder berücksichtigte hydrophobe Kontakte die Anzahl der internen Freiheitsgrade des Systems um zwei. Diese Parametrisierung wurde festgelegt, da sie für verschiedene Proteine die beste Korrelation der Flexibilität im nativen Zustand und der Lokalisierung von Faltungskernen mit dem Experiment ergab (Hespenheide *et al.* 2002; Rader und Bahar 2004).

2.3.2 Simulierte thermische Entfaltung von Proteinen

Die Rigiditätsperkolation von Proteinnetzwerken kann analog zur Rigiditätsperkolation von Netzwerken glasbildender Materialien untersucht werden. In Kap. 4.3 wird dies detailliert erläutert. Anders als bei den Netzwerken glasbildender Materialien wird die mittlere Koordinationszahl dabei aber nicht über die Variation der chemischen Zusammensetzung des Materials verändert, sondern über eine Reduzierung der Anzahl nichtkovalenter Wechselwirkungen. So kann die thermische Entfaltung der Struktur von Proteinen simuliert werden.

3 Aufgabenstellung

Hilfreich zur Unterstützung des Protein Engineering sind rechnerische Methoden, welche einerseits die thermische Stabilität einer Struktur sowie den Effekt einer Mutation abschätzen können, andererseits Hinweise darauf geben, in welchen Bereichen der Struktur thermostabilisierende Mutationen eingeführt werden können. Dies erlaubt das Testen von Mutanten sowie den Aufbau zielgerichteter Bibliotheken. Die Methoden sollten außerdem in der Lage sein die Auswirkung einer thermostabilisierenden Mutation auf die Aktivität vorherzusagen.

In dieser Arbeit soll ein Verfahren entwickelt werden, mit dem die Entfaltung von Proteinen simuliert werden kann. Anhand der Simulation soll die Schmelztemperatur bestimmt und damit die Thermostabilität einer Struktur untersucht werden können. Außerdem soll die Untersuchung struktureller Veränderungen möglich sein, die während der Entfaltung auftreten und letztendlich zum globalen Zerfall der Struktur führen. Die Bereiche, von denen der globale Zerfall der Struktur ausgeht, sollen bestimmt werden können. Es soll untersucht werden, inwieweit diese Entfaltungsregionen Strukturbereiche darstellen, deren Mutation die thermische Stabilität der Struktur beeinflusst. Das entwickelte Verfahren soll auf der Anwendung von Methoden aus der Rigiditätstheorie auf *bond-bending-* bzw. äquivalente *body-and-bar-*Netzwerke (*constraint networks*) von Proteinen basieren, was die Bestimmung rigider Cluster und flexibler Regionen in atomarer Auflösung erlaubt (Flexibilitätsanalyse). Die thermische Entfaltung der Strukturen soll über das sukzessive Aufbrechen nichtkovalenter Wechselwirkungen in den Netzwerken simuliert werden. Ansätze aus der Perkulations- und Netzwerktheorie sollen verwendet werden, um die Rigidität und Flexibilität in den Netzwerken während der Entfaltung zu untersuchen.

Diese unter dem Begriff der Analyse statischer Netzwerke (*constraint network analysis*, CNA) zusammengefassten Methoden sollen im ersten Teil dieser Arbeit auf einen Datensatz homologer meso- und thermophiler Proteine angewendet werden. Dabei soll untersucht werden, ob die thermophile Anpassung tatsächlich über eine Rigidisierung der Struktur erfolgt. Außerdem soll die Theorie der korrespondierenden Zustände getestet werden, die besagt, dass bei der thermophilen Anpassung trotz globaler Rigidisierung für die Bioaktivität wichtige flexible Bereiche konserviert sind. Mit Hilfe der CNA sollen die Entfaltungsregionen der Proteine aus dem Datensatz bestimmt und mit Regionen verglichen werden, in die thermostabilisierende Mutationen eingeführt wurden.

Darüber hinaus soll untersucht werden, inwieweit vorhergesagt werden kann, ob eine möglicherweise thermostabilisierende Mutation die Aktivität negativ beeinflusst.

Im zweiten Teil soll die CNA auf eine Serie von Phytasen unterschiedlicher Thermostabilitäten angewendet werden. Da es sich bei den Phytase-Strukturen um Homologiemodelle handelt, die nicht direkt mit Hilfe der Analyse statischer Netzwerke untersucht werden können, soll eine ensemblebasierte CNA etabliert werden. Dazu sollen kurze MD-Simulationen zur Verbesserung der homologiemodellierten Strukturen durchgeführt werden, aus denen dann ein konformationelles Strukturensemble extrahiert werden kann. Aus den konformationellen Ensembles sollen Thermostabilitäten vorhergesagt werden. Zudem sollen Wahrscheinlichkeiten berechnet werden, mit denen bestimmte Strukturbereiche Entfaltungsregionen darstellen, in denen Mutationen die thermische Stabilität der Struktur beeinflussen. Es soll gezeigt werden, dass mit der Entwicklung der ensemblebasierten CNA die methodischen Grundlagen für den Einsatz des Verfahrens zur Unterstützung des Protein Engineering von Phytasen geschaffen wurden.

4 Theorie und Methoden

Im folgenden werden die unter dem Begriff der Analyse statischer Netzwerke (*constraint network analysis*, CNA) zusammengefassten Methoden zur Untersuchung der Stabilität und der Flexibilität von Proteinen beschrieben. In Kap. 4.1 wird erläutert, wie die Proteine nach entsprechender Vorbereitung als *bond-bending*- bzw. äquivalente *body-and-bar*-Netzwerke modelliert wurden. Die mit Hilfe der Software FIRST (Jacobs *et al.* 2001) durchgeführten Flexibilitätsanalysen der Netzwerke werden in Kap. 4.2 beschrieben. In Kap. 4.3 wird erläutert, wie durch das sukzessive Aufbrechen nichtkovalenter Wechselwirkungen in den Netzwerken die thermische Entfaltung der Proteinstrukturen simuliert wurde. Verschiedene Parameter und Methoden zur Analyse der thermischen Entfaltungssimulationen werden in Kap. 4.4 beschrieben. In Kap. 4.5 wird der Ansatz der ensemblebasierten CNA erläutert. In Kap. 4.6 wird beschrieben, wie die für die ensemblebasierte CNA benötigten Strukturensamples mit Hilfe von MD-Simulationen erzeugt wurden. In Kap. 4.7 werden die Datensätze vorgestellt, auf denen die CNA bzw. ensemblebasierte CNA angewendet wurde. Abschließend findet sich in Kap. 4.8 eine Beschreibung der in dieser Arbeit entwickelten Computerprogramme.

4.1 Konstruktion der Netzwerke

Die Flexibilität von Proteinstrukturen kann mit Hilfe der Rigiditätstheorie analysiert werden. Dazu werden die Proteine nach entsprechender Vorbereitung als *bond-bending*- bzw. äquivalente *body-and-bar*-Netzwerke modelliert, wobei neben kovalenten Bindungen und Bindungswinkeln auch wichtige nichtkovalente Wechselwirkungen als Beschränkungen modelliert werden müssen. Im *body-and-bar*-Netzwerk werden Atome als Körperchen (*bodies*) modelliert, die jeweils sechs Freiheitsgrade besitzen. Bindungen zwischen den Atomen werden in Form mehrerer Verstrebungen (*bars*) repräsentiert, die jeweils eine Beschränkung darstellen. Für eine drehbare kovalente Bindung werden fünf Verstrebungen eingefügt, für eine nicht drehbare Bindung sechs. Für eine H- oder Salzbrücke werden vier Verstrebungen eingefügt, für hydrophobe Kontakte dagegen nur zwei. In der vorliegenden Arbeit wurden die Proteinnetzwerke mit Hilfe der Software FIRST (Version 6.2) (Jacobs *et al.* 2001) konstruiert. Als Eingabe benötigt FIRST eine Proteinstruktur im Proteindatenbank(PDB)-Format.

Die Proteinstrukturen wurden durch manuelles Editieren der PDB-Datei für die Konstruktion des Netzwerks vorbereitet. Basierend auf bereits veröffentlichten Studien,

die einen vernachlässigbaren Einfluss struktureller Wassermoleküle auf die mit Hilfe der Rigiditätstheorie vorhergesagte Flexibilität von Proteinen (Hespenheide *et al.* 2002) und einem Protein-Protein-Komplex (Gohlke *et al.* 2004) zeigten, wurden alle Wassermoleküle entfernt. Ebenso werden auch Lösungsmittel-, Puffer-, Substrat- und Kofaktormoleküle entfernt. Von der Proteinstruktur koordinierte Metallionen wurden bei der Konstruktion des Netzwerks dagegen berücksichtigt, da sie (wie im Fall thermolysinähnlicher Protease) eine sehr wichtige Rolle für die Stabilität der Struktur spielen können. Im Fall von Kristallstrukturen wurde die Qualität der Proteinstrukturen mit Hilfe des PDBREPORT-Servers (Hooft *et al.* 1996) überprüft. Zudem wurden nur Kristallstrukturen einer Auflösung von unter 2,6 Å verwendet. Wasserstoffatome wurden mit Hilfe der Software REDUCE (Word *et al.* 1999) hinzugefügt. Diese Software wurde zudem verwendet, um die Ausrichtung von Aspartat- (Asp), Glutamat- (Glu) und Histidinseitenketten (His) zu überprüfen und gegebenenfalls zu ändern.

Energieminimierungen einzelner Proteinstrukturen wurden mit Hilfe des SANDER-Moduls aus dem AMBER-Programmpakets (Version 9) (Case *et al.* 2005) unter Verwendung des Kraftfelds parm99SB (Hornak *et al.* 2006) durchgeführt. Fehlende Atome wurden mit dem TLEAP-Modul hinzugefügt. Der Einfluss des Lösungsmittels wurde implizit mit Hilfe des generalisierten Born-Modells von Onufried, Bashford und Case (GB^{OBC}) (Onufriev *et al.* 2004) unter Verwendung der *mbondi2* GB-Radien berücksichtigt. Während der Minimierung wurden die nichtbindenden Wechselwirkungen auf einen Bereich von 16 Å beschränkt. Die Atom- und Restnamen der mit Software REDUCE protonierten Ausgangsstrukturen wurden mit Hilfe eines Perl-Skripts (`reduce2amber.pl`) an die von AMBER verwendeten Bezeichnungen angepasst. Die Minimierung wurde in 500 Schritten durchgeführt, wobei zuerst die Methode des steilsten Abstiegs (*steepest descent*) und dann die des konjugierten Gradienten (*conjugated gradient*) verwendet wurde. Nach der Minimierung wurden die AMBER-spezifischen Restbezeichnungen (Hie, Hid, Hip und Cyx) wieder in das PDB-Format (His und Cys) übertragen.

Generell wurden nur globuläre Proteine zur Konstruktion der Netzwerke verwendet, da die Standardbehandlung von H- und Salzbrücken sowie hydrophoben Kontakten nur für lösliche Proteine Gültigkeit hat (Jacobs 1998; Rader *et al.* 2002). Membranproteine können zwar ebenfalls untersucht werden, allerdings erfordert dies gegebenenfalls die manuelle Editierung des Netzwerks. Die Software FIRST identifiziert H- und Salzbrücken sowie hydrophobe Kontakte wie in Kap. 2.3.1 beschrieben. Grundsätzlich wurden

nur H- und Salzbrücken einer variablen Mindeststärke von $E \text{ kcal mol}^{-1}$ zur Konstruktion des Netzwerks herangezogen. Hydrophobe Kontakte zwischen Kohlenstoff- oder Schwefelatomen wurden berücksichtigt, wenn der Abstand zwischen diesen Atomen kleiner als die Summe ihrer van der Waals-Radien (C: 1,7 Å, S: 1,8 Å) plus 0,25 Å war. Um dies zu erreichen, wurde FIRST grundsätzlich mit den Optionen „H1“ und „c0.25“ gestartet. Wechselwirkungen zwischen gestapelten aromatischen Ringen (*stacking*) werden von FIRST ebenfalls identifiziert. Das Ergebnis der Konstruktion des Netzwerks sind vier Dateien, die Listen der identifizierten kovalenten Bindungen (cov.out), H- und Salzbrücken (hbonds.out), hydrophoben Kontakte (hphobes.out) und Wechselwirkungen zwischen gestapelten aromatischen Ringen (stacked.out) enthalten.

4.2 *Flexibilitätsanalyse*

Mit Hilfe der Software FIRST können rigide Cluster und flexible Regionen in den Netzwerken bestimmt werden (Flexibilitätsanalyse) (Jacobs *et al.* 2001). Dabei kommen Methoden aus der Rigiditätstheorie zum Einsatz. Das Ergebnis der Flexibilitätsanalyse ist die Zerlegung des Netzwerks in rigide Cluster und flexible Regionen (*rigid cluster decomposition*, RCD). Dazu wird für jede Bindung bestimmt, ob sie drehbar ist oder nicht. Anschließend wird für jedes einzelne Atom bestimmt, in welchem rigiden Cluster es enthalten ist. Das Ergebnis der Anwendung von FIRST sind drei Dateien: bond.txt, data.txt und results.txt. Die Datei bond.txt enthält eine Liste aller Bindungen sowie einen dazugehörigen Flexibilitätsindex. Die Datei results.txt enthält eine Zusammenfassung der Ergebnisse der Flexibilitätsanalyse und eine Liste möglicherweise aufgetretener Fehler. Die Datei data.txt enthält eine Liste aller Atome sowie die Nummer des dazugehörigen rigiden Clusters. Zur Visualisierung der RCD wurde zudem basierend auf den in data.txt enthaltenen Informationen ein Pymol-Skript (RCD.pml) sowie eine PDB-Datei (RCD.pdb) erzeugt. Ursprünglich von FIRST erzeugte Pymol-Skripte (ebenefalls RCD.pml bzw. RCD.pdb genannt) wurden dabei überschrieben, da sie in der vorliegenden Arbeit nicht benötigt wurden.

4.3 *Simulation der thermischen Entfaltung*

Die Rigiditätsperkolation von Proteinnetzwerken kann analog zur Rigiditätsperkolation von Netzwerken glasbildender Materialien untersucht werden. Anders als bei den Netzwerken glasbildender Materialien wird die mittlere Koordinationszahl dabei aber nicht über die Variation der chemischen Zusammensetzung des Materials verändert, sondern über die zunehmende Verdünnung ausschließlich nichtkovalenter Wechselwir-

kungen. Durch das sukzessive Aufbrechen dieser Wechselwirkungen in einer bestimmten Reihenfolge wird eine Temperaturerhöhung und somit die thermische Entfaltung der Poteinstruktur (Kap. 1.2.3) simuliert.

Grundsätzlich ähnelt das Vorgehen in der vorliegenden Arbeit dem Vorgehen in den Arbeiten von Rader *et al.* (Rader *et al.* 2002), in denen die Entfaltung als eine Veränderung in der Anzahl der *floppy modes* im Netzwerk beschrieben wurde, dessen mittlere Koordinationszahl durch das Aufbrechen nichtkovalenter Wechselwirkungen verringert wurde. In der vorliegenden Arbeit wird die Entfaltung dagegen nicht als Funktion der mittleren Koordinationszahl untersucht, sondern als Funktion einer relativen Temperatur (Kap. 4.4.5). Dadurch können thermische Entfaltungssimulationen von Proteinen unterschiedlicher Thermostabilität verglichen werden.

Die Simulation der thermischen Entfaltung wird bei einer niedrigen Temperatur gestartet. Unter diesen Bedingungen sind sehr viele nichtkovalente Wechselwirkungen im Netzwerk vorhanden. Wegen der großen Anzahl an Beschränkungen ist das Netzwerk rigide, d.h. es befindet sich in der „rigiden Phase“. Bei einer hohen Temperatur sind sehr viele nichtkovalente Wechselwirkungen gebrochen, und das Netzwerk ist weitgehend flexibel, d.h. es befindet sich in der „flexiblen Phase“. Die rigide Phase entspricht dem gefalteten Zustand des Proteins, die flexible Phase dem ungefalteten Zustand. Die Temperaturerhöhung wird durch das schrittweise Aufbrechen von H- und Salzbrücken in der Reihenfolge abnehmender Stärke simuliert. Dies wurde erreicht, indem zuerst eine sortierte Liste der Energiewerten aller möglichen H- und Salzbrücken erzeugt wurde. Dabei wird angenommen, dass die Energie einer H- oder Salzbrücke (Abschnitt 2.3.1.1) eine Funktion der Temperatur ist, bei der diese Wechselwirkung bricht (Kap. 4.4.5). Für jeden Energiewert E aus der sortierten Liste wurde ein Netzwerk konstruiert, wobei nur H- und Salzbrücken einer Mindeststärke von E_{CUT} herangezogen wurden. So wurde eine Serie von Netzwerken erzeugt, die sich jeweils um eine (oder wenige) H- und Salzbrücken und somit in ihrer Temperatur unterscheiden. Für jedes Netzwerk wurde die Flexibilitätsanalyse durchgeführt und die thermische Entfaltung simuliert.

Die Anzahl der hydrophoben Kontakte wurde während der Entfaltungssimulation nicht verändert. Dies ist damit begründet, dass der den hydrophoben Kontakten zugrunde liegende hydrophobe Effekt mit steigender Temperatur erst einmal zunimmt, bis er ab einer Temperatur von ungefähr 75 °C wieder abnimmt (Kap. 1.2.2). Van der Waals-Wechselwirkungen, die zur Stabilisierung hydrophober Kontakte beitragen, werden

dagegen mit zunehmender Temperatur schwächer. In erster Näherung kann somit davon ausgegangen werden, dass hydrophobe Kontakte bis etwa 75 °C gleich stark bleiben und danach schwächer werden.

Die Entfaltungssimulation kann untersucht werden, indem die Zerlegung des Netzwerks in rigide Cluster und flexible Regionen (RCD) bei verschiedenen Temperaturen verglichen wird. Dafür wurden die während der Flexibilitätsanalyse erzeugten Pymol-Skripte zur Visualisierung der RCD verwendet. Die gesamte Entfaltungssimulation kann allerdings auch in Form einer zweidimensionalen Stabilitätskarte abgebildet werden. Dazu muss untersucht werden, bei welcher Temperatur ein „rigider Kontakt“ zwischen zwei Resten gebrochen oder ausgebildet wird. Ein Rest kann dabei durch sein C α - (Rückgradrigidität) oder C β -Atom (Seitenkettenrigidität) repräsentiert werden. Ein rigider Kontakt zwischen zwei Resten besteht, wenn die beiden Reste zum selben rigiden Cluster gehören. Der Kontakt wird gebrochen, sobald dieser Cluster zerfällt. Die Temperatur, bei der ein rigider Kontakt gebrochen wird, gibt einen Hinweis auf seine mikroskopischen Stabilität. Die Stabilitätswerte werden in kcal mol⁻¹ und relativ zur makroskopischen Thermostabilität angegeben: Bricht ein rigider Kontakt bei der Schmelztemperatur des Proteins, so beträgt die mikroskopische Stabilität dieses Kontakts 0 kcal mol⁻¹. Bricht ein rigider Kontakt 20 K unter- oder oberhalb der Schmelztemperatur des Proteins, so beträgt die mikroskopische Stabilität dieses Kontakts 1 bzw. -1 kcal mol⁻¹. Instabile Kontakte werden auf der Karte weiß eingefärbt, stabile Kontakte rot. Zur Erstellung der Stabilitätskarten wurde die von der Software FIRST ausgegebene Datei data.txt ausgelesen und verarbeitet. Informationen über den Atomtyp wurden aus der ursprünglichen PDB-Datei ausgelesen.

Stabilitätskarten zweier homologer Proteine können verglichen werden, indem die Hälfte der Stabilitätskarte des einen mit der des anderen Proteins aligniert wird. Dazu wurden zuvor mit Hilfe des DaliLite-Servers (Holm und Park 2000) Strukturalignments berechnet. Diese Alignments wurden in Form von Binärworten bei der Erzeugung einer „kombinierten“ Stabilitätskarte mit eingelesen. Die Ähnlichkeit zweier alignierter Karten kann mit Hilfe des Korrelationskoeffizienten nach Pearson quantifiziert werden. Die Rigiditätskarten mit Hilfe der Statistiksoftware R erstellt und ausgewertet.

4.4 Analyse der Entfaltungssimulationen

Ähnlich wie bei der Rigiditätsperkolation von Netzwerken glasbildender Materialien kann während der thermischen Entfaltungssimulation von Proteinen ein Übergang

von einer rigiden zu einer flexiblen Phase beobachtet werden (Rader *et al.* 2002). Anders als bei den kontinuierlichen zufälligen Netzwerken glasbildender Materialien, wo die das Netzwerk durchziehende („perkolierende“) Rigidität durch das Entfernen einer einzigen Bindung zerfällt, ist der Zerfall der Rigidität in Proteinnetzwerken wegen deren Heterogenität aber weniger scharf und zeichnet sich durch eine „Rigiditätsperkolati-on in mehreren Schritten“ aus. Der Phasenübergang von rigide nach flexibel wird als dem Übergang vom gefalteten zum ungefalteten Zustand eines Proteins entsprechend angenommen. Die Temperatur des Phasenübergangs wird T_p bezeichnet und muss mit Hilfe geeigneter Parameter aus der Perkolationstheorie und Netzwerktheorie identifiziert werden. Neben der Identifizierung globaler Eigenschaften des Netzwerks müssen aber auch lokale Struktur- oder Flexibilitätseigenschaften der Proteine während der Entfaltung analysiert werden.

4.4.1 Rigiditätsordnungsparameter

Das Hauptmerkmal der Entfaltungssimulationen ist die Abnahme der Rigidität im Netzwerk, d.h. der Zerfall von rigiden Clustern. Mit Hilfe von Parametern aus der Perkolationstheorie können die Zerfalleigenschaften einzelner Cluster untersucht werden, was einer Beschreibung mikroskopischer Eigenschaften des Netzwerks entspricht (Stauffer und Aharony 1994). Besonders interessant ist dabei die Untersuchung der Eigenschaften des jeweils größten Clusters im Netzwerk, der auch „Riesencluster“ (*giant cluster*) bezeichnet wird. Bei niedrigen Temperaturen gehören nahezu alle Atome zum Riesencluster, der bei hohen Temperaturen vollständig zerfallen ist. Der Zerfall des Riesenclusters geschieht dabei aber nicht kontinuierlich, sondern plötzlich, ähnlich einem Phasenübergang erster Ordnung, wenn auch (wie im Fall von Proteinen) in mehreren Schritten. Da der Riesencluster das Netzwerk dominiert und auch beim Phasenübergang von rigide nach flexibel eine große Rolle spielt, wird die normalisierte Größe des Riesenclusters während der Entfaltungssimulation als Ordnungsparameter aufgezeichnet (Stauffer und Aharony 1994). Der so genannte Rigiditätsordnungsparameter P_∞ beschreibt die Wahrscheinlichkeit, dass ein zufällig ausgewähltes Atom zum Riesencluster gehört. In kontinuierlich zufälligen Netzwerken ist $P_\infty = 1$ in der rigiden Phase und $P_\infty = 0$ in der flexiblen Phase.

Der Rigiditätsordnungsparameter wird berechnet, indem die Anzahl der Atome im Riesencluster bestimmt und durch die Gesamtzahl der Atome im Protein (N) geteilt wird. In der vorliegenden Arbeit wurde dies für jeden Temperaturschritt (d.h. für jedes erzeugte Netzwerk) während der Entfaltungssimulation gemacht. Dazu wurde die In-

formation aus der von der Software FIRST ausgegebenen Datei data.txt ausgelesen und verarbeitet. Der Rigiditätsordnungsparameter wurde anschließend graphisch als Funktion der Temperatur aufgetragen. Im Fall der homogenen Netzwerke glasbildender Materialien kann aus dem Kurvenverlauf eindeutig ein Perkolationsschwellenwert bestimmt werden, der die Temperatur bzw. die kritische mittlere Koordinationszahl bezeichnet, die über- oder unterschritten werden muss, so dass es zum Phasenübergang kommt. Wegen der Heterogenität von Proteinnetzwerken kann dieser kritische Wert nicht so einfach bestimmt werden, da der Phasenübergang in mehreren Schritten passiert. Hier dient der Rigiditätsordnungsparameter also nicht der Identifizierung der genauen Phasenübergangstemperatur, sondern der Beschreibung des Perkulationsverhaltens des Netzwerks.

4.4.2 Perkulationsindex

Neben dem Rigiditätsordnungsparameter, der ein globales Maß zur Beschreibung des Perkulationsverhaltens darstellt, kann auch ein lokaler Perkulationsindex ermittelt werden, der für jedes Atom angibt, ab welcher Temperatur es nicht mehr zum Riesencluster gehört. Der Perkulationsindex wurde als ein lokaler Parameter für die Beschreibung des Perkulationsverhaltens eines Proteins entwickelt, im späteren Verlauf dieser Arbeit aber nicht weiter angewendet. Der Perkulationsindex wird berechnet, indem untersucht wird, welche Atome zum Riesencluster gehören. In der vorliegenden Arbeit wurde dies für jeden Temperaturschritt (d.h. für jedes erzeugte Netzwerk) während der Entfaltungssimulation gemacht. Dazu wurde die Information aus der von der Software FIRST ausgegebenen Datei data.txt ausgelesen und verarbeitet. Der Perkulationsindex kann per Rest aufgetragen werden, indem der Index des jeweiligen Ca -Atoms verwendet wird. Informationen über den Atomtyp wurden aus der ursprünglichen PDB-Datei ausgelesen. Der Perkulationsindex ist entgegen des in der Software FIRST implementierten Flexibilitätsindex (Jacobs *et al.* 2001) oder des von Elena Schmidt (Goethe-Universität, Frankfurt am Main) und Christopher Pflieger (Heinrich-Heine-Universität, Düsseldorf) entwickelten Rigiditätsindex kein Maß für die Flexibilität eines Atoms, sondern ein Parameter, der eine globale Stabilitätseigenschaft lokal abbildet.

4.4.3 Clusterkonfigurationsentropie

Zusätzlich zur Beschreibung mikroskopischer Eigenschaften des Netzwerks können makroskopische Eigenschaften untersucht werden, die mit der Verteilung aller Cluster und insbesondere der Clustergrößen in Zusammenhang stehen (Tsang und Tsang 1999).

Hier wird also die Diversität oder eine Entropie im Netzwerk untersucht. Andraud *et al.* führten in diesem Zusammenhang die Clusterkonfigurationsentropie S als morphologischen Deskriptor heterogener Materialien ein (Andraud *et al.* 1994). In der vorliegenden Arbeit wird die Clusterkonfigurationsentropie erstmals auf Proteinnetzwerke angewendet, die ebenfalls heterogene Materialien darstellen. In Anlehnung an informationstheoretische Grundsätze von Shannon ist diese Entropie definiert als Funktion der Wahrscheinlichkeit w , dass ein Atom zu einem rigiden Clusters der Größe s (s -Cluster) gehört (Andraud *et al.* 1994):

$$S = - \sum_s w \ln w \quad (10)$$

Die Wahrscheinlichkeit w wird dabei folgendermaßen berechnet:

$$w = \frac{sn_s}{\sum_s sn_s} \quad (11)$$

Hier bedeutet n_s die normalisierte Anzahl der s -Cluster:

$$n_s = \frac{\text{Anzahl der } s\text{-Cluster}}{N} \quad (12)$$

Hier bedeutet N die Anzahl der Atome. Die Clusterkonfigurationsentropie ist ein Maß für die „Unordnung“ im Netzwerk bei einer bestimmten Temperatur. Solange der Riesencluster das Netzwerk dominiert, ist die Entropie niedrig: Die Unordnung ist gering, da es nur wenige Möglichkeiten gibt, das Netzwerk mit einem sehr großen Cluster zu konfigurieren. Mit dem beginnenden Zerfall des Riesenclusters in mehrere kleinere Cluster nimmt die Entropie sprunghaft hohe Werte an: Die Unordnung ist groß, da es viele Möglichkeiten gibt, ein Netzwerk mit vielen rigiden Clustern verschiedener Größe zu konfigurieren. Der sprunghafte Anstieg der Entropie ist kennzeichnend für den Übergang während der Rigiditätsperkolation. Im Fall von Proteinen, wo die Rigiditätsperkolation in mehreren Schritten abläuft, weist die Entropie auf einen sehr frühen Schritt des Phasenübergangs hin.

In der vorliegenden Arbeit soll die Clusterkonfigurationsentropie dazu verwendet werden, die Temperatur T_p zu identifizieren, bei der der Riesencluster aufhört, das System zu dominieren. Dieses Ereignis entspricht dem Zerfall eines Faltungskerns und somit dem Übergang zwischen dem gefalteten und ungefalteten Protein, wird aber nicht zwingend durch den ersten Schritt des Phasenübergangs von rigide nach flexibel beschrieben. Die ursprüngliche Definition der Entropie nach Andraud *et al.* (Andraud *et al.* 1994) ist deswegen nicht dazu geeignet, Proteinnetzwerke auf die Temperatur T_p hin zu untersuchen.

Aus diesem Grund wurde in der vorliegenden Arbeit eine veränderte Version der Clusterkonfigurationsentropie entwickelt, die als Funktion der Wahrscheinlichkeit definiert ist, dass ein Atom zu einem Clusters der Größe s^2 (s^2 -Cluster) gehört. Die Wahrscheinlichkeit w wird dabei folgendermaßen berechnet:

$$w = \frac{s^2 n_s}{\sum_s s^2 n_s} \quad (13)$$

Die Verwendung von s^2 anstatt von s hat keinen Einfluss auf das generelle Verhalten der Clusterkonfigurationsentropie. Der sprunghafte Anstieg der Entropie weist jetzt allerdings auf spätere Zerfallsereignisse hin, bei denen ein letzter Rest des Riesenclusters zerfällt und kein großer Cluster mehr die Struktur dominiert. Dieser Zerfall eines „rigiden Kerns“ entspricht dem Zerfall eines Faltungskerns und somit dem Übergang zwischen dem gefalteten und ungefalteten Protein (Hespenheide *et al.* 2002; Rader und Bahar 2004). Die Temperatur, bei der dieses Ereignis stattfindet, ist T_p und kann mit der Schmelztemperatur T_m des Proteins verglichen werden.

Die Clusterkonfigurationsentropie wird berechnet, indem zuerst die Verteilung der Clustergrößen bestimmt wird. In der vorliegenden Arbeit wurde dies für jeden Temperaturschritt (d.h. für jedes erzeugte Netzwerk) während der Entfaltungssimulation gemacht. Dazu wurde die Information aus der von der Software FIRST ausgegebenen Datei data.txt ausgelesen und verarbeitet. Die Clusterkonfigurationsentropie wurde anschließend graphisch als Funktion der Temperatur aufgetragen. Aus dem Verlauf kann die Phasenübergangstemperatur T_p visuell bestimmt werden. Mit Hilfe der Statistiksoftware R kann der Verlauf aber auch automatisch ausgewertet werden, wobei T_p numerisch bestimmt wird. Zur automatischen Bestimmung der Phasenübergänge wurden die Entropiewerte mit Hilfe eines Splines angepasst, wobei Methoden aus dem Paket

pspline verwendet wurden. T_p wurde als die Temperatur bestimmt, bei der die erste Ableitung des Splines ein Maximum hat. Die automatische Bestimmung der Übergänge ist zuverlässig, wenn die Entfaltungssimulationen zuvor mit einem Temperaturschritt von 2 K (entsprechend $-0,1 \text{ kcal mol}^{-1}$, Kap. 4.4.5) durchgeführt wurden.

4.4.4 Entfaltungsregionen

Entfaltungsregionen entsprechen den labilsten Bereichen des Riesenclusters. Sie werden bestimmt, indem diejenigen Atome des Riesenclusters identifiziert werden, die unmittelbar nach dem Phasenübergang nicht mehr einem rigiden Cluster angehören. In der vorliegenden Arbeit wurde dies über den Vergleich der RCD unmittelbar vor und nach dem Phasenübergang gemacht. Dazu wurde die Information aus der von der Software FIRST ausgegebenen Datei data.txt ausgelesen und verarbeitet. Die Entfaltungsregionen können per Rest ausgegeben werden, indem die Information des jeweiligen C α -Atoms verwendet wird. Informationen über den Atomtyp wurden aus der ursprünglichen PDB-Datei ausgelesen. Zur Visualisierung der Entfaltungsregionen wurde ein Pymol-Skript (un.pml) sowie eine PDB-Datei (un.pdb) erzeugt.

4.4.5 Anpassung der Temperaturskala

Die während der Simulation verwendete Temperatur ist eine relative Größe, die erst anhand eines Datensatzes experimentell bestimmter Schmelztemperaturen grob an eine absolute Temperaturskala angepasst werden muss. Dazu muss eine Funktion bestimmt werden, die E_{CUT} auf T abbildet. In einem ersten Schritt wurde der Energieschwellenwert E_p bestimmt, bei dem der Phasenübergang stattfindet. Anschließend wurde $E_{\text{CUT}} = E_p$ anhand experimentell bestimmter T_m für einen Datensatz homologer meso- und thermophiler Proteine (Kap. 4.7.1) kalibriert. Dabei wurden nur diejenigen Paare meso- und thermophiler Proteine berücksichtigt, für die der höhere Energieschwellenwert E_p des mesophilen Proteins korrekt vorhergesagt wurde. In den Fällen, in denen keine experimentell bestimmte T_m verfügbar waren, wurden diese abgeschätzt aus den optimalen Wachstumstemperaturen (T_{og}) der Organismen, aus denen die Proteine stammen: Die T_m liegt rund 25 K über der T_{og} (Dehouck *et al.* 2008). Die Anpassung führte zu folgender Beziehung:

$$T = -20 \frac{\text{K}}{\text{kcal mol}^{-1}} \cdot E_{\text{CUT}} + 300 \text{ K} \quad (14)$$

Die mit Hilfe dieser Beziehung angepasste Temperatur kann nur dazu verwendet werden, um Differenzen in der T_p homologer Proteine mit Schmelz- oder optimalen Wachstumstemperaturdifferenzen (ΔT_m oder ΔT_{og}) zu vergleichen, da der absolute Wert von T_p durch die Größe des Netzwerks und die Topologie der Proteinstruktur beeinflusst wird. Die Proteine müssen zudem von „gleicher Herkunft“ sein, da die Methode, mit der die Struktur bestimmt oder bearbeitet wurde, ebenfalls einen Einfluss auf generelle Eigenschaften des Netzwerks und somit auf den absoluten Wert von T_p hat.

Um Enzyme in einem aktiven (funktionierenden) Zustand miteinander vergleichen zu können, wurden Arbeitstemperaturen bestimmt, indem (a) davon ausgegangen wurde, dass diese Temperaturen niedriger sein müssen als der T_p Wert des entsprechenden Proteins, und (b) dass Strukturbereiche, die sich während der Funktion des Enzyms bewegen, bei dieser Temperatur als flexibel (beweglich) vorhergesagt werden müssen. So wurde die Arbeitstemperatur für thermolysinähnliche Protease (TLP) als 8 K unter der T_p bestimmt, für 3-Isopropylmalatdehydrogenase (IPMDH) als 3 K unter der T_p .

4.5 Ensemblebasierte Analyse statischer Netzwerke

Die Analyse statischer Netzwerke (CNA) kann auf Proteinstrukturen angewendet werden, von denen einzelne Kristallstrukturen guter Qualität vorhanden sind. Diese Einzelstrukturanalyse ist aber sehr empfindlich gegenüber kleinen Defekten in der Struktur, die einen großen Einfluss auf die Flexibilitätsanalyse haben können. Um die CNA dennoch auch auf Strukturen geringerer Qualität oder Homologiemodelle anwenden zu können, wurde in der vorliegenden Arbeit ein Ansatz entwickelt, der die Anwendung der CNA auf konformationelle Ensembles der Struktur erlaubt (ensemblebasierte Analyse statischer Netzwerke). So kann eine Struktur schlechter Qualität in einem ersten Schritt mit geeigneten Verfahren (wie beispielsweise MD-Simulationen) verbessert („verfeinert“) werden (Fan und Mark 2004). Auf das resultierende Strukturensemble kann dann in einem zweiten Schritt die so genannte ensemblebasierte CNA angewendet werden. Dabei wurden für alle Strukturen aus dem Ensemble Einzelstrukturanalysen durchgeführt, d.h. es wurden Netzwerke konstruiert (Kap. 4.1), die Flexibilität analysiert (Kap. 4.2) und die thermische Denaturierung simuliert (Kap. 4.3). Theoretisch kann dann auch jede einzelne Simulation analysiert werden. Generell sind Aussagen über die Struktur aber nur möglich, wenn über alle Einzelstrukturen im Ensemble gemittelte Parameter bestimmt werden.

Zur Abbildung der Entfaltungssimulation des Ensembles in Form einer gemittelten Stabilitätskarte wurden die Stabilitätskarten der Einzelstrukturen erzeugt, addiert und anschließend durch die Anzahl der Einzelstrukturen (*snapshots*) im Ensemble dividiert. Die Einzelstrukturdaten über die Clusterkonfigurationsentropie können dagegen nicht so einfach gemittelt werden. In diesem Fall wurde für jede Struktur aus dem Ensemble die Phasenübergangstemperatur automatisch mit Hilfe der Statistiksoftware R bestimmt. Aus der Verteilung der T_p der Einzelstrukturen wurde dann die „gemittelte“ Phasenübergangstemperatur des Ensembles T_p' bestimmt, indem der Median der Verteilung berechnet wurde. Der Median wurde verwendet, da dieser unempfindlicher gegenüber Ausreißern in einer Verteilung ist als das arithmetische Mittel.

Als „repräsentative Einzelstrukturen“ wurden Strukturen aus dem Ensemble ausgewählt, deren T_p dem T_p' entsprach. Über das Ensemble gemittelte Entfaltungsregionen werden bestimmt, indem Entfaltungsregionen für jede Einzelstruktur bestimmt wurden und anschließend die Wahrscheinlichkeit (w_{ER}) berechnet wurde, dass ein Atom zu einer Entfaltungsregion gehört. Zur Visualisierung dieser „Entfaltungsregionswahrscheinlichkeit“ wurde eine PDB-Datei (*un_statistics.pdb*) erzeugt, indem der Wert dieser Wahrscheinlichkeit für jedes Atom im Feld des B-Faktors abgelegt wurde. Darüber hinaus wurden „Entfaltungskerne“ bestimmt. Diese entsprechen Seitenkettenclustern weniger Reste, deren w_{ER} höher ist als die der umgebenden Reste. Die Bestimmung der Entfaltungskerne erfolgte manuell.

4.6 Erzeugung von Strukturensamples

In der vorliegenden Arbeit wurden Strukturensamples mit Hilfe von MD-Simulationen unter Verwendung eines Molekularmechanik(MM)-Kraftfelds erzeugt. In der Molekularmechanik werden Wechselwirkungen zwischen kovalent gebundenen Atomen durch Gleichgewichtslängen, -winkel und Torsionswinkel charakterisiert. So genannte Kraftkonstanten beschreiben, wie stark die Gleichgewichtswerte gegenüber Auslenkungen bevorzugt werden. MM-Kraftfelder wurden empirisch parametrisiert und bestehen aus einem Satz von Atomtypen, Gleichgewichtsbindungs-längen, -winkel und Torsionswinkel sowie dazugehörigen Kraftkonstanten. Auch elektrostatische und van der Waals-Wechselwirkungen werden durch die Kraftfelder beschrieben. Gegenwärtig existieren mehrere etablierte, verschieden parametrisierte Kraftfelder. Durch ein Kraftfeld wird eine hochdimensionale Potentialfunktion beschrieben, aus der die potentielle Energie einer bestimmten Konformation eines Moleküls in Abhängigkeit von den

räumlichen Koordinaten aller Atome berechnet werden kann. In MD-Simulationen werden mittels dieser Potentialfunktionen die Kräfte berechnet, die in einer bestimmten Ausgangskonformation auf jedes einzelne Atom wirken. Dabei wird die Newton'sche Bewegungsgleichung iterativ gelöst. Man erhält einen Satz von Konformationen (Trajektorie), der die Bewegung aller Atome über einen bestimmten Zeitraum beschreibt.

In der vorliegenden Arbeit wurden alle Simulationen mit Hilfe des AMBER-Programmpakets (Version 9) (Case *et al.* 2005) unter Verwendung des Kraftfelds parm99SB (Hornak *et al.* 2006) und des TIP3P Wassermodells (Jorgensen 1982) durchgeführt. Die Ausgangsstrukturen wurden vor der Simulation mit der Software REDUCE (Word *et al.* 1999) protoniert. Die Atom- und Restnamen wurden anschließend mit Hilfe eines Perl-Skripts (`reduce2amber.pl`) an die von AMBER verwendeten Bezeichnungen angepasst. Bei der Simulation wurden periodische Randbedingungen verwendet, mit Simulationszellen in Form eines *truncated octahedron*. Die Dimensionen der periodischen Zellen wurden so gewählt, dass die Minimaldistanz zwischen dem Protein und dem Rand der Zelle 11 Å betrug. Die Systeme wurden mit Hilfe der Methode des steilsten Abstiegs (*steepest descent*) und des konjugierten Gradienten (*conjugated gradient*) in 500 Schritten minimiert und anschließend für 50 ps mit einem Integrations-Zeitschritt von 2 fs bei konstantem Volumen auf 300 K aufgeheizt. Das Volumen des Systems wurde dann unter Beibehaltung konstanten Drucks (1 atm) und konstanter Temperatur (300 K) für 50 ps mit einem Integrations-Zeitschritt von 2 fs eingestellt, wobei die Beweglichkeit der Schweratome des Proteins durch ein angelegtes harmonisches Potential beschränkt wurde. Anschließend wurde das System bei konstantem Volumen und konstanter Temperatur (300 K) für weitere 200 ps equilibriert, wobei die harmonischen Potentiale nach und nach auf Null gesetzt wurden.

Alle Simulationen wurden bei konstantem Volumen und einer konstanten Temperatur von 300 K durchgeführt (NVT-Ensemble). Während der Simulation wurden die nichtbindenden Wechselwirkungen auf einen Bereich von 8 Å beschränkt. Elektrostatische Wechselwirkungen jenseits dieses Radius wurden mit Hilfe der Particle-Mesh-Ewald(PME)-Methode (Darden *et al.* 1993) berücksichtigt. Alle Längen von Bindungen, an denen Wasserstoff-Atome beteiligt sind, wurden mit dem SHAKE-Algorithmus (Ryckaert *et al.* 1977) fixiert. Der Integrations-Zeitschritt bei der Simulation betrug 2 fs. Die Systeme wurden durch schwache Kopplung an ein Temperaturbad bei einer Temperatur von 300 K gehalten (Berendsen *et al.* 1984). Nach jeweils 1000 Integrations-Schritten wurden Translations- und Rotationsbewegungen des Proteins entfernt.

Zur Analyse der Konformationen aus der Trajektorie, die während der Simulation in Intervallen von 1 ps ausgegeben wurden, wurde das PTRAJ-Modul aus dem AMBER-Programmpaket verwendet. Es wurden generelle Eigenschaften der Konformationen aus der Trajektorie berechnet, insbesondere wurde die Wurzel der mittleren quadratischen Abweichung (*root mean square deviation, RMSD*) zwischen zwei Konformationen (in Å) bestimmt. Dazu wurden eine Konformation, die während der Simulation erhalten wurde, und die Startstruktur anhand der Position der C α - oder aller Schweratome bestmöglich überlagert. Die RMS-Abweichung wurde folgendermaßen berechnet:

$$RMSD = \left[\frac{1}{N} \sum_{i=1}^N (\mathbf{r}_i - \mathbf{r}_i^0)^2 \right]^{\frac{1}{2}} \quad (15)$$

Hier bedeuten \mathbf{r}_i und \mathbf{r}_i^0 die kartesischen Koordinaten von Atom i nach der Überlagerung bzw. die Startkoordinaten von Atom i und N die Anzahl der Atome im Molekül. Das PTRAJ-Modul wurde auch dazu verwendet, alle 40 ps eine Einzelstruktur im PDB-Format aus der Trajektorie für die anschließende ensemblebasierte CNA zu extrahieren. Aus den ersten 2 ns der Simulation wurden keine Einzelstrukturen extrahiert.

4.7 Datensätze

Im ersten Teil der vorliegenden Arbeit wurde die CNA auf einen Datensatz von Kristallstrukturen homologer Proteine von mesophilen und thermophilen Organismen angewendet. Im zweiten Teil der Arbeit wurde mit Hilfe der ensemblebasierten CNA eine Serie von Phytasen untersucht, für die fast ausschließlich nur Homologiemodelle verfügbar waren. Die beiden Datensätze werden in den folgenden Kapiteln vorgestellt.

4.7.1 Homologe meso- und thermophile Proteine

Kristallstrukturen homologer Proteine von meso- und thermophilen Organismen (d.h. meso- und thermophile Proteine) wurden aus der PDB (Bourne *et al.* 2004) gewonnen. Als Ausgangspunkt für die Zusammenstellung des Datensatzes dienten Arbeiten von Brinda *et al.* (Brinda und Vishveshwara 2005) und Sadeghi *et al.* (Sadeghi *et al.* 2006). In den Datensatz wurden nur Strukturen aufgenommen, die bestimmte Qualitätskriterien erfüllten. So musste die Auflösung geringer als 2,6 Å sein. Es wurden nur Strukturen berücksichtigt, deren kristallographische Einheit der biologischen Einheit entsprach. Die Qualität der Strukturen wurde außerdem mit Hilfe des PDBREPORT-

Servers (Hooft *et al.* 1996) überprüft. Fehlerhafte Strukturen wurden (wenn möglich) mit einer anderen Struktur desselben Proteins ersetzt. Um homologe meso- und thermophile Proteine untersuchen zu können, wurden nur Paare von Strukturen aufgenommen, deren Sequenzen zu mindestens 35 % identisch waren. Auch die strukturelle Ähnlichkeit der jeweiligen Strukturen wurde überprüft. Sowohl die Sequenz- als auch die Strukturähnlichkeit wurde mit Hilfe des DaliLite-Servers (Holm und Park 2000) bestimmt. Paare von Strukturen wurden nur in den Datensatz mit aufgenommen, wenn die jeweiligen Strukturen im selben oligomeren und ligandenbindenden Zustand waren und ähnliche Auflösung (Unterschied in der Auflösung kleiner als 1 Å sein) hatten. Der endgültige Datensatz bestand aus 19 Paaren homologer Proteine aus verschiedenen Proteinfamilien. Die Strukturen aus dem Datensatz und deren Eigenschaften sind in Tabelle 1 aufgelistet.

Trotz sorgfältiger Suche war es in einigen Fällen nicht möglich, Strukturen zu finden, die allen Qualitätskriterien entsprachen. Im Fall der Proteinfamilien 5 (Elongationsfaktor TU), 14 (Zellulase) und 18 (Histidintransportprotein) war die kristallographische Einheit größer als die biologische. Hier wurde nur eine Untermenge der Oligomere aus der kristallographischen Einheit für die Konstruktion der Netzwerke verwendet. Im Fall der Proteinfamilien 4 (Xyloseisomerase), 12 (Cyclodextrin-Glycosyltransferase), 15 (thermolysinähnliche Protease) und 17 (Endozellulase) war entweder die meso- oder die thermophile Struktur ligandengebunden. Hier wurde der Ligand aus der entsprechenden Kristallstruktur entfernt. Im Fall der Proteinfamilie 17 (Endozellulase) wurde zudem eine Abfolge von elf terminalen Histidinresten (Reste 68–77) sowie ein loser Terminus (Reste 78–87) entfernt.

Wenn mehrere Konformere für eine Seitenkette angegeben waren, dann wurde nur das Konformer mit dem höchsten relativen Besetzungsrat oder bei gleich häufiger Besetzung das zuerst angegebene Konformer verwendet. Strukturen mit fehlenden Atomen wurden nur dann in den Datensatz mit aufgenommen, wenn es sich bei den fehlenden Atomen um Seitenkettenatome an der Proteinoberfläche handelte. Diese entsprechen losen Enden (*dangling ends*), die nicht zur Rigidität im Netzwerk beitragen (Rader *et al.* 2002) und deswegen weggelassen werden können. Wasserstoffatome wurden mit Hilfe der Software REDUCE (Word *et al.* 1999) hinzugefügt. Diese Software wurde zudem verwendet, um die Ausrichtung von Aspartat-, Glutamat- und Histidinseitenketten zu überprüfen und gegebenenfalls zu ändern.

Die Proteinstrukturen wurden durch manuelles Editieren der PDB-Datei für die Konstruktion der Netzwerke vorbereitet. Alle Wasser-, Lösungsmittel-, Puffer-, Substrat- und Kofaktormoleküle wurden entfernt. Von der Proteinstruktur koordinierte Metallionen wurden bei der Konstruktion des Netzwerks dagegen berücksichtigt. Dazu wurden kovalente Bindungen zwischen dem Metallion und den koordinierenden Gruppen des Proteins manuell zum Netzwerk hinzugefügt. Die koordinierenden Gruppen des Proteins wurden mit Hilfe der Software Pymol identifiziert, indem Proteinatome im Umkreis von 2 Å vom Metallion gesucht werden. Im Fall der Proteinfamilie 10 (3-Isopropylmalatdehydrogenase) wurden die Kristallstrukturen zusätzlich kurzen Energieminimierungen unterzogen, um durch Kristallkontakte verursachte Defekte (Deformationen) in der mesophilen Struktur zu beheben. Die Minimierungen wurden mit Hilfe des AMBER-Programmpakets (Version 9) (Case *et al.* 2005) wie in Kap. 4.1 beschrieben unter Verwendung des Kraftfelds parm99SB (Hornak *et al.* 2006) durchgeführt.

Die Entfaltungssimulation wurde mit dem kleinstmöglichen variablen Temperaturschritt durchgeführt, d.h. H- oder Salzbrücken wurden einzeln entfernt. Im Fall eines großen Proteins mit rund 700 Resten, wie beispielsweise der 3-Isopropylmalatdehydrogenase, resultierte dies in rund 800 Temperaturschritten. Die Rechenzeit der CNA eines solchen Proteins betrug 20 Minuten (nur Konstruktion der Netzwerke und Flexibilitätsanalysen) auf einem Knoten eines Quadcore-Prozessors mit einer Taktfrequenz von 2,4 GHz. Die Berechnung der Stabilitätskarte sowie der Parameter für die Analyse der Simulation (Rigiditätsordnungsparameter, Clusterkonfigurationsentropie, Stabilitätskarten) betrug zusätzliche 9 Stunden. Diese Zeit beinhaltet auch die relativ zeitaufwändige statistische Analyse der Stabilitätskarten.

Den Strukturen wurden optimale Wachstumstemperaturen (T_{og}) der Organismen zugewiesen, aus denen sie stammen (Querol *et al.* 1996; Vogt *et al.* 1997). Diese Daten wurden aus der Literatur und/oder aus der *prokaryotic growth temperature database* (PGTdb) (Huang *et al.* 2004) gewonnen. Die optimale Wachstumstemperaturen dienen als Deskriptor für die Thermostabilität einer Struktur, da sie bis zu einem gewissen Grad mit Schmelztemperaturen korreliert sind (Gromiha *et al.* 1999; Dehouck *et al.* 2008). In wenigen Fällen konnten auch Schmelztemperaturen aus der Literatur oder der ProTherm-Datenbank (Bava *et al.* 2004) gewonnen werden. Detaillierte Daten über thermostabilisierende Mutationen und die Aktivität der Proteinfamilien zehn (3-Isopropylmalatdehydrogenase) und 15 (thermolysinähnliche Protease) stammen aus der Literatur. Die Daten zur Thermostabilität sind in Tabelle 1 aufgelistet.

4.7.2 Phytasen

Homologiemodellierte Strukturen von sechs Phytasen wurden durch Dr. Wolfgang Höffken und Dr. Stefan Häfner (BASF SE, Ludwigshafen) zur Verfügung gestellt. Darüber hinaus wurden die Kristallstrukturen der *Escherichia coli* AppA, *Aspergillus ficum* PhyA und *Aspergillus fumigatus* PhyA Phytasen aus der PDB (Bourne *et al.* 2004) gewonnen. Die Homologiemodelle wurden mit Hilfe der Software Prime (Schrödinger LLC) erzeugt. Als Template diente die Kristallstruktur der *E. coli* AppA (PDB-Code 1dkl) bzw. der *A. ficum* PhyA Phytase (PDB-Code 1ihp). Die Kristallstrukturen lagen alle im nicht ligandengebundenen Zustand vor und hatten eine Auflösung von unter 2,6 Å. Im Fall der Phytasen von *A. ficum* und *A. fumigatus* entsprach die kristallographische Einheit der biologischen Einheit. Ein fehlender Loop (Reste 249–252) der *A. ficum* PhyA Phytase wurde mit Hilfe des ModLoop-Servers (Fiser und Sali 2003) hinzugefügt. Im Fall der *E. coli* AppA Phytase wurde ein zweites Monomer aus der kristallographischen Einheit gelöscht. Fehlende Reste (181–185) wurden mit Hilfe des ModLoop-Servers (Fiser und Sali 2003) hinzugefügt. Die Qualität der Kristallstrukturen wurde mit Hilfe des PDBREPORT-Servers (Hooft *et al.* 1996) überprüft. Sowohl die Sequenz- als auch die Strukturähnlichkeit wurde mit Hilfe des DaliLite-Servers (Holm und Park 2000) bestimmt. Ein Sequenzalignment der Consensus-Phytase 10, der *E. coli* AppA und der *Yersinia mollaretii* Wildtyp (WT) Phytasen wurde mit Hilfe der Software T-Coffee (Notredame *et al.* 2000) durchgeführt und manuell bearbeitet. Die Strukturen aus dem Datensatz und deren Eigenschaften sind in Tabelle 6 aufgelistet.

Da bereits kleine strukturelle Defekte die Ergebnisse der CNA einer Einzelstruktur verfälschen können, wurden die Homologiemodelle verbessert, indem MD-Simulationen durchgeführt wurden. Zuvor wurden Wasserstoffatome mit Hilfe der Software REDUCE (Word *et al.* 1999) hinzugefügt. Diese Software wurde zudem verwendet, um die Ausrichtung von Aspartat-, Glutamat- und Histidinseitenketten zu überprüfen und gegebenenfalls zu ändern. Alle Simulationen wurden mit Hilfe des AMBER-Programmpakets (Version 9) (Case *et al.* 2005) wie in Kap. 4.6 beschrieben unter Verwendung des Kraftfelds parm99SB (Hornak *et al.* 2006) und des TIP3P Wassermolekells (Jorgensen 1982) durchgeführt. Durch das Hinzufügen von negativ (Cl⁻) bzw. positiv geladener Gegenionen (Na⁺) wurden die Systeme neutralisiert. Vorhandene Disulfidbrücken wurden explizit berücksichtigt. Vor der Simulation wurden die Atom- und Restnamen der mit der Software REDUCE protonierten Ausgangsstrukturen mit Hilfe eines Perl-Skripts (reduce2amber.pl) an die von AMBER verwendeten Bezeichnungen

angepasst. Die Systeme wurden für 7 ns im NVT-Ensemble bei einer konstanten Temperatur von 300 K simuliert. Die Rechenzeit für 1 ns der Simulation betrug rund 43 Stunden auf vier Knoten eines Quadcore-Prozessors mit einer Taktfrequenz von 2,4 GHz. Nach der Simulation wurden die AMBER-spezifischen Restbezeichnungen (Hie, Hid, Hip und Cyx) wieder in das PDB-Format (His und Cys) übertragen. Auch für die Kristallstrukturen wurden auf diese Weise Ensembles erzeugt.

Mit Hilfe des PTRAJ-Moduls aus dem AMBER-Programmpaket wurde alle 40 ps eine Einzelstruktur im PDB-Format aus der Trajektorie von der zweiten bis zur siebten ns für die anschließende ensemblebasierte CNA extrahiert. Aus den ersten 2 ns der Simulation wurden keine Einzelstrukturen extrahiert. Dies ergab 125 Einzelstrukturen. Wassermoleküle und Gegenionen wurden entfernt. Die Entfaltungssimulationen wurden mit einem Temperaturschritt von 2 K (entsprechend $-0,1 \text{ kcal mol}^{-1}$) durchgeführt. Dies resultierte in rund 70 Temperaturschritten. Die Rechenzeit der CNA einer Einzelstruktur betrug 1 Minute (nur Konstruktion der Netzwerke und Flexibilitätsanalysen) auf einem Knoten eines Quadcore-Prozessors mit einer Taktfrequenz von 2,4 GHz. Die Berechnung der Stabilitätskarte sowie der Parameter für die Analyse der Simulation (Rigiditätsordnungsparameter, Clusterkonfigurationsentropie) betrug zusätzliche 11 Minuten. Den Strukturen wurden Halbwertszeittemperaturen (T_{50}) und Temperaturen der maximalen Aktivität (T_{opt}) zugeordnet. Diese Daten wurden durch Dr. Wolfgang Höffken und Dr. Stefan Häfner (BASF SE, Ludwigshafen) zur Verfügung gestellt. Halbwertszeittemperaturen wurden so gemessen, dass ihr Wert dem Schmelzpunkt der Struktur entsprach. Für die Phytasen von *A. ficum* und *A. fumigatus* wurden T_{opt} aus der Literatur gewonnen. Halbwertszeittemperaturen und Temperaturen der maximalen Aktivität sind in Tabelle 6 aufgelistet.

Anhand der aus der thermischen Entfaltungssimulation der *Y. mollaretii* WT Phytase berechneten Entfaltungsregionwahrscheinlichkeit wurden drei Entfaltungskerne bestimmt. Der Effekt von Mutationen in den Entfaltungskernen wurde sowohl rechnerisch als auch experimentell untersucht. Für die rechnerische Untersuchung wurden drei repräsentative Einzelstrukturen aus dem Ensemble ausgewählt und deren Netzwerke im Bereich der Entfaltungskerne verändert. Dabei wurden zusätzliche, nichtnative hydrophobe Kontakte zwischen den Resten der Entfaltungskerne eingefügt, ohne tatsächlich Aminosäureaustausche durchzuführen. Zur Untersuchung des Effekts einer solchen „Mutation“ wurden die mutierten Einzelstrukturen erneut mit Hilfe der CNA untersucht. Gemittelte Phasenübergangstemperaturen wurden in diesem Fall über die Berechnung

des arithmetischen Mittels der Phasenübergangstemperaturen der drei repräsentativen Einzelstrukturen bestimmt. Die experimentelle Untersuchung der Entfaltungskerne erfolgte durch das gezielte Mutieren einzelner Reste in den Entfaltungsregionen. Die Experimente wurden durch Alexander Denning und Amol V. Shivange (Rheinisch-Westfälische Technische Hochschule, Aachen) sowie Dr. Stefan Häfner (BASF SE, Ludwigshafen) durchgeführt. Dabei wurde die Veränderung in der T_{50} zwischen Wildtyp und Mutante gemessen.

4.8 Computerprogramme und programmiertechnische Hilfsmittel

Die Computerprogramme zur Simulation der thermischen Entfaltung sowie zur Berechnung der Parameter aus der Perkulations- und Netzwerktheorie zur Analyse der Simulation wurden in der Programmiersprache Python (Version 2.5) erstellt. Skripte zur Auswertung der Parameter aus der Perkulations- und Netzwerktheorie sowie für das Erstellen von (kombinierten) Stabilitätskarten wurden in der Sprache R verfasst. Die Entwicklung erfolgte unter Linux auf einem Quadcore-PC mit einer Taktfrequenz von 2,40 GHz pro Knoten. Das CNAnalysis Programmpaket besteht aus mehreren Klassen und Skripten:

- Die Klassen `first.py` und `first_analysis.py` implementieren die Berechnung und Analyse eines einzelnen Temperaturschritts während einer Simulation der thermischen Entfaltung. In `first.py` erfolgt dazu der Aufruf der Software FIRST (Version 6.2).
- Die Klasse `dilution.py` implementiert die Berechnung einer Simulation der thermischen Entfaltung. Die Klassen `dilution_analysis.py`, `dilution_pi.py` und `dilution_un.py` implementieren die Analyse einer Simulation der thermischen Entfaltung, d.h. die Berechnung der Parameter aus der Perkulations- und Netzwerktheorie, des Perkulationsindex bzw. der Entfaltungsregionen.
- Die Klassen `fluctnet.py` und `fluctnet_analysis.py` implementieren die Berechnung und Analyse einer Simulation der thermischen Entfaltung für ein Strukturensemble.
- Die Klassen `methods.py`, `pdb.py`, `hbonds.py` und `hphobes.py` implementieren verschiedene Methoden zum Einlesen verschiedener Dateien, zum Schreiben von PDB-Dateien bzw. zur effizienten Analyse der H- und Salzbrücken oder hydrophoben Kontakte in den Netzwerken.

- Die Skripte `cce.R`, `rop.R` und `rigidity_map.R` dienen der grafischen Darstellung der Verlaufs des Rigiditätsordnungsparameters oder der Clusterkonfigurationsentropie bzw. der Stabilitätskarte. Die Skripte erlauben zudem die statistische Auswertung der Parameter und die Erzeugung kombinierter Stabilitätskarten.

Eine frühe Version des Programmpakets wurde in Zusammenarbeit mit Elena Schmidt (Goethe-Universität, Frankfurt am Main) entwickelt. Spätere Entwicklungen, insbesondere die Entwicklung der Klassen `fluctnet.py` und `fluctnet_analysis.py` erfolgten in Zusammenarbeit mit Christopher Pflieger (Heinrich-Heine-Universität, Düsseldorf). Alle in dieser Arbeit gezeigten Abbildungen von Proteinen und Alignments wurden unter Microsoft Windows Vista mit Hilfe der Software Pymol (Version 0.97) (DeLano 2002) bzw. der Software Texshade (Version 1.20a) (Beitz 2000) auf einem Dualcore-PC mit einer Taktfrequenz von 2,4 GHz pro Knoten erstellt.

5 Ergebnisse und Diskussion

In dieser Arbeit wird ein rechnerisches Verfahren vorgestellt, mit dem die Simulation der thermischen Entfaltung von Proteinen und die Untersuchung ihrer Thermostabilität möglich ist. Das Verfahren beruht auf der Modellierung der Strukturen als eine bestimmte Art von *bar-and-joint*-Netzwerke, in denen die Atome durch Knoten (*joints*) und die Wechselwirkungen zwischen den Atomen durch Beschränkungen (*bars*) beschrieben werden. Neben kovalenten und nichtkovalenten Bindungsbeschränkungen werden auch Winkelbeschränkungen in das Netzwerk eingefügt. Diese *bond-bending*- bzw. äquivalente *body-and-bar*-Netzwerke können hinsichtlich ihrer Flexibilität analysiert werden, d.h. die direkte Identifizierung rigider Cluster und flexibler Bereiche in atomarer Auflösung ist möglich. Dabei kommen Methoden aus der Rigiditätstheorie zum Einsatz.

Die thermische Entfaltung des Proteins wird durch sukzessives Aufbrechen nichtkovalenter Wechselwirkungen und anschließender erneuter Flexibilitätsanalyse simuliert. Weitergehende Analysen ermöglichen die Beschreibung des globalen Zerfalls, die Quantifizierung der Thermostabilität des Proteins, die Identifizierung von Bereichen, von denen der globale Zerfall ausgeht, und somit die Vorhersage von Entfaltungsregionen, in denen thermostabilisierende Mutationen eingeführt werden können. Darüber hinaus ist auch die Untersuchung der für die Bioaktivität wichtigen lokalen Flexibilität möglich. Diese unter dem Begriff der Analyse statischer Netzwerke (*constraint network analysis*, CNA) zusammengefassten Methoden werden auf einen Datensatz homologer meso- und thermophiler Proteine sowie auf eine Serie von Phytasen angewendet. Die Ergebnisse werden in Kap. 5.1 bzw. Kap. 5.2 beschrieben und diskutiert. In Kap. 5.3 folgt eine abschließende Diskussion des verwendeten Verfahrens sowie ein Ausblick.

5.1 Analyse statischer Netzwerke thermophiler Proteine

Der Vergleich von Proteinen aus mesophilen Organismen (optimale Wachstumstemperatur von 15–45 °C) mit natürlich vorkommenden thermostabilen Proteinen aus thermophilen oder hyperthermophilen Organismen (optimale Wachstumstemperatur von 50–80 bzw. mehr als 80 °C) erlaubt die Aufklärung der molekularen Grundlagen der Thermostabilität (Jaenicke 1998). Experimentelle Untersuchungen haben gezeigt, dass die Anpassung thermophiler Proteine durch gezielte Veränderungen der nichtkovalenter Wechselwirkungen erfolgt, die sowohl die elektrostatischen Kräfte (H- und Salz-

brücken), als auch die Packung der Aminosäureseitenketten und die Stärke des hydrophoben Effekts betreffen (Querol *et al.* 1996; Vogt und Argos 1997; Vogt *et al.* 1997; Szilágyi und Závodszy 2000; Gianese *et al.* 2001; Gromiha 2001; Kumar und Nussinov 2001; Gianese *et al.* 2002; Robinson-Rechavi *et al.* 2006; Sadeghi *et al.* 2006; Folch *et al.* 2008; Saelensminde *et al.* 2009). Diese Veränderungen im Netzwerk nicht-kovalenter Wechselwirkungen führen zu einer Verringerung der Flexibilität und somit zu einer Rigidisierung thermophiler Proteine (Vieille und Zeikus 2001; Sterner und Brunner 2008). Wenn die Untersuchung homologer meso- und thermophiler Proteine bei der jeweiligen optimalen Wachstumstemperatur des Organismus, aus dem sie stammen, erfolgt, dann befinden sie sich in korrespondierenden Zuständen ähnlicher Bioaktivität, was trotz globaler Rigidisierung auf eine Konservierung wichtiger flexibler Bereiche schließen lässt (Somero 1978; Jaenicke und Böhm 1998; Sterner und Brunner 2008).

In diesem Teil der Arbeit werden homologe meso- und thermophile Proteine mit Hilfe der Analyse statischer Netzwerke untersucht. Der verwendete Datensatz wird in Kap. 5.1.1 beschrieben. Die Anwendung des rechnerischen Verfahrens soll zeigen, ob die thermophile Anpassung der Proteine aus dem Datensatz tatsächlich über eine Rigidisierung der Struktur erfolgt, und ob sich die homologen meso- und thermophilen Proteine bei ihrer jeweiligen Arbeitstemperatur in Zuständen ähnlicher Flexibilität und Bioaktivität befinden. Darüber hinaus soll insbesondere auch die Eignung der CNA zur Vorhersage der Thermostabilität sowie wichtiger Strukturbereiche für die Thermostabilität und damit die Eignung des Verfahrens zur Unterstützung des Protein Engineering getestet werden.

In Kap. 5.1.2 werden globale Eigenschaften der mit Hilfe der CNA durchgeführten thermischen Entfaltungssimulationen beschrieben. Vorhergesagte thermische Stabilitäten der homologen meso- und thermophilen Proteine aus dem Datensatz werden miteinander verglichen. In Kap. 5.1.3 werden die Ergebnisse eingehender Analysen der thermischen Entfaltungssimulationen sowie mögliche molekulare Grundlagen der Thermostabilität beschrieben. Dabei werden Entfaltungsregionen vorhergesagt und mit Strukturbereichen verglichen, in denen thermostabilisierende Mutationen experimentell eingeführt wurden. In Kap. 5.1.4 wird die mikroskopische Stabilität der meso- und thermophilen Proteine aus dem Datensatz bestimmt und quantitativ miteinander verglichen. Dies geschieht über die Berechnung der Korrelation so genannter Stabilitätskarten. Anschließend werden in Kap. 5.1.5 vorhergesagte mikroskopische Stabilitätswerte kataly-

tischer sowie weiterer für die Katalyse, die Substrat- oder Kofaktorbindung wichtiger Reste bestimmt und mit Daten aus der Literatur verglichen.

5.1.1 Datensatz homologer meso- und thermophiler Proteine

Der in diesem Teil der Arbeit verwendete Datensatz besteht aus 19 Paaren homologer meso- und thermophiler Proteine, die verschiedenen Proteinfamilien angehören. Die Strukturen sowie deren Eigenschaften sind in Tabelle 1 aufgelistet. Bei allen Proteinen handelt es sich um Kristallstrukturen, die insbesondere hinsichtlich ihrer Struktur, ihrer

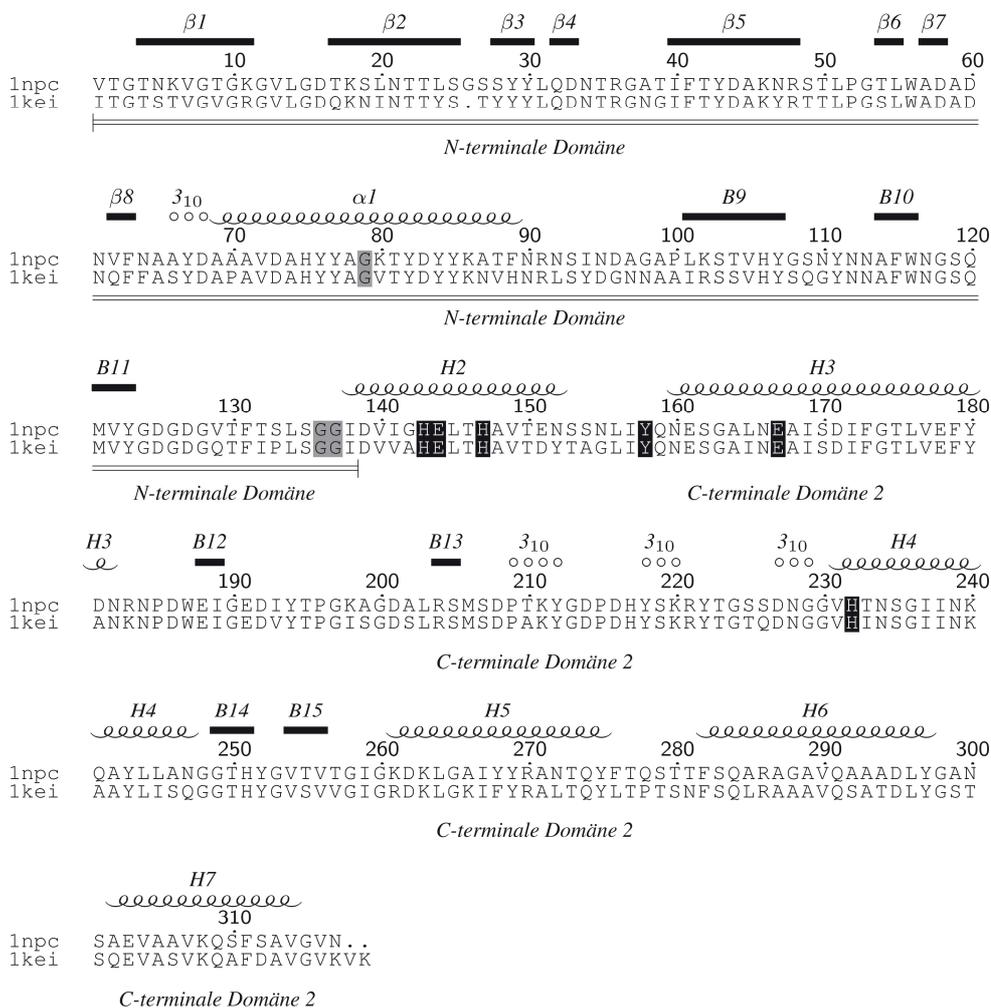


Abbildung 7. Alignment meso- und thermophiler thermolysinähnlicher Protease (TLP). Bei dem Alignment handelt es sich um ein Strukturalignment von *B. cereus* TLP (1npc) und Thermolysin (1kei), das mit Hilfe des DaliLite Servers (Holm und Park 2000) berechnet wurde. Katalytische Reste sind schwarz schattiert, weitere für die Katalyse, die Substrat- oder Kofaktorbindung wichtige Reste (siehe Text) grau. Die das Zinkion koordinierenden Reste werden zu den katalytischen Resten gezählt.

Auflösung und ihres oligomeren sowie ligandenbindenden Zustands bestimmte Qualitätskriterien erfüllen. Bei der Zusammenstellung des Datensatzes wurde darauf geachtet,

dass zusammengehörige meso- und thermophile Proteine „vergleichbar“ waren, d.h. die „gleiche Herkunft“ (Röntgenstrukturanalyse) haben, eine ähnliche Auflösung und den gleichen oligomeren sowie ligandenbindenden Zustand hatten. Die hohe Qualität und

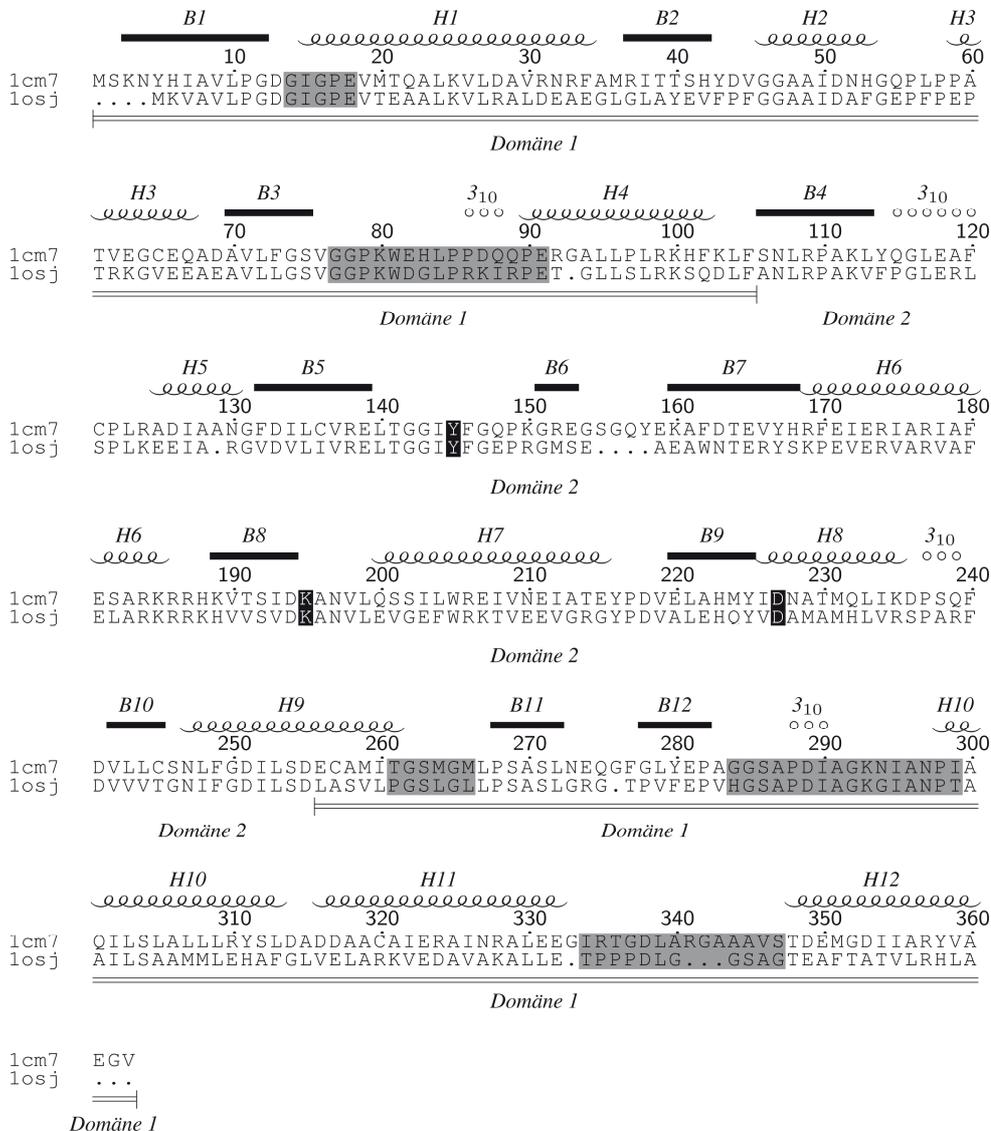


Abbildung 8. Alignment meso- und thermophiler 3-Isopropylmalatdehydrogenase (IPMDH). Bei dem Alignment handelt es sich um ein Strukturalignment von *E. coli* (1cm7) und *T. thermophilus* IPMDH (1osj), das mit Hilfe des DaliLite Servers (Holm und Park 2000) berechnet wurde. Es ist nur das Alignment der ersten Untereinheiten gezeigt. Katalytische Reste sind schwarz schattiert, weitere für die Katalyse, die Substrat- oder Kofaktorbindung wichtige Reste (siehe Text) grau.

die Vergleichbarkeit zusammengehöriger meso- und thermophiler Proteine ist von großer Wichtigkeit, da die während der CNA mit Hilfe der Software FIRST (Jacobs *et al.* 2001) durchgeführte Flexibilitätsanalyse der einzelnen Kristallstrukturen sehr empfindlich gegenüber strukturellen Defekten ist. Zudem können thermische Entfaltungssimula-

tionen nur dann auf der selben Temperaturskala verglichen werden, wenn sie auf Strukturen ähnlicher Größe und Topologie sowie „gleicher Herkunft“ basieren (Kap. 4.4.5).

Die Sequenzen zusammengehöriger Proteine aus dem Datensatz sind zu mindestens 37 % identisch, weswegen die Proteine eindeutig homolog sind. Ihre Strukturen haben eine *RMS*-Abweichung zwischen 0,6 und 2,5 Å für die $C\alpha$ -Atome. Eine höhere *RMS*-Abweichung wurde nur für die Proteine der Familie 5 (Elongationsfaktor TU) und 11 (Phosphoglyceratkinase) gemessen, kann aber auf Unterschiede in der Anordnung der einzelnen Domänen zurückgeführt werden. Wenn die Domänen separat verglichen werden, dann haben sie eine *RMS*-Abweichung von unter 2,5 Å für die $C\alpha$ -Atome. Die hohe Strukturähnlichkeit der Proteine belegt, dass sie die gleiche Topologie haben. Sequenzen und Strukturen zusammengehöriger Proteine können somit eindeutig aligniert werden. Dies gilt insbesondere auch für die Enzyme aus dem Datensatz. Das Vorhandensein konservierter katalytischer sowie weiterer für die Katalyse, die Substrat- oder Kofaktorbindung wichtiger Reste deutet darauf hin, dass diese Proteine auch von einem mechanistischen Standpunkt aus eng miteinander verwandt, wenn nicht sogar identisch sind. In Abbildung 7 und Abbildung 8 sind die Alignments meso- und thermophiler thermolysinähnlicher Protease (TLP) bzw. 3-Isopropylmalatdehydrogenase (IPMDH) gezeigt. Die katalytischen sowie weitere für die Katalyse, die Substrat- oder Kofaktorbindung wichtigen Reste und Sekundärstrukturelemente sind gekennzeichnet.

Tabelle 1. Paare homologer meso- und thermophiler Proteine und ihre Eigenschaften.

| Proteinfamilie | Organismus | PDB ID | T_{og}^a | Auflösung ^b | Sequenzlänge (Monomer) | Oligomerer Zustand | Sequenzidentität ^c | <i>RMS</i> -Abweichung ^b | |
|----------------|---|---------------------------|------------|------------------------|------------------------|--------------------|-------------------------------|-------------------------------------|-----|
| 1 | Pyrrolidon-Carboxylpeptidase | <i>Bacillus</i> | 1aug | 30 | 2,00 | 215 | Tetramer | 40 | 1,2 |
| | | <i>amylolique-faciens</i> | | | | | | | |
| | | <i>Thermococcus</i> | 1a2z | 83 | 1,73 | 220 | | | |
| | | <i>litoralis</i> | | | | | | | |
| 2 | aspartatamino-transferase-ähnliches Protein | <i>Trypanosoma</i> | 1bw0 | 37 | 2,50 | 416 | Dimer | 37 | 1,9 |
| | | <i>cruzi</i> | | | | | | | |
| | | <i>Thermus ther-</i> | 1bjw | 70–75 | 1,80 | 382 | | | |
| | | <i>mophilus</i> | | | | | | | |

| | | | | | | | | | |
|----|---|-----------------------------------|------|--------|------|-----|-----------|----|------|
| 3 | β-Glucanase | <i>Bacillus agar-adhearens</i> | 1a3h | 30 | 1,57 | 300 | Mono-mer | 39 | 2,5 |
| | | <i>Thermomonospora fusca</i> | 1bqc | 50 | 1,50 | 302 | | | |
| 4 | Xyloseisomerase | <i>Actinoplanes missouriensis</i> | 2xim | 38 | 2,30 | 393 | Tetra-mer | 66 | 1,6 |
| | | <i>T. thermophilus</i> | 1bxb | 70–75 | 2,20 | 387 | | | |
| 5 | Elongationsfaktor TU | <i>Escherichia coli</i> | 1efc | 37 | 2,05 | 393 | Mono-mer | 84 | 10,5 |
| | | <i>T. thermophilus</i> | 1exm | 70–75 | 1,70 | 405 | | | |
| 6 | Signalerkennungspartikel (SRP) | <i>E. coli</i> | 1fts | 37 | 2,20 | 295 | Mono-mer | 73 | 2,4 |
| | | <i>Thermus aquaticus</i> | 1ffh | 70–72 | 2,05 | 294 | | | |
| 7 | Glyceraldehyd-3-Phosphatdehydrogenase (GAPDH) | <i>Homo sapiens</i> | 1znq | 37 | 2,50 | 338 | Tetra-mer | 70 | 0,8 |
| | | <i>B. stearothermophilus</i> | 1gd1 | 52,5 | 1,80 | 334 | | | |
| 8 | Methylguanin-DNA-Methyltransferase | <i>E. coli</i> | 1sfe | 37 | 2,10 | 180 | Mono-mer | 40 | 2,1 |
| | | <i>Pyrococcus kodakaraensis</i> | 1mgt | 55 | 1,80 | 174 | | | |
| 9 | Carboxypeptidase | <i>Bos taurus</i> | 1m4l | 37 | 1,25 | 307 | Mono-mer | 37 | 1,4 |
| | | <i>Thermoactinomyces vulgaris</i> | 1obr | 55 | 2,30 | 326 | | | |
| 10 | 3-Isopropylmalatdehydrogenase (IPMDH) | <i>E. coli</i> | 1cm7 | 37 | 2,06 | 363 | Dimer | 63 | 2,1 |
| | | <i>T. thermophilus</i> | 1osj | 75 | 2,35 | 349 | | | |
| 11 | Phosphoglyceratkinase | <i>T. brucei</i> | 16pk | 37 | 1,60 | 415 | Mono-mer | 60 | 3,1 |
| | | <i>Thermatoga maritima</i> | 1vpe | 80 | 2,00 | 398 | | | |
| 12 | Cyclodextrin-Glycosyltransferase | <i>Bacillus circulans</i> | 1cgt | 30 | 2,00 | 684 | Mono-mer | 64 | 1,4 |
| | | <i>B. stearothermophilus</i> | 1qho | 55 | 1,70 | 688 | | | |
| 13 | Carboanhydrase | <i>E. coli</i> | 1lxa | 37 | 2,60 | 262 | Mono-mer | 41 | 2,3 |
| | | <i>Methanosarcina fusca</i> | 1qq0 | 50 | 1,76 | 247 | | | |
| 14 | Zellulase | <i>Clostridium cellolyticum</i> | 1g87 | (56,4) | 1,60 | 614 | Mono-mer | 75 | 1,6 |
| | | <i>T. fusca</i> | 1tf4 | 55 | 1,90 | 605 | | | |
| | | | | (72,2) | | | | | |

| | | | | | | | | | |
|----|--|---|------|--------|------|-----|--------------|----|-----|
| 15 | thermolysin- ähnliche Protease (TLP) | <i>Bacillus cereus</i> | 1npc | 30 | 2,00 | 317 | Mono- mer | 73 | 0,9 |
| | | <i>Bacillus thermo- proteolyticus</i> | 1kei | 55 | 1,70 | 316 | | | |
| 16 | Subtilase | <i>Bacillus lentus</i> | 1st3 | 26 | 1,40 | 269 | Mono- mer | 53 | 1,3 |
| | | <i>T. vulgaris</i> | 1thm | 60 | 1,37 | 279 | | | |
| 17 | Endozellulase | <i>Mycobacterium tuberculosis</i> | 1uoz | 30 | 1,10 | 309 | Mono- mer | 50 | 1,5 |
| | | <i>T. fusca</i> | 1tml | 50 | 1,80 | 286 | | | |
| 18 | Histidin- transportprotein | <i>B. subtilis</i> | 2hpr | 37 | 2,00 | 87 | Mono- mer | 79 | 0,6 |
| | | | | (74,4) | | | | | |
| | | <i>B. stearo- thermophilus</i> | 1y4y | 55 | 2,00 | 88 | | | |
| | | | | (88,9) | | | | | |
| 19 | Adenylatkinase | <i>B. subtilis</i> | 1p3j | 37 | 1,90 | 217 | Mono- mer | 88 | 1,6 |
| | | | | (43,3) | | | | | |
| | | <i>B. stearo- thermophilus</i> | 1zio | 55 | 1,96 | 217 | | | |
| | | | | (74,5) | | | | | |

^a Optimale Wachstumstemperatur des Organismus, aus dem ein Protein stammt, in °C; Die Schmelztemperatur (T_m) ist in Klammern angegeben, wenn verfügbar; ^b In Å; ^c In %.

Den Strukturen aus dem Datensatz wurden Daten über ihre thermische Stabilität zugeordnet. Allerdings waren nur für einige wenige Proteine Schmelztemperaturen (T_m) verfügbar. Schmelztemperaturen sind ein sehr guter Deskriptor für die Thermostabilität von Proteinen, können aber nur gemessen werden, wenn der Prozess der Denaturierung der Struktur vollständig reversibel ist. Allen Proteinen aus dem Datensatz konnten dagegen optimale Wachstumstemperaturen (T_{og}) der Organismen zugeordnet werden, aus denen sie stammen. Optimale Wachstumstemperaturen wurden bereits in anderen Vergleichsstudien meso- und thermophiler Proteine als Deskriptor für die Thermostabilität verwendet (Gromiha *et al.* 1999). Neuere Untersuchungen zeigen aber, dass die Korrelation zwischen T_m und T_{og} schwächer ist als ursprünglich vermutet (Dehouck *et al.* 2008). Dies muss bei der späteren Auswertung der Ergebnisse berücksichtigt werden.

5.1.2 Thermostabilitäten homologer meso- und thermophiler Proteine

Für alle meso- und thermophilen Proteine aus dem Datensatz wurde die thermische Entfaltung mit Hilfe der CNA simuliert. Globale Eigenschaften dieser Simulationen werden in Abschnitt 5.1.2.1 am Beispiel einer thermolysinähnlicher Protease (TLP) aus dem mesophilen Organismus *Bacillus cereus* beschrieben. Mit Hilfe von Parametern

aus der Perkulations- und Netzwerktheorie können Veränderungen der Rigidität und Flexibilität in den Strukturen der meso- und thermophilen Proteine während der Entfaltung untersucht und hinsichtlich der Thermostabilität interpretiert werden. Insbesondere können Phasenübergangstemperaturen (T_p) bestimmt werden, die experimentell bestimmten Schmelztemperaturen (T_m) entsprechen. Vorhergesagte thermische Stabilitäten der homologen meso- und thermophilen Proteine werden in Abschnitt 5.1.2.2 miteinander verglichen. Außerdem wird die Gültigkeit der mit Hilfe der CNA durchgeführten thermischen Entfaltungssimulation besprochen. In Abschnitt 5.1.2.3 wird untersucht, welchen Einfluss Energieminimierungen der Kristallstruktur auf die Ergebnisse der Simulation haben, und ob dieses Verfahren verwendet werden kann, um kleinere strukturelle Defekte zu beheben.

5.1.2.1 Globale Eigenschaften der Entfaltungssimulation

Einzelne „Snapshots“ aus der thermischen Entfaltungssimulation von *B. cereus* TLP sind in Abbildung 9 gezeigt. Insbesondere sind hier bei jeweils fünf verschiedenen Temperaturen die Zerlegung des Netzwerks in rigide Cluster und flexible Regionen (*rigid cluster decomposition*, RCD) dargestellt. Bei niedrigen Temperaturen trägt eine hohe Anzahl nichtkovalenter Wechselwirkungen zur Rigidität im Netzwerk bei. Bei 307 K (Abbildung 9 a) wird die Struktur deswegen fast vollständig von einem einzelnen großen rigiden Cluster durchzogen. Mit zunehmender Temperatur werden immer mehr nichtkovalente Wechselwirkungen gebrochen, und die Rigidität im Netzwerk nimmt ab. Dies führt zum Verlust der strukturellen Spezifität und hat die Entfaltung des Proteins zur Folge. Bei 320 K sind bereits erste Bereiche vom großen rigiden Cluster abgebrochen (Abbildung 9 b). Bei 348 K ist nur noch ein kleiner Rest des großen rigiden Clusters vorhanden (Abbildung 9 d). Bei 362 K ist das Netzwerk schließlich bis auf wenige einzelne Sekundärstrukturelemente oder Seitenketten umfassende Bereiche vollständig flexibel (Abbildung 9 e).

Eine genauere Analyse der thermischen Entfaltungssimulation von *B. cereus* TLP zeigt, dass die Abnahme der Rigidität im Netzwerk nicht kontinuierlich, sondern schrittweise geschieht. Jeder Schritt stellt ein „katastrophales“, mit einem Phasenübergang erster Ordnung vergleichbares Ereignis dar, bei dem durch das Entfernen einer einzigen H- oder Salzbrücke große Bereiche von einem großen rigiden Clusters im Netzwerk abbrechen. Dieses Verhalten kann mit der Rigiditätsperkolation in Netzwerken glasbildender Materialien verglichen werden (Thorpe 1983). In den homogenen Netzwerken glasbildender Materialien kann ein scharfer Übergang von einer rigiden zu

einer flexiblen Phase beobachtet werden. Der Phasenübergang wird durch den plötzlichen Zerfall eines das gesamte Netzwerk durchziehenden („perkolierenden“) großen rigiden Clusters verursacht und entspricht dem Schmelzen des Glases, d.h. dem Phasenübergang vom festen zum flüssigen Material. Der perkolierende große rigide Cluster im Netzwerk wird auch „Riesencluster“ (*giant cluster*) bezeichnet (Stauffer und Aharony 1994).

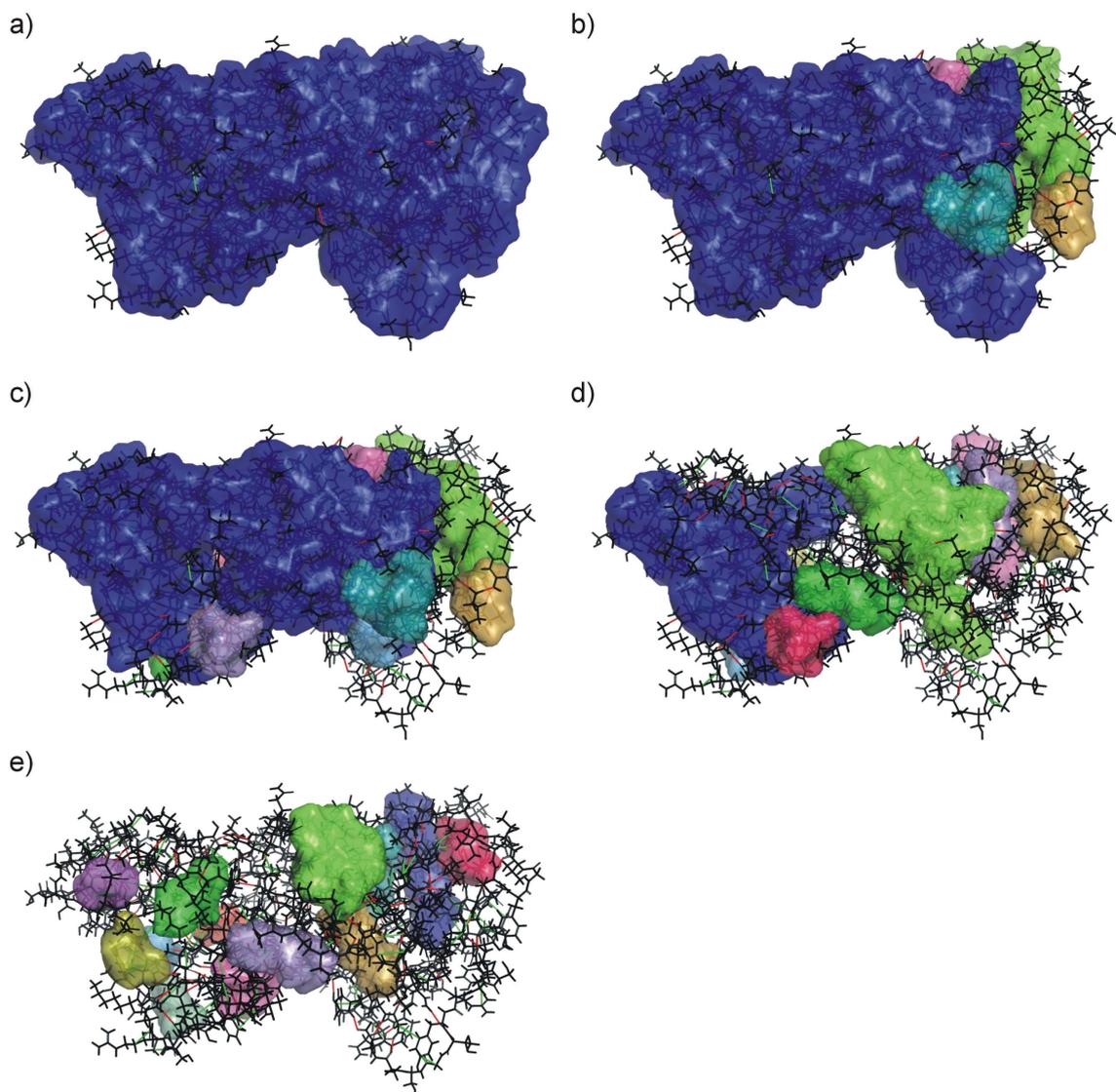


Abbildung 9. „Snapshots“ aus der thermischen Entfaltungssimulation von thermolysinähnlicher Protease (TLP) aus dem mesophilen Organismus *B. cereus* bei 306 (a), 320 (b), 334 (c), 348 (d) und 362 K (e). Es ist jeweils die Zerlegung des Netzwerks in rigide Cluster (RCD) gezeigt. Die rigiden Cluster sind als einheitlich gefärbte Bereiche dargestellt, wobei der jeweils größte rigide Cluster im Netzwerk blau und die übrigen Cluster willkürlich eingefärbt sind. H- und Salzbrücken im Netzwerk sind als rote, hydrophobe Kontakte als grüne Balken zwischen den wechselwirkenden Atomen dargestellt.

Ein Phasenübergang kann offensichtlich auch während der Simulation der thermischen Entfaltung von Proteinen beobachtet werden. Allerdings hat man es hier mit einer Rigiditätsperkolation in mehreren Schritten zu tun, die die Heterogenität von Proteinnetzwerken widerspiegelt. Die Heterogenität von Proteinnetzwerken ist dadurch bedingt, dass Proteinstrukturen aus mehreren Modulen (Untereinheiten, Domänen oder Subdomänen) bestehen, deren Kontaktflächen Bereiche unterschiedlichen Vernetzungsgrads darstellen. Die schrittweise Rigiditätsperkolation reflektiert diesen modularen Aufbau von Proteinen. Jeder Schritt entspricht dem Abbrechen eines Strukturelements vom Riesencluster und stellt ein Zerfallsereignis dar, das zum Phasenübergang vom gefalteten zum ungefalteten Protein beiträgt.

5.1.2.2 Vergleich vorhergesagter Thermostabilitäten

Charakteristisch für die thermische Entfaltungssimulation von Proteinen ist offenbar der schrittweise Zerfall des Riesenclusters im Netzwerk. Für die Analyse der thermischen Entfaltungssimulationen wurde in dieser Arbeit dementsprechend ein Parameter verwendet, mit dem dieser Zerfall des Riesenclusters beobachtet werden kann: Der einheitslose Rigiditätsordnungsparameters (P_∞) ist definiert als die normalisierte Größe des größten Clusters im Netzwerk (Stauffer und Aharony 1994). Mit Hilfe dieses Parameters kann das Perkolationsverhalten der Netzwerke homologer meso- und thermophiler Proteine aus dem Datensatz verglichen werden, ohne dass die RCD oder die Netzwerkarchitektur im Detail untersucht werden müssen. Für die Analyse wird P_∞ als Funktion der Temperatur aufgetragen.

Der Verlauf des Rigiditätsordnungsparameters für zwei Paare meso- und thermophiler Proteine (TLP und IPMDH) ist in Abbildung 10 a) und b) gezeigt. Für das mesophile Protein ist der Verlauf als durchgezogene Linie dargestellt, für das thermophile Protein als gestrichelte Linie. Der P_∞ spiegelt die schrittweise Rigiditätsperkolation in den Proteinnetzwerken sehr gut wieder. Der Verlauf des Rigiditätsordnungsparameters für homologe Proteine ist trotz der Komplexität des Perkolationsverhaltens sehr ähnlich. Im Fall von TLP beispielsweise zerfällt der Riesencluster in drei Schritten (Abbildung 10 a). Dies ist sowohl für das meso- als auch für das thermophile Protein der Fall. Auch für meso- und thermophile IPMDH kann ein solches korrespondierendes Verhalten beobachtet werden (Abbildung 10 b). Der einzige deutliche Unterschied zwischen den Verläufen des P_∞ für die homologen meso- und thermophilen Proteine besteht in einer horizontalen Verschiebung der Werte entlang der Temperaturachse. Die Zerfallsereignisse im Netzwerk des thermophilen Proteins finden also bei höheren

Temperaturen statt als die im Netzwerk des mesophilen Proteins. Der Vergleich des Perkulationsverhaltens homologer Proteine gibt damit einen ersten Hinweis auf die höhere Thermostabilität des thermophilen Proteins.

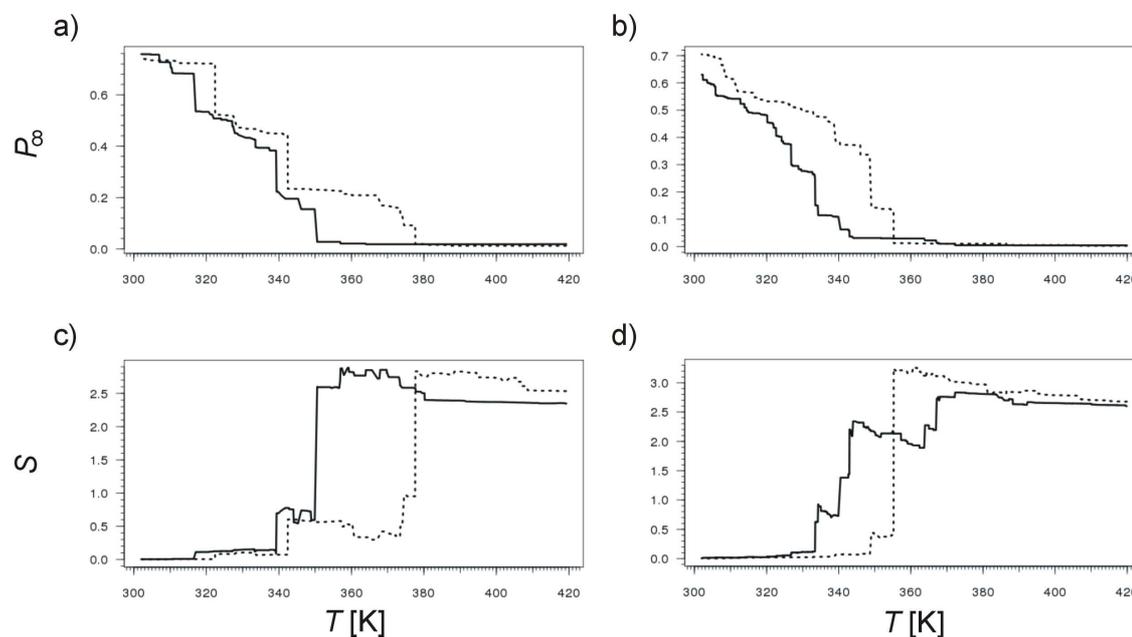


Abbildung 10. Verlauf des Rigiditätsordnungsparameters (P_∞) (a, b) und der Clusterkonfigurationsentropie (S) (c, d) für meso- und thermophile thermolysinähnliche Protease (TLP) (a, c) und 3-Isopropylmalatdehydrogenase (IPMDH) (b, d). Der Verlauf ist für das mesophile Protein als durchgezogene Linie dargestellt, für das thermophile als gestrichelte Linie. Die Parameter wurden aus den thermischen Entfaltungssimulationen der Kristallstrukturen von TLP und IPMDH berechnet.

Ein quantitativer Vergleich der vorhergesagten Thermostabilitäten homologer meso- und thermophiler Proteine erfordert die Bestimmung des Zerfallsereignisses, das dem Phasenübergang zwischen dem gefalteten und ungefalteten Protein entspricht. Zur Bestimmung dieses Phasenübergangs wurde in dieser Arbeit ein Parameter aus der Netzwerktheorie eingeführt und verwendet: Die einheitslose Clusterkonfigurationsentropie (S) ist definiert als eine Funktion der Wahrscheinlichkeit, dass sich ein Atom in einem Cluster einer bestimmten Größe befindet (Andraud *et al.* 1994). Diese Entropie wird dazu verwendet, die Temperatur T_p zu identifizieren, bei der der Riesencluster aufhört, das System zu dominieren.

Für die Analyse wird S als Funktion der Temperatur aufgetragen. Ein sprunghafter Anstieg der Entropie weist auf das Zerfallsereignis hin, bei dem ein letzter Rest des Riesenclusters zerfällt und kein großer Cluster mehr die Struktur dominiert. Dieser Zerfall eines „rigiden Kerns“ entspricht dem Verlust eines Faltungskerns in der Struktur und

somit dem gesuchten Übergang zwischen dem gefalteten und ungefalteten Protein (Hespenheide *et al.* 2002; Rader und Bahar 2004). Die Temperatur, bei der dieses Ereignis stattfindet, ist die gesuchte T_p und kann mit der Schmelztemperatur T_m des Proteins verglichen werden.

Der Verlauf der Clusterkonfigurationsentropie für zwei Paare meso- und thermophiler Proteine (TLP und IPMDH) ist in Abbildung 10 c) und d) gezeigt. Für das mesophile Protein ist der Verlauf als durchgezogene Linie dargestellt, für das thermophile Protein als gestrichelte Linie. Der Phasenübergang ist jeweils eindeutig zu erkennen. Abbildung 10 ermöglicht den direkten Vergleich der Clusterkonfigurationsentropie und des Rigiditätsordnungsparameters. Der aus dem Verlauf der Clusterkonfigurationsentropie bestimmte Phasenübergang entspricht tatsächlich dem Zerfall eines letzten Rests des Riesenclusters.

Aus dem Verlauf der Clusterkonfigurationsentropie können neben dem Phasenübergang auch noch weitere Übergänge bestimmt werden, die mit weniger drastischen Veränderungen der Entropie verbunden sind und für die thermische Stabilität des Proteins keine oder nur eine untergeordnete Rolle spielen. Sie entsprechen Zerfallsereignissen, bei denen nur ein bestimmter Teil des Riesenclusters zerfällt und die mikroskopische Stabilität einer bestimmten Region im Protein verändert wird, ohne dass dies einen Einfluss auf den globalen Zerfall der Struktur hätte.

Im Fall der zwei näher betrachteten Paare meso- und thermophiler Proteine (TLP und IPMDH) kann für die thermophile Struktur eine höhere Phasenübergangstemperatur (T_p) bestimmt werden (373 bzw. 355 K) als für die entsprechende mesophile Struktur (350 bzw. 340 K). Absolute T_p -Werte sollen an dieser Stelle aber nicht mit experimentellen Daten über die Thermostabilität verglichen werden, da es sich bei den verfügbaren experimentellen Daten fast ausschließlich um optimale Wachstumstemperaturen (T_{og}) der Organismen handelt, aus denen die Proteine stammen, und nicht um Schmelztemperaturen (T_m). Zudem können auf der verwendeten Temperaturskala nur homologe Proteine miteinander verglichen werden (Kap. 4.4.5). Aus diesen Gründen werden hier lediglich die Vorzeichen von ΔT_p und ΔT_{og} verglichen.

Differenzen der optimalen Wachstumstemperaturen der Organismen und Phasenübergangstemperaturdifferenzen aller Paare meso- und thermophiler Proteine aus dem Datensatz sind in Tabelle 2 aufgelistet. Für rund zwei Drittel der Paare kann für die thermophile Struktur in Übereinstimmung mit den experimentellen Daten eine höhere Phasenübergangstemperatur (T_p) bestimmt werden als für die entsprechende mesophile

Struktur. Dieses Ergebnis ist ein klarer Hinweis darauf, dass die thermophile Anpassung tatsächlich über eine Rigidisierung der Struktur erfolgt, und dass die CNA dazu verwendet werden kann, die thermische Stabilität einer Struktur vorherzusagen.

Tabelle 2. Differenzen der optimalen Wachstumstemperaturen (ΔT_{og}) der Organismen und Phasenübergangstemperaturdifferenzen (ΔT_p) aller Paare meso- und thermophiler Proteine.

| Proteinfamilie ^a | ΔT_{og} ^b | ΔT_p ^c |
|-----------------------------|------------------------------|---------------------------|
| 1 | 53,0 | 18,0 |
| 2 | 35,5 | -7,2 |
| 3 | 20,0 | -11,3 |
| 4 | 34,5 | 11,5 |
| 5 | 35,5 | 9,7 |
| 6 | 34,0 | 15,3 |
| 7 | 15,5 | -0,4 |
| 8 | 58,0 | 10,9 |
| 9 | 18,0 | 7,2 |
| 10 | 38,0 | 15,3 |
| 11 | 43,0 | 1,1 |
| 12 | 25,0 | 29,8 |
| 13 | 13,0 | 4,9 |
| 14 | 20,0 | -16,8 |
| 15 | 25,0 | 23,6 |
| 16 | 34,0 | 2,8 |
| 17 | 20,0 | -25,9 |
| 18 | 18,0 | 16,2 |
| 19 | 18,0 | -4,7 |

^a Nummern entsprechen den Nummern der Proteinfamilien in Tabelle 1; ^b Differenz der optimalen Wachstumstemperaturen der Organismen, aus denen die Proteine stammen, in K; ^c In K; ΔT_p ist fett und kursiv markiert, wenn die Vorzeichen von ΔT_p und ΔT_{og} übereinstimmen.

Für das restliche Drittel der Paare meso- und thermophiler Proteine kann die erwartete höhere Phasenübergangstemperatur (T_p) der thermophilen Struktur nicht erfolgreich vorhergesagt werden. Eine mögliche Ursache dieser nicht erfolgreichen Vorhersagen liegt darin, dass die thermophile Anpassung *in vivo* auch mit Hilfe extrinsischer Faktoren erfolgen kann. Solche Faktoren sind beispielsweise Glykosylierungen der Proteinstruktur oder der Einfluss von Lösungsmitteln (*compatible solutes*). So wurde für Lysozym in einem Sucrose-Wasser- (Cornicchi *et al.* 2006) und Glycerol-Wasser-Gemisch (de Francesco *et al.* 2004) eine Verringerung der thermischen Stabilität und Rigidität

mit zunehmendem Wasseranteil und damit abnehmender Lösungsmittelviskosität beobachtet. Ein solcher Einfluss des Lösungsmittels kann mit Hilfe der CNA nicht untersucht werden, da Lösungsmittelmoleküle bei der Konstruktion der Netzwerke nicht explizit berücksichtigt werden. Zudem ist die Verfügbarkeit experimenteller Daten über die Abhängigkeit der Thermostabilität von extrinsischen Faktoren begrenzt. Eine weitere mögliche Ursache der nicht erfolgreichen Vorhersagen liegt in der Verwendung von optimalen Wachstumstemperaturen anstelle von Schmelztemperaturen. Tatsächlich gibt es mesophile Proteine, deren Schmelztemperatur höher ist als die eines homologen thermophilen Proteins, wie das für die α -Amylasen aus *Bacillus licheniformis* (einem mesophilen Organismus) und *Bacillus stearothermophilus* (einem thermophilen Organismus) gezeigt werden konnte (Fitter *et al.* 2001).

Letztendlich kann die Ursache der nicht erfolgreichen Vorhersagen aber auch darin liegen, dass die Berechnung der Energie von H- und Salzbrücken, sowie die Modellierung der Temperaturabhängigkeit dieser Wechselwirkungen und der hydrophoben Kontakte Schwächen aufweist. So werden beispielsweise bei der Berechnung der Energie von H- und Salzbrücken kooperative Effekte völlig außer Acht gelassen. Die Tatsache, dass Salzbrücken bei höheren Temperaturen stärker werden, wird ebenfalls nicht berücksichtigt. Auch die Modellierung hydrophober Wechselwirkungen als temperaturabhängig stellt nur eine Näherung dar. An dieser Stelle müssen zukünftige Arbeiten ansetzen.

5.1.2.3 Einfluss von Energieminimierungen der Kristallstruktur

Eine genauere Betrachtung von Abbildung 10 d) zeigt, dass der aus dem Verlauf der Clusterkonfigurationsentropie für IPMDH bestimmte Phasenübergang des mesophilen Proteins weniger scharf ist als der für das thermophile Protein. Für das mesophile Protein kann zudem ein Übergang bei 364 K bestimmt werden, der keine Entsprechung im Verlauf der Clusterkonfigurationsentropie für das thermophile Protein hat. Diese Beobachtungen deuten auf einen möglichen Defekt der Kristallstruktur von *Escherichia coli* IPMDH hin. Eingehende Untersuchungen der Struktur zeigen, dass tatsächlich ein durch Kristallkontakte verursachter Defekt (Deformation) besteht (Abschnitt 5.1.3.2).

Um zu testen, ob ein solcher Defekt der Kristallstrukturen mit Hilfe einfacher rechnerischer Ansätze behoben werden können, wurden die Strukturen von meso- und thermophiler IPMDH kurzen Energieminimierungen unterzogen. Die kurzen Minimierungen bringen nur minimale Veränderungen der Struktur mit sich. So betrug die RMS-Abweichung der minimierten von den originalen *E. coli* IPMDH Strukturen weniger als

0,3 Å für die $C\alpha$ -Atome. Unter Verwendung der minimierten Strukturen wurde die thermische Entfaltung erneut simuliert und mit den Entfaltungssimulationen der originalen Strukturen verglichen.

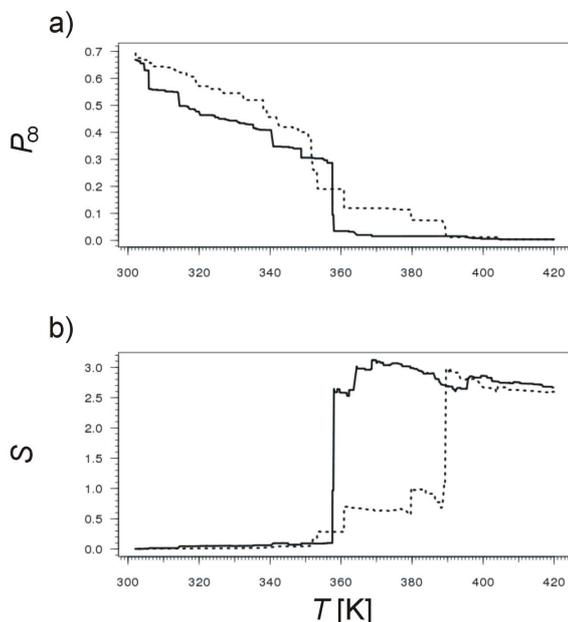


Abbildung 11. Verlauf des Rigiditätsordnungsparameters (P_∞) (a) und der Clusterkonfigurationsentropie (S) (b) für meso- und thermophile 3-Isopropylmalatdehydrogenase (IPMDH). Der Verlauf ist für das mesophile Protein als durchgezogene Linie dargestellt, für das thermophile als gestrichelte Linie. Die Parameter wurden aus den thermischen Entfaltungssimulationen der energieminierten Strukturen von IPMDH berechnet.

Der Verlauf des Rigiditätsordnungsparameters und der Clusterkonfigurationsentropie für die minimierte meso- und thermophile IPMDH ist in Abbildung 11 a) und b) gezeigt. Für das mesophile Protein ist der Verlauf als durchgezogene Linie dargestellt, für das thermophile Protein als gestrichelte Linie. Der Verlauf beider Parameter zeigt, dass die Strukturen durch die Energieminimierung um rund 20–30 K thermostabiler geworden sind als die originalen Strukturen. Dieser stabilisierende Effekt der Minimierung kann darauf zurückgeführt werden, dass insbesondere die elektrostatischen Wechselwirkungen im Protein (H- und Salzbrücken) verbessert werden und somit eine Struktur unnatürlich hoher Qualität erzeugt wird. Die Phasenübergangstemperaturdifferenz zwischen dem meso- und thermophilen Protein bleibt trotz dieses Effekts aber erhalten. Diese Beobachtung bestätigt, dass thermische Entfaltungssimulationen nur dann auf der selben Temperaturskala verglichen werden können, wenn sie auf Strukturen „gleicher Herkunft“ basieren.

Der Verlauf der Clusterkonfigurationsentropie für die minimierte Struktur von *E. coli* IPMDH unterscheidet sich vom Verlauf dieses Parameters für die originale Struktur. So ist der Phasenübergang nach der Minimierung deutlich schärfer als vor der Minimierung. Zudem ist der im Verlauf der Clusterkonfigurationsentropie für die originale Struktur des mesophilen Proteins zusätzlich erkennbare Übergang bei 364 K nach der Minimierung nicht mehr vorhanden. Insgesamt scheint das Perkulationsverhalten von meso- und thermophiler IPMDH nach der Minimierung ähnlicher zu sein als vor der Minimierung, d.h. kleine strukturelle Defekte können mit Hilfe rechnerischer Ansätze behoben werden. Allerdings werden durch die Energieminimierung Strukturen erzeugt, die eine unnatürlich hohe Qualität haben, was offensichtlich einen großen Einfluss auf die vorhergesagte Thermostabilität hat. Im weiteren Verlauf der Arbeit wird daher auf die Verwendung energieminimierter Strukturen verzichtet, zumal der Defekt der Struktur von *E. coli* IPMDH nur Ereignisse betrifft, die jenseits des Phasenübergangs und im ungefalteten Zustand des Proteins stattfinden (Abschnitt 5.1.3.2).

5.1.3 Molekulare Ursachen erhöhter Thermostabilität

Im vorangehenden Kapitel wurde gezeigt, dass für rund zwei Drittel der Paare meso- und thermophiler Proteine aus dem Datensatz für die thermophile Struktur eine höhere Phasenübergangstemperatur (T_p) bestimmt werden kann als für die entsprechende mesophile Struktur. Dieses Ergebnis deutet klar darauf hin, dass die thermophile Anpassung tatsächlich über eine Rigidisierung der Struktur erfolgt. Um aber die molekularen Ursachen dieser Rigidisierung thermophiler Proteine und der damit verbundenen erhöhten Thermostabilität verstehen zu können, muss das beobachtete makroskopische Perkulationsverhalten der Proteinnetzwerke mit mikroskopischen Eigenschaften der Struktur verknüpft werden. In den Abschnitten 5.1.3.1 und 5.1.3.2 werden die Entfaltungssimulationen von meso- und thermophiler thermolysinähnlicher Protease (TLP) bzw. 3-Iso-propylmalatdehydrogenase (IPMDH) eingehend analysiert. Dabei werden (lokale) Strukturbereiche bestimmt, an denen der (globale) Zerfall der Struktur beginnt, und die deswegen für die thermische Stabilität von großer Wichtigkeit sind. Diese Entfaltungsregionen werden für die untersuchten Proteine mit Regionen verglichen, in denen thermostabilisierende Mutationen experimentell eingeführt wurden.

5.1.3.1 Entfaltungsregionen von thermolysinähnlichen Proteasen

Thermolysinähnliche Proteasen sind Mitglieder der Peptidasefamilie M4 (Adekoya und Sylte 2009). Der Prototyp dieser Familie ist Thermolysin, das aus dem thermophi-

len Organismus *Bacillus thermoproteolyticus* stammt. Zahlreiche Mitglieder der Familie werden biotechnologisch genutzt, beispielsweise bei der Lederproduktion zur Entfernung von Haaren, Haut- und Fleischresten (Braun *et al.* 2006). Thermolysin selbst wird für die Synthese des künstlichen Süßstoffs Aspartam verwendet (Adekoya und Sylte 2009). Alle TLP benötigen ein Zinkion, um aktiv zu sein. Außerdem enthalten sie mehrere Calciumionen, die für ihre Stabilität wichtig sind (Corbett und Roche 1983). Im Datensatz meso- und thermophiler Proteine ist die TLP aus dem mesophilen Organismus *B. cereus* und Thermolysin enthalten. *B. cereus* TLP und Thermolysin haben eine hohe Sequenzidentität und sind auch strukturell sehr ähnlich (Tabelle 1). Die Strukturen bestehen aus einer N-terminalen Domäne (Reste 1–146), die eine α -Helix sowie ein β -Faltblatt enthält, und einer vorrangig α -helikalen C-terminalen Domäne (Reste 147–317). Die beiden Domänen sind über eine zentrale α -Helix (Reste 139–154) verbunden, die den Boden der aktiven Tasche bildet und mehrere für die Aktivität wichtige Reste (das so genannte aktive Zentrum) enthält. Diese Reste sowie Sekundärstrukturelemente der TLP sind in dem in Abbildung 7 gezeigten Alignment des meso- und thermophilen Enzyms gekennzeichnet. Grundsätzlich entspricht die in diesem Abschnitt verwendete Nummerierung der Reste der Nummerierung in diesem Alignment.

Die Organismen, aus denen mesophile TLP und Thermolysin stammen (*B. cereus* und *B. thermoproteolyticus*) haben optimale Wachstumstemperaturen von 30 bzw. 55 °C. Schmelztemperaturen können nicht bestimmt werden, da der Vorgang der Denaturierung für diese Proteine irreversibel ist (Eijssink *et al.* 1995). Unter denaturierenden Bedingungen (wie erhöhter Temperatur) führen lokale Entfaltungseignissen zur Autolyse und damit zur irreversiblen Zerstörung des Enzyms. Generell können irreversible, den oligomeren Zustand oder die chemischen Struktur eines Proteins verändernde Prozesse (Ahern und Klivanov 1988; Vieille und Zeikus 2001) nicht mit Hilfe der CNA modelliert werden. Somit wird die reversible thermische Entfaltung simuliert, obwohl diese nicht experimentell zugänglich ist. Dies erschwert einen späteren Vergleich der Simulationsergebnisse mit experimentellen Daten, wobei aber davon ausgegangen werden kann, dass die zur reversiblen Entfaltung führenden Vorgänge den experimentell untersuchten Ereignissen entsprechen, die zur irreversiblen Denaturierung führen (Kap. 1.2.4) (Lumry und Eyring 1954).

Veränderungen der Rigidität und Flexibilität im Netzwerk meso- und thermophiler TLP während der Simulation der thermischen Entfaltung wurden in Kap. 5.1.2 mit Hilfe des Rigiditätsordnungsparameters und der Clusterkonfigurationsentropie untersucht.

Die Ergebnisse dieser Analysen deuteten klar auf ein korrespondierendes Perkolationsverhalten von *B. cereus* TLP und Thermolysin hin. Der Verlauf des Rigiditätsordnungsparameters zeigte, dass während der Entfaltung der *B. cereus* TLP und des Thermolysins das Riesencluster in jeweils drei Schritten zerfällt (Abbildung 10 a). Die Analyse

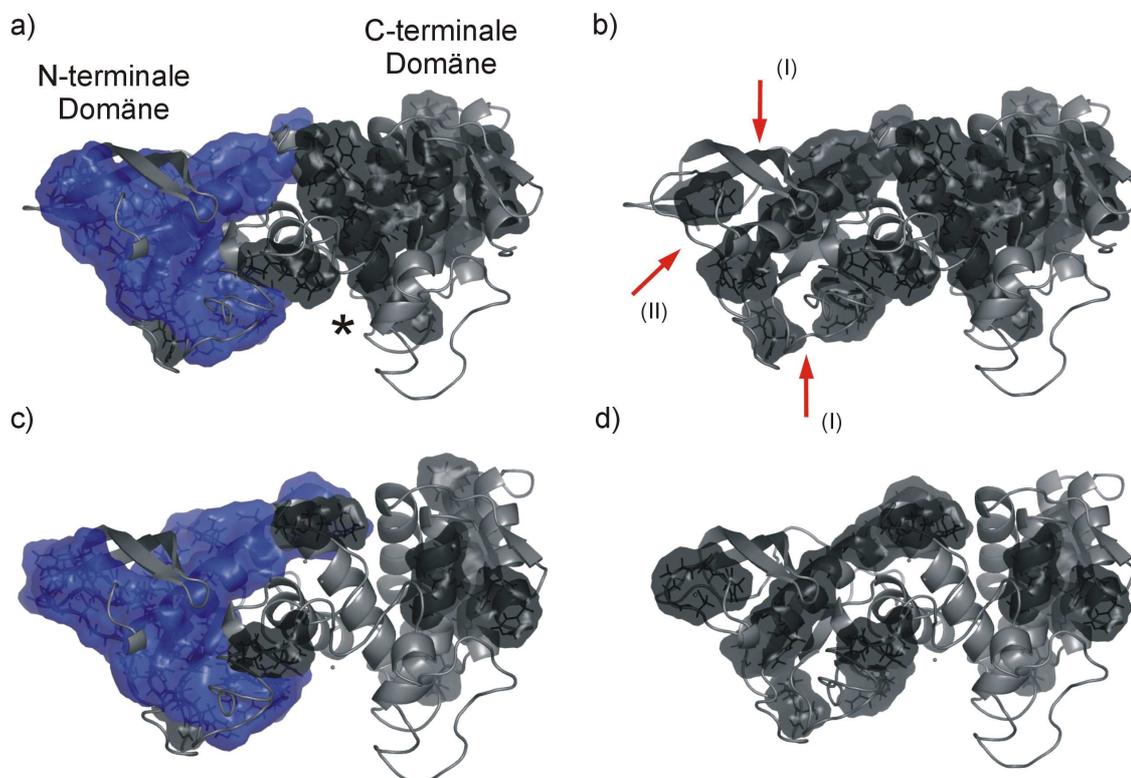


Abbildung 12. Snapshots aus der thermischen Entfaltungssimulation von meso- (a, b) und thermophiler (c, d) thermolysinähnlicher Protease (TLP) unmittelbar vor (a, c) und nach dem Phasenübergang (b, d) bei 350 bzw. 373 K. Es ist jeweils die Zerlegung des Netzwerks in rigide Cluster (RCD) gezeigt. Die rigiden Cluster sind als einheitlich gefärbte Bereiche dargestellt, wobei der Riesencluster blau eingefärbt ist. Die übrigen Cluster sind dunkel schattiert. Von der Proteinstruktur ist nur das Rückgrat gezeigt und dunkel-grau schattiert. Die Pfeile kennzeichnen die für das mesophile Protein identifizierten Entfaltungsregionen. Der Stern kennzeichnet die aktive Tasche.

des Verlaufs der Clusterkonfigurationsentropie ergab, dass jeweils das letzte Zerfallereignis dem Phasenübergang zwischen dem gefalteten und ungefalteten Protein entsprach (Abbildung 10 c). Dieses makroskopische Perkolationsverhalten kann nun mit mikroskopischen Eigenschaften der Struktur verknüpft werden, indem die RCD für das meso- und thermophile Protein unmittelbar vor und nach dem Phasenübergang bei 350 bzw. 373 K verglichen werden (Abbildung 12).

Aus Abbildung 12 a) und c) ist ersichtlich, dass die Strukturen von *E. coli* TLP und Thermolysin unmittelbar vor dem jeweiligen Phasenübergang von einem rigiden Kern

dominiert werden. Dieser Kern stellt einen rigiden Cluster dar, der die gesamte N-terminale Domäne umfasst, mit der Ausnahme einer β -Schleife (Reste 4–23) und weniger Loopregionen. Auch die α -Helix in der N-terminalen Domäne (Reste 67–89) ist in diesem Cluster enthalten. In Abbildung 12 a) und c) ist dieser Cluster blau eingefärbt. Der Zerfall des Kerns verursacht den Phasenübergang. Aus Abbildung 12 b) und d) ist ersichtlich, dass unmittelbar nach dem Phasenübergang nur noch wenige, einzelne Sekundärstrukturelemente umfassende rigide Cluster in den Netzwerken vorhanden sind. *E. coli* TLP und Thermolysin verhalten sich während des Phasenübergangs sehr ähnlich, obwohl die jeweiligen Phasenübergänge bei deutlich verschiedenen Temperaturen stattfinden.

Unterschiede im Verhalten der homologen Proteine finden sich lediglich in der Anzahl der im Netzwerk enthaltenen Cluster und deren Größe. Die geringere Anzahl und Größe der im Netzwerk des thermophilen Proteins enthaltenen Cluster dient möglicherweise der Kompensation hoher thermischer Energie bei hohen Temperaturen. Trotz dieser Unterschiede befinden sich die homologen Proteine bei Temperaturen um ihre jeweilige Phasenübergangstemperatur in korrespondierenden Zuständen ähnlicher Flexibilität. Die Ähnlichkeit der Größe und Lage der unmittelbar vor dem jeweiligen Phasenübergang in den Strukturen enthaltenen Kerne deutet zudem darauf hin, dass für die Thermostabilität von meso- und thermophiler TLP ähnliche Strukturbereiche von Bedeutung sein müssen.

Die Analyse der Veränderungen der Rigidität und Flexibilität im Netzwerk unmittelbar vor und nach dem Phasenübergang erlaubt die Bestimmung von (lokalen) Strukturbereichen, an denen der (globale) Zerfall der Struktur beginnt, und die deswegen für die thermische Stabilität von großer Wichtigkeit sind. Solche Strukturbereiche sind in der Vergangenheit vielfach mit verschiedenen experimentellen oder rechnerischen Methoden bestimmt worden, was zur Einführung der Begriffe „Schwachstelle“ (*weak point*) (Siddiqui *et al.* 1993; Finkelstein 1997), „kritische Region“ (*critical region*) (Eijsink *et al.* 1995), „Erstentfaltungsregion“ (*first unfolding region*) (Creveld *et al.* 1998) oder „Entfaltungskern“ (*unfolding nucleation site*) (Gaseidnes *et al.* 2003) führte. Methoden, mit denen solche Regionen vorhergesagt werden können, sind hilfreich zur Unterstützung des Protein Engineering, da sie Hinweise geben, in welchen Strukturbereichen thermostabilisierende Mutationen eingeführt werden können.

In dieser Arbeit sind Entfaltungsregionen als die labilsten Bereiche des unmittelbar vor dem Phasenübergang in der Struktur enthaltenen Kerns definiert. Die Entfaltungsre-

gionen werden bestimmt, indem die Reste identifiziert werden, die unmittelbar vor dem Phasenübergang im Riesencluster enthalten und nach dem Übergang flexibel geworden sind. Auf diese Weise können für meso- und thermophile TLP jeweils zwei identische Entfaltungsregionen bestimmt werden. Die erste beinhaltet viele Reste des β -Faltblatts in der N-terminalen Domäne. Die zweite beinhaltet eine Calciumbindestelle (Reste 56, 58, 60, 68), die hydrophobe Region um Phe 63 sowie die N-terminalen Reste der α -Helix in der N-terminalen Domäne (Reste 67–89). In der C-terminalen Domäne können keine Entfaltungsregionen bestimmt werden.

Tabelle 3. Vergleich der vorhergesagten Entfaltungsregionen mesophiler thermolysinähnlicher Protease (TLP) mit Resten aus Strukturbereichen, in denen thermostabilisierende Mutationen experimentell eingeführt wurden.

| Strukturbereich | Reste in vorhergesagten Entfaltungsregionen^{a, b} | Experimentell eingeführte thermostabilisierende Mutationen^{a, b, c} |
|---|---|---|
| I β -Faltblatt in der N-terminalen Domäne | <i>21–24</i> , 29, 31–34, <i>39–42</i> , 44, 101–107, 114–115, <i>122–123</i> | <i>19, 38, 104, 120</i> , 134, 142, 183 |
| II Calciumbindestelle, hydrophobe Region um Phe 63 und N-terminale Reste der α -Helix in der N-terminalen Domäne | 54, <i>56–62, 68–70</i> | 4, 8, <i>57, 59, 61, 64, 66–70</i> |

^a Die Nummerierung der Reste entspricht der Nummerierung in dem in Abbildung 7 gezeigten Alignment meso- und thermophiler TLP; ^b Reste, für die die Vorhersage mit den experimentellen Daten übereinstimmt, sind fett und kursiv markiert. Dies ist dann der Fall, wenn die Reste nicht um mehr als zwei Positionen in der Proteinsequenz voneinander entfernt liegen; ^c Erläuterungen im Text. Nach (Imanaka *et al.* 1986; van den Burg *et al.* 1991; Hardy *et al.* 1993; van den Burg *et al.* 1994; Kidokoro *et al.* 1995; Veltman *et al.* 1996; Mansfeld *et al.* 1997; Veltman *et al.* 1997; van den Burg *et al.* 1998).

Die Entfaltungsregionen des mesophilen Proteins sind in Abbildung 12 d) durch Pfeile gekennzeichnet. Die enthaltenen Reste sind in Tabelle 3 aufgelistet. In Tabelle 3 sind den Entfaltungsregionen zudem die Reste aus Strukturbereichen von meso- und thermophiler TLP gegenübergestellt, in denen thermostabilisierende Mutationen experimentell eingeführt wurden (Imanaka *et al.* 1986; van den Burg *et al.* 1991; Hardy *et al.* 1993; van den Burg *et al.* 1994; Kidokoro *et al.* 1995; Veltman *et al.* 1996; Mansfeld *et al.* 1997; Veltman *et al.* 1997; van den Burg *et al.* 1998).

Die vorhergesagten Entfaltungsregionen stimmen offensichtlich sehr gut mit den experimentellen Daten überein (Tabelle 3). Insbesondere gilt dies für die aus den

N-terminalen Resten 56–69 bestehende Entfaltungsregion. Mit Hilfe verschiedener Experimente konnte gezeigt werden, dass die Bedeutung dieses Strukturbereichs für die thermische Stabilität von meso- und thermophiler TLP darin begründet ist, dass er unter denaturierenden Bedingungen entfaltet und zur Autolyse und damit zur irreversiblen Zerstörung des Enzyms führt (Eijsink *et al.* 1995). Der irreversible Prozess kann nicht mit Hilfe der CNA modelliert werden. Es wird also die reversible thermische Entfaltung simuliert, obwohl diese nicht experimentell zugänglich ist. Der Vergleich der Simulationsergebnisse mit den experimentellen Daten bestätigt aber, dass die zur reversiblen Entfaltung führenden Vorgänge den experimentell untersuchten, zur irreversiblen Denaturierung führenden Vorgängen entsprechen.

Eine geringere Übereinstimmung zwischen der Vorhersage und den experimentellen Daten findet sich für die aus den Resten des β -Faltblatts in der N-terminalen Domäne bestehende Entfaltungsregion. Eine mögliche Ursache dafür liegt darin, dass das Einführen von Mutationen in β -Faltblättern aufgrund deren Größe sehr schwierig ist und deswegen nur selten unternommen wird. Tatsächlich fanden sich in der Literatur keinerlei Berichte über Mutationen im β -Faltblatts in der N-terminalen Domäne, unabhängig davon, ob sie einen Einfluss auf die thermische Stabilität von TLP haben oder nicht. Möglicherweise kann die geringe Übereinstimmung in diesem Fall also auf fehlende experimentelle Daten zurückgeführt werden. Anders sieht es für die Reste in der C-terminalen Domäne aus, wo die Vorhersage wieder sehr gut mit den experimentellen Daten übereinstimmt. Mutationen verschiedener Reste in dieser Domäne hatten nur einen marginalen oder keinen Einfluss auf die thermische Stabilität von TLP (Eijsink *et al.* 1992; Eijsink *et al.* 1992; Eijsink *et al.* 1992; Eijsink *et al.* 1992; Margarit *et al.* 1992; Veltman *et al.* 1996). Entsprechend werden in dieser Domäne auch keine Entfaltungsregionen vorhergesagt.

Insgesamt sprechen diese Ergebnisse für ein Modell, nach dem die (reversible oder irreversible) Entfaltung von Proteinen an definierten Regionen des Moleküls beginnt (Schellenberger und Ulbrich 1989; Ulbrich-Hofmann *et al.* 1999; Arnold *et al.* 2005). Über die gezielte Stabilisierung dieser Entfaltungsregionen kann die Entfaltung hinausgezögert und die Thermostabilität von Proteinen erhöht werden (Mansfeld *et al.* 1999; Arnold *et al.* 2005). Untersuchungen mit der Methode der limitierten Proteolyse (Price und Johnson 1990) zeigten, dass es sich bei den Entfaltungsregionen meist um Strukturbereiche (Loops) handelt, die nicht innerhalb von Sekundärstrukturelementen liegen (Arnold *et al.* 2005). Diese Loops stellen den Kontakt zwischen Sekundärstrukturele-

menten her, deren Integrität nach der Entfaltung der Loops verloren geht. Auch mit Hilfe von MD-Entfaltungssimulationen konnte gezeigt werden, dass Loops und die Kontaktflächen zwischen Sekundärstrukturelementen Entfaltungsregionen darstellen (Finkelstein 1997; Creveld *et al.* 1998). Bei der in mit Hilfe der CNA vorhergesagten, aus den N-terminalen Resten 56–69 bestehenden Entfaltungsregion handelt es sich ebenfalls um einen solchen Loop, dessen Entfaltung zum Verlust der Integration der α -Helix (Reste 67–89) sowie des β -Faltblatts in der N-terminalen Domäne und somit zur (reversiblen oder irreversiblen) Entfaltung führt (Eijsink *et al.* 1995).

5.1.3.2 Entfaltungsregionen von 3-Isopropylmalatdehydrogenasen

3-Isopropylmalatdehydrogenasen sind Enzyme im Biosyntheseweg der Aminosäure Leucin und katalysieren die Dehydrierung und gleichzeitige Decarboxylierung von 3-Isopropylmalat (IPM) zu 2-Oxo-4-Methylvalerat unter Verwendung eines Nikotinamid-Adenindinucleotids (NAD^+) als Kofaktor (Závodszy *et al.* 1998). Im Datensatz meso- und thermophiler Proteine sind die IPMDH aus dem mesophilen Organismus *E. coli* und dem thermophilen Organismus *Thermus thermophilus* enthalten. *E. coli* und *T. thermophilus* IPMDH haben eine Sequenzidentität von 63 % und sind auch strukturell sehr ähnlich (Tabelle 1).

Beide Enzyme liegen als Homodimer vor, wobei jede Untereinheit des meso- und thermophilen Proteins aus 345 bzw. 363 Aminosäuren besteht (Wallon *et al.* 1997). Die Polypeptidkette einer jeden Untereinheit faltet sich in zwei Domänen ähnlicher Topologie, die jeweils aus parallelen α/β -Motiven bestehen. Die erste Domäne enthält den N- und C-Terminus und besteht aus den Resten 1–105 und 256–363 (Wallon *et al.* 1997). Die zweite Domäne enthält das Untereinheits-Interface und besteht aus den Resten 106–255 (Wallon *et al.* 1997). Das aktive Zentrum befindet sich in einer Tasche zwischen den beiden Domänen und besteht aus Resten beider Untereinheiten (Kadono *et al.* 1995), weswegen die IPMDH auch nur im dimeren Zustand aktiv sein kann. Grundsätzlich entspricht die in diesem Abschnitt verwendete Nummerierung der Reste der Nummerierung in dem in Abbildung 8 gezeigten Alignment.

Meso- und thermophile IPMDH unterscheiden sich in der optimalen Wachstumstemperatur der Organismen, aus denen sie stammen. Die T_{og} beträgt für *E. coli* 37 °C, für *T. thermophilus* 75 °C. Zudem unterscheiden sich die Proteine in ihrer Schmelztemperatur: Der T_{m} von *E. coli* IPMDH beträgt 73,5 °C, der von *T. thermophilus* IPMDH 90 °C. Die in dieser Arbeit für die CNA verwendete *T. thermophilus* IPMDH ist eine Mutante (A172L), die sich in ihrer Struktur und Thermostabilität aber nur geringfügig

vom Wildtyp unterscheidet (Qu *et al.* 1997). Die Struktur der Mutante wurde gewählt, da diese eine bessere Qualität aufwies als die Struktur des Wildtyps.

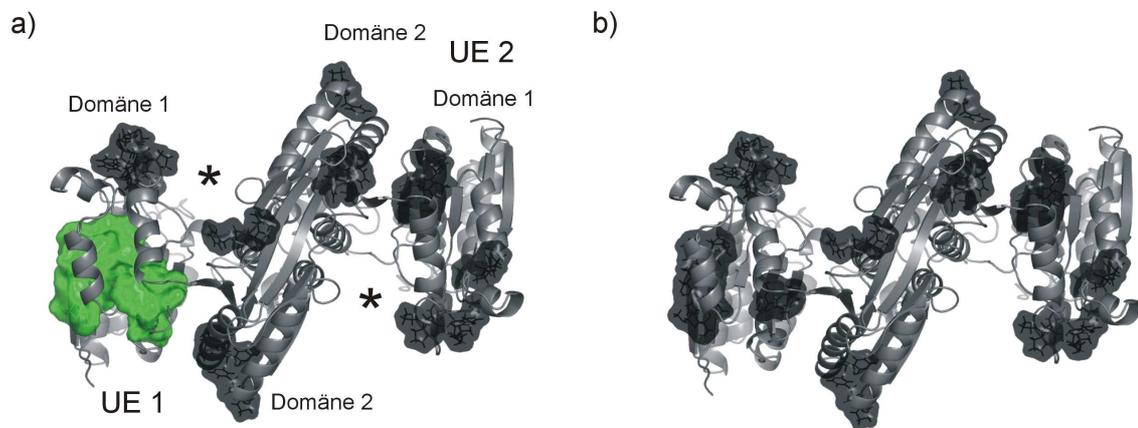


Abbildung 13. Snapshots aus der thermischen Entfaltungssimulation von mesophiler 3-Isopropylmalatdehydrogenase (IPMDH) unmittelbar vor (a) und nach dem Übergang (b) bei 364 K. Es ist jeweils die Zerlegung des Netzwerks in rigide Cluster (RCD) gezeigt. Die rigiden Cluster sind als einheitlich gefärbte Bereiche dargestellt, wobei der im Text beschriebene, auch noch bei hohen Temperaturen rigide Cluster in der ersten Domäne der ersten Untereinheit grün eingefärbt ist. Die übrigen Cluster sind dunkel schattiert. Von der Proteinstruktur ist nur das Rückgrat gezeigt und dunkelgrau schattiert. Die Sterne kennzeichnen die beiden aktiven Taschen.

Die thermische Entfaltung von meso- und thermophiler IPMDH wurde in der Vergangenheit mit Hilfe experimenteller Methoden untersucht (Akanuma *et al.* 1998; Motono *et al.* 2001). Die Daten über die Entfaltung von *E. coli* IPMDH konnten gegen ein Modell mit drei Zuständen (und zwei Übergängen) gefittet werden (Motono *et al.* 2001). Neben dem gefalteten und ungefalteten Zustand tritt während der Entfaltung von IPMDH also ein intermediärer Zustand auf. Die Daten über die Entfaltung von *T. thermophilus* IPMDH konnten ähnlich interpretiert werden, wenn auch mit geringerer Bestimmtheit (Akanuma *et al.* 1998). Weitere experimentelle Untersuchungen zeigten, dass es sich bei dem Intermediat um ein Dimer handelte. Im Fall von *E. coli* IPMDH wies das Intermediat einen Verlust von 50 % seiner Sekundär- und Tertiärstruktur auf. Dies wurde dahingegen interpretiert, dass sich die erste Domäne während des ersten Übergangs entfaltet, die zweite Domäne während des zweiten Übergangs. Im Fall von *T. thermophilus* IPMDH wies das Intermediat nur einen Verlust von 10 % seiner Sekundärstruktur auf. In diesem Fall entfaltet sich möglicherweise nur ein kleiner Teil der ersten Domäne während des ersten Übergangs, der Rest der Struktur während des zweiten Übergangs (Motono *et al.* 1999; Motono *et al.* 2001).

Veränderungen der Rigidität und Flexibilität im Netzwerk meso- und thermophiler IPMDH während der Simulation der thermischen Entfaltung wurden in Kap. 5.1.2 mit Hilfe des Rigiditätsordnungsparameters und der Clusterkonfigurationsentropie untersucht. Die Ergebnisse dieser Analysen deuteten klar auf ein korrespondierendes Perkolationsverhalten der homologen Proteine hin, obgleich im Verlauf der Clusterkonfigurationsentropie des mesophilen Proteins ein Übergang bei 364 K zu erkennen war, der keine Entsprechung im Verlauf der Clusterkonfigurationsentropie des thermophilen

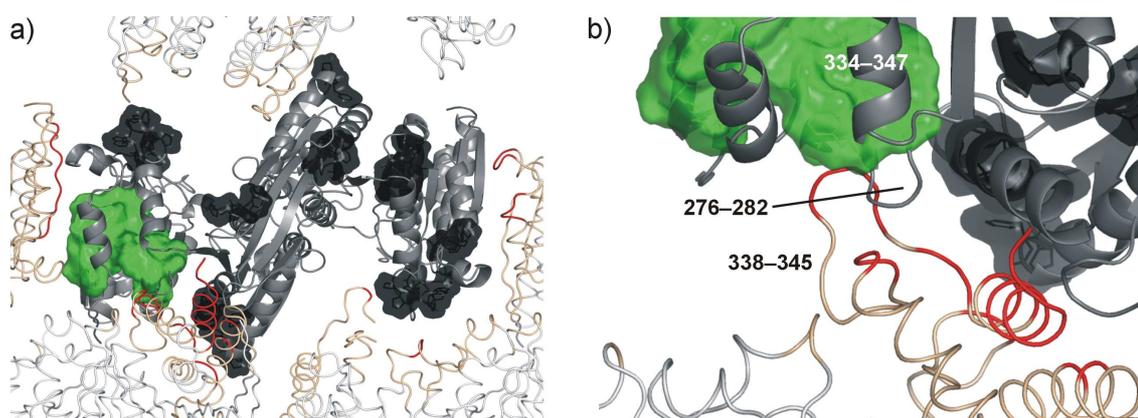


Abbildung 14. Kristallkontakte der mesophilen 3-Isopropylmalatdehydrogenase (IPMDH) (a, b). Es ist jeweils die Zerlegung des Netzwerks in rigide Cluster (RCD) unmittelbar vor dem Übergang bei 364 K gezeigt. Die rigiden Cluster sind als einheitlich gefärbte Bereiche dargestellt, wobei der im Text beschriebene, auch noch bei hohen Temperaturen rigide Cluster in der ersten Domäne der ersten Untereinheit grün eingefärbt ist. Die übrigen Cluster sind dunkel schattiert. Von der Proteinstruktur ist nur das Rückgrat gezeigt und dunkelgrau schattiert. Von den Kristallnachbarn ist ebenfalls nur das Rückgrat gezeigt und hellgrau schattiert, beige oder rot eingefärbt. Die Färbung kodiert wie nahe bestimmte Bereiche der Kristallnachbarn der Proteinstruktur kommen, wobei rot 5 Å und beige 10 Å bedeuten. In (b) sind die Kristallkontakte in der Nähe der ersten Domäne der ersten Untereinheit mesophiler IPMDH vergrößert dargestellt. Reste 276–282 und 299–313 sind in direktem Kontakt mit einem Loop (338–345) des Kristallnachbarn.

Proteins hatte (Abbildung 10 d). Die RCD für das mesophile Protein unmittelbar vor und nach diesem Übergang bei 364 K sind in Abbildung 13 gezeigt. Der Übergang wird offensichtlich durch den Zerfall eines auch noch bei hohen Temperaturen rigiden Clusters in der ersten Domäne der ersten Untereinheit verursacht. In Abbildung 13 a) ist dieser Cluster grün eingefärbt. Außer diesem auch noch bei hohen Temperaturen rigiden Cluster sind nur wenige, einzelne Sekundärstrukturelemente umfassende rigide Cluster im Netzwerk vorhanden, d.h. bis auf die erste Domäne der ersten Untereinheit befindet sich die Struktur bei 364 K im ungefalteten Zustand.

Eingehende Untersuchungen der Struktur von *E. coli* IPMDH zeigen, dass Kristallkontakte zwischen einem Loop in der ersten Domäne der zweiten Untereinheit (Reste 338–345) sowie Sekundärstrukturelementen der ersten Domäne der ersten Untereinheit existieren, die möglicherweise die Konformation von Resten im Interface zwischen einer α -Helix und einem β -Faltblatt in der ersten Domäne der ersten Untereinheit (Reste 299–313 bzw. 276–282) beeinflussen. Diese Kristallkontakte sind in Abbildung 14 gezeigt. Weitere Kristallkontakte finden sich zwischen der N-terminalen α -Helix in der

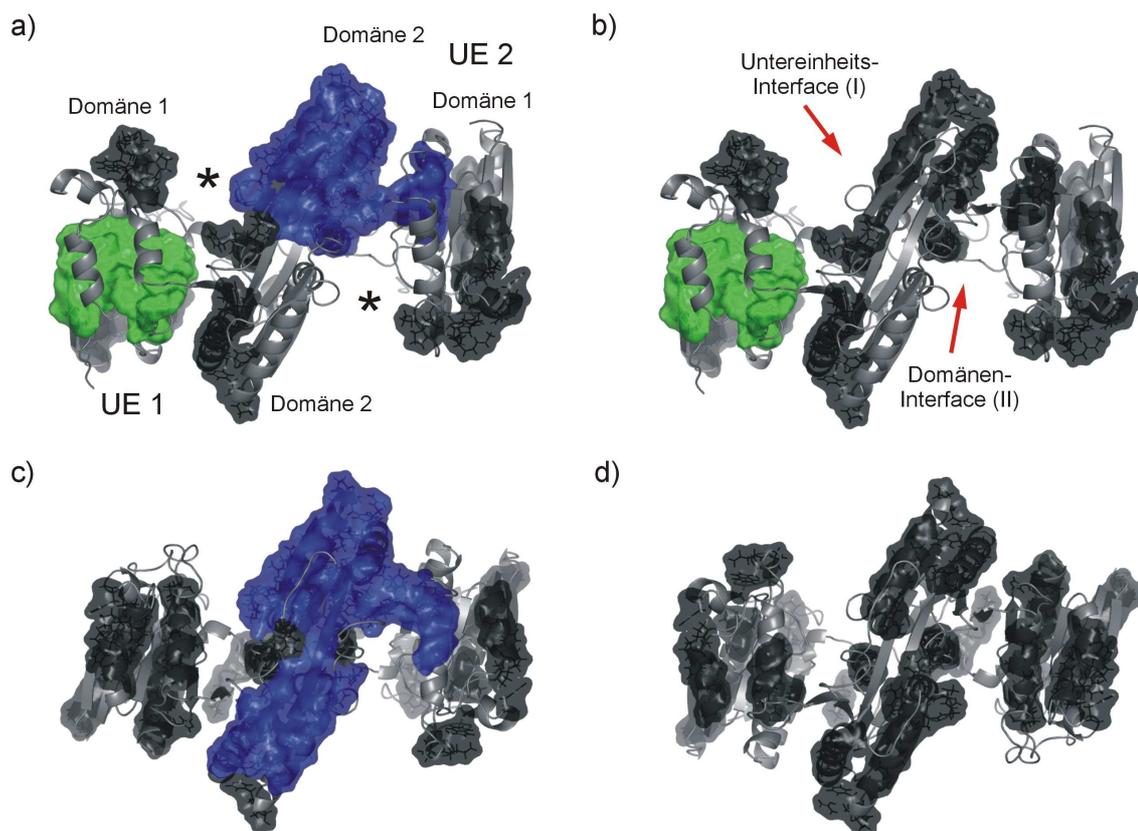


Abbildung 15. Snapshots aus der thermischen Entfaltungssimulation von meso- (a, b) und thermophiler (c, d) 3-Isopropylmalatdehydrogenase (IPMDH) unmittelbar vor (a, c) und nach dem Phasenübergang (b, d) bei 340 bzw. 355 K. Es ist jeweils die Zerlegung des Netzwerks in rigide Cluster (RCD) gezeigt. Die rigiden Cluster sind als einheitlich gefärbte Bereiche dargestellt, wobei der Riesencuster blau und der auch noch bei hohen Temperaturen rigide Cluster in der ersten Domäne der ersten Untereinheit der mesophilen IPMDH (a, b) grün eingefärbt ist. Die übrigen Cluster sind dunkel schattiert. Von der Proteinstruktur ist nur das Rückgrat gezeigt und dunkelgrau schattiert. Die Pfeile kennzeichnen die für das mesophile Protein identifizierten Entfaltungsregionen. Die Sterne kennzeichnen die beiden aktiven Taschen.

ersten Domäne der zweiten Untereinheit und dem Untereinheits-Interface in der ersten Untereinheit. Die auch noch bei hohen Temperaturen vorhandene und in diesem Sinne „extreme“ Rigidität in der ersten Domäne der ersten Untereinheit lässt sich also auf ei-

nen sehr kleinen, von Kristallkontakten verursachten strukturellen Defekt zurückführen. In Kap. 2.1.2 wurde gezeigt, dass dieser Defekt mit Hilfe einer Energieminimierung der Kristallstruktur behoben werden kann. Da in dieser Arbeit wegen der in Kap. 2.1.2 genannten Gründe von der Verwendung energieminimierter Strukturen abgesehen wird, darf die erste Domäne der ersten Untereinheit bei den weiteren Analysen von *E. coli* IPMDH nicht berücksichtigt werden.

Tabelle 4. Vergleich der vorhergesagten Entfaltungsregionen mesophiler 3-Isopropylmalatdehydrogenase (IPMDH) mit Resten aus Strukturbereichen, in denen thermostabilisierende Mutationen experimentell eingeführt wurden.

| Strukturbereich | Reste in vorhergesagten Entfaltungsregionen ^{a, b} | Experimentell eingeführte thermostabilisierende Mutationen ^{a, b, c} |
|---------------------------|--|---|
| I Untereinheits-Interface | 151–163, 165–167, 173, 185–200 , 215, 219–224, 241–248, 255, 256 | 191, 194, 200 , 204, 256 , 259 |
| II Domänen-Interface | 105–112, 133–143 , 174–175, 301–305 | 98, 112, 114–116 , 118, 120, 131, 140 , 182, 271, 295, 311 |
| III Erste Domäne | | 20, 24, 42, 57, 60–62, 84, 86, 92, 350, 356–360 |

^a Die Nummerierung der Reste entspricht der Nummerierung in dem in Abbildung 8 gezeigten Alignment meso- und thermophiler IPMDH; ^b Reste, für die die Vorhersage mit den experimentellen Daten übereinstimmt, sind fett und kursiv markiert. Dies ist dann der Fall, wenn die Reste nicht um mehr als zwei Positionen in der Proteinsequenz voneinander entfernt liegen; ^c Erläuterungen im Text. Nach (Kirino *et al.* 1994; Sakurai *et al.* 1995; Tamakoshi *et al.* 1995; Akanuma *et al.* 1996; Kotsuka *et al.* 1996; Akanuma *et al.* 1997; Aoshima und Oshima 1997; Akanuma *et al.* 1998; Akanuma *et al.* 1999; Hori *et al.* 2000; Németh *et al.* 2000; Nurachman *et al.* 2000; Numata *et al.* 2001; Tamakoshi *et al.* 2001; Ohkuri und Yamagishi 2003; Watanabe *et al.* 2006; Ohkuri und Yamagishi 2007).

Die Analyse des Verlaufs der Clusterkonfigurationsentropie in Kap. 5.1.2 zeigte, dass der Phasenübergang zwischen dem gefalteten und ungefalteten Protein bei 340 bzw. 355 K stattfindet (Abbildung 10 d). Wie im Fall von TLP kann das makroskopische Perkulationsverhalten von IPMDH nun mit mikroskopischen Eigenschaften der Struktur verknüpft werden, indem die RCD für das meso- und thermophile Protein unmittelbar vor und nach dem Phasenübergang verglichen werden (Abbildung 15). Aus Abbildung 15 a) und c) ist ersichtlich, dass die Strukturen von *E. coli* und *T. thermophilus* IPMDH unmittelbar vor dem jeweiligen Phasenübergang von einem rigiden Kern

dominiert werden. Dieser Kern stellt einen rigiden Cluster dar, der die zweite Domäne und somit das Untereinheits-Interface enthält. In Abbildung 15 a) und c) ist dieser Cluster blau eingefärbt. Im Fall des thermophilen Enzyms ist dieser Cluster größer als im Fall des mesophilen Enzyms. Tatsächlich konnte auch experimentell gezeigt werden, dass thermophile IPMDH bei Temperaturen leicht unterhalb des Schmelzpunkts deutlich mehr Sekundär- und Tertiärstruktur aufwies als mesophile IPMDH (Akanuma *et al.* 1998; Motono *et al.* 1999; Motono *et al.* 2001).

Unmittelbar nach dem Phasenübergang (Abbildung 15 b) und d) sind nur noch wenige, einzelne Sekundärstrukturelemente umfassende rigide Cluster im Netzwerk vorhanden. Eine Ausnahme bildet dabei der „extrem“ rigide Cluster in der ersten Domäne der ersten Untereinheit von *E. coli* IPMDH, der nicht weiter berücksichtigt werden soll. Dieser Cluster ist in Abbildung 15 a) grün eingefärbt. Meso- und thermophile IPMDH verhalten sich während des Phasenübergangs sehr ähnlich, obwohl die jeweiligen Phasenübergänge bei deutlich verschiedenen Temperaturen stattfinden. Die Unterschiede im Verhalten der homologen Proteine nach dem Phasenübergang können auf den durch Kristallkontakte verursachten Defekt in der mesophilen Struktur zurückgeführt werden.

Die homologen Proteine befinden sich bei Temperaturen um ihre jeweilige Phasenübergangstemperatur offensichtlich in korrespondierenden Zuständen ähnlicher Flexibilität. Die Ähnlichkeit der Größe und Lage des in den Netzwerken enthaltenen Riesencusters deutet zudem darauf hin, dass für die Thermostabilität von meso- und thermophiler IPMDH ähnliche Strukturbereiche von Bedeutung sein müssen. Diese Bereiche können bestimmt werden, indem Veränderungen der Rigidität und Flexibilität im Netzwerk unmittelbar vor und nach dem Phasenübergang analysiert werden. Die auf diese Weise für meso- und thermophile IPMDH bestimmten Entfaltungsregionen (Abbildung 15) stimmen sehr gut mit Strukturbereichen überein, in denen Mutationen eingeführt wurden, um die Thermostabilität verschiedener mesophiler (Kirino *et al.* 1994; Aoshima und Oshima 1997; Akanuma *et al.* 1998; Németh *et al.* 2000; Tamakoshi *et al.* 2001; Ohkuri und Yamagishi 2003) und chimärer (Numata *et al.* 1995; Sakurai *et al.* 1995; Kotsuka *et al.* 1996; Numata *et al.* 1999; Hori *et al.* 2000; Numata *et al.* 2001) IPMDH zu erhöhen.

Die experimentelle Untersuchung thermophiler IPMDH hat außerdem gezeigt, dass sich die meisten für die Thermostabilität wichtigen Reste entweder im Untereinheits-Interface (Imada *et al.* 1991; Kirino *et al.* 1994; Németh *et al.* 2000), im Domänen-Interface (Tamakoshi *et al.* 1995; Kotsuka *et al.* 1996; Akanuma *et al.* 1997; Qu *et al.*

1997) oder in der ersten Domäne befinden (Akanuma *et al.* 1996; Nurachman *et al.* 2000). Eben diese Strukturbereiche wurden auch als Entfaltungsregionen vorhergesagt (Tabelle 4), abgesehen von der ersten Domäne, in der keine Entfaltungsregion bestimmt werden konnte, obwohl thermostabilisierende Mutationen auch in dieser Domäne eingeführt wurden (Akanuma *et al.* 1996; Nurachman *et al.* 2000). Aus Tabelle 4 ist ersichtlich, dass die Übereinstimmung zwischen den Simulationsergebnissen und experimentellen Daten dennoch sehr gut ist. Diese Ergebnisse zeigen, dass die CNA hilfreich zur Unterstützung des Protein Engineering ist, da mit ihrer Hilfe die thermische Stabilität einer Struktur abgeschätzt werden kann, andererseits Hinweise darauf gegeben werden, in welchen Bereichen der Struktur thermostabilisierende Mutationen eingeführt werden können.

5.1.4 Konserviertheit der mikroskopischen Stabilität

Im vorangehenden Kapitel wurden (lokale) Strukturbereiche bestimmt, an denen der (globale) Zerfall der Struktur beginnt, und die deswegen für die thermische Stabilität von großer Wichtigkeit sind. Zur Bestimmung dieser Entfaltungsregionen wurden ausschließlich die das Riesencoluster betreffenden Ereignisse bei der Phasenübergangstemperatur betrachtet. Die CNA erlaubt aber auch die Untersuchung von lokalen Veränderungen der Rigidität und Flexibilität im Netzwerk, die nur für die mikroskopische Stabilität einer bestimmten Region von Bedeutung sind. Lokale Veränderungen der Rigidität und Flexibilität finden bei Temperaturen unterhalb der Phasenübergangstemperatur statt und können insbesondere für die Bioaktivität des Proteins von Bedeutung sein. Experimentelle Untersuchungen zeigten, dass thermophile Proteine trotz globaler Rigidisierung wichtige flexible Bereiche (und damit ihre mikroskopische Stabilität) konserviert haben (Somero 1978; Jaenicke und Böhm 1998; Sterner und Brunner 2008). In diesem Kapitel wird die mikroskopische Stabilität der meso- und thermophilen Proteine aus dem Datensatz aus der Simulation der thermischen Entfaltung dieser Proteine bestimmt und quantitativ verglichen.

Zur Bestimmung der mikroskopischen Stabilität von Proteinen aus der Simulation der thermischen Entfaltung muss untersucht werden, bei welcher Temperatur ein „rigider Kontakt“ zwischen zwei Resten gebrochen oder ausgebildet wird. Ein Rest wird dabei durch sein C α -Atom repräsentiert. Ein rigider Kontakt zwischen zwei Resten besteht, wenn die beiden Reste zum selben Cluster gehören. Der Kontakt ist gebrochen, sobald dieser Cluster zerfällt. Die Temperatur, bei der ein rigider Kontakt gebrochen wird, ist proportional zu seiner mikroskopischen Stabilität. Die Stabilitätswerte sind in

kcal mol^{-1} und relativ zur makroskopischen Thermostabilität angegeben: Bricht ein rigider Kontakt unterhalb der Schmelztemperatur des Proteins, so ist er instabil. Bricht er dagegen oberhalb der Schmelztemperatur, so ist er stabil. Die auf diese Weise bestimm-

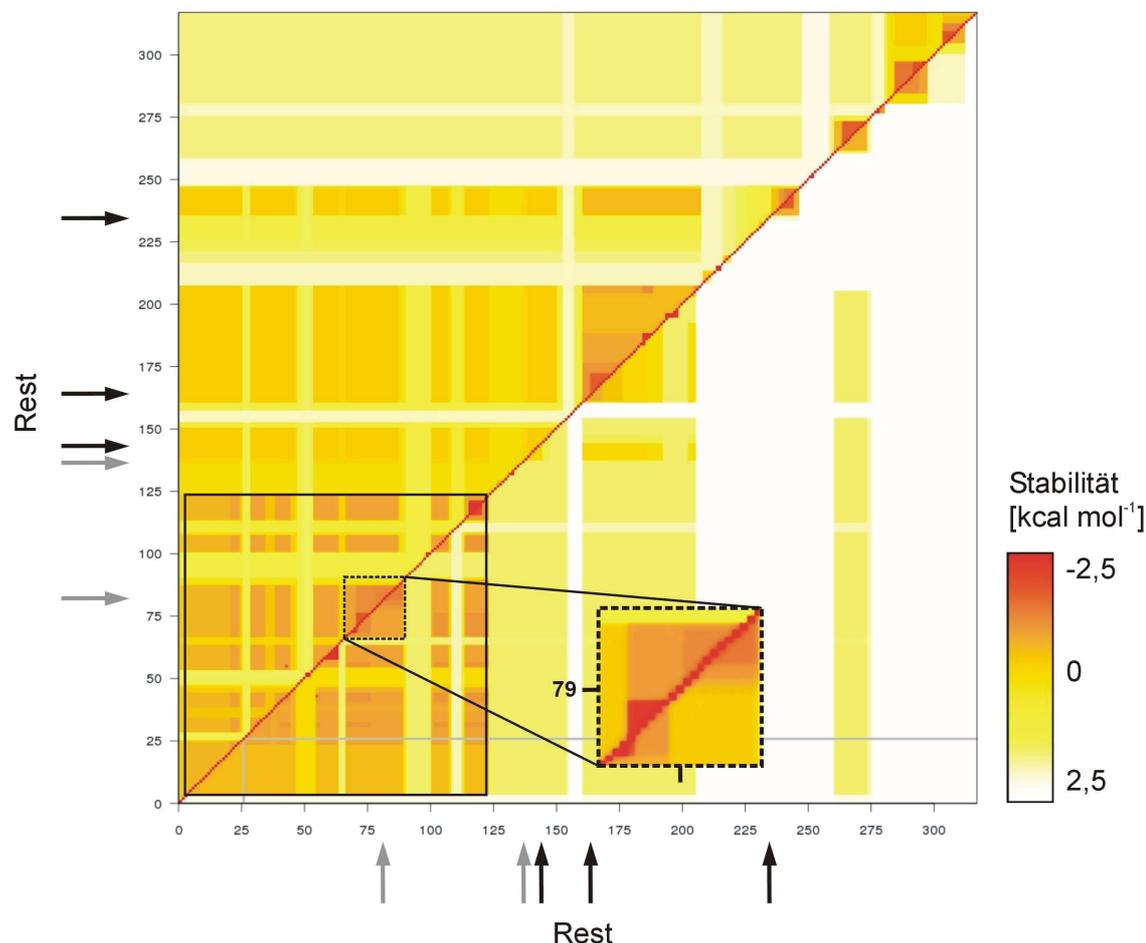


Abbildung 16. Kombinierte Stabilitätskarte für meso- und thermophile thermolysinähnliche Protease (TLP). Die Stabilitätswerte sind relativ zur makroskopischen Thermostabilität angegeben. Die mikroskopischen Stabilitätswerte der Kontakte sind farblich kodiert, wobei rot stabile und weiß instabile Kontakte bedeuten. In der oberen Hälfte über der Diagonalen sind die mikroskopischen Stabilitätswerte für das mesophile, in der unteren Hälfte die für das thermophile Protein aufgetragen. Die beiden Kartenhälften wurden unter Verwendung des in Abbildung 7 gezeigten Strukturalignments kombiniert. Gaps im Alignment sind auf der Karte grau dargestellt. Katalytische Reste werden durch schwarze Pfeile, weitere für die Katalyse, die Substrat- oder Kofaktorbindung wichtige Reste (siehe Text) durch graue Pfeile gekennzeichnet. Die Kontakte zwischen den Resten im Riesencuster sind mit einer durchgezogenen Linie umrandet. Die Kontakte zwischen den Resten der langen α -Helix in der N-terminalen Domäne sind mit einer gestrichelten Linie umrandet. Dieser Bereich der Stabilitätskarte ist zudem vergrößert dargestellt.

ten mikroskopischen Stabilitätswerte aller Kontakte zwischen zwei Resten können auf einer zweidimensionalen Stabilitätskarte abgebildet werden.

Stabilitätskarten können hinsichtlich ihres Aufbaus mit den von Livesay und Jacobs verwendeten Kooperationskorrelationskarten (Livesay und Jacobs 2006; Livesay *et al.* 2008) oder mit Fluktuationskorrelationskarten verglichen werden. Die Stabilitätskarten

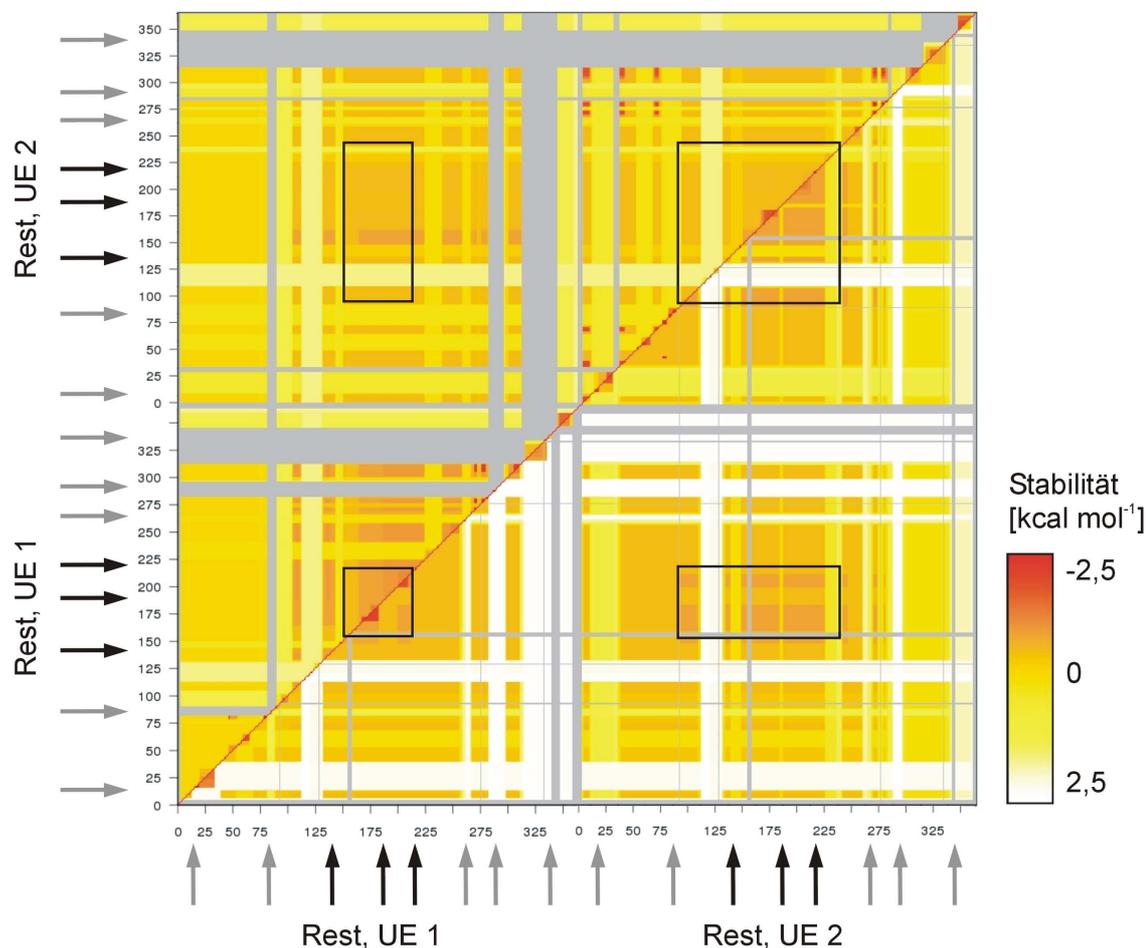


Abbildung 17. Kombinierte Stabilitätskarte für meso- und thermophile 3-Isopropylmalatdehydrogenase (IPMDH). Die Stabilitätswerte sind relativ zur makroskopischen Thermostabilität angegeben. Die mikroskopischen Stabilitätswerte der Kontakte sind farblich kodiert, wobei rot stabile und weiß instabile Kontakte bedeuten. In der oberen Hälfte über der Diagonalen sind die mikroskopischen Stabilitätswerte für das mesophile, in der unteren Hälfte die für das thermophile Protein aufgetragen. Die beiden Kartenhälften wurden unter Verwendung des in Abbildung 8 gezeigten Strukturalignments kombiniert. Gaps im Alignment sind auf der Karte grau dargestellt. Katalytische Reste werden durch schwarze Pfeile, weitere für die Katalyse, die Substrat- oder Kofaktorbindung wichtige Reste (siehe Text) durch graue Pfeile gekennzeichnet. Die Kontakte zwischen den Resten im Riesencuster sind mit einer durchgezogenen Linie umrandet. Die Abkürzung UE steht für Untereinheit.

unterscheiden sich aber dahingehend von Korrelationskooperations- oder Fluktuationskorrelationskarten, dass sie direkte Informationen über die Verteilung der Rigidität und Flexibilität in der Proteinstruktur liefern und nur aus der Simulation der thermischen

Entfaltung berechnet werden. Korrelationskooperations- oder Fluktuationskorrelationskarten werden dagegen aus Ensembles berechnet, die entweder mit Hilfe von Methoden aus der statistischen Thermodynamik oder MD-Simulationen erzeugt werden müssen.

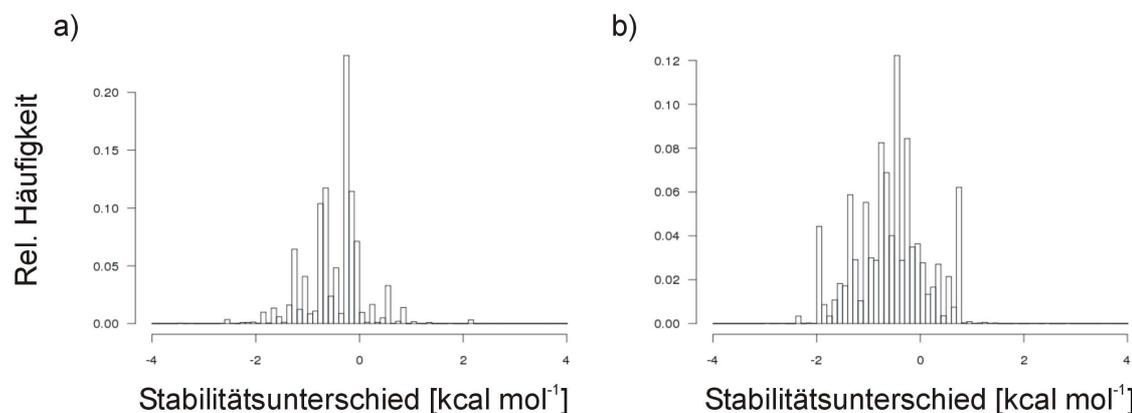


Abbildung 18. Histogramme der Unterschiede zwischen den mikroskopischen Stabilitäten der rigiden Kontakte in thermophiler thermolysinähnlicher Protease (TLP) (a) und 3-Isopropylmalatedehydrogenase (IPMDH) (b) und den mikroskopischen Stabilitäten der entsprechenden rigiden Kontakte im jeweiligen mesophilen Protein. Insgesamt sind 49.928 (a) und 238.050 rigide Kontakte (b) berücksichtigt. Ein negativer Stabilitätsunterschied deutet darauf hin, dass der rigide Kontakt im thermophilen Protein stabiler ist als im mesophilen Protein.

Stabilitätskarten für TLP und IPMDH sind in Abbildung 16 bzw. Abbildung 17 gezeigt. Die mikroskopischen Stabilitätswerte der rigiden Kontakte sind farblich kodiert, wobei rot stabile und weiß instabile Kontakte bedeuten. Bei den gezeigten Karten handelt es sich um „kombinierte“ Stabilitätskarten, die entlang der Diagonalen zweigeteilt sind. In der oberen Hälfte sind die mikroskopischen Stabilitätswerte der rigiden Kontakte im mesophilen Protein, in der unteren Hälfte die im thermophilen Protein aufgetragen. Die beiden Kartenhälften wurden entsprechend eines zuvor durchgeführten Strukturalignments kombiniert. Im Fall von TLP und IPMDH wurde dafür das in Abbildung 7 bzw. Abbildung 8 gezeigte Alignment verwendet. Gaps im Alignment sind auf der Karte grau dargestellt.

Aus den Stabilitätskarten ist ersichtlich, dass rigide Kontakte zwischen den im jeweiligen Riesencluster enthaltenen Resten eine hohe mikroskopische Stabilität haben. Diese Kontakte zwischen den Resten im jeweiligen Riesencluster sind in Abbildung 16 und Abbildung 17 zur Orientierung mit einer durchgezogenen Linie umrandet. In Übereinstimmung mit den bisherigen Ergebnissen reflektiert also auch die Stabilitätskarte, dass die Größe und Lage des Riesenclusters im Netzwerk des meso- und thermophilen

Proteins korrespondierend ist. Neben der Konserviertheit des Riesenclusters lässt sich in der Stabilitätskarte aber auch erkennen, dass viele weitere rigide und flexible Bereiche im meso- und thermophilen Protein konserviert sind.

Tabelle 5. Quantitativer Vergleich der mikroskopischen Stabilitäten homologer Strukturen über die Korrelation der Stabilitätskarten des meso- und thermophilen Proteins.

| Proteinfamilie ^a | r^b |
|-----------------------------|-----------------|
| 1 | 0,46 |
| 4 | 0,30 |
| 5 | 0,48 |
| 6 | 0,55 |
| 8 | 0,69 |
| 9 | 0,35 |
| 10 | 0,65 |
| 11 | 0,66 |
| 12 | 0,33 |
| 13 | 0,53 |
| 15 | 0,74 |
| 16 | 0,25 |
| 18 | 0,90 |
| Mittelwert | $0,53 \pm 0,19$ |

^a Nummern entsprechen den Nummern der Proteinfamilien in Tabelle 1; ^b Korrelationskoeffizient nach Pearson.

In Abbildung 18 sind Histogramme der Unterschiede zwischen den mikroskopischen Stabilitäten der rigiden Kontakte in thermophiler TLP und IPMDH und den mikroskopischen Stabilitäten der entsprechenden rigiden Kontakte im jeweiligen mesophilen Protein gezeigt. Ein negativer Stabilitätsunterschied deutet darauf hin, dass der rigide Kontakt im thermophilen Protein stabiler ist als im mesophilen Protein. Sowohl für TLP als auch IPMDH beträgt der Stabilitätsunterschied der meisten rigiden Kontakte zwischen 0 und $-2,0 \text{ kcal mol}^{-1}$, wobei ungefähr 55 bzw. 75 % der Kontakte einen Stabilitätsunterschied zwischen 0 und $-1,0 \text{ kcal mol}^{-1}$ aufweisen. Somit ist klar, dass im thermophilen Protein die meisten rigiden Kontakte stabiler sind als im mesophilen Protein. Allerdings ist die Zunahme in der Stabilität pro rigiden Kontakt nicht besonders groß.

Die mikroskopische Stabilität homologer Strukturen aus dem Datensatz kann quantitativ verglichen werden, indem die Korrelation der Stabilitätskarten des meso- und

thermophilen Proteins berechnet wird. Korrelationskoeffizienten der Paare homologer Proteine aus dem Datensatz sind in Tabelle 5 aufgelistet. Die Korrelation wurde nur für die zwei Drittel der Paare von Proteinen aus dem Datensatz berechnet, für die das thermophile Protein thermostabiler als das mesophile Protein vorhergesagt wurde. Die Werte reichen von 0,25 bis 0,90, mit einem Mittelwert von $0,53 \pm 0,19$. Alle Korrelationen sind signifikant ($p \leq 0,05$). Als Beispiel sind in Abbildung 19 Streudiagramme gezeigt, in denen die mikroskopischen Stabilitätswerte der rigiden Kontakte in mesophiler TLP und IPMDH gegen die mikroskopischen Stabilitätswerte der entsprechenden rigiden Kontakte im jeweiligen thermophilen Protein aufgetragen sind.

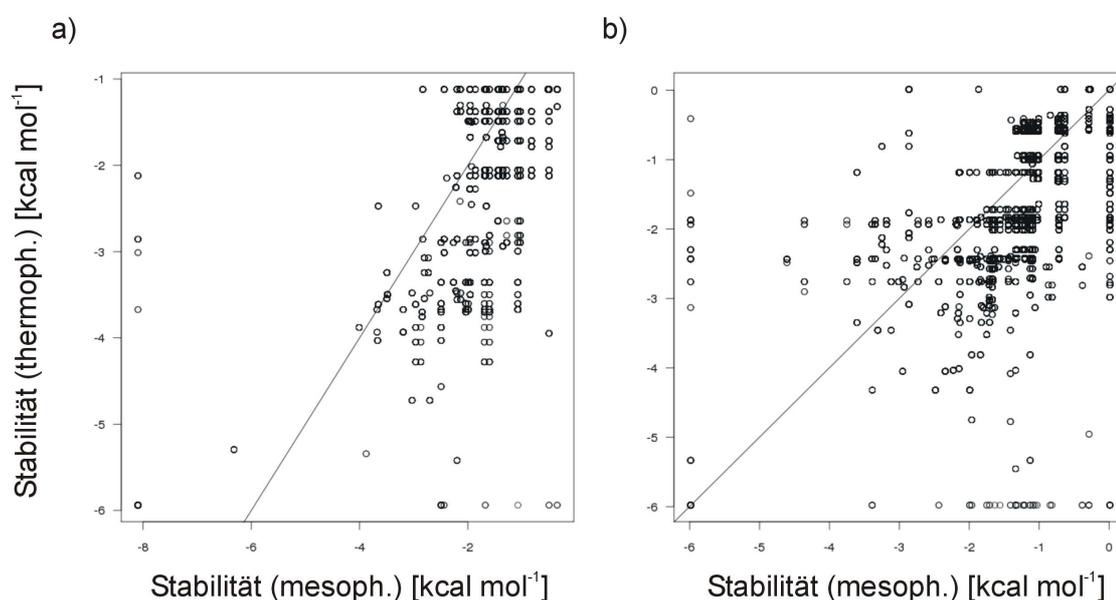


Abbildung 19. Streudiagramme, in denen die mikroskopischen Stabilitätswerte der rigiden Kontakte in mesophiler thermolysinähnlicher Protease (TLP) (a) und 3-Isopropylmalatedehydrogenase (IPMDH) (b) gegen die mikroskopischen Stabilitätswerte der entsprechenden rigiden Kontakte im jeweiligen thermophilen Protein aufgetragen sind. Zudem ist die Gerade der optimalen Übereinstimmung gezeigt. Insgesamt sind 49.928 (a) und 238.050 rigide Kontakte (b) berücksichtigt.

Die mikroskopischen Stabilitäten der homologen meso- und thermophilen Proteine aus dem Datensatz sind also sehr ähnlich. Diese Ähnlichkeit ist ein klarer Hinweis auf die Bedeutung der durch die mikroskopischen Stabilität bedingten lokalen Flexibilität für die Bioaktivität. Diese Ergebnisse sprechen für die Theorie der korrespondierenden Zustände, die besagt, dass bei der thermophilen Anpassung trotz globaler Rigidisierung für die Bioaktivität wichtige flexible Bereiche konserviert sind (Somero 1978; Jaenicke und Böhm 1998; Sterner und Brunner 2008).

5.1.5 Mikroskopische Stabilität und Bioaktivität

In den vorangehenden Kapiteln konnte gezeigt werden, dass die thermophile Anpassung der thermophilen Proteine aus dem Datensatz über eine globale Rigidisierung der Struktur erfolgt. Andererseits konnte eine große Ähnlichkeit der mikroskopischen Stabilitäten homologer meso- und thermophiler Proteine festgestellt werden, was auf die Bedeutung lokaler Flexibilität für die Bioaktivität hindeutet. Experimentell ist das Zusammenspiel zwischen Flexibilität, Stabilität und Aktivität vielfach mit Hilfe kinetischer Parameter für die Enzymfunktion und Katalyse (K_m und k_{cat}) untersucht worden (Fields 2001). Diese Parameter sind allerdings nur ein indirektes Maß für die Flexibilität und liefern keine Informationen über die Bedeutung der Flexibilität einzelner Reste (Georlette *et al.* 2004).

Hilfreich sind hier physikalische Methoden, mit denen die Dynamik der Struktur in atomarer Auflösung untersucht werden kann, wie beispielsweise H/D-Austauschexperimente, Neutronenstreuung und NMR-Spektroskopie (Sternier und Brunner 2008). Die experimentelle Beschreibung und Quantifizierung der Flexibilität bleibt in vielen Fällen aber insbesondere hinsichtlich der verschiedenen Arten und Zeitskalen der gemessenen Bewegungen schwierig (Kamerzell und Middaugh 2008; Sternier und Brunner 2008). Die gemessene Dynamik (Bewegung) ist nur ein indirektes Maß für die tatsächlich vorhandene Flexibilität (Beweglichkeit). Diese Schwierigkeiten bei der Interpretation experimenteller Daten sind möglicherweise die Ursache dafür, dass einige thermophile Enzyme ähnlich flexibel (Hollien und Marqusee 1999; Hollien und Marqusee 2002) oder sogar flexibler (Fitter und Heberle 2000; Hernández *et al.* 2000; Fitter *et al.* 2001; Hernández und LeMaster 2001) als homologe mesophile Proteine zu sein scheinen (Jaenicke 2000).

Wenn Veränderungen der Rigidität und Flexibilität dagegen mit Hilfe der CNA bestimmt werden, dann müssen keine Annahmen über die Arten und Zeitskalen der zugrunde liegenden Bewegungen gemacht werden, denn in diesem Fall wird die Flexibilität im Netzwerk meso- und thermophiler Proteine direkt und in atomarer Auflösung bestimmt. Diese Stärken der CNA werden in diesem Kapitel ausgenutzt, um die lokale Flexibilität und die mikroskopischen Stabilitätswerte katalytischer sowie weiterer für die Katalyse, die Substrat- oder Kofaktorbindung wichtiger Reste von meso- und thermophiler TLP (Abschnitt 5.1.5.1) und IPMDH (Abschnitt 5.1.5.2) vorherzusagen. Diese Vorhersagen werden mit experimentellen Daten aus der Literatur verglichen.

5.1.5.1 Lokale Flexibilität von thermolysinähnlichen Proteasen

Das aktive Zentrum von thermolysinähnlichen Proteasen befindet sich in einer Vertiefung zwischen der N- und der C-terminalen Domäne, wo auch ein katalytisches Zinkion gebunden ist. Für die katalytische Funktionsweise von TLP wurden zwei Mechanismen vorgeschlagen. Demnach fungiert entweder Glu 144 (Hangauer *et al.* 1984; Matthews 1988) oder His 232 als Protonenakzeptor (Mock und Aksamawati 1994; Mock und Stanford 1996). In jedem Fall spielen Glu 144, His 232, Tyr 158 und ein an das Zinkion gebundenes Wassermolekül bei der Katalyse eine wichtige Rolle (Adekoya und Sylte 2009). Das Zinkion wird von His 143 und 147, Glu 167 sowie einem Wassermolekül koordiniert. Die meisten der katalytischen und das Zinkion koordinierenden Reste sind in der zentralen α -Helix (Reste 139–154) enthalten, die den Boden der aktiven Tasche bildet (Adekoya und Sylte 2009). Die katalytischen und das Zinkion koordinierenden Reste sind in dem in Abbildung 7 gezeigten Alignment meso- und thermophiler TLP durch eine schwarze Schattierung sowie in der in Abbildung 16 gezeigten Stabilitätskarte durch schwarze Pfeile gekennzeichnet.

Untersuchungen mit rechnerischen Methoden zeigten, dass die Substratbindung und Katalyse von TLP eine Scharnierbewegung (*hinge bending motion*) erfordert, bei der sich die beiden Domänen des Enzyms über der aktiven Tasche schließen (van Aalten *et al.* 1995). Die zentrale α -Helix oder Strukturbereiche in unmittelbarer Nachbarschaft fungieren dabei als flexibles Scharnier. Mit Hilfe von Mutationsstudien konnte gezeigt werden, dass insbesondere Gly 136 und 137 am N-terminalen Ende der zentralen α -Helix für die Scharnierbewegung von Bedeutung sind (Veltman *et al.* 1998). Der Austausch dieser Reste beeinflusst die Bewegung des Enzyms bei der Katalyse, was zu einer Änderung von k_{cat} führt (Veltman *et al.* 1998). Darüber hinaus ist ein weiteres Gly (Rest 79) in der Mitte der langen α -Helix in der N-terminalen Domäne (Reste 67–89) für die Bewegung von Bedeutung (Holland *et al.* 1992; Holland *et al.* 1995; Veltman *et al.* 1998). Der Austausch von Gly 79 verändert die Flexibilität der aktiven Tasche und hat einen Einfluss auf die Affinität des Enzyms zum Substrat. Die für die Funktion wichtigen Reste sind in dem in Abbildung 7 gezeigten Alignment meso- und thermophiler TLP durch eine graue Schattierung sowie in der in Abbildung 16 gezeigten Stabilitätskarte durch graue Pfeile gekennzeichnet.

Mikroskopische Stabilitätswerte der für die Funktion wichtigen Reste 79, 136 und 137 können aus der Simulation der thermischen Entfaltung bestimmt werden. Dazu wird die in Abbildung 16 gezeigte kombinierte Stabilitätskarte von meso- und ther-

mophiler TLP untersucht. Es ist zu erwarten, dass die Kontakte zwischen den für die Funktion wichtigen Resten und ihren Nachbarn relativ instabil sind. Dies ist die Voraussetzung für ihre lokale Flexibilität. Für Gly 79 kann dies insbesondere für *B. cereus* TLP beobachtet werden. So sind die Kontakte zwischen Rest 79 und anderen Resten in der langen α -Helix in der N-terminalen Domäne (Reste 67–89) relativ instabil, wenn sie mit den Kontakten zwischen den anderen Resten dieser α -Helix verglichen werden. Die Kontakte zwischen Resten in der langen α -Helix sind in Abbildung 16 zur Orientierung mit einer gestrichelten Linie umrandet. Der entsprechende Bereich der Stabilitätskarte ist zudem vergrößert dargestellt. Rest 79 stellt möglicherweise eine „Sollbruchstelle“ der langen α -Helix dar. Dieses Ergebnis zeigt, dass die Bedeutung von Gly 79 für die Funktion auf einen kleinen Strukturbereich beschränkt ist.

Die Bedeutung von Gly 136 und 137 als Reste im Scharnier kann mit Hilfe der Stabilitätskarten ebenfalls gut beobachtet werden. Während die Kontakte zwischen den Resten des jeweiligen Riesenclusters eine hohe mikroskopische Stabilität aufweisen, wie auch die Kontakte zwischen den Resten des rigiden Clusters in der C-terminalen Domäne (138–208), sind die Kontakte zwischen den Resten 125–137 relativ instabil. Bei Temperaturen leicht unterhalb der jeweiligen Phasenübergangstemperatur stellen die Reste 125–137 somit einen flexiblen Bereich dar, der unmittelbar an die zentrale α -Helix grenzt und dafür sorgt, dass das jeweilige Riesencluster nicht mit dem die Reste 138–208 enthaltenden rigiden Cluster verschmelzt. Die Reste 125–137 werden somit in Übereinstimmung mit den experimentellen Daten als Scharnier vorhergesagt, das die wichtige Bewegung der TLP bei der Katalyse ermöglicht. Die katalytischen und das Zinkion koordinierenden Reste sind nur zum Teil in diesem flexiblen Bereich enthalten.

Aus den in Abbildung 20 a) und d) gezeigten RCD der aktiven Tasche mesophiler TLP bzw. von Thermolysin bei 342 bzw. 364 K (entsprechend der Arbeitstemperatur des jeweiligen Enzyms) ist ersichtlich, dass manche der katalytischen und das Zinkion koordinierenden Reste zu dem aus den Resten 138–208 bestehenden rigiden Cluster in der C-terminalen Domäne gehören. Diese Beobachtung deutet darauf hin, dass die katalytischen und das Zinkion koordinierenden Reste in einem lokal stabilen Strukturbereich enthalten sind, der global aufgrund der Anwesenheit des Scharniers flexibel ist. Diese lokale Stabilität und globale Flexibilität konnte auch mit Hilfe von H/D-Austauschexperimenten gezeigt werden (Liu und Konermann 2008).

B. cereus TLP und Thermolysin unterscheiden sich in der Temperatur ihrer maximalen Aktivität (de Kreij *et al.* 2000). Diese beträgt für das mesophile Enzym 60 °C

und für das thermophile 75 °C. Das thermophile Enzym ist zudem bei der Temperatur maximaler Aktivität des mesophilen Enzyms nur noch marginal aktiv. Mit Hilfe der CNA kann gezeigt werden, dass dieses Verhalten auf einen Verlust der Flexibilität der

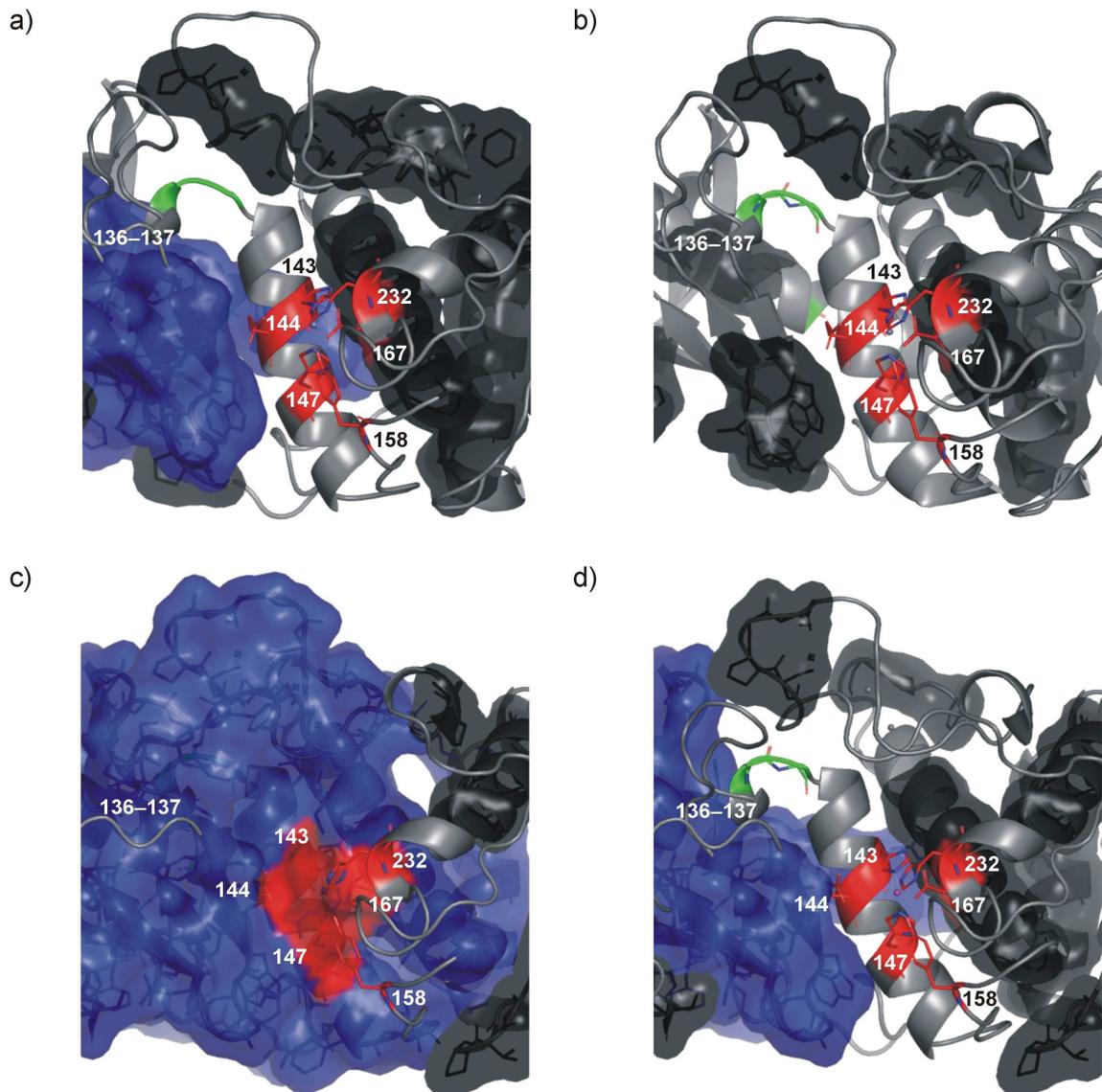


Abbildung 20. Aktive Tasche mesophiler (a, b) und thermophiler thermolysinähnlicher Protease (TLP) (b, c) bei der Arbeitstemperatur des mesophilen Enzyms (342 K) (a, c) und der Arbeitstemperatur des thermophilen Enzyms (364 K) (b, d). Es ist jeweils die Zerlegung des Netzwerks in rigide Cluster (RCD) gezeigt. Die rigiden Cluster sind als einheitlich gefärbte Bereiche dargestellt, wobei der Riesencluster blau eingefärbt ist. Die übrigen Cluster sind dunkel schattiert. Von der Proteinstruktur ist nur das Rückgrat gezeigt und dunkelgrau schattiert. Die katalytischen Reste sind rot, weitere für die Katalyse, die Substrat- oder Kofaktorbindung wichtige Reste grün eingefärbt.

für die Funktion wichtigen Strukturbereiche zurückzuführen ist. So werden bei Temperaturen um 20–30 K unterhalb der Phasenübergangstemperatur Kontakte zwischen den

Resten des Scharniers (125–136), den Resten des Riesenclusters und den Resten des rigiden Clusters in der C-terminalen Domäne (137–208) ausgebildet, was eine Rigidisierung der gesamten aktiven Tasche zur Folge hat. In Abbildung 20 c) ist die RCD der aktiven Tasche für das thermophile Enzym bei 342 K gezeigt. Diese Temperatur liegt leicht unterhalb der Phasenübergangstemperatur für *E. coli* TLP (350 K) und entspricht der Arbeitstemperatur dieses Enzyms. Die Rigidisierung der aktiven Tasche ist klar zu erkennen. Auf der anderen Seite ist die aktive Tasche des mesophilen Enzyms bei 364 K (entsprechend der Arbeitstemperatur von Thermolysin) vollständig flexibel (Abbildung 20 b), was die Entfaltung und Inaktivierung des Enzyms zur Folge hat.

Bei der Entwicklung neuer thermostabiler Proteine wird immer wieder versucht, die Aktivität von der Thermostabilität zu entkoppeln und damit durch geeignete Mutationen Enzyme zu erzeugen, die sowohl thermostabil als auch über einen großen Temperaturbereich aktiv sind. Die Schwierigkeit dabei besteht darin, dass thermostabilisierende Mutationen die Aktivität negativ beeinflussen können, wenn sie außer den für die globale Stabilität wichtigen Bereichen auch die aktive Tasche rigidisieren, oder umgekehrt, dass aktivierende Mutationen die Thermostabilität verringern, wenn sie außer der Flexibilität der aktiven Tasche auch die Flexibilität der für die globale Stabilität wichtigen Bereiche erhöhen. So führte der Austausch von Leu 145 durch Ser zu einer Erhöhung der Flexibilität der aktiven Tasche und somit der Aktivität, ohne die Thermostabilität des Proteins zu beeinflussen (Yasukawa und Inouye 2007). Der Austausch von Gln 120 durch Glu führte ebenfalls zu einer Verbesserung der Aktivität, allerdings wurde in diesem Fall auch die Thermostabilität beeinflusst (Kidokoro *et al.* 1995). Mit Hilfe der CNA können diese Beobachtungen verstanden werden. Beide Reste befinden sich in unmittelbarer Nachbarschaft des Scharniers, dessen Flexibilität für die Bioaktivität von Bedeutung ist. Rest 120 befindet sich darüber hinaus in unmittelbarer Nähe einer vorhergesagten Entfaltungsregion (Tabelle 3). Mutationen in den Entfaltungsregionen beeinflussen die Thermostabilität des Proteins (Abschnitt 5.1.3.1).

Diese Ergebnisse zeigen, dass die CNA dahingehend hilfreich zur Unterstützung des Protein Engineering ist, dass der Effekt einer Mutation auf die Thermostabilität und auf die Aktivität abgeschätzt werden kann. Zudem können die Simulationsergebnisse möglicherweise Hinweise darauf geben, in welchen Bereichen der Struktur thermostabilisierende Mutationen eingeführt werden können, die keinen negativen Einfluss auf die Aktivität haben.

5.1.5.2 Lokale Flexibilität von 3-Isopropylmalatdehydrogenasen

Das aktive Zentrum von 3-Isopropylmalatdehydrogenasen befindet sich in einer Vertiefung zwischen der ersten und zweiten Domäne und besteht aus Resten beider Untereinheiten des homodimeren Enzyms (Kadono *et al.* 1995). Die katalytischen Reste Lys 195 und Asp 227 gehören zu der einen Untereinheit, Tyr 145 zu der anderen. Die katalytischen Reste sind in dem in Abbildung 8 gezeigten Alignment meso- und thermophiler IPMDH durch eine schwarze Schattierung sowie in der in Abbildung 17 gezeigten Stabilitätskarte durch schwarze Pfeile gekennzeichnet. Aus der in Abbildung 21 a) und d) gezeigten RCD der aktiven Tasche von *E. coli* bzw. *T. thermophilus* IPMDH bei 337 bzw. 352 K (entsprechend der Arbeitstemperatur des jeweiligen Enzyms) ist ersichtlich, dass sich die katalytischen Reste 145, 195 und 227 in unmittelbarer Nachbarschaft des Riesencluster befinden, selbst aber nicht zu diesem Cluster gehören. Die Untersuchung der in Abbildung 17 gezeigten kombinierten Stabilitätskarte von meso- und thermophiler IPMDH zeigt, dass die Kontakte zwischen den katalytischen Resten und ihren unmittelbaren Nachbarn relativ instabil sind, wenn sie mit den Kontakten der im Riesencluster enthaltenen Reste verglichen werden. Diese Ergebnisse deuten darauf hin, dass die katalytischen Reste von IPMDH anders als die katalytischen Reste von TLP in einem lokal flexiblen Strukturbereich enthalten sind, der global aufgrund seiner unmittelbaren Nachbarschaft zum Riesencluster stabilisiert wird.

Eine genauere Betrachtung der mikroskopischen Stabilitätswerte des thermophilen Enzyms zeigt, dass bei Temperaturen von 7–10 K unterhalb der Phasenübergangstemperatur Kontakte zwischen den katalytischen Resten und den Resten des Riesenclusters ausgebildet werden. In Abbildung 21 c) ist die RCD der aktiven Tasche für *T. thermophilus* IPMDH bei 337 K gezeigt. Diese Temperatur entspricht der Arbeitstemperatur des mesophilen Enzyms. Der Verlust der lokalen Flexibilität der katalytischen Reste ist klar zu erkennen. Möglicherweise trägt dieser Verlust der lokalen Flexibilität dazu bei, dass das thermophile Enzym bei der Arbeitstemperatur des mesophilen Enzyms nur noch marginal aktiv sein kann. Auf der anderen Seite ist die aktive Tasche des mesophilen Enzyms bei 352 K (entsprechend der Arbeitstemperatur des thermophilen Enzyms) vollständig flexibel (Abbildung 21 b), was die Entfaltung und Inaktivierung des Enzyms zur Folge hat. Tatsächlich konnte experimentell gezeigt werden, dass die Temperatur für die optimale Aktivität von *E. coli* IPMDH 50 °C, die für *T. thermophilus* IPMDH 74 °C beträgt, und dass *T. thermophilus* IPMDH bei der Temperatur maximaler Aktivität des mesophilen Enzyms nur noch marginal aktiv (Závodszy *et al.* 1998).

Neben den mikroskopischen Stabilitätswerten der katalytischen Reste sind in erster Linie auch Veränderungen der Rigidität und Flexibilität der substrat- und kofaktorbindenden Reste für die Bioaktivität der IPMDH von Bedeutung. Eingehende Analysen der

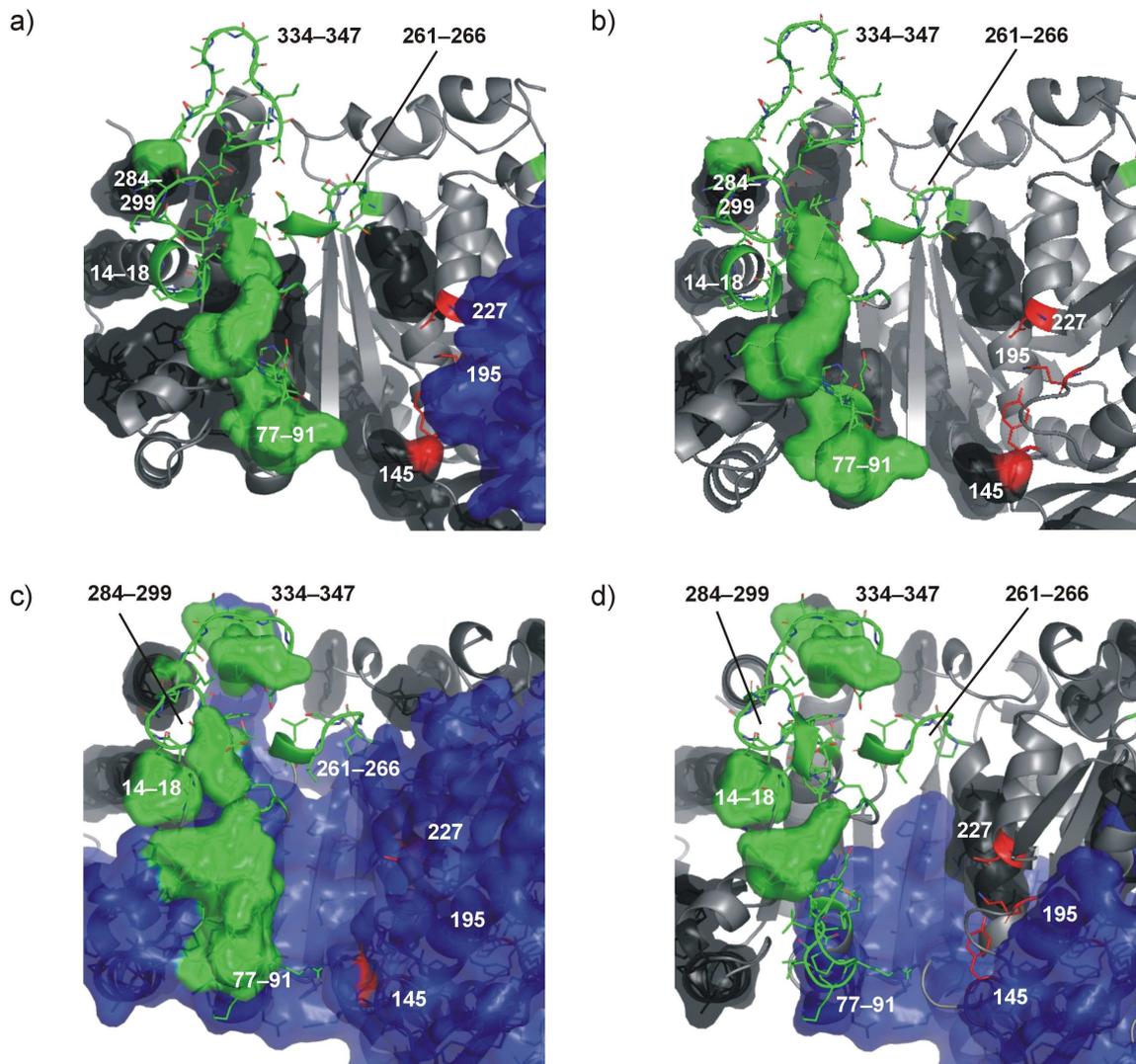


Abbildung 21. Aktive Tasche mesophiler (a, b) und thermophiler 3-Isopropylmalatdehydrogenase (IPMDH) (c, d) bei der Arbeitstemperatur des mesophilen Enzyms (337 K) (a, c) und der Arbeitstemperatur des thermophilen Enzyms (352 K) (b, d). Es ist jeweils die Zerlegung des Netzwerks in rigide Cluster (RCD) gezeigt. Die rigiden Cluster sind als einheitlich gefärbte Bereiche dargestellt, wobei der Riesencuster blau eingefärbt ist. Die übrigen Cluster sind dunkel schattiert. Von der Proteinstruktur ist nur das Rückgrat gezeigt und dunkelgrau schattiert. Die katalytischen Reste sind rot, weitere für die Katalyse, die Substrat- oder Kofaktorbindung wichtige Reste grün eingefärbt.

verfügbaren Kristallstrukturen von IPMDH zeigten, dass bei der Substrat- und Kofaktorbindung eine induzierte Anpassung (*induced fit*) der Tasche erfolgt (Hurley und Dean 1994; Kadono *et al.* 1995). Insbesondere die Kofaktorbindung löst Konformationsveränderungen in unmittelbarer Nähe der Adenosin- und der Nikotinamidbindestelle aus.

So bewegen sich die C α -Atome der Reste 14–18, 284–299 und 334–347 in der ersten Domäne sowie der Reste 261–266 am Domänen-Interface um bis zu 2,5 Å bei der Bindung der Adenosineinheit (Hurley und Dean 1994; Kadono *et al.* 1995). Die Reste 77–91 in der ersten Domäne bewegen sich dagegen bei der Bindung der Nikotinamideinheit (Hurley und Dean 1994). Alle diese Reste befinden sich in Loops, mit der Ausnahme der Reste 16–18 und 90–91, die an den N-terminalen Enden zweier α -Helices liegen. Untersuchungen der konformationellen Veränderungen bei der Bindung des NAD⁺-Analogons ADP-Ribose zeigten, dass sich der aus den Resten 284–299 bestehende Loop über der ADP-Ribose schließt (Kadono *et al.* 1995). Diese Schließbewegung erlaubt es den Resten Asp 289 und Ile 290 die Adenosineinheit spezifisch zu binden.

Bei der Bestimmung der mikroskopischen Stabilitätswerte der Reste in den wichtigen Loops ist zu erwarten, dass Kontakte zwischen diesen Resten und ihren Nachbarn relativ instabil sind, da dies die Voraussetzung für ihre lokale Flexibilität ist. Aus der in Abbildung 17 gezeigten kombinierten Stabilitätskarte ist ersichtlich, dass die mikroskopischen Stabilitätswerte dieser Reste tatsächlich sehr gering sind. Die aus den Resten 14–18 und 77–91 bestehenden Loops verlieren ihre lokale Flexibilität bei Temperaturen um 10–20 K unterhalb der Phasenübergangstemperatur. Möglicherweise trägt dieser Verlust der lokalen Flexibilität wiederum dazu bei, dass das thermophile Enzym bei der Temperatur der maximalen Aktivität des mesophilen Enzyms nur noch marginal aktiv sein kann. Ein Austausch von Resten in diesen Strukturbereichen kann die lokale Flexibilität und die Aktivität des thermophilen Enzyms bei niedrigen Temperaturen erhöhen. Tatsächlich konnte durch das Einführen von Mutationen in diesen Bereichen eine Form der *T. thermophilus* IPMDH entwickelt werden, die bei niedrigen Temperaturen aktiv und trotzdem thermostabil ist (Hirose *et al.* 2001; Yasugi *et al.* 2001; Yasugi *et al.* 2001; Suzuki *et al.* 2002).

5.2 Analyse statischer Netzwerke von Phytasen

Phytasen sind Phytinsäure zersetzende Enzyme, die in der Schweine- und Geflügelzucht als Zusatzstoff in pflanzlichem Futter eingesetzt werden (Haefner *et al.* 2005). Im Gegensatz zu Wiederkäuern können Schweine und Geflügel den zu 80 % in Form von Phytinsäure gebundenen Phosphor in pflanzlicher Nahrung nicht nutzen. Dies macht die Zufütterung anorganischen Phosphats erforderlich, die bei falscher Dosierung zu erheblichen Umweltbelastungen führen kann (Braun *et al.* 2006). Durch den alternativen Zu-

satz von Phytase wird der in den Futterpflanzen vorhandene Phosphor als Nährstoff erschlossen. Auf die sonst übliche Zufütterung von Phosphat kann dann mit positivem Effekt für die Umwelt verzichtet werden: Allein in der Schweinezucht konnte der Phosphataustrag über die Gülle nach Einsatz von Phytasen um 30 % verringert werden (Braun *et al.* 2006). Die bei der Produktion des Futters (Pellettierung) auftretenden hohen Temperaturen (70–90 °C) machen es erforderlich, dass die eingesetzte Phytase thermostabil ist. Gleichzeitig muss sie aber bei niedrigeren Temperaturen im Magen der Tiere (30–40 °C) aktiv sein.

Derzeit sind verschiedene in der EU als Futtermittelzusatzstoff zugelassene Phytasen auf dem Markt erhältlich. So auch die unter der Handelsbezeichnung Natuphos von der BASF SE, Ludwigshafen, erhältliche Phytase aus dem Organismus *Aspergillus ficum* (Simon und Igbasan 2002). Aufgrund der hohen Anforderungen an das Produkt besteht auch weiterhin die Notwendigkeit, Phytasen mit verbesserten thermischen Eigenschaften, die sowohl die Stabilität als auch die Aktivität betreffen, zu entwickeln (Simon und Igbasan 2002). Dabei kommen verschiedene Ansätze des Protein Engineering zum Einsatz, wie beispielsweise die gerichtete Evolution, bei der das natürliche Evolutionsprinzip kopiert wird: Sequenzen des ursprünglichen Proteins werden zufällig mutiert und exprimiert. Die Bibliothek der so erzeugten Sequenzen kann bis zu 10^6 Mutanten enthalten. Anschließend wird nach gewünschter Kriterien (Stabilität, Aktivität, Selektivität) *gescreent* oder selektiert (Eijsink *et al.* 2005). In der Praxis stellt dabei die Größe der Bibliothek einen limitierenden Faktor dar. Daher sind Ansätze von Interesse, die bei dem Aufbau „zielgerichteter“ Bibliotheken helfen.

Im ersten Teil dieser Arbeit konnte anhand der Analyse homologer meso- und thermophiler Proteine gezeigt werden, dass der rechnerische Ansatz der Analyse statischer Netzwerke die thermische Stabilität einer Struktur abschätzen kann, andererseits Hinweise darauf gibt, in welchen Bereichen der Struktur thermostabilisierende Mutationen eingeführt werden können. Zudem konnte gezeigt werden, dass der Einfluss einer möglicherweise thermostabilisierenden Mutation auf die Bioaktivität des Enzyms bei niedrigen Temperaturen vorhergesagt werden kann. Damit ist die CNA möglicherweise hilfreich zur Unterstützung des Protein Engineering von Phytasen, indem sie die experimentell zu testenden Mutationen auf wenige Strukturbereiche begrenzt und dadurch den Aufbau einer zielgerichteten Bibliothek ermöglicht. Um dies zu überprüfen, werden in diesem Teil der Arbeit thermische Entfaltungssimulationen einer Serie von Phytasen durchgeführt und ausgewertet. Das Ziel dabei besteht nicht darin, tatsächlich eine Phy-

tase mit verbesserten Eigenschaften zu entwickeln, sondern zu zeigen, dass mit der Entwicklung der CNA die methodischen Grundlagen geschaffen wurden, ein solche Aufgabe durchzuführen. Der verwendete Datensatz wird in Kap. 5.2.1 beschrieben.

Von den meisten Phytasen aus dem Datensatz sind nur Homologiemodelle verfügbar, deren Qualität in der Regel geringer ist als die von Kristallstrukturen. Um diese Strukturen niedriger Qualität dennoch untersuchen zu können, muss die CNA auf konformationelle Ensembles von Strukturen angewendet werden, die zuvor zur Verbesserung („Verfeinerung“) der Homologiemodelle erzeugt wurden. Die mit Hilfe von MD-Simulationen erzeugten konformationellen Ensembles der Phytasen werden in Kap. 5.2.2 beschrieben. Für die Phytasen aus dem Datensatz werden thermische Stabilitäten vorhergesagt und in Kap. 5.2.3 mit experimentellen Daten verglichen. In Kap. 5.2.4 erfolgt die Beschreibung der Ergebnisse eingehender Analysen der thermischen Entfaltungssimulationen. Dabei werden mögliche molekulare Grundlagen der Thermostabilität zweier Phytasen aus dem Datensatz diskutiert und Entfaltungsregionen vorhergesagt, die mit Strukturbereichen verglichen werden, in denen thermostabilisierende Mutationen experimentell eingeführt wurden. In Kap. 5.2.5 werden die Vorhersagen für eine weitere Phytase experimentell getestet.

5.2.1 Datensatz

Der in diesem Teil der Arbeit verwendete Datensatz besteht aus sechs homologiemodellierten sowie drei Kristallstrukturen von 3-Phytasen (EC 3.1.3.8). Die Strukturen sowie deren Eigenschaften sind in Tabelle 6 aufgelistet. Die Phytasen aus dem Datensatz können in zwei Gruppen eingeteilt werden. Die erste Gruppe enthält Phytasen aus Askomyzeten (*Aspergillus*) und Mutanten davon. In ihr sind die Kristallstrukturen der PhyA Phytasen aus *A. ficum* (Natuphos) und *A. fumigatus* enthalten. Zudem gehören die anhand der Kristallstruktur der *A. ficum* PhyA Phytase homologiemodellierten Strukturen der Consensus-Phytase 10 und der Phytase DSM-1 zu dieser Gruppe. Die zweite Gruppe enthält Phytasen aus Enterobakterien (*Escherichia* und *Yersinia*) und Mutanten davon. In ihr sind die Kristallstruktur der *E. coli* AppA Phytase sowie die anhand des *E. coli* Enzyms homologiemodellierte Struktur der Phytase Quantum enthalten. Auch drei Phytasen aus *Y. mollaretii*, der Wildtyp (WT) sowie zwei Mutanten (M1 und M2), gehören zu dieser Gruppe.

Die Strukturen innerhalb einer Gruppe weisen eine hohe Sequenz- und Strukturähnlichkeit auf. Die Sequenzidentität von Phytasen aus verschiedenen Gruppen ist dagegen gering. So sind die Sequenzen der *A. ficum* PhyA und *E. coli* AppA Phytase nur zu

16 % identisch (Lim *et al.* 2000). In Abbildung 22 ist das Sequenzalignment der Consensus-Phytase 10, der *E. coli* AppA und der *Y. mollaretii* WT Phytase gezeigt. Die

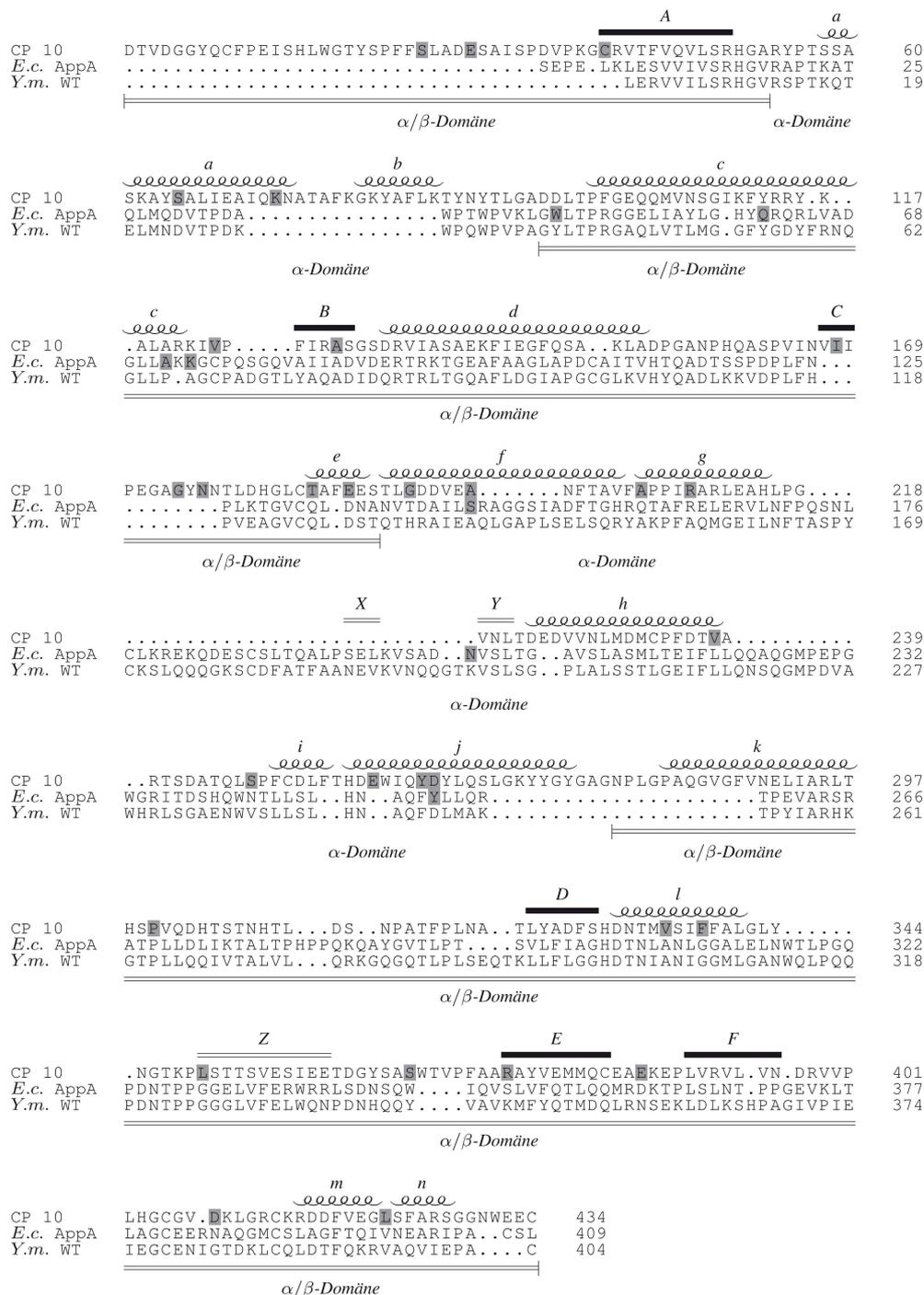


Abbildung 22. Sequenzalignment der Consensus-Phytase (CP) 10, der *E. coli* (E.c.) AppA und der *Y. mollaretii* (Y.m.) WT Phytase. Das Alignment wurde mit Hilfe der Software T-Coffee (Notredame *et al.* 2000) durchgeführt und manuell bearbeitet. Die für die Thermostabilität wichtigen Reste (siehe Text) sind grau schattiert. Die Kennzeichnung der Sekundärstrukturelemente folgt der Nomenklatur für die *A. ficum* PhyA Phytase (Kostrewa *et al.* 1997). Die β -Stränge X, Y und Z sind nur in den Strukturen der *E. coli* AppA und der *Y. mollaretii* WT Phytase vorhanden.

für die Thermostabilität wichtigen Reste und Sekundärstrukturelemente sind gekennzeichnet. Die Kennzeichnung der Sekundärstrukturelemente folgt der Nomenklatur für die *A. ficum* PhyA Phytase (Kostrewa *et al.* 1997). Trotz der geringen Sequenzidentität sind die Strukturen der Phytasen aus dem Datensatz sehr ähnlich. Sie werden durch eine große Tasche in zwei Domänen geteilt. Die größere α/β -Domäne enthält ein zentrales

Tabelle 6. Phytasen und ihre Eigenschaften.

| Struktur | | PDB ID | T_{50} bei pH4 ^{a,b} | T_{opt} bei pH 4 ^{b,c} | T_{opt} bei pH 5,5 ^b |
|----------|--------------------------|-------------------|---------------------------------|-----------------------------------|-----------------------------------|
| Nr. | Name | | | | |
| 1 | Consensus-Phytase 10 | 1ihp ^e | 82 | – ^f | – ^f |
| 2 | DSM-1 | 1ihp ^e | 74 | – ^f | – ^f |
| 3 | <i>A. ficum</i> PhyA | 1ihp ^d | – ^f | – ^f | 58 ^g |
| 4 | <i>A. fumigatus</i> PhyA | 1ska ^d | – ^f | – ^f | 60 ^g |
| 5 | Quantum | 1dkl ^e | 65 | 62 | 64 ^c |
| 6 | <i>E. coli</i> AppA | 1dkl ^d | 57 | 57 | 60 ^c |
| 7 | <i>Y. mollaretii</i> WT | 1dkl ^e | 55 | 52 | 55 ^c |
| 8 | <i>Y. mollaretii</i> M1 | 1dkl ^e | 48 | 48 | 50 ^c |
| 9 | <i>Y. mollaretii</i> M2 | 1dkl ^e | – ^f | 41 | 43 ^c |

^a Halbwertszeittemperatur, in °C; ^b Temperatur der maximalen Aktivität, in °C; ^c Nach Mitteilung durch Dr. Stefan Häfner (BASF SE, Ludwigshafen); ^d PDB ID; ^e PDB ID der Kristallstruktur, nach der homologi modelliert wurde; ^f Kein Wert verfügbar; ^g Nach (Wyss *et al.* 1998).

β -Faltblatt (*A–F*), das an jeder Seite von zwei langen α -Helices (*c* und *d* sowie *k* und *l*) begrenzt wird. Die kleinere α -Domäne besteht aus zwei oder dreilangen sowie mehreren kurzen α -Helices. Der Aufbau der α/β -Domäne ist weitgehend konserviert, bis auf die Tatsache, dass die enterobakteriellen Phytasen einen zusätzlichen Strang (*Z*) im β -Faltblatt besitzen. Die Strukturen aus der ersten und zweiten Gruppe unterscheiden sich durch den Aufbau der α -Domäne. So sind bei den Strukturen aus der zweiten Gruppe die α -Helices *a* und *b* verkürzt bzw. nicht vorhanden. Zudem besitzen diese Strukturen zwischen den α -Helices *g* und *h* ein zusätzliches zweisträngiges β -Faltblatt (*X* und *Y*). Insgesamt beträgt die RMS-Abweichung der *A. ficum* PhyA Phytase von der *E. coli* AppA Phytase 2,8 Å für die C α -Atome (Lim *et al.* 2000). Trotz der Unterschiede kann davon ausgegangen werden, dass es sich um homologe Proteine handelt.

Den Strukturen aus dem Datensatz wurden Daten über ihre thermische Stabilität zugeordnet. Für die Consensus-Phytase 10, die Phytase DSM-1 sowie für die *E. coli* AppA, die Quantum und die *Y. mollaretii* WT und M1 Phytasen standen Halbwertszeit-

temperaturen (T_{50}) der Strukturen bei pH 4 zur Verfügung, die laut persönlicher Mitteilung von Dr. Stefan Häfner (BASF SE, Ludwigshafen) so gemessen wurden, dass ihr Wert dem Schmelzpunkt der Struktur entspricht. Für die Phytase Quantum, die *E. coli*

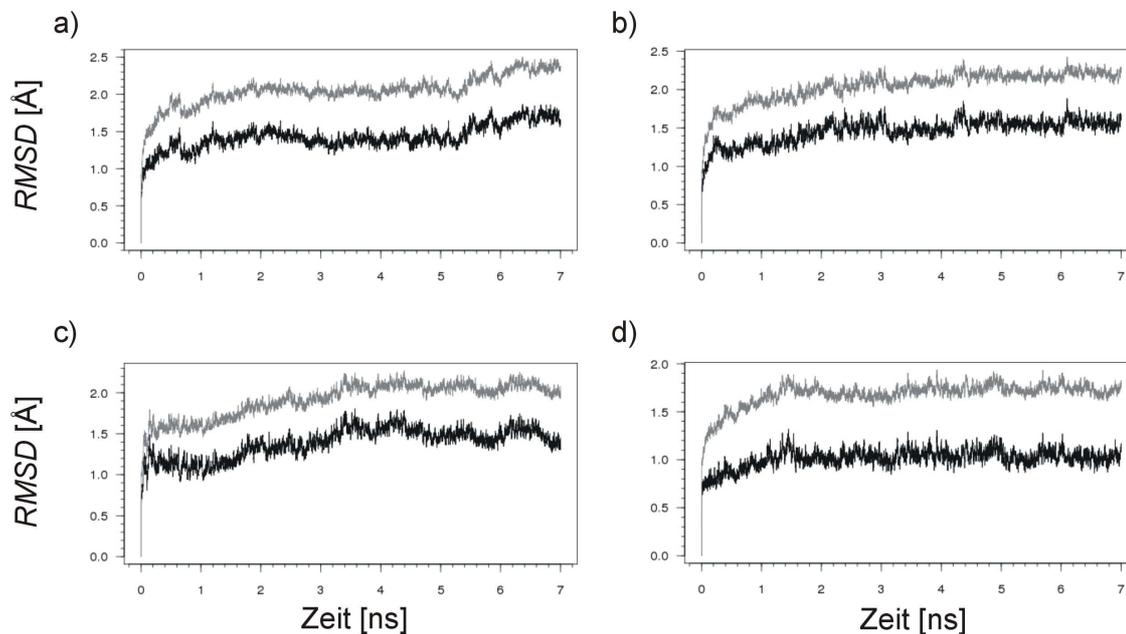


Abbildung 23. RMS-Abweichungen zwischen den Konformationen aus den Trajektorien und den jeweiligen Startstrukturen der Phytasen aus der ersten Gruppe, der Consensus-Phytase 10 (a), der Phytase DSM-1 (b) sowie der PhyA Phytase aus *A. ficum* (c) und *A. fumigatus* (d). Die RMS-Abweichungen für die C α -Atome sind schwarz dargestellt, für alle Schweratome grau.

AppA und die *Y. mollaretii* WT und M1 Phytasen wurden zudem Temperaturen der maximalen Aktivität dieser Enzyme (T_{opt}) bei pH 4 und 5,5 bestimmt. Für die PhyA Phytasen aus *A. ficum* und *A. fumigatus* sind T_{opt} bei pH 5,5 aus der Literatur vorhanden (Wyss *et al.* 1998). Diese Daten deuten darauf hin, dass die PhyA Phytasen von *A. ficum* und *A. fumigatus* ähnlich thermostabil sind wie die *E. coli* AppA Phytase. Die relativ niedrige T_{opt} der *Y. mollaretii* M2 Phytase deutet zudem darauf hin, dass die Thermostabilität dieses Enzyms niedriger ist als die der *Y. mollaretii* WT und M1 Phytasen.

5.2.2 Erzeugung konformationeller Ensembles

Erste vorläufige Untersuchungen der Phytasen aus dem Datensatz wurden durchgeführt, indem die CNA direkt auf die homologiemodellierten Strukturen angewendet wurde. Vorhergesagte Thermostabilitäten stimmten dabei nicht mit den experimentellen Daten überein. Zudem deuteten Analysen der thermischen Entfaltungssimulationen auf das Vorhandensein struktureller Defekte hin. Im ersten Teil dieser Arbeit wurde am

Beispiel der 3-Isopropylmalatdehydrogenase (IPMDH) aus *E. coli* gezeigt, dass kleinere, von Kristallkontakten verursachte strukturelle Defekte mit Hilfe kurzer Energieminimierungen behoben werden können (Abschnitt 5.1.2.3). Analysen der thermischen

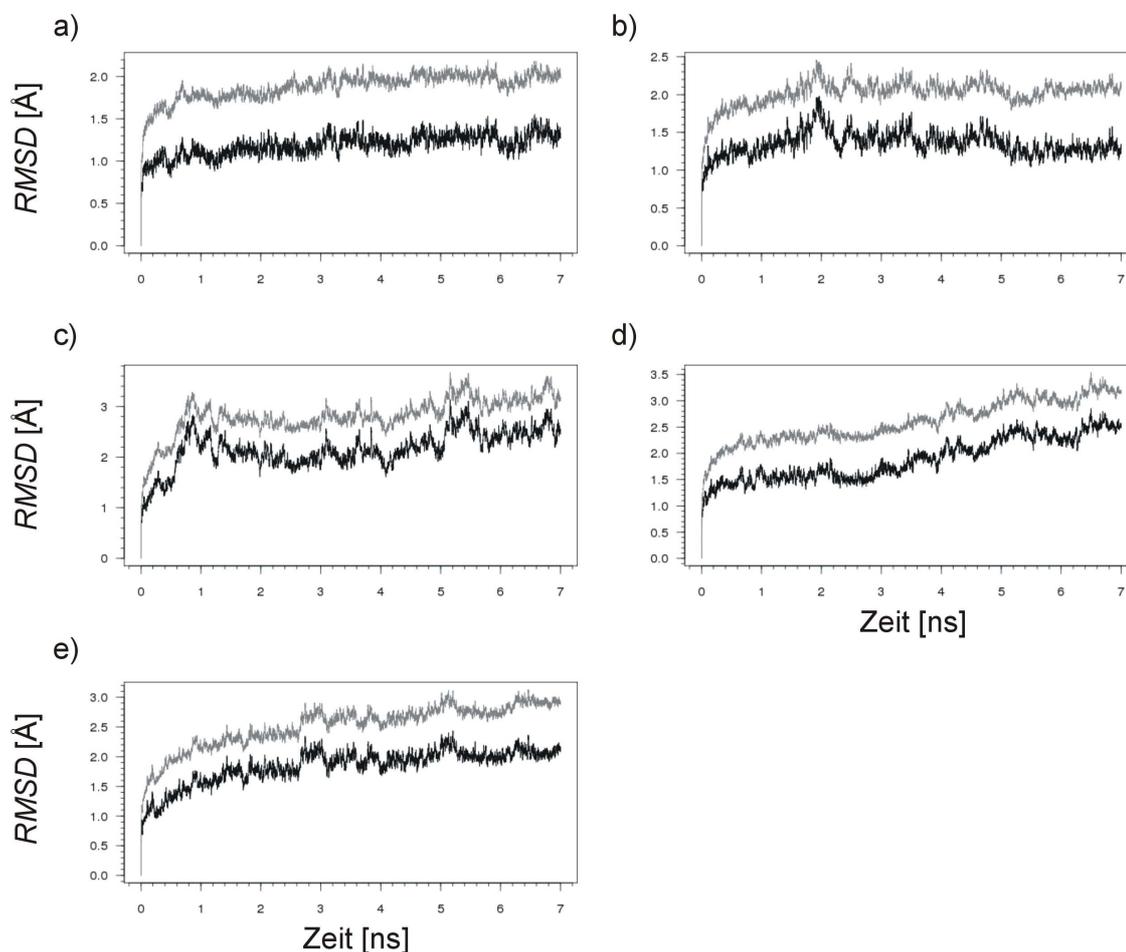


Abbildung 24. RMS-Abweichungen zwischen den Konformationen aus den Trajektorien und den jeweiligen Startstrukturen der Phytasen aus der zweiten Gruppe, der Phytase Quantum (a), der *E. coli* AppA Phytase (b) sowie der *Y. mollaretii* WT (c), M1 (d) und M2 Phytase (e). Die RMS-Abweichungen für die C α -Atome sind schwarz dargestellt, für alle Schweratome grau.

Entfaltungssimulationen von minimierten Strukturen der Phytasen aus dem Datensatz zeigten aber, dass die in Homologiemodellen vorhandenen strukturellen Defekte nicht mit Hilfe kurzer Energieminimierungen behoben werden können. In der Regel werden zur Verbesserung („Verfeinerung“) von Homologiemodellen oder anderen Strukturen schlechter Qualität MD-Simulationen verwendet (Fan und Mark 2004). Dabei erhält man einen Satz von Konformationen (Trajektorie), der die Bewegung aller Atome über einen bestimmten Zeitraum beschreibt und dem den nativen Zustand des Proteins repräsentierende konformationellen Ensemble entspricht.

Für die Strukturen der Phytasen aus dem Datensatz wurden Trajektorien von 7 ns Länge erzeugt. Da mit Hilfe der CNA erzeugte thermische Entfaltungssimulationen nur dann auf der selben Temperaturskala verglichen werden können, wenn sie auf Strukturen „gleicher Herkunft“ basieren (Abschnitt 5.1.2.3), wurden außer den homologiemo- dellierten auch die Kristallstrukturen MD-Simulationen unterzogen. Die Bewertung der MD-Simulationen erfolgte über die Untersuchung der *RMS*-Abweichungen zwischen den Konformationen aus den Trajektorien und den jeweiligen Startstrukturen. Für die Phytasen aus der ersten und zweiten Gruppe sind diese in Abbildung 23 bzw. Abbildung 24 gezeigt. Durchweg betragen die *RMS*-Abweichungen weniger als 2,0 und 2,5 Å für die C α - bzw. alle Schweratome. Während der Simulation ändern sich die Werte zudem nur geringfügig, was auf die Stabilität der Simulationen hindeutet. Eine Ausnahme bilden die Phytasen aus *Y. mollaretii*. Hier werden *RMS*-Abweichungen bis zu 3,0 und 3,5 Å für die C α - bzw. alle Schweratome erreicht. Insbesondere die MD-Simulation der *Y. mollaretii* M1 Phytase scheint nicht stabil zu sein. Selbst nach 7 ns ist noch ein Zunehmen der *RMS*-Abweichung von der Startstruktur zu beobachten. Bei der späteren Auswertung der thermischen Entfaltungssimulation muss dies berücksichtigt werden.

Aus den jeweils letzten 5 ns der Trajektorien wurden alle 40 ps unabhängige Einzelstrukturen (*snapshots*) extrahiert. Gohlke *et al.* konnten für die Untersuchung von Veränderungen der Rigidität und Flexibilität eines Proteinkomplexes während der Bindung zeigen, dass bereits ein Intervall von 20 ps ausreicht, um unabhängige Einzelstrukturen zu extrahieren (Gohlke *et al.* 2004). Der untersuchte Proteinkomplex hatte dabei eine ähnliche Größe wie eine Phytase. Die resultierenden konformationellen Ensembles umfassen jeweils 125 Einzelstrukturen.

5.2.3 Thermostabilitäten von Phytasen

Für alle Phytasen aus dem Datensatz wurde die thermische Entfaltung mit Hilfe der ensemblebasierten CNA simuliert. Um die Rechenzeit zu begrenzen wurden die Entfaltungssimulationen mit einem Temperaturschritt von 2 K (entsprechend $-0,1 \text{ kcal mol}^{-1}$) durchgeführt. Mit Hilfe von Parametern aus der Perkulations- und Netzwerktheorie können Veränderungen der Rigidität und Flexibilität in den Einzelstrukturen aus den konformationellen Ensembles untersucht und hinsichtlich der Thermostabilität interpretiert werden. Insbesondere können Phasenübergangstemperaturen (T_p) bestimmt und über das jeweilige Ensembles gemittelt werden. Die gemittelte Phasenübergangstemperatur des Ensembles (T_p') entspricht der experimentell bestimmten Schmelztemperatur

(T_m). In diesem Kapitel werden die gemittelten Phasenübergangstemperaturen der Phytasen aus dem Datensatz bestimmt und mit experimentellen bestimmten Halbwertszeittemperaturen (T_{50}) verglichen.

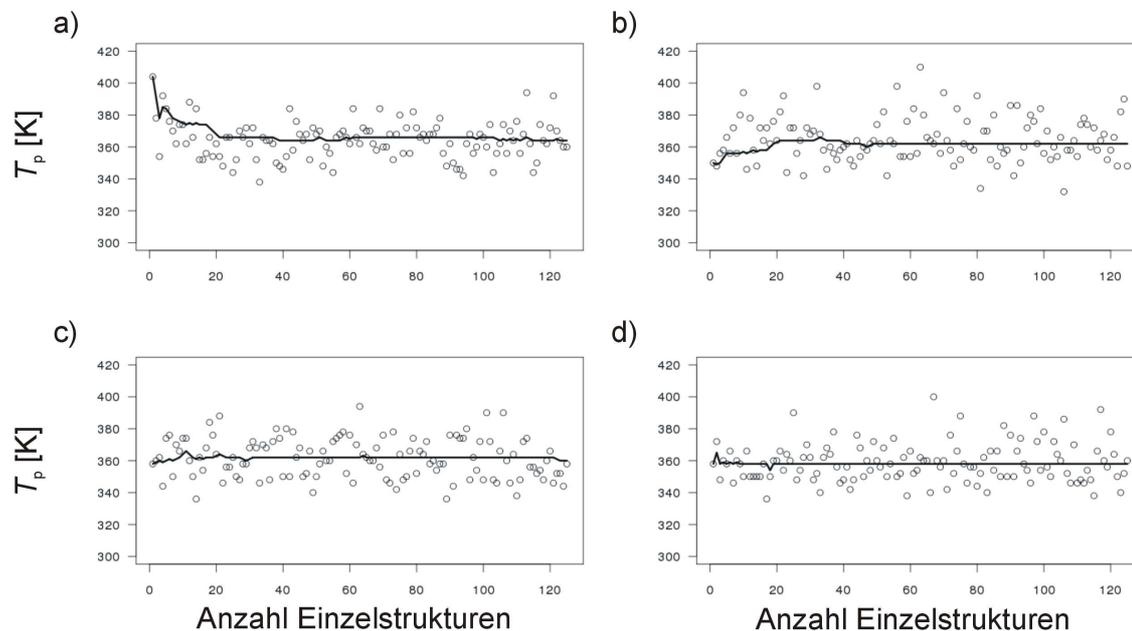


Abbildung 25. Phasenübergangstemperaturen der Einzelstrukturen aus den konformationellen Ensembles der Phytasen aus der ersten Gruppe, der Consensus-Phytase 10 (a), der Phytase DSM-1 (b) sowie der PhyA Phytase aus *A. ficum* (c) und *A. fumigatus* (d). Zudem ist die Entwicklung der über das Ensemble gemittelten Phasenübergangstemperatur (entsprechend dem Median der Verteilung) als Funktion der Anzahl der verwendeten Einzelstrukturen aufgetragen.

Phasenübergangstemperaturen der Einzelstrukturen aus den konformationellen Ensembles der Phytasen wurden bestimmt, indem die Clusterkonfigurationsentropie berechnet und ihr Verlauf automatisch ausgewertet wurde. In Abbildung 25 und Abbildung 26 sind die Phasenübergangstemperaturen der Einzelstrukturen aus den konformationellen Ensembles der Phytasen aus der ersten bzw. zweiten Gruppe gezeigt. Zudem ist die Entwicklung der über das Ensemble gemittelten Phasenübergangstemperatur als Funktion der Anzahl der verwendeten Einzelstrukturen aufgetragen. Die „gemittelte“ Phasenübergangstemperatur wird bestimmt, indem der Median der Verteilung berechnet wird. Der Median wird verwendet, da dieser unempfindlicher gegenüber Ausreißern in einer Verteilung ist als das arithmetische Mittel, obgleich weitere Untersuchungen zeigen, dass sich das arithmetische Mittel der Verteilung um höchstens 2 K vom Median unterscheidet. Für die Phytasen aus der ersten Gruppe ist die gemittelte Phasenübergangstemperatur bereits nach der Verwendung von 40 Einzelstrukturen kon-

vergiert. Für die Phytasen der zweiten Gruppe ist dies erst nach der Verwendung von 60 Einzelstrukturen der Fall. Für die *Y. mollaretii* M1 Phytase sind auch nach der Verwendung von 120 Einzelstrukturen noch Veränderungen in der gemittelten Phasenübergangstemperatur zu beobachten. Bei der weiteren Auswertung der thermischen Entfaltungssimulation muss dies berücksichtigt werden.

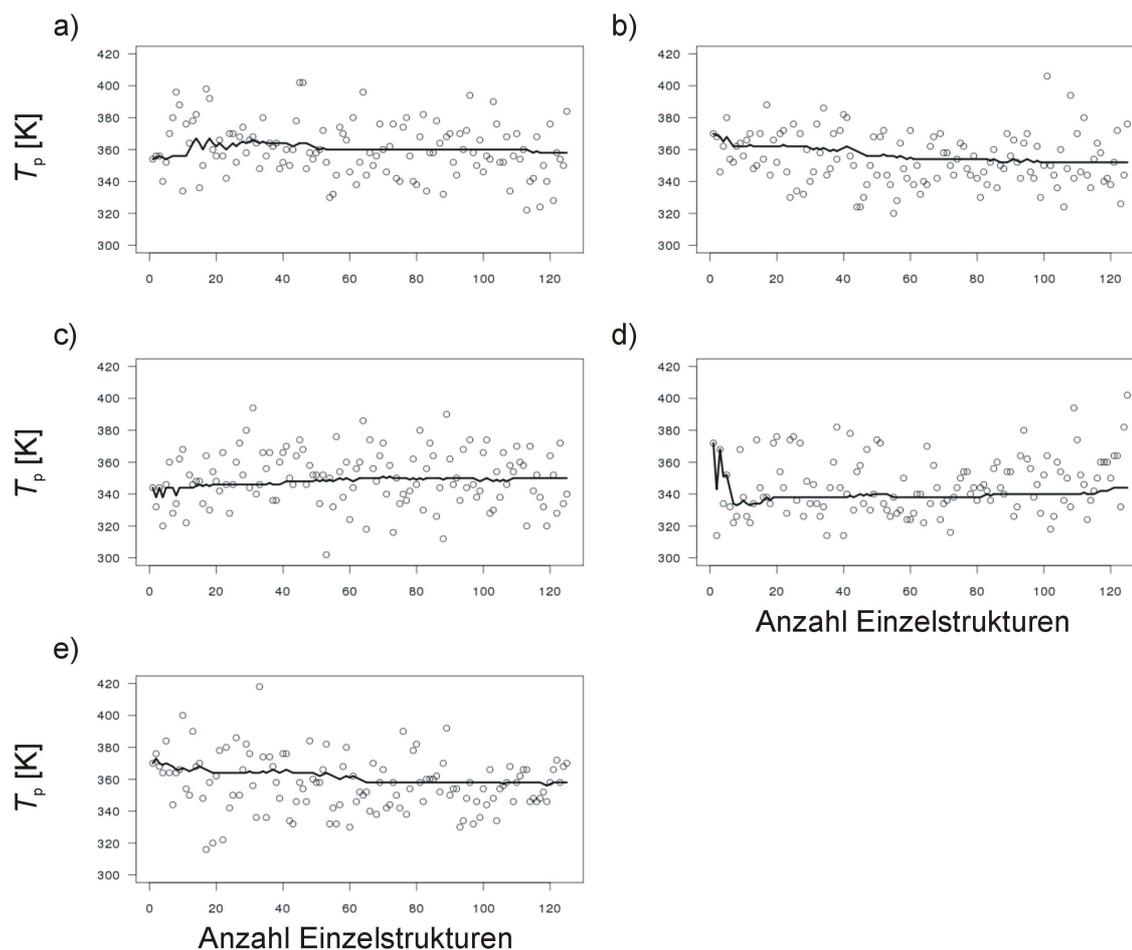


Abbildung 26. Phasenübergangstemperaturen der Einzelstrukturen aus den konformationellen Ensembles der Phytasen aus der ersten Gruppe, der Phytase Quantum (a), der *E. coli* AppA Phytase (b) sowie der *Y. mollaretii* WT (c), M1 (d) und M2 Phytase (e). Zudem ist die Entwicklung der über das Ensemble gemittelten Phasenübergangstemperatur (entsprechend dem Median der Verteilung) als Funktion der Anzahl der verwendeten Einzelstrukturen aufgetragen.

Die relativen gemittelten Phasenübergangstemperaturen der Phytasen stimmen sehr gut mit den experimentellen Daten überein (Tabelle 7). Die Korrelation zwischen den Halbwertszeit- und Phasenübergangstemperaturen ist hoch und signifikant. Der Korrelationskoeffizient beträgt 0,91 für die Phytasen aus dem Datensatz, für die experimentelle Daten vorhanden waren. Die Phasenübergangstemperaturen für die *A. ficum* und

A. fumigatus PhyA Phytasen sowie die *Y. mollaretii* M2 Phytase können nicht unmittelbar mit experimentellen Daten verglichen werden. Für die *A. ficum* und *A. fumigatus* PhyA Phytasen wurde eine ähnliche Thermostabilität wie für die *E. coli* AppA Phytase

Tabelle 7. Halbwertszeit- (ΔT_{50}) und Phasenübergangstemperaturdifferenzen (ΔT_p) von der Halbwertszeit- bzw. Phasenübergangstemperatur der Consensusphytase 10 aller Phytasen.

| Struktur ^a | ΔT_{50} bei pH 5,5 ^b | ΔT_p ^c |
|-----------------------|---|---------------------------|
| 1 | 0 | 0 |
| 2 | -8 | -2 |
| 3 | - ^d | -4 |
| 4 | - ^d | -6 |
| 5 | -17 | -6 |
| 6 | -25 | -12 |
| 7 | -27 | -14 |
| 8 | -34 | -20 |
| 9 | - ^d | -6 |

^a Nummern entsprechen den Nummern der Strukturen in Tabelle 6; ^b Halbwertszeittemperaturdifferenz von der Halbwertszeit der Struktur 1, in K; ^c Phasenübergangstemperaturdifferenz von der Phasenübergangstemperatur der Struktur 1, in K; ^d Kein Wert verfügbar.

vermutet (Kap. 5.2.1). Die vorhergesagte Thermostabilität ist jedoch deutlich höher. Auch für die *Y. mollaretii* M2 Phytase ist die vorhergesagte Thermostabilität deutlich höher als vermutet (Kap. 5.2.1). Ob die Thermostabilitäten dieser Phytasen tatsächlich falsch vorhergesagt wurden lässt sich aber erst klären, wenn die Halbwertszeittemperaturen dieser Proteine unter den selben experimentellen Bedingungen wie für die anderen Phytasen aus dem Datensatz bestimmt werden.

5.2.4 Molekulare Ursachen der Thermostabilität

Die hohe Korrelation zwischen der vorhergesagten und experimentell bestimmten Thermostabilität der Phytasen aus dem Datensatz ist ein Hinweis darauf, dass die für die Thermostabilität wichtigen mikroskopischen Eigenschaften der Struktur korrekt beschrieben werden. In den folgenden Abschnitten werden die Entfaltungssimulationen der Consensus-Phytase 10 und der *E. coli* AppA Phytase eingehend analysiert. Dafür werden zunächst für repräsentative Einzelstrukturen (lokale) Strukturbereiche bestimmt, an denen der (globale) Zerfall der Struktur beginnt, und die deswegen für die thermische Stabilität von großer Wichtigkeit sind. Unter Verwendung aller Einzelstrukturen aus den konformationellen Ensembles werden dann Wahrscheinlichkeiten berechnet,

mit denen bestimmte Strukturbereiche Entfaltungsregionen darstellen. Bereiche mit hoher „Entfaltungsregionwahrscheinlichkeit“ (w_{ER}) werden für die untersuchten Proteine mit Regionen verglichen, in denen thermostabilisierende Mutationen experimentell eingeführt wurden.

5.2.4.1 Entfaltungsregionen der Consensus-Phytase 10

Die Consensus-Phytase 10 ist eine thermostabile Variante der Phytasen aus der ersten Gruppe, die mit Hilfe des so genannten Consensus-Ansatzes entwickelt wurde (Lehmann *et al.* 2000; Lehmann und Wyss 2001; Lehmann *et al.* 2002). Dieser Ansatz beruht auf der Hypothese, dass Consensus-Aminosäuren aus Sequenzalignments homologer Proteine mehr zur Stabilität beitragen als Nichtconsensus-Aminosäuren (Steipe *et al.* 1994; Lehmann *et al.* 2000). Die Sequenz der Consensus-Phytase 10 wurden aus dem Alignment von insgesamt 19 Sequenzen von Phytasen aus Askomyzeten berechnet, darunter auch die PhyA Phytasen aus *A. ficum* und *A. fumigatus* (Lehmann *et al.* 2002). Die Sequenzen der Consensus-Phytase 10 und der *A. ficum* PhyA Phytase sind zu 75 % identisch. Die anhand der *A. ficum* PhyA Phytase homologiemodellierte Struktur der Consensus-Phytase 10 zeigt, dass die beiden Enzyme auch strukturell sehr ähnlich sind. Aufgrund des Vorhandenseins neutraler Mutationen ist nicht davon auszugehen, dass alle zwischen der *A. ficum* PhyA Phytase und der Consensus-Phytase 10 ausgetauschten Reste zur Thermostabilisierung beitragen.

Veränderungen der Rigidität und Flexibilität im Netzwerk der Consensus-Phytase 10 während der Simulation der thermischen Entfaltung können unter Verwendung einer repräsentativen Einzelstruktur aus dem konformationellen Ensemble untersucht werden. Die für die Thermostabilität wichtigen mikroskopischen Eigenschaften der Struktur werden bestimmt, indem die RCD für die Consensus-Phytase 10 unmittelbar vor und nach dem Phasenübergang verglichen werden (Abbildung 27). Aus Abbildung 27 a) ist ersichtlich, dass die Struktur der Consensus-Phytase 10 unmittelbar vor dem Phasenübergang von einem rigiden Kern dominiert wird. Dieser Kern stellt einen rigiden Cluster dar, der das β -Faltblatt und die α -Helices *c* und *d* in der α/β -Domäne sowie Teile des Domänen-Interface und der α -Domäne umfasst. In Abbildung 27 a) ist dieser Cluster blau eingefärbt. Der Zerfall des Kerns verursacht den Phasenübergang. Aus Abbildung 27 b) ist ersichtlich, dass unmittelbar nach dem Phasenübergang nur noch wenige, einzelne Sekundärstrukturelemente umfassende rigide Cluster im Netzwerk vorhanden sind.

Unter Verwendung aller Einzelstrukturen aus dem konformationellen Ensemble können nun Wahrscheinlichkeiten berechnet werden, mit denen bestimmte Strukturbereiche der Consensus-Phytase 10 Entfaltungsregionen darstellen. In Abbildung 28 ist eine Struktur der Consensus-Phytase 10 gezeigt, die entsprechend der Entfaltungsregionswahrscheinlichkeit (w_{ER}) jedes Rests unterschiedlich angefärbt ist, wobei rot eine hohe w_{ER} bedeutet und blau eine niedrige. Eine hohe w_{ER} haben insbesondere die Reste des β -Faltblatts und der α -Helices *c*, *d* und *l* in der α/β -Domäne. Aber auch Teile des

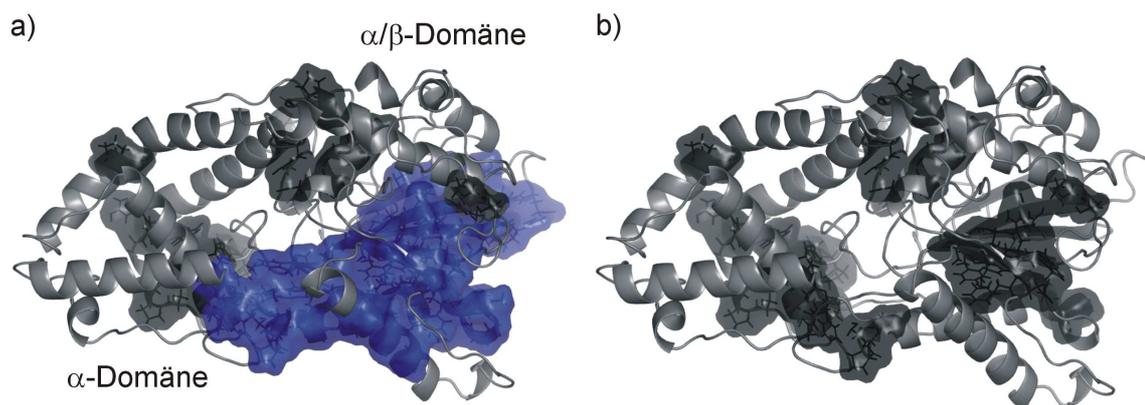


Abbildung 27. Snapshots aus der thermischen Entfaltungssimulation einer repräsentativen Einzelstruktur aus dem konformationellen Ensemble der Consensus-Phytase 10 unmittelbar vor (a) und nach dem Phasenübergang (b) bei 364 K. Es ist jeweils die Zerlegung des Netzwerks in rigide Cluster (RCD) gezeigt. Die rigiden Cluster sind als einheitlich gefärbte Bereiche dargestellt, wobei der Riesencuster blau eingefärbt ist. Die übrigen Cluster sind dunkel schattiert. Von der Proteinstruktur ist nur das Rückgrat gezeigt und dunkelgrau schattiert.

N-Terminus, des Domänen-Interface und der α -Helices *h* in der α -Domäne stellen Entfaltungsregionen dar. Ein Vergleich der Vorhersage mit experimentellen Daten ist schwierig, da der Effekt auf die Thermostabilität nur für wenige der zwischen der *A. ficum* PhyA Phytase und der Consensus-Phytase 10 ausgetauschten Reste gemessen oder veröffentlicht wurde. In der in Abbildung 22 gezeigten Sequenz der Consensus-Phytase 10 sind Reste gekennzeichnet, deren Austausch einen positiven oder negativen Effekt auf die der Consensus-Phytase 10 sehr ähnlichen Consensus-Phytase 1 hatte (Lehmann *et al.* 2002). In Abbildung 28 sind die $C\alpha$ -Atome dieser Reste als Kugeln dargestellt.

Die vorhersagten Entfaltungsregionen stimmen mit diesen experimentellen Daten dahingehend überein, dass viele die Thermostabilität verändernde Mutationen in der α/β -Domäne eingeführt wurden, insbesondere in der unmittelbaren Umgebung der β -Stränge *A*, *B*, *C* und *E* sowie den α -Helices *c* und *l*. Auch die hohe w_{ER} der Reste am

Domänen-Interface und der C-terminalen Reste der α -Helix *h* stimmt mit den experimentellen Daten überein. Eine ganze Reihe die Thermostabilität verändernde Mutationen wurden aber auch im Bereich der α -Helices *a*, *e–g* sowie *i* und *j* sowie *a* in der α -Domäne eingeführt. In diesem Bereich der Struktur konnten keine Entfaltungsregio-

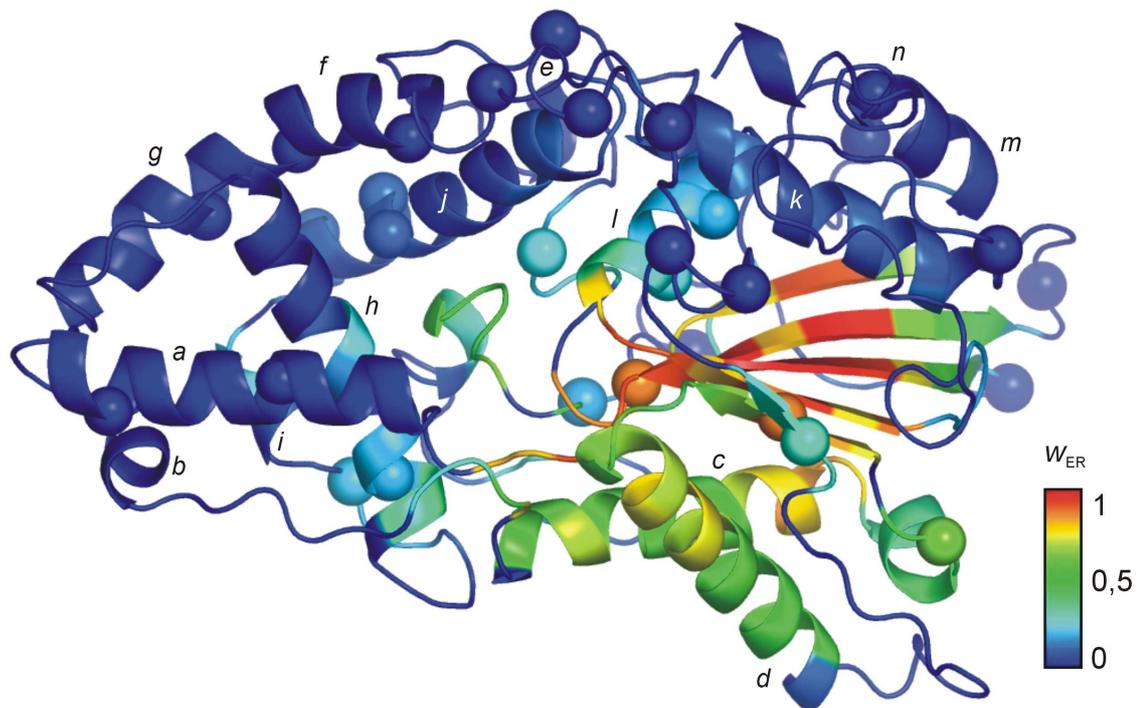


Abbildung 28. Struktur der Consensus-Phytase 10. Es ist nur das Rückgrat gezeigt und entsprechend der Entfaltungsregionwahrscheinlichkeit (w_{ER}) jedes Rests unterschiedlich angefarbt. Rot bedeutet eine hohe w_{ER} und blau eine niedrige. Die Kennzeichnung der α -Helices entspricht der Kennzeichnung in dem in Abbildung 22 gezeigten Alignment. Die Ca -Atome der Reste, deren Austausch einen positiven oder negativen Effekt auf die der Consensus-Phytase 10 sehr ähnlichen Consensus-Phytase 1 hatte (Lehmann *et al.* 2002), sind als Kugeln dargestellt.

nen bestimmt werden. Eine mögliche Ursache dieser nicht erfolgreichen Vorhersage liegt darin, dass die Thermostabilisierung von Phytasen *in vivo* möglicherweise auch mit Hilfe von Glykosylierungen erfolgt (Han und Lei 1999; Wyss *et al.* 1999). Glykosylierungsstellen der *A. ficum* und *A. fumigatus* PhyA Phytasen konnten unter anderem im Bereich der α -Helices *e–g* und *a* identifiziert werden (Kostrewa *et al.* 1997; Wyss *et al.* 1999; Xiang *et al.* 2004). Der Einfluss dieser Glykosylierungen kann mit Hilfe der CNA nicht untersucht werden. Es ist somit nicht auszuschließen, dass die Nichtberücksichtigung der Glykosylierungen Ursache dafür ist, dass in diesen Bereichen keine Entfaltungsregionen vorhergesagt werden konnten.

5.2.4.2 Entfaltungsregionen der *E. coli* AppA Phytase

Die *E. coli* AppA Phytase ist der Prototyp der Phytasen aus der zweiten Gruppe und diente als Ausgangspunkt für die Entwicklung der thermostabilen Phytase Quantum. Im Gegensatz zu der PhyA Phytase aus *A. ficum* kann das *E. coli* Enzym nicht direkt als Futtermittelzusatzstoff verwendet werden, da es die bei der Pellettierung auftretenden hohen Temperaturen nicht übersteht (Igbasan *et al.* 2000). Der Vorteil der *E. coli* AppA Phytase gegenüber der Phytasen aus Askomyceten besteht in einer höheren Resistenz gegenüber Pepsin und Pankreatin im Magen der Tiere und in einer achtfach höheren Aktivität (Igbasan *et al.* 2001). Die Entwicklung thermostabiler Varianten der *E. coli* AppA Phytase ist deswegen sehr attraktiv.

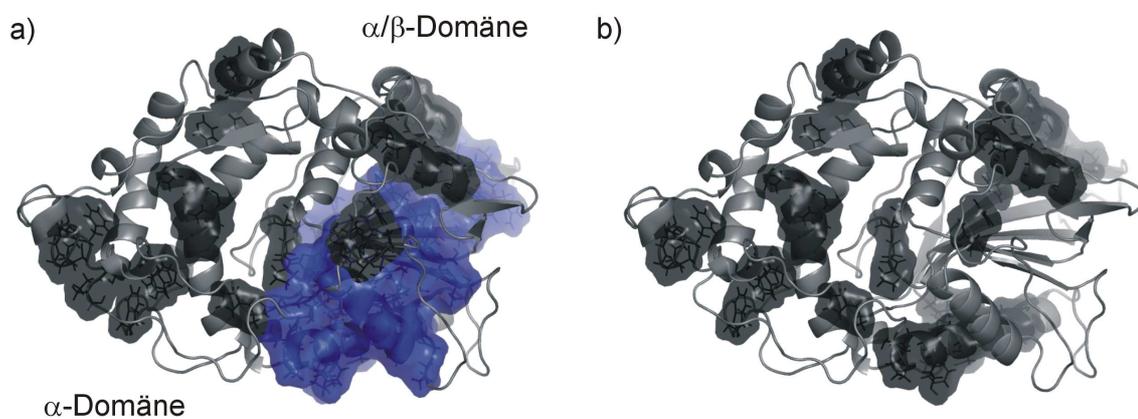


Abbildung 29. Snapshots aus der thermischen Entfaltungssimulation einer repräsentativen Einzelstruktur aus dem konformationellen Ensemble der *E. coli* AppA Phytase unmittelbar vor (a) und nach dem Phasenübergang (b) bei 352 K. Es ist jeweils die Zerlegung des Netzwerks in rigide Cluster (RCD) gezeigt. Die rigiden Cluster sind als einheitlich gefärbte Bereiche dargestellt, wobei der Riesencluster blau eingefärbt ist. Die übrigen Cluster sind dunkel schattiert. Von der Proteinstruktur ist nur das Rückgrat gezeigt und dunkelgrau schattiert.

Wie für die Consensus-Phytase 10, können auch Veränderungen der Rigidität und Flexibilität im Netzwerk der *E. coli* AppA Phytase während der Simulation der thermischen Entfaltung unter Verwendung einer repräsentativen Einzelstruktur aus dem konformationellen Ensemble untersucht werden. Die für die Thermostabilität wichtigen mikroskopischen Eigenschaften der Struktur werden wieder bestimmt, indem die RCD für die *E. coli* AppA Phytase unmittelbar vor und nach dem Phasenübergang verglichen werden (Abbildung 29). Aus Abbildung 29 a) ist ersichtlich, dass der Kern, der die Struktur der *E. coli* AppA Phytase unmittelbar vor dem Phasenübergang dominiert, einen rigiden Cluster darstellt, der das β -Faltblatt und die α -Helices *c* und *d* in der

α/β -Domäne umfasst. Auch Teile des Domänen-Interface sind im Kern enthalten, aber weniger als im Fall der Consensus-Phytase 10 (Abbildung 27 a). Unmittelbar nach dem Phasenübergang sind nur noch wenige, einzelne Sekundärstrukturelemente umfassende rigide Cluster im Netzwerk vorhanden (Abbildung 29 b).

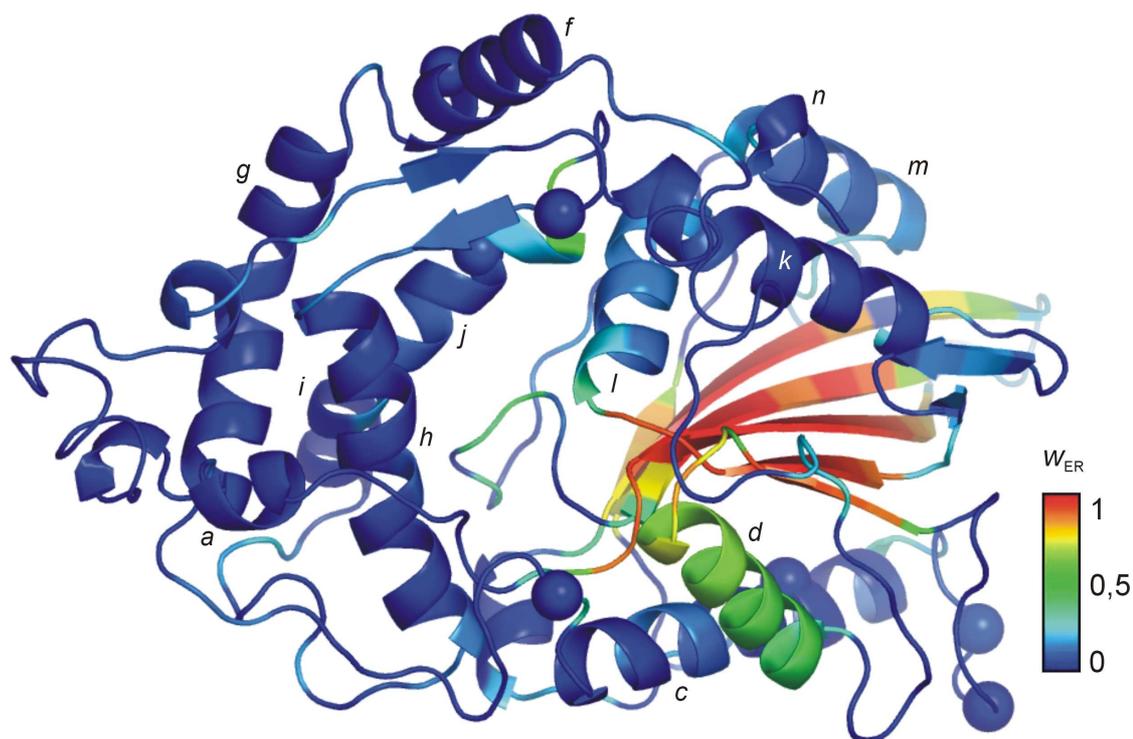


Abbildung 30. Struktur der *E. coli* AppA Phytase. Es ist nur das Rückgrat gezeigt und entsprechend der Entfaltungsregionwahrscheinlichkeit (w_{ER}) jedes Rests unterschiedlich angefärbt. Rot bedeutet eine hohe w_{ER} und blau eine niedrige. Die Kennzeichnung der α -Helices entspricht der Kennzeichnung in dem in Abbildung 22 gezeigten Alignment. Die $C\alpha$ -Atome der Reste, deren Austausch zur Entwicklung der Phytase Quantum führten, sind als Kugeln dargestellt.

Unter Verwendung aller Einzelstrukturen aus dem konformationellen Ensemble können wieder Wahrscheinlichkeiten berechnet werden, mit denen bestimmte Strukturbereiche der *E. coli* AppA Phytase Entfaltungsregionen darstellen. In Abbildung 30 ist eine Struktur der *E. coli* AppA Phytase gezeigt, die entsprechend der Entfaltungsregionwahrscheinlichkeit (w_{ER}) jedes Rests unterschiedlich angefärbt ist, wobei rot eine hohe w_{ER} bedeutet und blau eine niedrige. Eine hohe w_{ER} haben insbesondere die Reste des β -Faltblatts und der α -Helix *d* in der α/β -Domäne. Auch die α -Helices *c* und *l* in der α/β -Domäne sowie die C-terminalen Reste der α -Helix *h* und die N-terminalen Reste der α -Helix *j* in der α -Domäne am Domänen-Interface stellen Entfaltungsregionen dar, wenn auch mit geringer Wahrscheinlichkeit.

Die vorhergesagten Entfaltungsregionen können mit Regionen verglichen werden, in denen thermostabilisierende Mutationen experimentell eingeführt wurden und zur Entwicklung der Phytase Quantum führten. Dabei handelt es sich um sieben Mutationen im Bereich der α -Helix *c* in der α/β -Domäne sowie der α -Helices *f*, *h* und *j* in der α -Domäne (Abbildung 22). In Abbildung 30 sind die Ca-Atome der Reste, deren Austausch zur Entwicklung der Phytase Quantum führten, als Kugeln dargestellt. Alle diese Strukturbereiche liegen in der unmittelbaren Umgebung von Bereichen mit einer hohen vorhergesagten Entfaltungsregionswahrscheinlichkeit, mit der Ausnahme einer Mutation in der α -Helix *f* in der α -Domäne. In eben diesem Strukturbereich fand sich bereits für Consensus-Phytase 10 keine Übereinstimmung zwischen der Simulation und den experimentellen Daten. Insgesamt ist die Übereinstimmung geringer als im Fall der Consensus-Phytase 10, was auch daran liegt, dass nicht genügend experimentelle Daten zur Validierung der Simulation zur Verfügung standen.

5.2.5 Vorhersage und Testung thermostabilisierender Mutationen

Im vorangehenden Kapitel wurde gezeigt, dass die für die Consensus-Phytase 10 und die *E. coli* AppA Phytase vorhergesagten Entfaltungsregionen gut mit Bereichen übereinstimmen, in denen thermostabilisierende Mutationen experimentell eingeführt wurden. Dies deutet darauf hin, dass die CNA tatsächlich hilfreich zur Unterstützung des Protein Engineering von Phytasen ist. Um dies genauer zu überprüfen, wurden Entfaltungsregionen der *Y. mollaretii* WT Phytase vorhergesagt und durch Arbeitsgruppen an der Rheinisch-Westfälischen Technischen Hochschule, Aachen, und der BASF SE, Ludwigshafen, experimentell getestet. Yersinien sind Enterobakterien, die entfernt mit *E. coli* verwandt sind. Die Sequenzen der *Y. mollaretii* WT und der *E. coli* AppA Phytase sind zu 45 % identisch. Die *Y. mollaretii* Phytase stellt einen neuen Ausgangspunkt für die Entwicklung von Phytasen mit verbesserten thermischen Eigenschaften dar.

Zur Vorhersage von Regionen, in denen thermostabilisierende Mutationen eingeführt werden können, müssen Veränderungen der Rigidität und Flexibilität im Netzwerk der *Y. mollaretii* WT Phytase während der Simulation der thermischen Entfaltung unter Verwendung einer repräsentativen Einzelstruktur aus dem konformationellen Ensemble untersucht werden. Dies erfolgt genau so, wie das bereits für die Consensus-Phytase 10 und die *E. coli* AppA Phytase gemacht wurde. Die für die Thermostabilität wichtigen mikroskopischen Eigenschaften der Struktur werden dabei bestimmt, indem die RCD für die *Y. mollaretii* WT Phytase unmittelbar vor und nach dem Phasenübergang verglichen werden (Abbildung 31). Aus Abbildung 31 a) ist ersichtlich, dass der Kern,

der die Struktur der *Y. mollaretii* WT Phytase unmittelbar vor dem Phasenübergang dominiert, einen rigiden Cluster darstellt, der das β -Faltblatt und die α -Helix *d* und Teile der α -Helix *c* in der α/β -Domäne umfasst. Unmittelbar nach dem Phasenübergang sind nur noch wenige, einzelne Sekundärstrukturelemente umfassende rigide Cluster im Netzwerk vorhanden (Abbildung 31 b).

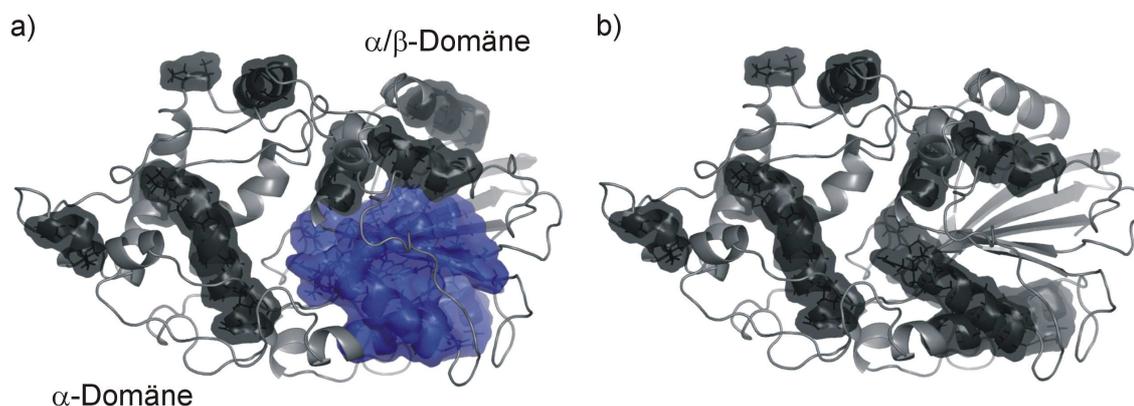


Abbildung 31. Snapshots aus der thermischen Entfaltungssimulation einer repräsentativen Einzelstruktur aus dem konformationellen Ensemble der *Y. mollaretii* WT Phytase unmittelbar vor (a) und nach dem Phasenübergang (b) bei 350 K. Es ist jeweils die Zerlegung des Netzwerks in rigide Cluster (RCD) gezeigt. Die rigiden Cluster sind als einheitlich gefärbte Bereiche dargestellt, wobei der Riesencluster blau eingefärbt ist. Die übrigen Cluster sind dunkel schattiert. Von der Proteinstruktur ist nur das Rückgrat gezeigt und dunkelgrau schattiert.

Unter Verwendung aller Einzelstrukturen aus dem konformationellen Ensemble können wieder Wahrscheinlichkeiten berechnet werden, mit denen bestimmte Strukturbereiche der *Y. mollaretii* WT Phytase Entfaltungsregionen darstellen. In Abbildung 32 ist eine Struktur der *Y. mollaretii* WT Phytase gezeigt, die entsprechend der Entfaltungsregionswahrscheinlichkeit (w_{ER}) jedes Rests unterschiedlich angefärbt ist, wobei rot eine hohe w_{ER} bedeutet und blau eine niedrige. Eine hohe w_{ER} haben insbesondere das β -Faltblatt und die α -Helix *d* in der α/β -Domäne. Auch die α -Helices *c* und *l* in der α/β -Domäne sowie die C-terminalen Reste der α -Helix *h* und die N-terminalen Reste der α -Helix *j* in der α -Domäne am Domänen-Interface stellen Entfaltungsregionen dar, wenn auch mit geringer Wahrscheinlichkeit. Anhand der Verteilung der Bereiche mit hoher w_{ER} können drei „Entfaltungskerne“ (I, II und III) bestimmt werden. Die Entfaltungskerne und die jeweiligen zentralen Reste sind in Tabelle 8 aufgelistet. In Abbildung 30 sind die $C\alpha$ -Atome der zentralen Reste als Kugeln dargestellt.

Der Effekt der Veränderung der zentralen Reste der Entfaltungskerne wurde zunächst mit Hilfe des Computers untersucht. Dazu wurden zusätzliche, nicht native hyd-

rophobe Kontakte zwischen den zentralen Resten der Entfaltungskerne in das Netzwerk von drei repräsentativen Einzelstrukturen eingefügt. Die Atome, zwischen denen die zusätzlichen Kontakte eingefügt wurden, sind in Tabelle 8 aufgelistet. Mit Hilfe der zusätzlichen Kontakte wird der Effekt einer Mutation modelliert, die zu einer Verbesse-

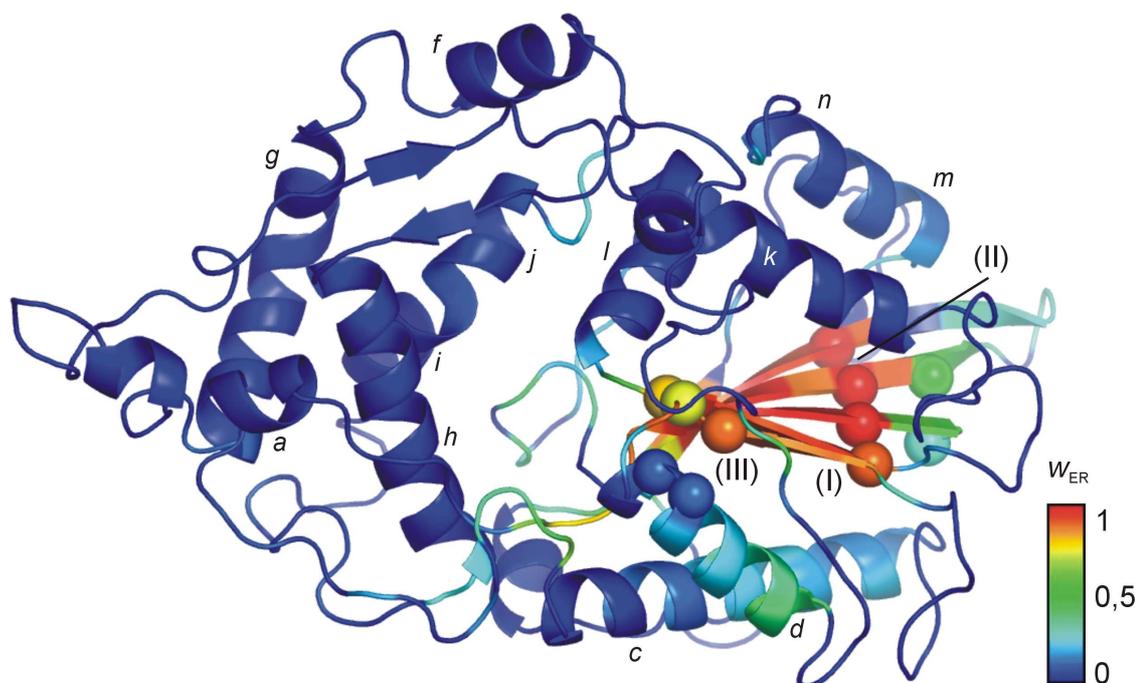


Abbildung 32. Struktur der *Y. mollaretii* WT Phytase. Es ist nur das Rückgrat gezeigt und entsprechend der Entfaltungsregionswahrscheinlichkeit (w_{ER}) jedes Rests unterschiedlich angefärbt. Rot bedeutet eine hohe w_{ER} und blau eine niedrige. Die Kennzeichnung der α -Helices entspricht der Kennzeichnung in dem in Abbildung 22 gezeigten Alignment. Die Ca -Atome der zentralen Reste der Entfaltungskerne (Tabelle 8) sind als Kugeln dargestellt.

rung der Packung hydrophober Reste führt. Eine solche „Mutation“ kann einerseits dem Austausch eines Rests mit einer kleinen unpolaren Seitenkette (Ala) durch einen Rest mit einer großen unpolaren Seitenkette (Val, Leu, Ile) entsprechen, andererseits dem Entfernen einer großen „störenden“ Seitenkette.

Die veränderten Netzwerke wurden mit Hilfe der CNA untersucht. In Abbildung 33 sind die Verläufe der Clusterkonfigurationsentropie für die drei Einzelstrukturen vor und nach der Einführung der Mutationen gezeigt. Über die drei Einzelstrukturen gemittelt erhöht sich die Thermostabilität bei Einführung der „Mutationen“ in den Entfaltungskernen I, II oder III um 3, 4 bzw. 3 K (Abbildung 33 b), c) und d). Werden die Mutationen in den drei Entfaltungskernen gleichzeitig eingeführt, so erhöht sich die Thermostabilität um 8 K (Abbildung 33 e). Bei den Entfaltungskernen handelt es sich

also tatsächlich um Strukturbereiche, in denen das Einführen zusätzlicher Kontakte zu einer Erhöhung der Thermostabilität führt.

Tabelle 8. Entfaltungskerne der *Y. mollaretii* WT Phytase.

| Entfaltungskern | Zentrale Reste | Reste und Atome, zwischen denen zusätzliche hydrophobe Kontakte eingefügt wurden | |
|-----------------|---|--|--|
| I | N-terminale Reste des β -Strangs <i>A</i> , N-terminale Reste des β -Strangs <i>D</i> und C-terminale Reste des β -Strangs <i>Z</i> am Interface zwischen dem β -Faltblatt und der α -Helix <i>c</i> in der α/β -Domäne | Leu 1, Val 4, Leu 292, Glu 331 | Val 4 (C β) und Leu 1 (C γ), Val 4 (C β) und Leu 292 (C γ), Val 4 (C β) und Glu 331 (C β) |
| II | C-terminale Reste des β -Strangs <i>Z</i> , N-terminale Reste des β -Strangs <i>E</i> und Reste im Loop zwischen der α -Helix <i>m</i> und dem β -Strang <i>F</i> | Glu 331, Trp 333, Ala 344, Leu 386 | Ala 344 (C β) und Trp 333 (C δ 1), Ala 344 (C β) und Glu 331 (C β), Ala 344 (C β) und Leu 386 (C γ) |
| III | C-terminale Reste des β -Strangs <i>B</i> , N-terminale Reste des α -Helix <i>d</i> und C-terminale Reste des β -Strangs <i>D</i> | Ala 79, Asp 80, Thr 85, Arg 86, Gly 297 | Arg 86 (C β) und Ala 79 (C β), Asp 80 (C γ) und Thr 85 (C β), Asp 80 (C γ) und Gly 297 (C α) |

Das experimentelle Testen der Entfaltungskerne erfolgte durch das Einführen verschiedenen Mutationen durch Arbeitsgruppen an der Rheinisch-Westfälischen Technischen Hochschule, Aachen, und der BASF SE, Ludwigshafen. Zum gegenwärtigen Zeitpunkt wurde dabei aber noch nicht systematisch vorgegangen. Statt dessen wurde nur der Effekt einzelner Veränderungen in den Entfaltungskernen I und II untersucht. Der Austausch von Val 4 durch Ala führte zu einer Erhöhung der Halbwertszeittemperatur um 0,3–0,7 K. Der Austausch von Ala 344 durch Ile und Arg führte dagegen zu einer deutlichen Verringerung der Halbwertszeittemperatur um 2,8 bzw. $2,3 \pm 0,5$ K. Auch wenn der Effekt der Mutationen auf die Thermostabilität gering ist, so zeigen die experimentellen Daten dennoch, dass Mutationen in den Entfaltungskernen einen signifikanten Einfluss auf die Halbwertszeittemperatur haben können.

Damit ist gezeigt, dass mit der Entwicklung der ensemblebasierten CNA die methodischen Grundlagen zur Vorhersage von Regionen gelegt sind, in denen thermostabilisierende Mutationen eingeführt werden können. Sobald weitere Test durchgeführt worden sind und mehr Daten über Mutationen und deren Effekt auf die Halbwertszeittemperatur vorhanden sind, kann die CNA zudem dazu verwendet werden, den Effekt

nachgewiesenermaßen thermostabilisierender Mutationen auf die Aktivität der *Y. mollaretii* Phytase zu untersuchen.

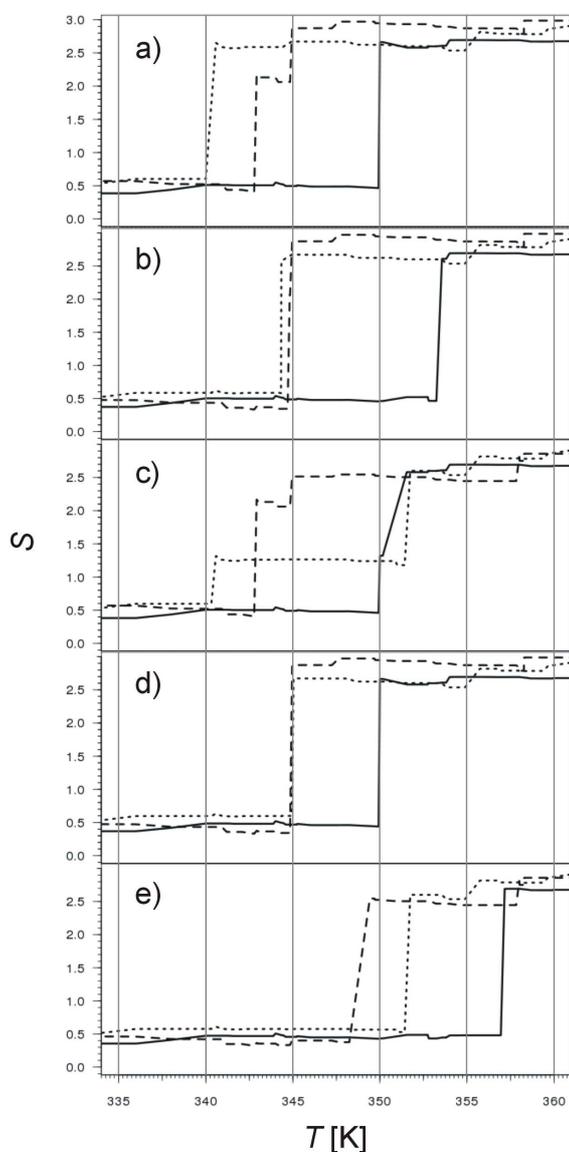


Abbildung 33. Verläufe der Clusterkonfigurationsentropie für drei repräsentative Einzelstrukturen aus dem konformationellen Ensemble der *Y. mollaretii* WT Phytase vor (a) und nach (b-e) der Einführung von Mutationen in die Entfaltungskerne. Die Verläufe der Clusterkonfigurationsentropie nach Einführung von Mutationen in den Entfaltungskern I sind in (b), nach Einführung von Mutationen in den Entfaltungskern II in (c), nach Einführung von Mutationen in den Entfaltungskern III in (d) und nach Einführung von Mutationen in alle drei Entfaltungskerne in (e) gezeigt. Die Nummerierung der Entfaltungskerne folgt der in Tabelle 8 verwendeten Nummerierung.

5.3 Abschließende Diskussion und Ausblick

Die Anwendung der CNA auf einen Datensatz homologer meso- und thermophiler Proteine (Kap. 5.1) sowie eine Serie von Phytasen unterschiedlicher Thermostabilitäten

(Kap. 5.2) hat gezeigt, dass das Verfahren hilfreich zur Unterstützung des Protein Engineering ist, da die thermische Stabilität einer Struktur abgeschätzt werden kann, andererseits Hinweise darauf gegeben werden, in welchen Bereichen der Struktur thermostabilisierende Mutationen eingeführt werden können.

Im Gegensatz zu anderen rechnerischen Methoden zur Unterstützung des Protein Engineering (Dahiyat und Mayo 1997; Malakauskas und Mayo 1998; Kuhlman und Baker 2000; Korkegian *et al.* 2005) basiert die CNA dabei nicht auf der Berechnung relativer Gesamtstabilitäten, sondern auf der Simulation der thermischen Entfaltung der Struktur, anhand derer die Schmelztemperatur vorhergesagt und damit die Thermostabilität untersucht werden kann. Zwar ist Malakauskas und Mayo mit Hilfe energiebasierter Methoden der Entwurf eines hyperthermostabilen Proteins gelungen (Malakauskas und Mayo 1998), grundsätzlich ist aber nicht immer davon auszugehen, dass mit diesen Verfahren zur Vorhersage der Gesamtstabilität auch Aussagen über die Thermostabilität eines Proteins gemacht werden können. Eine Erhöhung der Thermostabilität muss nicht zwangsläufig durch eine Erhöhung der Gesamtstabilität des Proteins erfolgen. Zur Untersuchung der Thermostabilität ist die CNA somit besser geeignet als die energiebasierten Verfahren.

Die bei den energiebasierten Verfahren verwendeten Scoringfunktionen berücksichtigen mehrere empirische Komponenten, die als wichtig für die Thermostabilisierung angesehen werden, wie beispielsweise die Packung, die Stärke des hydrophoben Effekts und das Vermögen eines Rests zur Sekundärstrukturausbildung betreffen. Während der Erfolg energiebasierter Methoden stark von der Auswahl der einzelnen Komponenten abhängt, werden bei der Anwendung der CNA alle möglichen Mechanismen der Thermostabilisierung implizit berücksichtigt. Das Ergebnis der Anwendung der CNA hängt damit nur von der Qualität der verwendeten Proteinstruktur ab und vom Netzwerk, das anhand dieser Struktur konstruiert wird. Ein Nachteil der CNA besteht darin, dass sie sehr empfindlich gegenüber kleinen Defekten in der Struktur ist, die einen großen Einfluss auf die Flexibilitätsanalyse haben können. Um das Verfahren dennoch auch auf Strukturen geringerer Qualität oder Homologiemodelle anwenden zu können, wurde in der vorliegenden Arbeit die ensemblebasierte CNA entwickelt, die die Anwendung der CNA auf konformationelle Ensembles der Struktur erlaubt. Eine weitere Möglichkeit der Verbesserung des Ansatzes besteht darin, die Konstruktion des Netzwerks zu verändern. Insbesondere können andere Funktionen zur Bestimmung der H- und Salzbrücken sowie der hydrophoben Kontakte verwendet werden. Denkbar ist auch die Berücksich-

tigung der unterschiedlichen Stärke von H- und Salzbrücken oder die Veränderung der Stärke dieser Wechselwirkungen mit zunehmender Temperatur.

Mit Hilfe von Simulationen der Denaturierung von Proteinen können strukturelle Veränderungen untersucht werden, die während der Entfaltung auftreten und letztendlich zum globalen Zerfall der Struktur führen. Diese Ansätze liefern keine Informationen über die Gesamtstabilität von Proteinen (ΔG) oder den Stabilitätsunterschied ($\Delta\Delta G$) zwischen einem Wildtyp und seiner Mutante, können aber direkte Hinweise auf die thermische Stabilität von Proteinen (T_m oder ΔT_m) geben. Zudem können neben dem globalen Zerfall auch lokale Entfaltungsereignisse untersucht werden, welche die Flexibilität bestimmter Bereiche der Struktur ermöglichen und mit Hilfe der mikroskopischen Stabilität der jeweiligen Region beschrieben werden können. Die durch die Mikro-stabilität bedingte Flexibilität ist die Voraussetzung für die Bioaktivität von Enzymen (Henzler-Wildman und Kern 2007). So kann mit Hilfe der CNA der Effekt einer thermostabilisierenden Mutation auf die Aktivität eines Enzyms vorhergesagt werden. Darin liegt ein weiterer Vorteil der CNA gegenüber energiebasierter Methoden. Die Vorhersage des Effekts einer thermostabilisierenden Mutation auf die Aktivität erfolgt bislang allerdings nur qualitativ. Weitere Arbeiten sind notwendig um zu bestimmen, wie der Effekt solcher Mutationen auch quantitativ vorhergesagt werden kann. Dazu muss der Effekt von Mutationen im Bereich der aktiven Tasche anhand eines großen Datensatzes von Enzymen systematisch untersucht werden. Ein wichtiges Werkzeug werden dabei die in dieser Arbeit entwickelten Stabilitätskarten sein, mit denen die mikroskopische Stabilität einzelner Regionen der Enzymstrukturen untersucht werden kann.

Klassische Simulationen der thermischen Denaturierung sind MD-Simulationen bei hohen Temperaturen. Hier wird vorausgesetzt, dass die Entfaltung weitestgehend temperaturunabhängig ist und die gleichen Eigenschaften aufweist wie die Faltung, was für mehrere Beispiele bestätigt werden konnte (Cavalli *et al.* 2002; Day *et al.* 2002), aber nicht allgemein gilt (Finkelstein 1997). Mit Hilfe solcher Methoden wurde beispielsweise die Entfaltung einer Cutinase untersucht (Creveld *et al.* 1998). Über den Vergleich von Simulationen unter verschiedenen Bedingungen konnte zwischen flexiblen Regionen, die für die Aktivität des Enzyms wichtig waren, und flexiblen Regionen, von denen die Entfaltung ausging, unterschieden werden. So konnten Schwachstellen (*weak spots*) vorhergesagt werden, deren Mutation die thermische Stabilität, nicht aber die Aktivität beeinflussen dürfte. Der Nachteil von MD-Simulationen besteht darin, dass diese Rechnungen sehr viel Zeit kosten. Zudem ist unklar, ob die verwendeten Kraftfelder und

insbesondere Lösungsmittelmodelle geeignet sind, das Verhalten des Systems bei hohen Temperaturen korrekt wiederzugeben.

Netzwerkbasierende Methoden (wie die CNA) stellen eine Alternative zu Dynamiksimulationen dar. Sie sind bereits vielfach verwendet worden, um die Flexibilität und Stabilität von Proteinen zu untersuchen (Dokholyan *et al.* 2002; Vendruscolo *et al.* 2002; Böde *et al.* 2007). Heringa und Argos verwendeten Strukturnetzwerke von Proteinen, um Cluster dicht gepackter Seitenketten zu identifizieren (Heringa und Argos 1991; Heringa *et al.* 1995). Brinda *et al.* verwendeten einen ähnlichen Ansatz (Kannan und Vishveshwara 1999), um Unterschiede zwischen meso- und thermophilen Proteinen zu finden (Brinda und Vishveshwara 2005). Auch Robinson-Rechavi *et al.* verwendeten ein solches Verfahren (Greene und Higman 2003) zur Untersuchung struktureller Eigenschaften thermophiler Proteine (Robinson-Rechavi *et al.* 2006). Diesen Arbeiten ist gemeinsam, dass bestimmte Atome oder Reste identifiziert werden können, die für die Topologie des Netzwerks bestimmend sind. Die CNA unterscheidet sich von diesen Methoden, indem nicht topologische Eigenschaften des Netzwerks untersucht werden, sondern direkt die Flexibilität und Rigidität im Netzwerk vorhergesagt und mit der makro- und mikroskopischen Stabilität der Proteinstruktur verknüpft wird.

Bereits in der Vergangenheit sind Informationen über die Flexibilität und Rigidität einer Proteinstruktur erfolgreich zur Unterstützung des Protein Engineering eingesetzt worden. So wurden beispielsweise Informationen über die Beweglichkeit einzelner Reste im nativen Zustand (aus B-Faktoren) für die Vorhersage thermostabilisierender Mutationen verwendet (Reetz *et al.* 2006). Ebenso wie die mit Hilfe von MD-Simulationen oder NMA-ähnlichen Verfahren gemessene Dynamik (Bewegung) ist die Information aus B-Faktoren aber nur ein indirektes Maß für die tatsächlich vorhandene Flexibilität (Beweglichkeit) einer Struktur. Nur über eine direkte Bestimmung rigider und flexibler Regionen kann eine Aussage über die makroskopische und mikroskopische Stabilität der jeweiligen Region oder gesamten Struktur getroffen werden, da diese Eigenschaften beschreiben, ob ein bestimmter Strukturbereich deformiert werden kann und dann zur Destabilisierung des Proteins beiträgt. Tatsächlich handelt es sich bei den von Reetz *et al.* vorhergesagten Mutationen um Reste an der Proteinoberfläche (Reetz *et al.* 2006), wohingegen mit Hilfe der CNA auch Entfaltungsregionen innerhalb der Struktur und in Sekundärstrukturelementen bestimmt werden können.

6 Zusammenfassung

In dieser Arbeit wurde ein Verfahren entwickelt, mit dem die Entfaltung von Proteinen simuliert werden kann. Anhand der Simulation kann die Schmelztemperatur bestimmt und damit die Thermostabilität einer Struktur untersucht werden. Außerdem ist die Untersuchung struktureller Veränderungen möglich, die während der Entfaltung auftreten und letztendlich zum globalen Zerfall der Struktur führen. Die Bereiche, von denen der globale Zerfall der Struktur ausgeht, können bestimmt werden. Es wurde untersucht, inwieweit diese Entfaltungsregionen Strukturbereiche darstellen, deren Mutation die thermische Stabilität der Struktur beeinflusst. Das entwickelte Verfahren basiert auf der Anwendung von Methoden aus der Rigiditätstheorie auf *bond-bending*- bzw. äquivalente *body-and-bar*-Netzwerke (*constraint networks*) von Proteinen, was die direkte Vorhersage rigider Cluster und flexibler Regionen in atomarer Auflösung erlaubt (Flexibilitätsanalyse). Die thermische Entfaltung der Strukturen wird über das sukzessive Aufbrechen nichtkovalenter Wechselwirkungen in den Netzwerken simuliert. Ansätze aus der Perkulations- und Netzwerktheorie werden verwendet, um die Rigidität und Flexibilität in den Netzwerken während der Entfaltung zu untersuchen.

Diese unter dem Begriff der Analyse statischer Netzwerke (*constraint network analysis*, CNA) zusammengefassten Methoden wurden im ersten Teil dieser Arbeit auf einen Datensatz homologer meso- und thermophiler Proteine angewendet. Dabei wurde untersucht, ob die thermophile Anpassung tatsächlich über eine Rigidisierung der Struktur erfolgt. Außerdem wurde die Theorie der korrespondierenden Zustände getestet, die besagt, dass bei der thermophilen Anpassung trotz globaler Rigidisierung für die Bioaktivität wichtige flexible Bereiche konserviert sind. Mit Hilfe der CNA wurden Entfaltungsregionen der Proteine aus dem Datensatz bestimmt und mit Strukturbereichen verglichen, in die thermostabilisierende Mutationen eingeführt wurden. Außerdem wurde untersucht, inwieweit vorhergesagt werden kann, ob eine möglicherweise thermostabilisierende Mutation die Aktivität negativ beeinflusst.

Für zwei Drittel der thermophilen Proteine aus dem Datensatz konnte eine höhere Thermostabilität vorhergesagt werden als für das entsprechende mesophile Protein. Es konnte zudem gezeigt werden, dass die thermophile Anpassung dieser Proteine tatsächlich über eine Rigidisierung der Struktur erfolgt. Offensichtlich werden bei der Anwendung der CNA implizit alle möglichen Mechanismen der thermophilen Anpassung berücksichtigt. Die für zwei Paare meso- und thermophiler Proteine vorhergesagten Ent-

faltungsregionen stimmten sehr gut mit Bereichen überein, in die thermostabilisierende Mutationen eingeführt wurden. Damit wurde gezeigt, dass die CNA hilfreich zur Unterstützung des Protein Engineering ist, da die thermische Stabilität einer Struktur abgeschätzt werden kann, andererseits Hinweise darauf gegeben werden, in welchen Bereichen der Struktur thermostabilisierende Mutationen eingeführt werden können.

Die Untersuchung und der Vergleich mikroskopischer Stabilitäten der homologen meso- und thermophilen Proteine zeigte, dass bei der thermophilen Anpassung trotz globaler Rigidisierung für die Bioaktivität wichtige flexible Bereiche konserviert sind. Homologe meso- und thermophile Proteine befinden sich bei ihrer jeweiligen Arbeitstemperatur also tatsächlich in korrespondierenden Zuständen. Zum ersten mal ist damit die Theorie der korrespondierenden Zustände für einen so großen Datensatz mit Hilfe rechnerischer Methoden getestet worden. Die für zwei Paare meso- und thermophiler Proteine untersuchten mikroskopischen Stabilitäten ermöglichten es, das Zusammenspiel zwischen Flexibilität, Stabilität und Aktivität zu verstehen. Veränderungen in der Flexibilität dieser Enzyme, wie sie entweder durch Veränderungen der Temperatur oder durch das Einführen von Mutationen verursacht wurden, konnten mit experimentell bestimmten Veränderungen der Aktivität in Verbindung gebracht werden. Damit konnte gezeigt werden, dass die CNA ein Abschätzen des Effekts einer thermostabilisierenden Mutation auf die Aktivität erlaubt, was wiederum für den Einsatz der CNA zur Unterstützung des Protein Engineering spricht.

Im zweiten Teil dieser Arbeit wurde die CNA auf eine Serie von Phytasen unterschiedlicher Thermostabilitäten angewendet. Da es sich bei den Phytase-Strukturen um Homologiemodelle handelte, die nicht direkt mit der Analyse statischer Netzwerke untersucht werden konnten, wurde eine ensemblebasierte CNA etabliert. Dazu wurden kurze MD-Simulationen zur Verbesserung der homologiemodellierten Strukturen durchgeführt, aus denen dann ein konformationelles Strukturensemble extrahiert wurde. Aus den konformationellen Ensembles werden Thermostabilitäten vorhergesagt. Zudem wurden Wahrscheinlichkeiten berechnet, mit denen bestimmte Strukturbereiche Entfaltungsregionen darstellen, in denen Mutationen die thermische Stabilität der Struktur beeinflussen. Bei der Vorhersage der Thermostabilitäten ergab sich eine bemerkenswert gute Übereinstimmung mit experimentell bestimmten relativen Halbwertszeittemperaturen. Für zwei Phytasen aus dem Datensatz (Consensus-Phytase 10 und *E. coli* AppA Phytase) konnte gezeigt werden, dass Bereiche mit hoher Entfaltungsregionswahrscheinlichkeit gut mit Regionen übereinstimmten, in denen thermostabilisierende Mutationen

experimentell eingeführt wurden. Für eine weitere Phytase (*Y. mollaretii* WT Phytase) wurden vorhergesagte Entfaltungskerne (Bereiche mit hoher Entfaltungsregionwahrscheinlichkeit) durch Arbeitsgruppen an der Rheinisch-Westfälische Technische Hochschule, Aachen, und der BASF SE, Ludwigshafen, experimentell getestet. Dabei zeigte sich, dass Mutationen in den Entfaltungskernen tatsächlich einen signifikanten Einfluss auf die Halbwertszeittemperatur haben können. Diese Ergebnisse offenbaren, dass mit der Entwicklung der ensemblebasierten CNA die methodischen Grundlagen für den Einsatz des Verfahrens zur Unterstützung des Protein Engineering von Phytasen geschaffen wurden.

7 Anhang

7.1 Die Gibbs-Helmholtz-Gleichung

Unter der vereinfachenden Annahme einer temperaturunabhängigen Wärmekapazitätsdifferenz (ΔC_p) (Prabhu und Sharp 2005) kann die Stabilität eines Proteins mit der Gibbs-Helmholtz-Gleichung beschrieben werden:

$$\Delta G(T) = \Delta H(T_r) - T \cdot \Delta S(T_r) + \Delta C_p \cdot [(T - T_s)] - T \cdot \ln(T / T_r) \quad (16)$$

Hier bedeutet T_r eine beliebige Referenztemperatur. Wenn T_r gleich der Schmelztemperatur T_m gesetzt wird, bei der $\Delta G = 0$ und $\Delta S = \Delta H / T_m$ ist, dann folgt daraus die modifizierte Gibbs-Helmholtz-Gleichung.

7.2 Weitere Parameter zur Analyse der Entfaltungssimulationen

Mit Hilfe des Programmpakets CNAnalysis können neben den in Kap. 4.4 beschriebenen Parametern zur Analyse der Entfaltungssimulation auch die mittlere Koordinationszahl und die Anzahl der *floppy modes* berechnet werden. Die mittlere Koordinationszahl wird folgendermaßen berechnet (Rader *et al.* 2002):

$$\langle r \rangle = \frac{\sum_{r=2}^{r_{\max}} r N_r}{\sum_{r=2}^{r_{\max}} N_r} \quad (17)$$

Hier bedeutet N_r die Anzahl r -koordinierter Atome. Zur Berechnung wurde die Information aus den von der Software FIRST bei der Konstruktion des Netzwerks ausgegebenen Dateien cov.out, hbonds.out und hphobes.out ausgelesen und verarbeitet. Vor der Bestimmung der Koordinationszahl werden alle losen Enden (*dangling ends*) rekursiv entfernt (*pruning*). Lose Enden sind einfach koordinierte Atome, die nicht zur Rigidität im Netzwerk beitragen. Sie werden entfernt, um die Ergebnisse der Entfaltungssimulationen von Proteinen mit der Rigiditätsperkolation von Netzwerken glasbildender Materialien vergleichen zu können (Rader *et al.* 2002). Die Anzahl der *floppy modes* ist ein Maß für die im Netzwerk enthaltene Flexibilität. Zur Bestimmung der Anzahl der *floppy modes* wurde die Information aus der von der Software FIRST ausgegebenen Datei results.txt ausgelesen und verarbeitet. F wurde anschließend graphisch als Funktion der Temperatur aufgetragen.

8 Literatur

- Adekoya, O.A. und Sylte, I.: *The thermolysin family (M4) of enzymes: Therapeutic and biotechnological potential*. Chem. Biol. Drug Des. **2009**, 73, 7-16.
- Ahern, T.J. und Klibanov, A.M.: *Analysis of processes causing thermal inactivation of enzymes*. Method. Biochem. Anal. **1988**, 33, 91-127.
- Akanuma, S., Qu, C.X., Yamagishi, A., Tanaka, N. und Oshima, T.: *Effect of polar side chains at position 172 on thermal stability of 3-isopropylmalate dehydrogenase from Thermus thermophilus*. FEBS Lett. **1997**, 410, 141-144.
- Akanuma, S., Yamagishi, A., Tanaka, N. und Oshima, T.: *Spontaneous tandem sequence duplications reverse the thermal stability of carboxyl-terminal modified 3-isopropylmalate dehydrogenase*. J. Bacteriol. **1996**, 178, 6300-6304.
- Akanuma, S., Yamagishi, A., Tanaka, N. und Oshima, T.: *Serial increase in the thermal stability of 3-isopropylmalate dehydrogenase from Bacillus subtilis by experimental evolution*. Protein Sci. **1998**, 7, 698-705.
- Akanuma, S., Yamagishi, A., Tanaka, N. und Oshima, T.: *Further improvement of the thermal stability of a partially stabilized Bacillus subtilis 3-isopropylmalate dehydrogenase variant by random and site-directed mutagenesis*. Eur. J. Biochem. **1999**, 260, 499-504.
- Alsop, E., Silver, M. und Livesay, D.R.: *Optimized electrostatic surfaces parallel increased thermostability: A structural bioinformatic analysis*. Protein Eng. **2003**, 16, 871-874.
- Alvizo, O., Allen, B.D. und Mayo, S.L.: *Computational protein design promises to revolutionize protein engineering*. BioTechniques **2007**, 42, 31-39.
- Amadei, A., Linssen, A.B. und Berendsen, H.J.: *Essential dynamics of proteins*. Proteins **1993**, 17, 412-425.
- Andraud, C., Beghdadi, A. und Lafait, J.: *Entropic analysis of random morphologies*. Physica A **1994**, 207, 208-212.
- Anfinsen, C.B.: *Principles that govern the folding of protein chains*. Science **1973**, 181, 223-230.
- Antranikian, G.: *Industrial relevance of thermophiles and their enzymes. Thermophiles: Biology and technology at high temperatures*. Robb, F., Antranikian, G., Grogan, D. und Driessen, A. London, New York, CRC Press, **2008**, 113-160.

- Aoshima, M. und Oshima, T.: *Stabilization of Escherichia coli isopropylmalate dehydrogenase by single amino acid substitution*. Protein Eng. **1997**, 10, 249-254.
- Argos, P., Rossmann, M.G., Grau, U.M., Zuber, H., Frank, G. und Tratschin, J.D.: *Thermal stability and protein structure*. Biochemistry **1979**, 18, 5698-5703.
- Arnold, U., Köditz, J., Markert, Y. und Ulbrich-Hofmann, R.: *Local fluctuations vs. global unfolding of proteins investigated by limited proteolysis*. Biocatal. Bio-transform. **2005**, 23, 159-167.
- Arunachalam, J. und Gautham, N.: *Hydrophobic clusters in protein structures*. Proteins **2008**, 71, 2012-2025.
- Atomi, H.: *Recent progress towards the application of hyperthermophiles and their enzymes*. Curr. Opin. Chem. Biol. **2005**, 9, 166-173.
- Bahar, I., Atilgan, A.R. und Erman, B.: *Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential*. Fold. Des. **1997**, 2, 173-181.
- Baker, E.N. und Hubbard, R.E.: *Hydrogen-bonding in globular proteins*. Prog. Biophys. Mol. Biol. **1984**, 44, 97-179.
- Baldwin, R.L. und Zimm, B.H.: *Are denatured proteins ever random coils?* Proc. Natl. Acad. Sci. U. S. A. **2000**, 97, 12391-12392.
- Basu, S. und Sen, S.: *Turning a mesophilic protein into a thermophilic one: A computational approach based on 3D structural features*. J. Chem. Inf. Mod. **2009**, 49, 1741-1750.
- Bava, K.A., Gromiha, M.M., Uedaira, H., Kitajima, K. und Sarai, A.: *ProTherm, version 4.0: Thermodynamic database for proteins and mutants*. Nucleic Acids Res. **2004**, 32, D120-D121.
- Becktel, W.J. und Schellman, J.A.: *Protein stability curves*. Biopolymers **1987**, 26, 1859-1877.
- Beitz, E.: *TEXshade: shading and labeling of multiple sequence alignments using L^A-TEX2 epsilon*. Bioinformatics **2000**, 16, 135-139.
- Berendsen, H.J.C., Postma, J.P.M., Vangunsteren, W.F., Dinola, A. und Haak, J.R.: *Molecular-dynamics with coupling to an external bath*. J. Chem. Phys. **1984**, 81, 3684-3690.
- Böde, C., Kovács, I.A., Szalay, M.S., Palotai, R., Korcsmáros, T. und Csermely, P.: *Network analysis of protein dynamics*. FEBS Lett. **2007**, 581, 2776-2782.
- Bourne, P.E., Address, K.J., Bluhm, W.F., Chen, L., Deshpande, N., Feng, Z.K., Fleri, W., Green, R., Merino-Ott, J.C., Townsend-Merino, W., Weissig, H., Westbrook,

- J. und Berman, H.M.: *The distribution and query systems of the RCSB protein data bank*. Nucleic Acids Res. **2004**, 32, D223-D225.
- Braiuca, P., Buthe, A., Ebert, C., Linda, P. und Gardossi, L.: *Volsurf computational method applied to the prediction of stability of thermostable enzymes*. Biotechnol. J. **2007**, 2, 214-220.
- Braun, M., Teichert, O. und Zweck, A.: *Biokatalyse in der industriellen Produktion - Fakten und Potentiale der weißen Biotechnologie*. Düsseldorf, Zukünftige Technologien Consulting (ZTC) der VDI Technologiezentrum GmbH, **2006**.
- Brinda, K.V. und Vishveshwara, S.: *A network representation of protein structures: Implications for protein stability*. Biophys. J. **2005**, 89, 4159-4170.
- Bruins, M.E., Janssen, A.E.M. und Boom, R.M.: *Thermostables and their applications - A review of recent literature and patents*. Appl. Biochem. Biotechnol. **2001**, 90, 155-186.
- Cambillau, C. und Claverie, J.M.: *Structural and genomic correlates of hyperthermostability*. J. Biol. Chem. **2000**, 275, 32383-32386.
- Case, D.A.: *Normal mode analysis of protein dynamics*. Curr. Opin. Struct. Biol. **1994**, 4, 285-290.
- Case, D.A., Cheatham, T.E., Darden, T., Gohlke, H., Luo, R., Merz, K.M., Onufriev, A., Simmerling, C., Wang, B. und Woods, R.J.: *The Amber biomolecular simulation programs*. J. Comput. Chem. **2005**, 26, 1668-1688.
- Cavalli, A., Ferrara, P. und Caflisch, A.: *Weak temperature dependence of the free energy surface and folding pathways of structured peptides*. Proteins **2002**, 47, 305-314.
- Ceruso, M.A., Amadei, A. und Di Nola, A.: *Mechanics and dynamics of B1 domain of protein G: Role of packing and surface hydrophobic residues*. Protein Sci. **1999**, 8, 147-160.
- Chakravarty, S. und Varadarajan, R.: *Elucidation of determinants of protein stability through genome sequence analysis*. FEBS Lett. **2000**, 470, 65-69.
- Chaparro-Riggers, J.F., Polizzi, K.M. und Bommarius, A.S.: *Better library design: Data-driven protein engineering*. Biotechnol. J. **2007**, 2, 180-191.
- Chen, J.M. und Stites, W.E.: *Packing is a key selection factor in the evolution of protein hydrophobic cores*. Biochemistry **2001**, 40, 15280-15289.

- Chubynsky, M.V., Hesperheide, B.M., Jacobs, D.J., Kuhn, L.A., Lei, M., Rader, A.J., Thorpe, M.F., Whiteley, W. und Zavodszky, M.I.: *Constraint theory applied to proteins*. **2004**.
- Colombo, G., De Mori, G.M.S. und Roccatano, D.: *Interplay between hydrophobic cluster and loop propensity in beta-hairpin formation: A mechanistic study*. *Protein Sci.* **2003**, 12, 538-550.
- Cooper, A.: *Heat capacity of hydrogen-bonded networks: an alternative view of protein folding thermodynamics*. *Biophys. Chem.* **2000**, 85, 25-39.
- Corbett, R.J. und Roche, R.S.: *The unfolding mechanism of thermolysin*. *Biopolymers* **1983**, 22, 101-105.
- Cornicchi, E., Marconi, M., Onori, G. und Paciaroni, A.: *Controlling the protein dynamical transition with sugar-based bioprotectant matrices: A neutron scattering study*. *Biophys. J.* **2006**, 91, 289-297.
- Creveld, L.D., Amadei, A., van Schaik, R.C., Pepermans, H.A.M., de Vlieg, J. und Berendsen, H.J.C.: *Identification of functional and unfolding motions of cutinase as obtained from molecular dynamics computer simulations*. *Proteins* **1998**, 33, 253-264.
- Cui, Q. und Karplus, M.: *Allostery and cooperativity revisited*. *Protein Sci.* **2008**, 17, 1295-1307.
- Dahiyat, B.I., Gordon, D.B. und Mayo, S.L.: *Automated design of the surface positions of protein helices*. *Protein Sci.* **1997**, 6, 1333-1337.
- Dahiyat, B.I. und Mayo, S.L.: *De novo protein design: Fully automated sequence selection*. *Science* **1997**, 278, 82-87.
- Darden, T., York, D. und Pedersen, L.: *Particle Mesh Ewald - A $N \log(N)$ method for Ewald Sums in large systems*. *J. Chem. Phys.* **1993**, 98, 10089-10092.
- Das, R. und Gerstein, M.: *The stability of thermophilic proteins: a study based on comprehensive genome comparison*. *Funct. Integr. Genomics* **2000**, 1, 76-88.
- Day, R., Bennion, B.J., Ham, S. und Daggett, V.: *Increasing temperature accelerates protein unfolding without changing the pathway of unfolding*. *J. Mol. Biol.* **2002**, 322, 189-203.
- de Francesco, A., Marconi, M., Cinelli, S., Onori, G. und Paciaroni, A.: *Picosecond internal dynamics of lysozyme as affected by thermal unfolding in nonaqueous environment*. *Biophys. J.* **2004**, 86, 480-487.

- de Groot, B.L., van Aalten, D.M.F., Scheek, R.M., Amadei, A., Vriend, G. und Berendsen, H.J.C.: *Prediction of protein conformational freedom from distance constraints*. Proteins **1997**, 29, 240-251.
- de Kreijl, A., Venema, G. und van den Burg, B.: *Substrate specificity in the highly heterogeneous M4 peptidase family is determined by a small subset of amino acids*. J. Biol. Chem. **2000**, 275, 31115-31120.
- Dehouck, Y., Folch, B. und Rooman, M.: *Revisiting the correlation between proteins' thermoresistance and organisms' thermophilicity*. Protein Eng. Des. Sel. **2008**, 21, 275-278.
- DeLano, W.L. (2002). The PyMOL Molecular Graphics System. Palo Alto, CA, USA, DeLano Scientific.
- Demirjian, D.C., Moris-Varas, F. und Cassidy, C.S.: *Enzymes from extremophiles*. Curr. Opin. Chem. Biol. **2001**, 5, 144-151.
- Desiraju, G.R.: *Hydrogen bridges in crystal engineering: interactions without borders*. Acc. Chem. Res. **2002**, 35, 565-573.
- Dill, K.A.: *Dominant forces in protein folding*. Biochemistry **1990**, 29, 7133-7155.
- Dokholyan, N.V., Li, L., Ding, F. und Shakhnovich, E.I.: *Topological determinants of protein folding*. Proc. Natl. Acad. Sci. U. S. A. **2002**, 99, 8637-8641.
- Dougherty, R.C.: *Temperature and pressure dependence of hydrogen bond strength: A perturbation molecular orbital approach*. J. Chem. Phys. **1998**, 109, 7372-7378.
- Eijsink, V.G.H., Bjork, A., Gaseidnes, S., Sirevag, R., Synstad, B., van den Burg, B. und Vriend, G.: *Rational engineering of enzyme stability*. J. Biotechnol. **2004**, 113, 105-120.
- Eijsink, V.G.H., Dijkstra, B.W., Vriend, G., Vanderzee, J.R., Veltman, O.R., Vandervinne, B., Vandenburg, B., Kempe, S. und Venema, G.: *The effect of cavity-filling mutations on the thermostability of Bacillus stearothermophilus neutral protease*. Protein Eng. **1992**, 5, 421-426.
- Eijsink, V.G.H., Gaseidnes, S., Borchert, T.V. und van den Burg, B.: *Directed evolution of enzyme stability*. Biomol. Eng. **2005**, 22, 21-30.
- Eijsink, V.G.H., Veltman, O.R., Aukema, W., Vriend, G. und Venema, G.: *Structural determinants of the stability of thermolysin-like proteinases*. Nat. Struct. Biol. **1995**, 2, 374-379.

- Eijsink, V.G.H., Vriend, G., Vandenburg, B., Vanderzee, J.R., Veltman, O.R., Stulp, B.K. und Venema, G.: *Introduction of a stabilizing 10-residue beta-hairpin in Bacillus subtilis neutral protease*. Protein Eng. **1992**, 5, 157-163.
- Eijsink, V.G.H., Vriend, G., Vandenburg, B., Vanderzee, J.R. und Venema, G.: *Increasing the thermostability of a neutral protease by replacing positively charged amino acids in the N-terminal turn of alpha-helices*. Protein Eng. **1992**, 5, 165-170.
- Eijsink, V.G.H., Vriend, G., Vandervinne, B., Hazes, B., Vandenburg, B. und Venema, G.: *Effects of changing the interaction between subdomains on the thermostability of Bacillus neutral proteases*. Proteins **1992**, 14, 224-236.
- Elcock, A.H.: *The stability of salt bridges at high temperatures: Implications for hyperthermophilic proteins*. J. Mol. Biol. **1998**, 284, 489-502.
- Eriksson, A.E., Baase, W.A., Zhang, X.J., Heinz, D.W., Blaber, M., Baldwin, E.P. und Matthews, B.W.: *Response of a protein structure to cavity creating mutations and its relation to the hydrophobic effect*. Science **1992**, 255, 178-183.
- Espinosa, J.F., Munoz, V. und Gellman, S.H.: *Interplay between hydrophobic cluster and loop propensity in beta-hairpin formation*. J. Mol. Biol. **2001**, 306, 397-402.
- Facchiano, A.M., Colonna, G. und Ragone, R.: *Helix stabilizing factors and stabilization of thermophilic proteins: An X-ray based study*. Protein Eng. **1998**, 11, 753-760.
- Fan, H. und Mark, A.E.: *Refinement of homology-based protein structures by molecular dynamics simulation techniques*. Protein Sci. **2004**, 13, 211-220.
- Ferrer, M., Golyshina, O., Beloqui, A. und Golyshin, P.N.: *Mining enzymes from extreme environments*. Curr. Opin. Microbiol. **2007**, 10, 207-214.
- Fersht, A.R.: *The hydrogen bond in molecular recognition*. Trends Biochem. Sci. **1987**, 12, 301-304.
- Fields, P.A.: *Review: Protein function at thermal extremes: balancing stability and flexibility*. Comp. Biochem. Physiol., A: Comp. Physiol. **2001**, 129, 417-431.
- Finkelstein, A.V.: *Can protein unfolding simulate protein folding?* Protein Eng. **1997**, 10, 843-845.
- Fiser, A. und Sali, A.: *ModLoop: Automated modeling of loops in protein structures*. Bioinformatics **2003**, 19, 2500-2501.
- Fitter, J. und Heberle, J.: *Structural equilibrium fluctuations in mesophilic and thermophilic alpha-amylase*. Biophys. J. **2000**, 79, 1629-1636.

- Fitter, J., Herrmann, R., Dencher, N.A., Blume, A. und Hauss, T.: *Activity and stability of a thermostable alpha-amylase compared to its mesophilic homologue: Mechanisms of thermal adaptation*. *Biochemistry* **2001**, 40, 10723-10731.
- Folch, B., Rooman, M. und Dehouck, Y.: *Thermostability of salt bridges versus hydrophobic interactions in proteins probed by statistical potentials*. *J. Chem. Inf. Mod.* **2008**, 48, 119-127.
- Fulle, S. und Gohlke, H.: *Analyzing the flexibility of RNA structures by constraint counting*. *Biophys. J.* **2008**, 94, 4202-4219.
- Funahashi, J., Sugita, Y., Kitao, A. und Yutani, K.: *How can free energy component analysis explain the difference in protein stability caused by amino acid substitutions? Effect of three hydrophobic mutations at the 56th residue on the stability of human lysozyme*. *Protein Eng.* **2003**, 16, 665-671.
- Gaseidnes, S., Synstad, B., Jia, X.H., Kjellesvik, H., Vriend, G. und Eijsink, V.G.H.: *Stabilization of a chitinase from *Serratia marcescens* by Gly -> Ala and Xxx -> Pro mutations*. *Protein Eng.* **2003**, 16, 841-846.
- Georlette, D., Blaise, V., Collins, T., D'Amico, S., Gratia, E., Hoyoux, A., Marx, J.C., Sonan, G., Feller, G. und Gerday, C.: *Some like it cold: Biocatalysis at low temperatures*. *FEMS Microbiol. Rev.* **2004**, 28, 25-42.
- Gianese, G., Argos, P. und Pascarella, S.: *Structural adaptation of enzymes to low temperatures*. *Protein Eng.* **2001**, 14, 141-148.
- Gianese, G., Bossa, F. und Pascarella, S.: *Comparative structural analysis of psychrophilic and meso- and thermophilic enzymes*. *Proteins* **2002**, 47, 236-249.
- Gohlke, H., Kuhn, L.A. und Case, D.A.: *Change in protein flexibility upon complex formation: Analysis of Ras-Raf using molecular dynamics and a molecular framework approach*. *Proteins* **2004**, 56, 322-337.
- Greene, L.H. und Higman, V.A.: *Uncovering network systems within protein structures*. *J. Mol. Biol.* **2003**, 334, 781-791.
- Gribenko, A.V., Patel, M.M., Liu, J., McCallum, S.A., Wang, C. und Makhatadze, G.I.: *Rational stabilization of enzymes by computational redesign of surface charge-charge interactions*. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, 106, 2601-2606.
- Gromiha, M.M.: *Important inter-residue contacts for enhancing the thermal stability of thermophilic proteins*. *Biophys. Chem.* **2001**, 91, 71-77.

- Gromiha, M.M., Oobatake, M. und Sarai, A.: *Important amino acid properties for enhanced thermostability from mesophilic to thermophilic proteins*. Biophys. Chem. **1999**, 82, 51-67.
- Haefner, S., Knietsch, A., Scholten, E., Braun, J., Lohscheidt, M. und Zelder, O.: *Biotechnological production and applications of phytases*. Appl. Microbiol. Biotechnol. **2005**, 68, 588-597.
- Halle, B.: *Flexibility and packing in proteins*. Proc Natl Acad Sci USA **2002**, 99, 1274-1279.
- Han, Y. und Lei, X.G.: *Role of glycosylation in the functional expression of an Aspergillus niger phytase (phyA) in Pichia pastoris*. Arch. Biochem. Biophys. **1999**, 364, 83-90.
- Haney, P.J., Badger, J.H., Buldak, G.L., Reich, C.I., Woese, C.R. und Olsen, G.J.: *Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic Methanococcus species*. Proc. Natl. Acad. Sci. U. S. A. **1999**, 96, 3578-3583.
- Hangauer, D.G., Monzingo, A.F. und Matthews, B.W.: *An interactive computer-graphics study of thermolysin-catalyzed peptide cleavage and inhibition by N-carboxymethyl dipeptides*. Biochemistry **1984**, 23, 5730-5741.
- Hardy, F., Vriend, G., Veltman, O.R., Vandervinne, B., Venema, G. und Eijssink, V.G.H.: *Stabilization of Bacillus stearothermophilus neutral protease by introduction of prolines*. FEBS Lett. **1993**, 317, 89-92.
- Hendrickson, B.: *Conditions for unique graph realizations*. SIAM J. Comput. **1992**, 21, 65-84.
- Henzler-Wildman, K. und Kern, D.: *Dynamic personalities of proteins*. Nature **2007**, 450, 964-972.
- Heringa, J. und Argos, P.: *Side-chain clusters in protein structures and their role in protein folding*. J. Mol. Biol. **1991**, 220, 151-171.
- Heringa, J., Argos, P., Egmond, M.R. und Devlieg, J.: *Increasing thermal stability of subtilisin from mutations suggested by strongly interacting side-chain clusters*. Protein Eng. **1995**, 8, 21-30.
- Hernández, G., Jenney, F.E., Adams, M.W.W. und LeMaster, D.M.: *Millisecond time scale conformational flexibility in a hyperthermophile protein at ambient temperature*. Proc. Natl. Acad. Sci. U. S. A. **2000**, 97, 3166-3170.

- Hernández, G. und LeMaster, D.M.: *Reduced temperature dependence of collective conformational opening in a hyperthermophile rubredoxin*. *Biochemistry* **2001**, 40, 14384-14391.
- Hespenheide, B.M., Rader, A.J., Thorpe, M.F. und Kuhn, L.A.: *Identifying protein folding cores from the evolution of flexible regions during unfolding*. *J. Mol. Graph. Mod.* **2002**, 21, 195-207.
- Hirose, R., Suzuki, T., Moriyama, H., Sato, T., Yamagishi, A., Oshima, T. und Tanaka, N.: *Crystal structures of mutants of *Thermus thermophilus* IPMDH adapted to low temperatures*. *Protein Eng.* **2001**, 14, 81-84.
- Holland, D.R., Hausrath, A.C., Juers, D. und Matthews, B.W.: *Structural analysis of zinc substitutions in the active site of thermolysin*. *Protein Sci.* **1995**, 4, 1955-1965.
- Holland, D.R., Tronrud, D.E., Pley, H.W., Flaherty, K.M., Stark, W., Jansonius, J.N., McKay, D.B. und Matthews, B.W.: *Structural comparison suggests that thermolysin and related neutral proteases undergo hinge-bending motion during catalysis*. *Biochemistry* **1992**, 31, 11310-11316.
- Hollien, J. und Marqusee, S.: *Structural distribution of stability in a thermophilic enzyme*. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, 96, 13674-13678.
- Hollien, J. und Marqusee, S.: *Comparison of the folding processes of *T. thermophilus* and *E. coli* ribonucleases H*. *J. Mol. Biol.* **2002**, 316, 327-340.
- Holm, L. und Park, J.: *DaliLite workbench for protein structure comparison*. *Bioinformatics* **2000**, 16, 566-567.
- Hooft, R.W.W., Vriend, G., Sander, C. und Abola, E.E.: *Errors in protein structures*. *Nature* **1996**, 381, 272-272.
- Hopfner, K.P., Eichinger, A., Engh, R.A., Laue, F., Ankenbauer, W., Huber, R. und Angerer, B.: *Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius**. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, 96, 3600-3605.
- Hori, T., Moriyama, H., Kawaguchi, J., Hayashi-Iwasaki, Y., Oshima, T. und Tanaka, N.: *The initial step of the thermal unfolding of 3-isopropylmalate dehydrogenase detected by the temperature-jump Laue method*. *Protein Eng.* **2000**, 13, 527-533.
- Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. und Simmerling, C.: *Comparison of multiple Amber force fields and development of improved protein backbone parameters*. *Proteins* **2006**, 65, 712-725.

- Huang, S.L., Wu, L.C., Liang, H.K., Pan, K.T., Horng, J.T. und Ko, M.T.: *PGTdb: A database providing growth temperatures of prokaryotes*. *Bioinformatics* **2004**, 20, 276-278.
- Hurley, J.H. und Dean, A.M.: *Structure of 3-isopropylmalate dehydrogenase in complex with NAD(+) - Ligand-induced loop closing and mechanism for cofactor specificity*. *Structure* **1994**, 2, 1007-1016.
- Hvidt, A.: *Interactions of water with nonpolar solutes*. *Ann. Rev. Biophys. Bioeng.* **1983**, 12, 1-20.
- Ibrahim, B.S. und Pattabhi, V.: *Role of weak interactions in thermal stability of proteins*. *Biochem. Biophys. Res. Commun.* **2004**, 325, 1082-1089.
- Igbasan, F.A., Manner, K., Miksch, G., Borriss, R., Farouk, A. und Simon, O.: *Comparative studies on the in vitro properties of phytases from various microbial origins*. *Arch. Anim. Nutr.* **2000**, 53, 353-373.
- Igbasan, F.A., Simon, O., Miksch, G. und Manner, K.: *The effectiveness of an Escherichia coli phytase in improving phosphorus and calcium bioavailabilities in poultry and young pigs*. *Arch. Anim. Nutr.* **2001**, 54, 117-126.
- Imada, K., Sato, M., Tanaka, N., Katsube, Y., Matsuura, Y. und Oshima, T.: *3-dimensional structure of a highly thermostable enzyme, 3-isopropylmalate dehydrogenase of Thermus thermophilus at 2.2 Å resolution*. *J. Mol. Biol.* **1991**, 222, 725-738.
- Imanaka, T., Shibasaki, M. und Takagi, M.: *A new way of enhancing the thermostability of proteases*. *Nature* **1986**, 324, 695-697.
- Jacobs, D.J.: *Generic rigidity in three-dimensional bond-bending networks*. *J. Phys. A: Math. Gen.* **1998**, 31, 6653-6668.
- Jacobs, D.J. und Hendrickson, B.: *An algorithm for two-dimensional rigidity percolation: The pebble game*. *J. Comput. Phys.* **1997**, 137, 346-365.
- Jacobs, D.J., Rader, A.J., Kuhn, L.A. und Thorpe, M.F.: *Protein flexibility predictions using graph theory*. *Proteins* **2001**, 44, 150-165.
- Jacobs, D.J. und Thorpe, M.F.: *Generic rigidity percolation: The pebble game*. *Phys. Rev. Lett.* **1995**, 75, 4051-4054.
- Jaenicke, R.: *What ultrastable globular proteins teach us about protein stabilization*. *Biochemistry (Moscow, Russ. Fed.)* **1998**, 63, 312-321.
- Jaenicke, R.: *Do ultrastable proteins from hyperthermophiles have high or low conformational rigidity?* *Proc. Natl. Acad. Sci. U. S. A.* **2000**, 97, 2962-2964.

- Jaenicke, R. und Böhm, G.: *The stability of proteins in extreme environments*. Curr. Opin. Struct. Biol. **1998**, 8, 738-748.
- Jeffrey, G.A.: *An Introduction to Hydrogen Bonding*. New York, Oxford University Press, **1997**.
- Jenkins, N. und Curling, E.M.: *Glycosylation of recombinant proteins: Problems and prospects*. Enzyme Microb. Technol. **1994**, 16, 354-364.
- Jorgensen, W.L.: *Revised TIPS for simulations of liquid water and aqueous solutions*. J. Chem. Phys. **1982**, 77, 4156-4163.
- Kadono, S., Sakurai, M., Moriyama, H., Sato, M., Hayashi, Y., Oshima, T. und Tanaka, N.: *Ligand-induced changes in the conformation of 3-isopropylmalate dehydrogenase from *Thermus thermophilus**. J. Biochem. (Tokyo, Jpn.) **1995**, 118, 745-752.
- Kamerzell, T.J. und Middaugh, C.R.: *The complex inter-relationships between protein flexibility and stability*. J. Pharm. Sci. **2008**, 97, 3494-3517.
- Kannan, N. und Vishveshwara, S.: *Identification of side-chain clusters in protein structures by a graph spectral method*. J. Mol. Biol. **1999**, 292, 441-464.
- Kannan, N. und Vishveshwara, S.: *Aromatic clusters: A determinant of thermal stability of thermophilic proteins*. Protein Eng. **2000**, 13, 753-761.
- Karshikoff, A. und Ladenstein, R.: *Proteins from thermophilic and mesophilic organisms essentially do not differ in packing*. Protein Eng. **1998**, 11, 867-872.
- Karshikoff, A. und Ladenstein, R.: *Ion pairs and the thermotolerance of proteins from hyperthermophiles: A "traffic rule" for hot roads*. Trends Biochem. Sci. **2001**, 26, 550-556.
- Kauzmann, W.: *Some factors in the interpretation of protein denaturation*. Adv. Protein Chem. **1959**, 14, 1-63.
- Kidokoro, S., Miki, Y., Endo, K., Wada, A., Nagao, H., Miyake, T., Aoyama, A., Yoneya, T., Kai, K. und Ooe, S.: *Remarkable activity enhancement of thermolysin mutants*. FEBS Lett. **1995**, 367, 73-76.
- Kirino, H., Aoki, M., Aoshima, M., Hayashi, Y., Ohba, M., Yamagishi, A., Wakagi, T. und Oshima, T.: *Hydrophobic interaction at the subunit interface contributes to the thermostability of 3-isopropylmalate dehydrogenase from an extreme thermophile, *Thermus thermophilus**. Eur. J. Biochem. **1994**, 220, 275-281.
- Korkegian, A., Black, M.E., Baker, D. und Stoddard, B.L.: *Computational thermostabilization of an enzyme*. Science **2005**, 308, 857-860.

- Kostrewa, D., Leitch, F.G., D'Arcy, A., Broger, C., Mitchell, D. und vanLoon, A.P.G.M.: *Crystal structure of phytase from Aspergillus ficuum at 2.5 angstrom resolution*. Nat. Struct. Biol. **1997**, 4, 185-190.
- Kotsuka, T., Akanuma, S., Tomuro, M., Yamagishi, A. und Oshima, T.: *Further stabilization of 3-isopropylmalate dehydrogenase of an extreme thermophile, Thermus thermophilus, by a suppressor mutation method*. J. Bacteriol. **1996**, 178, 723-727.
- Kuhlman, B. und Baker, D.: *Native protein sequences are close to optimal for their structures*. Proc. Natl. Acad. Sci. U. S. A. **2000**, 97, 10383-10388.
- Kumar, S. und Nussinov, R.: *How do thermophilic proteins deal with heat?* Cell. Mol. Life Sci. **2001**, 58, 1216-1233.
- Kumar, S., Tsai, C.J. und Nussinov, R.: *Thermodynamic differences among homologous thermophilic and mesophilic proteins*. Biochemistry **2001**, 40, 14152-14165.
- Kundrot, C.E. und Richards, F.M.: *Crystal-structure of hen egg-white lysozyme at a hydrostatic pressure of 1000 atmospheres*. J. Mol. Biol. **1987**, 193, 157-170.
- Lagrange, J.L.: *Mécanique analytique*. Paris, Dessaint, **1788**.
- Laman, G.: *On graphs and the rigidity of plane skeletal structures*. J. Eng. Math. **1970**, 4, 331-340.
- Lee, M.R., Duan, Y. und Kollman, P.A.: *Use of MM-PB/SA in estimating the free energies of proteins: Application to native, intermediates, and unfolded villin head-piece*. Proteins **2000**, 39, 309-316.
- Lehmann, M., Kostrewa, D., Wyss, M., Brugger, R., D'Arcy, A., Pasamontes, L. und van Loon, A.P.G.M.: *From DNA sequence to improved functionality: Using protein sequence comparisons to rapidly design a thermostable consensus phytase*. Protein Eng. **2000**, 13, 49-57.
- Lehmann, M., Loch, C., Middendorf, A., Studer, D., Lassen, S.F., Pasamontes, L., van Loon, A.P.G.M. und Wyss, M.: *The consensus concept for thermostability engineering of proteins: Further proof of concept*. Protein Eng. **2002**, 15, 403-411.
- Lehmann, M. und Wyss, M.: *Engineering proteins for thermostability: The use of sequence alignments versus rational design and directed evolution*. Curr. Opin. Biotechnol. **2001**, 12, 371-375.
- Lim, D., Golovan, S., Forsberg, C.W. und Jia, Z.C.: *Crystal structures of Escherichia coli phytase and its complex with phytate*. Nat. Struct. Biol. **2000**, 7, 108-113.

- Liu, Y.H. und Konermann, L.: *Conformational dynamics of free and catalytically active thermolysin are indistinguishable by hydrogen/deuterium exchange mass spectrometry*. *Biochemistry* **2008**, 47, 6342-6351.
- Livesay, D.R., Huynh, D.H., Dallakyan, S. und Jacobs, D.J.: *Hydrogen bond networks determine emergent mechanical and thermodynamic properties across a protein family*. *Chem. Cent. J.* **2008**, 2, 1-20.
- Livesay, D.R. und Jacobs, D.J.: *Conserved quantitative stability/flexibility relationships (QSFR) in an orthologous RNase H pair*. *Proteins* **2006**, 62, 130-143.
- Lumry, R. und Eyring, H.: *Conformation changes of proteins*. *J. Phys. Chem.* **1954**, 58, 110-120.
- Luque, I. und Freire, E.: *Structural stability of binding sites: Consequences for binding affinity and allosteric effects*. *Proteins* **2000**, 63-71.
- Makhatadze, G.I. und Privalov, P.L.: *Energetics of protein structure*. *Adv Prot Chem* **1995**, 47, 307-425.
- Malakauskas, S.M. und Mayo, S.L.: *Design, structure and stability of a hyperthermophilic protein variant*. *Nat. Struct. Biol.* **1998**, 5, 470-475.
- Mansfeld, J., Vriend, G., Dijkstra, B.W., Veltman, O.R., Van den Burg, B., Venema, G., Ulbrich-Hofmann, R. und Eijsink, V.G.H.: *Extreme stabilization of a thermolysin-like protease by an engineered disulfide bond*. *J. Biol. Chem.* **1997**, 272, 11152-11156.
- Mansfeld, J., Vriend, G., Van den Burg, B., Eijsink, V.G.H. und Ulbrich-Hofmann, R.: *Probing the unfolding region in a thermolysin-like protease by site-specific immobilization*. *Biochemistry* **1999**, 38, 8240-8245.
- Margarit, I., Campagnoli, S., Frigerio, F., Grandi, G., Defilippis, V. und Fontana, A.: *Cumulative stabilizing effects of glycine to alanine substitutions in Bacillus subtilis neutral protease*. *Protein Eng.* **1992**, 5, 543-550.
- Matthews, B.W.: *Structural basis of the action of thermolysin and related zinc peptidases*. *Acc. Chem. Res.* **1988**, 21, 333-340.
- Maxwell, J.C.: *On the calculation of the equilibrium and stiffness of frames*. *Philos. Mag.* **1864**, 27, 294-299.
- McDonald, I.K. und Thornton, J.M.: *Satisfying hydrogen-bonding potential in proteins*. *J. Mol. Biol.* **1994**, 238, 777-793.

- Ming, D., Kong, Y., Wakil, S.J., Brink, J. und Ma, J.: *Domain movements in human fatty acid synthase by quantized elastic deformational model*. Proc. Natl. Acad. Sci. U. S. A. **2002**, 99, 7895-7899.
- Ming, D., Kong, Y., Wu, Y. und Ma, J.: *Substructure synthesis method for simulating large molecular complexes*. Proc. Natl. Acad. Sci. U. S. A. **2003**, 100, 104-109.
- Mock, W.L. und Aksamawati, M.: *Binding to thermolysin of phenolate-containing inhibitors necessitates a revised mechanism of catalysis*. Biochem. J. **1994**, 302, 57-68.
- Mock, W.L. und Stanford, D.J.: *Arazoformyl dipeptide substrates for thermolysin. Confirmation of a reverse protonation catalytic mechanism*. Biochemistry **1996**, 35, 7369-7377.
- Motono, C., Oshima, T. und Yamagishi, A.: *High thermal stability of 3-isopropylmalate dehydrogenase from Thermus thermophilus resulting from low delta Cp of unfolding*. Protein Eng. **2001**, 14, 961-966.
- Motono, C., Yamagishi, A. und Oshima, T.: *Urea-induced unfolding and conformational stability of 3-isopropylmalate dehydrogenase from the thermophile Thermus thermophilus and its mesophilic counterpart from Escherichia coli*. Biochemistry **1999**, 38, 1332-1337.
- Moukarzel, C.: *An efficient algorithm for testing the generic rigidity of graphs in the plane*. J. Phys. A: Math. Gen. **1996**, 29, 8079-8098.
- Myers, J.K. und Pace, C.N.: *Hydrogen bonding stabilizes globular proteins*. Biophys. J. **1996**, 71, 2033-2039.
- Naoki, K. und Shin-Ichi, T.: *A proof of the molecular conjecture*. Proceedings of the 25th annual symposium on computational geometry, Aarhus, Denmark, ACM, **2009**.
- Németh, A., Svingor, A., Pocsik, M., Dobó, J., Magyar, C., Szilágyi, A., Gál, P. und Zavódszky, P.: *Mirror image mutations reveal the significance of an intersubunit ion cluster in the stability of 3-isopropylmalate dehydrogenase*. FEBS Lett. **2000**, 468, 48-52.
- Nojima, H., Ikai, A., Oshima, T. und Noda, H.: *Reversible thermal unfolding of thermostable phosphoglycerate kinase - Thermostability associated with mean zero enthalpy change*. J. Mol. Biol. **1977**, 116, 429-442.
- Notredame, C., Higgins, D.G. und Heringa, J.: *T-Coffee: A novel method for fast and accurate multiple sequence alignment*. J. Mol. Biol. **2000**, 302, 205-217.

- Numata, K., Hayashi-Iwasaki, Y., Kawaguchi, J., Sakurai, M., Moriyama, H., Tanaka, N. und Oshima, T.: *Thermostabilization of a chimeric enzyme by residue substitutions: four amino acid residues in loop regions are responsible for the thermostability of Thermus thermophilus isopropylmalate dehydrogenase*. Biochim. Biophys. Acta **2001**, 1545, 174-183.
- Numata, K., Hayashi-Iwasaki, Y., Yutani, K. und Oshima, T.: *Studies on interdomain interaction of 3-isopropylmalate dehydrogenase from an extreme thermophile, Thermus thermophilus, by constructing chimeric enzymes*. Extremophiles **1999**, 3, 259-262.
- Numata, K., Muro, M., Akutsu, N., Nosoh, Y., Yamagishi, A. und Oshima, T.: *Thermal stability of chimeric isopropylmalate dehydrogenase genes constructed from a thermophile and a mesophile*. Protein Eng. **1995**, 8, 39-43.
- Nurachman, Z., Akanuma, S., Sato, T., Oshima, T. und Tanaka, N.: *Crystal structures of 3-isopropylmalate dehydrogenases with mutations at the C-terminus: Crystallographic analyses of structure-stability relationships*. Protein Eng. **2000**, 13, 253-258.
- Ohkuri, T. und Yamagishi, A.: *Increased thermal stability against irreversible inactivation of 3-isopropylmalate dehydrogenase induced by decreased van der Waals volume at the subunit interface*. Protein Eng. **2003**, 16, 615-621.
- Ohkuri, T. und Yamagishi, A.: *The effects of mutations at position 253 on the thermostability of the Bacillus subtilis 3-isopropylmalate dehydrogenase subunit interface*. J. Biochem. (Tokyo, Jpn.) **2007**, 141, 791-797.
- Onufriev, A., Bashford, D. und Case, D.A.: *Exploring protein native states and large-scale conformational changes with a modified generalized born model*. Proteins **2004**, 55, 383-394.
- Pace, C.N.: *Conformational stability of globular proteins*. Trends Biochem. Sci. **1990**, 15, 14-17.
- Pace, C.N.: *Polar group burial contributes more to protein stability than nonpolar group burial*. Biochemistry **2001**, 40, 310-313.
- Pace, C.N., Trevino, S., Prabhakaran, E. und Scholtz, J.M.: *Protein structure, stability and solubility in water and other solvents*. Philos. T. Roy. Soc. B **2004**, 359, 1225-1234.
- Pack, S.P. und Yoo, Y.J.: *Packing-based difference of structural features between thermophilic and mesophilic proteins*. Int. J. Biol. Macromol. **2005**, 35, 169-174.

- Paiardini, A., Sali, R., Bossa, F. und Pascarella, S.: *"Hot cores" in proteins: Comparative analysis of the apolar contact area in structures from hyper/thermophilic and mesophilic organisms*. *Bmc Struct. Biol.* **2008**, 8, 14.
- Pang, J. und Allemann, R.K.: *Molecular dynamics simulation of thermal unfolding of *Thermatoga maritima* DHFR*. *Phys. Chem. Chem. Phys.* **2007**, 9, 711-718.
- Pauling, L.: *The shared-electron chemical bond*. *Proc. Natl. Acad. Sci. U. S. A.* **1928**, 14, 359-362.
- Phillips, J.C.: *Topology of Covalent Non-Crystalline Solids (1) Short-Range Order in Chalcogenide Alloys*. *J. Non-Cryst. Solids* **1979**, 34, 153-181.
- Phillips, J.C.: *Topology of Covalent Non-Crystalline Solids (2) Medium-Range Order in Chalcogenide Alloys and *a*-Si(Ge)*. *J. Non-Cryst. Solids* **1981**, 43, 37-77.
- Prabhu, N.V. und Sharp, K.A.: *Heat capacity in proteins*. *Annu. Rev. Phys. Chem.* **2005**, 56, 521-548.
- Price, N.C. und Johnson, C.M.: *Proteinases as probes of conformation of soluble proteins. Proteolytic Enzymes: A practical approach*. Beynon, R. J. und Bond, J. S. Oxford, IRL Press, **1990**, 163-180.
- Purmonen, M., Valjakka, J., Takkinen, K., Laitinen, T. und Rouvinen, J.: *Molecular dynamics studies on the thermostability of family 11 xylanases*. *Protein Eng. Des. Sel.* **2007**, 20, 551-559.
- Qu, C.X., Akanuma, S., Moriyama, H., Tanaka, N. und Oshima, T.: *A mutation at the interface between domains causes rearrangement of domains in 3-isopropylmalate dehydrogenase*. *Protein Eng.* **1997**, 10, 45-52.
- Querol, E., Perez-Pons, J.A. und Mozo-Villarias, A.: *Analysis of protein conformational characteristics related to thermostability*. *Protein Eng.* **1996**, 9, 265-271.
- Rader, A.J. und Bahar, I.: *Folding core predictions from network models of proteins*. *Polymer* **2004**, 45, 659-668.
- Rader, A.J., Hespeneide, B.M., Kuhn, L.A. und Thorpe, M.F.: *Protein unfolding: Rigidity lost*. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, 99, 3540-3545.
- Razvi, A. und Scholtz, J.M.: *Lessons in stability from thermophilic proteins*. *Protein Sci.* **2006**, 15, 1569-1578.
- Rees, D.C. und Robertson, A.D.: *Some thermodynamic implications for the thermostability of proteins*. *Protein Sci.* **2001**, 10, 1187-1194.

- Reetz, M.T., Carballeira, J.D. und Vogel, A.: *Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability*. *Angew. Chem. Int. Ed.* **2006**, 45, 7745-7751.
- Richards, F.M.: *The interpretation of protein structures: total volume, group volume distributions and packing density*. *J. Mol. Biol.* **1974**, 82, 1-14.
- Richards, F.M.: *Protein stability: Still an unsolved problem*. *Cell. Mol. Life Sci.* **1997**, 53, 790-802.
- Robinson-Rechavi, M., Alibes, A. und Godzik, A.: *Contribution of electrostatic interactions, compactness and quaternary structure to protein thermostability: Lessons from structural genomics of *Thermotoga maritima**. *J. Mol. Biol.* **2006**, 356, 547-557.
- Rose, G.D. und Wolfenden, R.: *Hydrogen-bonding, hydrophobicity, packing, and protein folding*. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, 22, 381-415.
- Rozzell, J.D.: *Commercial scale biocatalysis: Myths and realities*. *Bioorg. Med. Chem.* **1999**, 7, 2253-2261.
- Russell, R.J., Ferguson, J.M., Hough, D.W., Danson, M.J. und Taylor, G.L.: *The crystal structure of citrate synthase from the hyperthermophilic archaeon *Pyrococcus furiosus* at 1.9 Å resolution*. *Biochemistry* **1997**, 36, 9983-9994.
- Russell, R.J.M. und Taylor, G.L.: *Engineering thermostability - Lessons from thermophilic proteins*. *Curr. Opin. Biotechnol.* **1995**, 6, 370-374.
- Ryckaert, J.P., Ciccotti, G. und Berendsen, H.J.C.: *Numerical-integration of cartesian equations of motion of a system with constraints - Molecular dynamics of n-alkanes*. *J. Comput. Phys.* **1977**, 23, 327-341.
- Sadeghi, M., Naderi-Manesh, H., Zarrabi, M. und Ranjbar, B.: *Effective factors in thermostability of thermophilic proteins*. *Biophys. Chem.* **2006**, 119, 256-270.
- Saelensminde, G., Halskau, O., Jr. und Jonassen, I.: *Amino acid contacts in proteins adapted to different temperatures: hydrophobic interactions and surface charges play a key role*. *Extremophiles* **2009**, 13, 11-20.
- Sakurai, M., Moriyama, H., Onodera, K., Kadono, S., Numata, K., Hayashi, Y., Kawaguchi, J., Yamagishi, A., Oshima, T. und Tanaka, N.: *The crystal-structure of thermostable mutants of chimeric 3-isopropylmalate dehydrogenase, 2T2M6T*. *Protein Eng.* **1995**, 8, 763-767.
- Sali, A., Shakhnovich, E. und Karplus, M.: *How does a protein fold?* *Nature* **1994**, 369, 248-251.

- Savage, H.J., Elliott, C.J., Freeman, C.M. und Finney, J.L.: *Lost hydrogen-bonds and buried surface-area - Rationalizing stability in globular proteins*. J. Chem. Soc., Faraday Trans. **1993**, 89, 2609-2617.
- Schell, D., Tsai, J., Scholtz, J.M. und Pace, C.N.: *Hydrogen bonding increases packing density in the protein interior*. Proteins **2006**, 63, 278-282.
- Schellenberger, A. und Ulbrich, R.: *Protein stabilization by blocking the native unfolding nucleus*. Biomed. Biochim. Acta **1989**, 48, 63-67.
- Scheraga, H.A., Nemethy, G. und Steinberg, I.Z.: *Contribution of hydrophobic bonds to thermal stability of protein conformations*. J. Biol. Chem. **1962**, 237, 2506-&.
- Sham, Y.Y., Ma, B., Tsai, C.J. und Nussinov, R.: *Thermal unfolding molecular dynamics simulation of Escherichia coli dihydrofolate reductase: Thermal stability of protein domains and unfolding pathway*. Proteins **2002**, 46, 308-320.
- Shi, Y.Y., Mark, A.E., Wang, C.X., Huang, F., Berendsen, H.J. und van Gunsteren, W.F.: *Can the stability of protein mutants be predicted by free energy calculations?* Protein Eng. **1993**, 6, 289-295.
- Shirley, B.A., Stanssens, P., Hahn, U. und Pace, C.N.: *Contribution of Hydrogen-Bonding to the Conformational Stability of Ribonuclease-T1*. Biochemistry **1992**, 31, 725-732.
- Siddiqui, K.S., Rangarajan, M., Hartley, B.S., Kitmitto, A., Panico, M., Blench, I.P. und Morris, H.R.: *Arthrobacter D-xylose isomerase - Partial proteolysis with thermolysin*. Biochem. J. **1993**, 289, 201-208.
- Simon, O. und Igbasan, F.: *In vitro properties of phytases from various microbial origins*. Int. J. Food Sci. Technol. **2002**, 37, 813-822.
- Singleton, M., Isupov, M. und Littlechild, J.: *X-ray structure of pyrrolidone carboxyl peptidase from the hyperthermophilic archaeon Thermococcus litoralis*. Structure **1999**, 7, 237-244.
- Somero, G.N.: *Temperature adaptation of enzymes - Biological optimization through structure-function compromises*. Annu. Rev. Ecol. Syst. **1978**, 9, 1-29.
- Stauffer, D. und Aharony, A.: *Introduction to percolation theory*. London, Taylor and Francis, **1994**.
- Steipe, B., Schiller, B., Pluckthun, A. und Steinbacher, S.: *Sequence statistics reliably predict stabilizing mutations in a protein domain*. J. Mol. Biol. **1994**, 240, 188-192.

- Sterner, R. und Brunner, E.: *The relationship between catalytic activity, structural flexibility and conformational stability as deduced from the analysis of mesophilic-thermophilic enzyme pairs and protein engineering studies. Thermophiles: Biology and technology at high temperatures.* Robb, F., Antranikian, G., Grogan, D. und Driessen, A. London, New York, CRC Press, **2008**, 25-38.
- Stickle, D.F., Presta, L.G., Dill, K.A. und Rose, G.D.: *Hydrogen bonding in globular proteins.* J. Mol. Biol. **1992**, 226, 1143-1159.
- Strickler, S.S., Gribenko, A.V., Gribenko, A.V., Keiffer, T.R., Tomlinson, J., Reihle, T., Loladze, V.V. und Makhatadze, G.I.: *Protein stability and surface electrostatics: a charged relationship.* Biochemistry **2006**, 45, 2761-2766.
- Su, J.G., Li, C.H., Hao, R., Chen, W.Z. und Wang, C.X.: *Protein unfolding behavior studied by elastic network model.* Biophys. J. **2008**, 94, 4586-4596.
- Suzuki, T., Yasugi, M., Arisaka, F., Oshima, T. und Yamagishi, A.: *Cold-adaptation mechanism of mutant enzymes of 3-isopropylmalate dehydrogenase from Thermus thermophilus.* Protein Eng. **2002**, 15, 471-476.
- Szilágyi, A. und Závodszky, P.: *Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: Results of a comprehensive survey.* Structure **2000**, 8, 493-504.
- Tama, F., Wriggers, W. und Brooks, C.L., 3rd: *Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory.* J. Mol. Biol. **2002**, 321, 297-305.
- Tamakoshi, M., Nakano, Y., Kakizawa, S., Yamagishi, A. und Oshima, T.: *Selection of stabilized 3-isopropylmalate dehydrogenase of Saccharomyces cerevisiae using the host-vector system of an extreme thermophile, Thermus thermophilus.* Extremophiles **2001**, 5, 17-22.
- Tamakoshi, M., Yamagishi, A. und Oshima, T.: *Screening of stable proteins in an extreme thermophile, Thermus thermophilus.* Mol. Microbiol. **1995**, 16, 1031-1036.
- Tanford, C.: *Protein denaturation.* Adv. Protein Chem. **1968**, 23, 121-282.
- Tay, T.-S. und Whiteley, W.: *Recent advances in generic rigidity of structures.* Struct Topol **1985**, 9, 31-38.
- Taylor, R. und Kennard, O.: *Crystallographic evidence for the existence of C-H...O, C-H...N, C-H...Cl hydrogen bonds.* J. Am. Chem. Soc. **1982**, 104, 5063-5070.
- Thorpe, M.F.: *Continuous deformations in random networks.* J. Non-Cryst. Solids **1983**, 57, 355-370.

- Tidor, B. und Karplus, M.: *Simulation analysis of the stability mutant R96H of T4 lysozyme*. *Biochemistry* **1991**, 30, 3217-3228.
- Tisi, L.C. und Evans, P.A.: *Conserved structural features on protein surfaces - Small exterior hydrophobic clusters*. *J. Mol. Biol.* **1995**, 249, 251-258.
- Tsang, I.R. und Tsang, I.J.: *Cluster size diversity, percolation, and complex systems*. *Phys. Rev. E* **1999**, 60, 2684-2698.
- Ulbrich-Hofmann, R., Arnold, U. und Mansfeld, J.: *The concept of the unfolding region for approaching the mechanisms of enzyme stabilization*. *J. Mol. Catal. B: Enzym.* **1999**, 7, 125-131.
- van Aalten, D.M.F., Amadei, A., Linssen, A.B.M., Eijsink, V.G.H., Vriend, G. und Berendsen, H.J.C.: *The essential dynamics of thermolysin: Confirmation of the hinge-bending motion and comparison of simulations in vacuum and water*. *Proteins* **1995**, 22, 45-54.
- van den Burg, B., Dijkstra, B.W., Vriend, G., Vandervinne, B., Venema, G. und Eijsink, V.G.H.: *Protein stabilization by hydrophobic interactions at the surface*. *Eur. J. Biochem.* **1994**, 220, 981-985.
- van den Burg, B., Enequist, H.G., Vanderhaar, M.E., Eijsink, V.G.H., Stulp, B.K. und Venema, G.: *A highly thermostable neutral protease from Bacillus caldolyticus: Cloning and expression of the gene in Bacillus subtilis and characterization of the gene product*. *J. Bacteriol.* **1991**, 173, 4107-4115.
- van den Burg, B., Vriend, G., Veltman, O.R., Venema, G. und Eijsink, V.G.: *Engineering an enzyme to resist boiling*. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, 95, 2056-2060.
- van Gunsteren, W.F.: *The role of computer simulation techniques in protein engineering*. *Protein Eng.* **1988**, 2, 5-13.
- van Gunsteren, W.F., Hünenberger, P.H., Kovacs, H., Mark, A.E. und Schiffer, C.A.: *Investigation of protein unfolding and stability by computer simulation*. *Protein folding*. Dobson, C. M. und Fersht, A. R. Cambridge, Cambridge University Press, **1995**, 49-60.
- Veltman, O.R., Eijsink, V.G.H., Vriend, G., de Kreij, A., Venema, G. und van den Burg, B.: *Probing catalytic hinge bending motions in thermolysin-like proteases by glycine -> alanine mutations*. *Biochemistry* **1998**, 37, 5305-5311.

- Veltman, O.R., Vriend, G., Hardy, F., Mansfeld, J., van den Burg, B., Venema, G. und Eijsink, V.G.H.: *Mutational analysis of a surface area that is critical for the thermal stability of thermolysin-like proteases*. Eur. J. Biochem. **1997**, 248, 433-440.
- Veltman, O.R., Vriend, G., Middelhoven, P.J., van den Burg, B., Venema, G. und Eijsink, V.G.H.: *Analysis of structural determinants of the stability of thermolysin-like proteases by molecular modelling and site-directed mutagenesis*. Protein Eng. **1996**, 9, 1181-1189.
- Vendruscolo, M., Dokholyan, N.V., Paci, E. und Karplus, M.: *Small-world view of the amino acids that play a key role in protein folding*. Phys. Rev. E **2002**, 65, 1-4.
- Vieille, C. und Zeikus, G.J.: *Hyperthermophilic enzymes: Sources, uses, and molecular mechanisms for thermostability*. Microbiol. Mol. Biol. Rev. **2001**, 65, 1-43.
- Vihinen, M.: *Relationship of protein flexibility to thermostability*. Protein Eng. **1987**, 1, 477-480.
- Vogt, G. und Argos, P.: *Protein thermal stability: Hydrogen bonds or internal packing?* Fold. Des. **1997**, 2, S40-S46.
- Vogt, G., Woell, S. und Argos, P.: *Protein thermal stability, hydrogen bonds, and ion pairs*. J. Mol. Biol. **1997**, 269, 631-643.
- Wallon, G., Kryger, G., Lovett, S.T., Oshima, T., Ringe, D. und Petsko, G.A.: *Crystal structures of Escherichia coli and Salmonella typhimurium 3-isopropylmalate dehydrogenase and comparison with their thermophilic counterpart from Thermus thermophilus*. J. Mol. Biol. **1997**, 266, 1016-1031.
- Watanabe, K., Ohkuri, T., Yokobori, S. und Yamagishi, A.: *Designing thermostable proteins: Ancestral mutants of 3-isopropylmalate dehydrogenase designed by using a phylogenetic tree*. J. Mol. Biol. **2006**, 355, 664-674.
- Whiteley, W.: *Realizability of polyhedra*. Struct. Topol. **1979**, 1, 46-58.
- Whiteley, W.: *Counting out to the flexibility of molecules*. Phys Biol **2005**, 2, S116-126.
- Wolfenden, R.: *Experimental measures of amino acid hydrophobicity and the topology of transmembrane and globular proteins*. J. Gen. Physiol. **2007**, 129, 357-362.
- Wooll, J.O., Wrabl, J.O. und Hilser, V.J.: *Ensemble modulation as an origin of denaturant-independent hydrogen exchange in proteins*. J. Mol. Biol. **2000**, 301, 247-256.
- Word, J.M., Lovell, S.C., Richardson, J.S. und Richardson, D.C.: *Asparagine and glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation*. J. Mol. Biol. **1999**, 285, 1735-1747.

- Wyss, M., Pasamontes, L., Friedlein, A., Remy, R., Tessier, M., Kronenberger, A., Middendorf, A., Lehmann, M., Schnoebelen, L., Rothlisberger, U., Kuszniir, E., Wahl, G., Muller, F., Lahm, H.W., Vogel, K. und van Loon, A.P.: *Biophysical characterization of fungal phytases (myo-inositol hexakisphosphate phosphohydrolases): Molecular size, glycosylation pattern, and engineering of proteolytic resistance*. Appl. Environ. Microbiol. **1999**, 65, 359-366.
- Wyss, M., Pasamontes, L., Remy, R., Kohler, J., Kuszniir, E., Gadiant, M., Muller, F. und van Loon, A.: *Comparison of the thermostability properties of three acid phosphatases from molds: Aspergillus fumigatus phytase, A. niger phytase, and A. niger PH 2.5 acid phosphatase*. Appl. Environ. Microbiol. **1998**, 64, 4446-4451.
- Xiang, T., Liu, Q., Deacon, A.M., Koshy, M., Kriksunov, I.A., Lei, X.G., Hao, Q. und Thiel, D.J.: *Crystal structure of a heat-resilient phytase from Aspergillus fumigatus, carrying a phosphorylated histidine*. J. Mol. Biol. **2004**, 339, 437-445.
- Xiao, L. und Honig, B.: *Electrostatic contributions to the stability of hyperthermophilic proteins*. J. Mol. Biol. **1999**, 289, 1435-1444.
- Yang, L.W. und Bahar, I.: *Coupling between catalytic site and collective dynamics: A requirement for mechanochemical activity of enzymes*. Structure **2005**, 13, 893-904.
- Yasugi, M., Amino, M., Suzuki, T., Oshima, T. und Yamagishi, A.: *Cold adaptation of the thermophilic enzyme 3-isopropylmalate dehydrogenase*. J Biochem **2001**, 129, 477-484.
- Yasugi, M., Suzuki, T., Yamagishi, A. und Oshima, T.: *Analysis of the effect of accumulation of amino acid replacements on activity of 3-isopropylmalate dehydrogenase from Thermus thermophilus*. Protein Eng. **2001**, 14, 601-607.
- Yasukawa, K. und Inouye, K.: *Improving the activity and stability of thermolysin by site-directed mutagenesis*. Biochim. Biophys. Acta **2007**, 1774, 1281-1288.
- Závodszky, P., Kardos, J., Svingor, A. und Petsko, G.A.: *Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins*. Proc. Natl. Acad. Sci. U. S. A. **1998**, 95, 7406-7411.
- Zhou, Y., Vitkup, D. und Karplus, M.: *Native proteins are surface-molten solids: application of the Lindemann criterion for the solid versus liquid state*. J. Mol. Biol. **1999**, 285, 1371-1375.

...laboraverimus.

Columella,

De re rustica