

Technische Universität München
Max-Planck-Institut für Entwicklungsbiologie
Abteilung für Proteinevolution

Evolution der β -Propeller Proteine

Indronil Chaudhuri

Vollständiger Abdruck der von der Fakultät für Chemie
der Technischen Universität München
zur Erlangung des akademischen Grades eines Doktors
der Naturwissenschaften genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. H. Kessler
Prüfer der Dissertation: 1. Hon.-Prof. Dr. W. Baumeister
2. Univ.-Prof. Dr. J. Buchner

Die Dissertation wurde am 28.09.2007 bei der Technischen Universität München
eingereicht und durch die Fakultät für Chemie am 29.11.2007 angenommen.

Kurzfassung/Überblick

β -Propeller bilden eine Gruppe von Proteinen, die in allen drei Ästen des Stammbaums der lebenden Organismen auftreten, überwiegend jedoch in Eukaryoten. β -Propeller erfüllen eine Vielzahl äußerst unterschiedlicher Funktionen, von der Vermittlung von Protein-Protein Interaktionen bis hin zu enzymatischen Reaktionen. Die Superfamilie der β -Propellerproteine ist durch eine ausgeprägte Sequenzdiversität charakterisiert. Alle Proteine haben jedoch ein gemeinsames strukturelles Merkmal: Ihre globuläre Struktur wird aus repetitiven Einheiten aufgebaut, den Propellerblättern. In den ca. 60 Strukturen in der Proteinstrukturdatenbank (PDB) sind Proteine mit vier bis acht Blättern hinterlegt. Es ist erstaunlich, dass die Natur eine so variable Zahl an strukturell ähnlichen Wiederholungseinheiten verwenden kann, um eine zirkuläre Proteinstruktur zu bilden.

Diese Arbeit versucht drei thematisch verwandte Fragen zu beantworten:

- Wie konnte diese Proteinfaltungsklasse (*Fold*) entstehen?
- Wie viele Propellerblätter können maximal in einer Domäne auftreten?
- Sind alle β -Propeller homolog?

Eine Möglichkeit für die Entstehung dieser Proteinfaltung ist, dass einzelne Blätter oligomerisierten und erst später zu einer durchgängigen Polypeptidkette fusionierten. Um dieses Szenario zu überprüfen haben wir oligomere Propeller aus kleineren Bruchstücken aufgebaut. Dabei konnten wir sogar, abhängig von der Größe der Ausgangsfragmente, Propeller unterschiedlicher Größe herstellen.

Um die maximale Anzahl der Blätter zu bestimmen, die in eine einzelne β -Propellerdomäne falten, haben wir ein Protein mit 14 identischen Wiederholungseinheiten kristallisiert. Dieses Protein faltet in zwei Propellerdomänen mit jeweils sieben Blättern. Diese Struktur war nicht vorhersehbar, da die hochrepetitive Sequenz keine offensichtliche Bruchstelle zwischen den Propellerblättern enthält.

β -Propeller sind eine auf der Sequenzebene sehr heterogene Gruppe, die sich durch große Diversität auszeichnet. Dennoch lassen sich Ähnlichkeiten zwischen verschiedenen Propellergruppen finden, die nicht nur strukturell begründet sind. Mit empfindlichen bioinformatischen Methoden konnten wir für 3/4 der bekannten Strukturen Sequenzverwandtschaften feststellen, die indikativ für eine gemeinsame Abstammung sind. Wir leiten daraus einen homologen Ursprung der β -Propeller ab.

Aus den Daten dieser Arbeit schließen wir, dass die β -Propeller einen gemeinsamen Ursprung haben und dass sich die Propellerdomäne in der Evolution aus kleineren Fragmenten gebildet haben könnte. Diese Bausteine sind in der Lage, β -Propeller mit vier bis zehn Blättern in einer Domäne zu bilden.

Abstract in English

β -Propeller proteins are a group of proteins which occur in all domains of life, with a strong preponderance among eukaryotes. β -propellers fulfill a range of very diverse functions, from mediating protein-protein interactions to enzymatic activities. The superfamily of β -propeller proteins is characterized by an enormous diversity at the sequence level. However, all proteins share a common structural feature: they build their globular structure from repetitive units, the propeller blades. In the 60-odd structures of β -propellers in the Protein Data Bank, the number of blades ranges from four to eight. It is remarkable that nature can use such diverse numbers of structurally similar repeats to construct compact protein structures.

This thesis tries to answer three related questions:

- How did this protein fold arise?
- What is the upper limit of blades that still fold into a single domain?
- Are all β -propellers of homologous origin?

One possible scenario how the β -propeller fold arose is that single blades oligomerized and were only later combined to a single polypeptide chain. To evaluate this scenario, we constructed full-sized, oligomeric propellers from smaller pieces. We could assemble propellers with different numbers of blades, depending on the size of the original building block.

To determine how many blades still fold into one propeller domain, a protein with 14 identical repeats was crystallized. This protein folds into two domains of seven blades each. This structure was not predictable as the highly repetitive sequence does not contain any obvious breaking points between the propeller blades.

On the sequence level β -propellers are a very heterogeneous group which shows considerable sequence diversity. Despite this diversity, they often show similarities across different groups, not only in structure but also in sequence. Of all 60 propellers in the SCOP25 database, 45 proteins show sequence similarities at a level indicative of homology if they are compared with sensitive bioinformatic methods. We propose that all β -propellers share a common origin.

We conclude that the β -propeller are indeed homologous and that the propeller domain may have evolved from smaller peptides. These building blocks are well capable of forming propellers with four to ten blades in one single domain.

Für meine Familie

Inhaltsverzeichnis

Kurzfassung/Überblick	1
Abstract in English	3
1 Einleitung	11
1.1 Proteinevolution	11
1.1.1 Proteindomänen	11
1.1.2 Evolution der Domänen aus Peptiden	13
1.1.3 Das Konzept der Homologie	15
1.1.4 Mechanismen der Proteinevolution	18
1.2 β -Propeller Proteine	20
1.2.1 Sequenzdiversität in der Gruppe der β -Propeller	22
1.2.2 Biologische Funktionen von β -Propellern	25
1.2.3 Biologische Verbreitung der β -Propeller Proteine	27
2 Homologie der β-Propeller	29
2.1 Sequenzdiversität bei β -Propellern	29
2.2 Bioinformatische Methoden	29
2.2.1 Strukturelle Überlagerung von β -Propellerblättern	29
2.2.2 Clusterkarte der All- β -Proteine	30
2.2.3 Matrix mit β -Propellerverbindungen, basierend auf HH- search	31
2.2.4 Phylogenien von β -Propellern	31
2.3 Ergebnisse/Diskussion Homologie-Untersuchungen	31
2.3.1 Ein Kontinuum von Sequenzmotiven	31
2.3.2 β -Propeller bilden eine Gruppe in einer Clusterkarte der All- β -Proteine	34
2.3.3 HMM-HMM Vergleiche von β -Propellern	35
2.4 Evolutionäre Szenarios	38
2.5 Zusammenfassung	40
3 Aufbau von β-Propellern aus Fragmenten	43
3.1 Einleitung	43
3.2 Methoden zur Charakterisierung von Proteinen	43
3.2.1 Statische Lichtstreuung	43
3.2.2 Circular dichroismus (CD)	45
3.2.3 Fluoreszenzspektroskopie	46
3.3 Material und Methoden	47

3.3.1	Klonierung des einzelnen Propellerblattes	47
3.3.2	Amplifikation der Propellerblätter	52
3.3.3	Alternative Propellerblätter	53
3.4	Expression und Reinigung der Konstrukte	53
3.4.1	GST-Fusionsproteine	53
3.4.2	Das verdoppelte Propellerblatt	54
3.4.3	Proteine mit einem His ₆ -Tag	54
3.4.4	Kristallisationsansätze der Konstrukte	55
3.4.5	Vorgehen bei den Mischungsexperimenten	55
3.5	Hintergründe der verwendeten Konstrukte	55
3.5.1	Informationen über das Ausgangsprotein für die einzelnen Blätter	56
3.5.2	Amplifikation der Propellerblätter	56
3.5.3	Alternative Propellerblätter	57
3.5.4	Zusammenfassung der Proteinreinigungen	59
3.6	Biochemische Charakterisierung der Konstrukte	59
3.6.1	Das einzelne Blatt	59
3.6.2	Alternative Propellerblätter	60
3.6.3	Das duplizierte Blatt - TINAR	60
3.6.4	Daten des entfalteten Doppelblattes	65
3.6.5	Das dreiblättrige Propellerkonstrukt - TriNAR	66
3.6.6	Das vierblättrige Propellerkonstrukt - FoNAR	68
3.6.7	Das fünfblättrige Propellerkonstrukt - FiNAR	72
3.6.8	Zusammenfassung der biochemischen Charakterisierung der Konstrukte	75
3.7	Mischungsexperimente	75
3.7.1	Inkorporation eines einzelnen Propellerblattes in einen bestehenden Propeller	75
3.7.2	Kofaltung verschiedener Propellerkonstrukte	76
3.7.3	Diskussion	78
3.8	Zusammenfassung	78
4	Struktur eines 14-blättrigen β-Propellers	81
4.1	Einleitung	81
4.1.1	Cyanobakterien	82
4.2	Kristallstrukturanalyse von Proteinen	83
4.3	Material und Methoden	85
4.3.1	Systematische bioinformatische Analyse der Nostoc-Genome	85
4.3.2	Klonierung und Reinigung der β -Propeller aus Nostoc . .	85
4.3.3	Kristallisation des 14-blättrigen β -Propellers	88
4.4	Ergebnisse und Diskussion	88
4.4.1	Repetitive Propeller in Nostoc sp.	88
4.4.2	Kristallisation und Datengewinnung	92
4.4.3	Lösung des Phasenproblems und Strukturlösung	92
4.5	Strukturbeschreibung	94
4.5.1	Überlegungen zur maximalen Anzahl von Propellerblät- tern in einer Domäne	98
4.5.2	Evolutionäre Implikationen der Struktur	100
4.6	Zusammenfassung	102

<i>INHALTSVERZEICHNIS</i>	9
5 Zusammenfassung und Ausblick	103
6 Abkürzungen und Erklärungen	107
Literaturverzeichnis	109
Mitwirkende und Danksagung	119

Kapitel 1

Einleitung

1.1 Proteinevolution

Diese Arbeit versucht einen Beitrag zu der grundlegenden Fragestellung zu leisten: Wie konnte Leben auf unserem Planeten entstehen? Das Leben in seiner heutigen Form ist ohne Proteine nicht denkbar¹. Im Rahmen dieser Arbeit werden Untersuchungen vorgestellt, die eine Möglichkeit beleuchten, wie Proteine entstanden sein könnten.

Proteine sind an einer Reihe von lebenswichtigen Prozessen beteiligt. So ist die enzymatische Katalyse für viele Stoffwechselprozesse notwendig, um Reaktionen im normalen Temperatur- und Druckbereich mit ausreichender Geschwindigkeit ablaufen zu lassen. Strukturelle Funktionen erfüllen Proteine vor allem in größeren Komplexen, sei es als reine Protein-Protein-Interaktionen wie im Zytoskelett oder durch Interaktion mit RNA im Ribosom.

Die meisten Proteine müssen gefaltet sein, um ihre Funktion erfüllen zu können. Die Proteinfaltung wird dadurch erreicht, dass sich die lineare Kette der Aminosäuren durch hydrophobe und elektrostatische Wechselwirkungen zu einer definierten Struktur zusammenlagert. Erst die richtig gefaltete Proteinstruktur bildet Bindungstaschen für Substrate aus und positioniert die Seitenketten von Aminosäuren im aktiven Zentrum so, dass eine Reaktion katalysiert werden kann. Dabei fällt auf, dass Proteine lange Primärsequenzen benötigen um richtig zu falten, so sind funktionelle Proteine in heute lebenden Organismen oft mehrere hundert Aminosäuren lang. Die Größe und Komplexität der Faltung machen es schwer vorstellbar, dass die Vorläufer der heutigen Proteine unter den Bedingungen der frühen Lebensentstehung spontan zusammengefügt wurden. Daraus ergibt sich die evolutionäre Frage, ob kleinere Einheiten als das komplette Protein denkbar sind.

1.1.1 Proteindomänen

Analysiert man Proteine in den Datenbanken, so erkennt man, dass die meisten großen Proteine (solche mit mehr als 200 Aminosäuren bzw. einem Molekulargewicht von über 20 kDa) aus unabhängig faltenden globulären Einheiten

¹Selbstverständlich ist das heutige Leben ohne D/RNA, Lipide und Kohlenhydrate genauso unmöglich, diese Stoffgruppen werden aber nicht weiter behandelt.



Abbildung 1.1: Die strukturelle Wiederholungseinheit des β -Propellers ist das einzelne Blatt, hier rot gefärbt. Für die in dieser Arbeit beschriebenen β -Propeller gilt: der ganze Propeller ist eine Domäne, ein einzelnes Blatt wird synonym als Wiederholungseinheit, Propellerfragment oder Baustein bezeichnet.

aufgebaut sind. Proteine, die aus solchen Domänen bestehen, lassen sich demnach in kleinere Struktureinheiten zerteilen. Dabei ist es von entscheidender Bedeutung, dass jede Domäne für sich alleine falten kann und stabil ist, eine weitere Zerteilung einer Domäne ist in der Regel nicht ohne weiteres möglich.

Insgesamt sind ca. 1600 unterschiedliche Domänen² beschrieben (SCOP Version 1.71 [5]). Interessanterweise reichen jedoch nur 200 Domänen aus, um mehr als 50% der vorhergesagten Gene in den bisher sequenzierten Genomen zu beschreiben [75]. Diese 200 Domänen treten in allen drei Ästen der lebenden Organismen auf: Bacteria, Archaea und Eukarya; es kann daher angenommen werden, dass diese Proteindomänen in der Evolution bereits vor der Aufspaltung in die drei Äste existierten. Mit der Zunahme der Komplexität der Organismen sind auch die Proteine komplexer und damit in erster Linie größer geworden, so dass in den Eukarya bis zu 90% der Gene für Multidomänenproteine codieren. Bakterielle und archaeale Proteine bestehen dagegen häufig nur aus einer Domäne [75].

Proteindomänen werden von der Natur nach einem Baukastenprinzip eingesetzt, die gleiche Domäne kann in verschiedenen Proteinen in verschiedenen Anordnungen mit anderen Proteindomänen auftreten [75]. Das meist ganze Domänen dupliziert und neu kombiniert werden macht deutlich, dass die Domäne die evolutionäre Einheit ist, die zur Proteinentstehung beiträgt. Dabei ist jedoch ungeklärt, wie die ursprünglichen Proteindomänen in der Evolution entstehen konnten, denn allein die Größe dieser Faltungseinheiten macht es unwahrscheinlich, dass Proteindomänen als Ganzes spontan entstanden sind.

In einigen Fällen kann jedoch ein Aufbau aus kleineren Einheiten zu einer symmetrischen Domäne postuliert werden. Auch wenn die heutigen Proteine nicht mehr aus identischen Wiederholungen solcher subdomänen-großen Fragmente aufgebaut sind, so ist doch eine Vorstufe denkbar, in der sich mehrere Kopien einer solchen Einheit zusammenlagern und erst im Zusammenschluss eine stabile Faltungseinheit bilden. Eine Proteinfamilie, für die ein solches Szenario denkbar ist, sind die β -Propeller, wie in Abb. 1.1 dargestellt. Ein Ziel der vorliegenden Arbeit war es experimentell zu untersuchen, ob sich am Beispiel der β -Propeller

²In der SCOP-Datenbank werden Domänen als *Folds* bezeichnet.

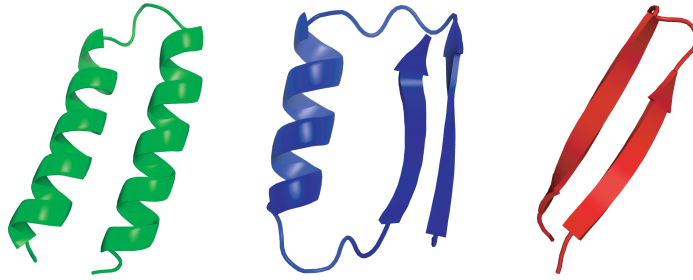


Abbildung 1.2: Beispiele für die drei wichtigsten Supersekundärstrukturelemente: in grün eine α -Haarnadel (α - α), in blau ein β - α - β -Element und in rot eine β -Haarnadel (β - β).

einzelne subdomänen-große Proteineinheiten durch spontane Selbstassemblierung zu einer kompakten Domäne zusammenlagern können.

1.1.2 Evolution der Domänen aus Peptiden

Die Möglichkeit, dass im Laufe der Evolution Proteindomänen aus einfachen Strukturen aufgebaut wurden, ist von zentralem Stellenwert für diese Arbeit. Daher soll hier die Theorie der ancestralen Peptide genauer vorgestellt werden [90, 61].

Grundsätzlich kann spekuliert werden, ob neben den β -Propellern nicht die Mehrzahl der Proteine in den heute lebenden Organismen aus einer kleinen Anzahl von Sekundärstrukturelementen aufgebaut wurde. Diese Theorie wird durch die Beobachtung unterstützt, dass die Strukturen in der Protein Data Bank (PDB) [11] nicht gleichmäßig auf die unterschiedlichen Faltungen verteilt sind. Vielmehr lassen sich die Proteinfaltungen in drei Gruppen unterteilen: (a) eine Vielzahl von Proteinen nehmen eine einmalige Struktur an, sogenannten Unifolds, von denen etwa 10.000 erwartet werden, (b) 80% der Proteine falten in eine von ca. 400 Strukturen (die Mesofolds), von denen die meisten bereits bekannt sind und (c) die 10 häufigsten Faltungen, die sogenannten Superfolds reichen aus, um etwa 25% der heute bekannten Proteinstrukturen zu beschreiben [22]. Es gibt demnach eine kleine Anzahl an Faltungen, die in der Natur sehr häufig verwendet werden (Meso- und Superfolds). Diese vielfach auftretenden Faltungen können aus wenigen Supersekundärstrukturelementen aufgebaut werden. Es wird angenommen, dass die drei in Abb. 1.2 dargestellten Einheiten leicht falten können und lokal kompakte, stabile Strukturen bilden [82].

Betrachtet man eine Vielzahl an unterschiedlichen Proteinen, so fällt auf, dass Wiederholungen ein häufig vorkommendes Phänomen sind. Die Einheiten, die wiederholt werden, sind dabei in ihrer Größe sehr variabel. Fibrilläre Proteine werden aus Abschnitten aufgebaut, die nur wenige Aminosäuren lang sind (Collagen, Coiled Coils oder β -Helices), solenoide Domänen werden aus einfachen Supersekundärstrukturelementen aufgebaut ($\alpha\alpha$ -Haarnadeln: Tetra-ricopeptide und HEAT Repeats; $\beta\beta$ -Haarnadeln: Cholin-bindende Domänen; $\beta\alpha$ -Elemente: Leucinreiche Wiederholungen) und globuläre Domänen werden

aus häufig verschachtelten komplexen Supersekundärstrukturelementen gebildet ($\beta\beta\alpha\beta$: Cradle-Loop-Barrel; $\beta\alpha\beta$: Ferredoxine). So werden auch die Superfolds, die keine interne Symmetrie besitzen zu über 70% aus Supersekundärstrukturelementen wie in Abb. 1.2 gebildet [82].

Davon ausgehend, dass eine geringe Anzahl von einfachen Strukturelementen ausreichend ist, um eine Vielzahl an komplexen Proteinen zu bilden, lässt sich ein kohärentes Konzept entwickeln, dass die Evolution der modernen Proteine erklärt. Dieses System macht eine Reihe von grundlegenden Aussagen:

- Die kombinatorische Komplexität verbietet es, Proteindomänen spontan entstehen zu lassen. So gibt es für eine Domäne von 100 Aminosäuren Länge theoretisch 20^{100} mögliche Sequenzen, von denen nur ein Bruchteil zu falten in der Lage sein wird. Für Einheiten von ca. 20 Aminosäuren Länge reduziert sich dieses Problem drastisch.
- Des Weiteren ist kein primitives (proteinfreies) System bekannt, das längere Aminosäureketten produzieren könnte.
- Die Verwendung von Modulen ist deutlich effizienter, um größere Strukturen zu schaffen als jede Proteindomäne vollständig neu zu generieren.

Die RNA-Welt

Wie bereits angeklungen, müssen auch die Supersekundärstrukturelemente erst einmal synthetisiert werden. Die ersten Peptide entstanden vermutlich in der sogenannten RNA-Welt, die aus der Ursuppe [65] bzw. den präbiotischen Reaktionen der frühen Erdgeschichte hervorgegangen ist, und wird als essentieller Schritt in der Entstehung des Lebens angesehen [49]. Für die RNA-Welt geht man davon aus, dass es vor dem DNA- und proteinbasierten Leben rudimentäre Lebensformen gab, die auf RNA basierten. Dabei könnte die RNA sowohl als Informationsträger aber auch als Katalysator einfacher Reaktionen gedient haben. Eventuell sehen wir heute noch Überreste dieser Welt, denn grundlegende Prozesse in der Zelle sind nach wie vor ohne RNA-abhängige Katalyse nicht denkbar: z. B. mRNA Prozessierung, tRNA Modifikationen und die Proteinsynthese, die doppelt RNA-abhängig ist, die mRNA als Informationsträger und das Ribosom, das in seiner zentralen Funktion ein Ribozym ist.

Die RNA wurde letztendlich schrittweise von Proteinen verdrängt, da diese chemisch stabiler sind und ein breiteres Spektrum an Reaktionen katalysieren können³. Dabei waren die ersten Peptide vermutlich sehr klein und nicht in der Lage selbstständig zu falten, sie dienten nur als Kofaktoren für die Ribozyme. Sobald jedoch die RNA-Maschinerie so weit entwickelt war, dass größere Peptide synthetisiert werden konnten und diese eigenständige Funktionen übernehmen konnten, begann die Verdrängung der RNA.

Ein Evolutionsszenario für die heutigen Proteine könnte daher so aussehen: In der Periode bevor Proteine wichtige Funktionen des Lebens katalysiert haben, wurden von den Ribozymen der RNA-Welt kurze Peptide synthetisiert, von denen sich die stabilsten, die Supersekundärstrukturelemente, zu größeren Komplexen zusammenlagerten und sich daraus ein Stabilitätsgewinn und neue Funktionalitäten ergeben haben. Diese zunächst nur zusammengelagerten

³So können Ribozyme ohne Zuhilfenahme von Kofaktoren keine Redoxreaktionen katalysieren [96].

Peptide fusionierten zu einfachen Proteindomänen. Aus diesen ersten Proteindomänen ist dann die gesamte moderne Proteinwelt entstanden.

1.1.3 Das Konzept der Homologie

Zwei oder mehr Proteine werden als homolog bezeichnet, wenn sie auf einen gemeinsamen Vorfahren zurückgeführt werden können. Im Gegensatz dazu werden Proteine mit ähnlicher Funktion aber ohne gemeinsame Abstammung als analog bezeichnet. Wenn zwei Proteine in demselben Organismus heutzutage verschiedene Aufgaben erfüllen, sich aber auf ein gemeinsames Vorläuferprotein (welches verdoppelt wurde) zurückführen lassen, so werden diese Proteine als paralog bezeichnet. Mit anderen Worten sind Paraloge also homologe Proteine, die im selben Organismus auftreten. Paraloge Proteine mit unterschiedlichen Funktionen entstehen durch die Duplikation ihres Gens, woraufhin eine Kopie nicht mehr dem selektiven Druck unterliegt, die ursprüngliche Funktion in der Zelle erfüllen zu müssen und daher relativ frei evolvieren kann und eventuell eine neue Aufgabe annimmt. Homologe in verschiedenen Organismen werden als ortholog bezeichnet, sie wurden also durch die Abspaltung einer neuen Spezies von einander getrennt, erfüllen aber typischer Weise nach wie vor dieselbe Funktion. Eine Sonderstellung nehmen xenologe Proteine ein, sie stammen ursprünglich von einem anderen Organismus, wurden aber durch einen lateralen Gentransfer und nicht durch Abspaltung von dem Ursprungsorganismus gewonnen.

In einigen Fällen haben sich ähnliche Strukturen oder Funktionen von unterschiedlichen, unabhängigen Vorfahren entwickelt, ein Prozess, der als konvergente Evolution bezeichnet wird⁴. So wurde für die weit verbreiteten Serinproteasen nachgewiesen, dass sie mindestens 16 mal unabhängig von einander entstanden sind [81]. Die Evidenz dafür ist, dass obwohl alle Serinproteasen die Seitenketten der katalytischen Triade (Ser-His-Asp) in derselben dreidimensionalen Anordnung positionieren, sich die Gesamtstrukturen erheblich unterscheiden. Zum Teil werden sie sogar in unterschiedliche SCOP-Folds klassifiziert. Für Proteine mit ähnlicher Funktion oder Struktur ist es also wichtig zwischen homologer und analoger Abstammung zu unterscheiden. Dabei sei festgehalten, dass Homologie ein absoluter Begriff ist: zwei Proteine sind entweder homolog oder nicht.

Um zwischen Homologie und Analogie zu unterscheiden, werden eine Reihe von Argumenten angeführt. Da es jedoch keine fossilen Überreste von Proteinen gibt, bleibt ein endgültiger Beweis eines homologen Ursprungs im Fall der Proteine aus. Jedoch erlauben Gemeinsamkeiten in Funktion, Struktur und Sequenz häufig sehr überzeugende Argumentationen [70, 23, 7].

Da eine gemeinsame Funktion auch analogen Ursprungs sein kann (wie gerade beschrieben), wird dieser Punkt meist nur als zusätzlicher Indikator angeführt.

Gemeinsamkeiten in der Struktur können dagegen sehr gute Indikatoren für einen gemeinsamen Ursprung sein, da im Laufe der Evolution lokale Strukturmerkmale viel stärker konserviert sind als Aminosäuresequenzen [16, 39]. Dass Proteinstrukturen sich viel langsamer verändern als Aminosäuresequenzen liegt daran, dass jedes Protein in der Zelle (mindestens) eine Funktion zu erfüllen hat und diese Funktion untrennbar mit seiner dreidimensionalen Struktur verknüpft

⁴Ein bekanntes Beispiel für konvergente Evolution sind die Flügel von Fledermäusen und Vögeln, diese sind trotz ihres vergleichbaren Aufbaus aus verschiedenen Vorläufern entstanden [33].

die Sequenzen von SUMO und Ubiquitin haben dagegen nur 13% Übereinstimmung, sie scheinen zunächst nicht miteinander verwandt zu sein. Ubiquitin wird in der Zelle verwendet um Proteine, die für den Abbau bestimmt sind, kovalent zu markieren. Da dieser Prozess streng reguliert werden muss, interagiert Ubiquitin mit einer ganzen Reihe von weiteren Proteinen. Dieses Interaktionsnetzwerk bestand bereits im letzten gemeinsamen Vorfahren zwischen Mensch und Hefe. Im weiteren Verlauf der Evolution konnten die Ubiquitine sich nicht verändern ohne dass die Funktion zerstört worden wäre. So lässt sich erklären, wie ferne Verwandte wie Mensch und Hefe nahezu identische Proteine verwenden. SUMO dagegen ist auch am Proteinabbau beteiligt, jedoch sind die beteiligten Proteinkomplexe andere, so dass sich die Sequenz von SUMO verändern konnte, ohne die Funktion zu beeinträchtigen. Die homologe Abstammung zu Ubiquitin ist aber nach wie vor in der Funktion und an der Struktur ersichtlich. Somit können strukturelle Ähnlichkeiten entfernte Homologien aufdecken [43]. Andererseits können solche Übereinstimmungen auch analogen Ursprungs sein und sich darin begründen, dass es im Vergleich zum Sequenzraum nur sehr beschränkte Möglichkeiten an Faltungen gibt. Somit ist strukturelle Übereinstimmung eine sehr sensitive Methode um Homologe zu finden, als alleiniges Argument jedoch nicht ausreichend.

Das dritte Argument für eine homologe Abstammung ist Sequenzähnlichkeit. Proteinesequenzen ergeben sich theoretisch aus der beliebigen Kombination von allen 20 proteinogenen Aminosäuren zu Ketten einer bestimmten Länge. Für ein Dipeptid aus zwei Aminosäuren ergeben sich bereits 400 Möglichkeiten und entsprechend für Proteine astronomisch viele Kombinationsmöglichkeiten⁶. Aus rein statistischen Überlegungen heraus ist es daher sehr unwahrscheinlich, dass zwei Proteine durch konvergente Evolution die gleiche Sequenz entwickeln. Im Umkehrschluss kann man Sequenzähnlichkeiten verwenden, um statistisch abgesichert für Proteine einen homologen Ursprung zu postulieren.

Verschiedene Methoden verwenden Sequenzvergleiche, um homologe Proteine in Datenbanken zu suchen, die bekanntesten sind sicherlich BLAST und PSI-BLAST [1, 2]. In neuerer Zeit sind deutlich empfindlichere Werkzeuge entwickelt worden, wie z. B. HHpred [88, 89]. Diese vergleichen nicht mehr einfache Sequenzen miteinander, sondern erweitern die Informationen, die in einer Sequenz enthalten sind, indem sie auch andere nah verwandte Proteinsequenzen mit berücksichtigen und die gesammelten Informationen in Sequenzprofilen oder Hidden Markov Modellen verarbeiten, die dann miteinander verglichen werden.

Sequenzähnlichkeiten sind statistisch das robusteste Argument für Homologie. Jedoch sind sequenzbasierte Methoden relativ unempfindlich, da Proteinsequenzen im Laufe der Zeit sehr starken Veränderungen durch Mutationen unterworfen sind - so können zwei homologe Sequenzen so stark verändert sein, dass sie nicht mehr als solche erkannt werden. In einem solchen Fall lassen sich konservierte Strukturelemente verwenden, um auch entfernte Verwandte zu ermitteln. Erst die Kombination aller drei Argumente (Funktion, Struktur und Sequenz) ermöglicht es, sicher von Homologie zu sprechen.

⁶Die Gesamtheit der möglichen Aminosäurekombinationen wird als der Sequenzraum bezeichnet.

1.1.4 Mechanismen der Proteinevolution

Proteine verändern sich im Laufe der Evolution. An dieser Stelle sollen Mechanismen beschrieben werden, die bei der Evolution von Proteinen eine Rolle spielen. Die Proteine, die wir heute sehen, sind nur die Produkte der erfolgreichen Veränderungen der ursprünglichen Proteine, da sich nur Proteine durchsetzen, die korrekt falten und eine gewinnbringende Funktion für die Zelle erfüllen. Die Häufigkeiten, mit der die unten beschriebenen Mechanismen auftreten, werden sich stark von den Frequenzen unterscheiden, mit denen sich die modifizierten Proteine durchsetzen.

Alle im Folgenden beschriebenen Veränderungen finden in der DNA statt, dem genetischen Material, das an die Folgegenerationen weitergegeben wird. Da die DNA für die Proteine einer Zelle codiert, haben Veränderungen im codierenden Bereich der DNA unmittelbar veränderte Proteine zur Folge.

Mutationen, Insertionen und Deletionen

Gendrift und natürliche Selektion verändern die Proteine in jedem Replikationszyklus. Aus der Anzahl der Veränderungen kann man in etwa den Zeitraum abschätzen, seit dem sich zwei Proteine auseinander entwickelt haben [24]. Dabei stellt man fest, dass Mutationen, Insertionen und Deletionen die häufigsten Mechanismen sind, die zu Veränderungen in der Proteinevolution führen. Aminosäuresubstitutionen treten in etwa eine Größenordnung häufiger auf als Insertionen und Deletionen [10, 77]. Des Weiteren ist es von Bedeutung, wo im Protein Veränderungen auftreten. So sind der Umgebung zugewandte *Loops* deutlich toleranter für Variationen als der Kern des Proteins. Werden katalytisch aktive Reste oder Teile einer Bindungstasche verändert, so können auch kleine Mutationen erhebliche Effekte haben.

Duplikationen

Duplikationen sind eine entscheidende Triebkraft in der Evolution [6]. Dabei finden Duplikationen auf verschiedenen Ebenen statt. Es können einzelne Gene oder nur Teile von einem Gen verdoppelt werden, aber auch Gengruppen in Operons, sogar Genomduplikationen haben stattgefunden [97]. Der große Vorteil bei Duplikationen ist, dass eine Genkopie nur noch einem reduzierten selektiven Druck unterliegt. Am Beispiel von Ubiquitin wurde in Abschnitt 1.1.3 beschrieben, dass die enge Einbindung in das Regulationsnetzwerk des Proteinabbaus Ubiquitin daran gehindert hat, sich seit der Aufspaltung von Hefe und Mensch zu verändern. Wird jedoch ein Gen dupliziert, so kann eine Kopie nach wie vor die ursprüngliche Funktion erfüllen, während die zweite evolvieren und neue Funktionen annehmen kann. Alternativ können sich die beiden neuen Genprodukte auch die ancestralen Aufgaben teilen und jeweils die Ausführung verfeinern oder erweitern. Eine Genduplikation muss aber nicht in jedem Fall erfolgreich sein. So wurde beschrieben, dass für die Mehrheit der duplizierten Gene die Stilllegung einer Kopie während der darauffolgenden ca. 4 Millionen Jahre die häufigste Folge einer Genduplikation ist [62].

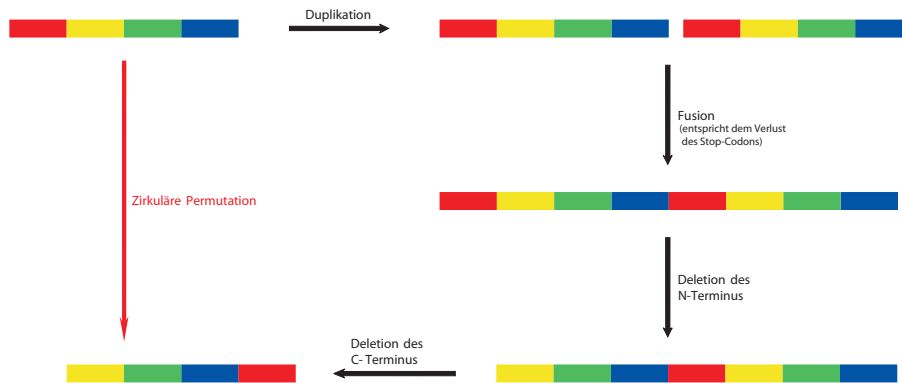


Abbildung 1.4: Schritte, die zu einer zirkulären Permutation in einem Protein führen. Alle beschriebenen Schritte finden auf der DNA-Ebene statt.

Zirkuläre Permutationen

Eine zirkuläre Permutation stellt eine Veränderung in der Topologie eines Proteins dar, bei der ein Teil des N-Terminus abgetrennt und an den C-Terminus angefügt wird [100]. Zirkuläre Permutationen in β -Propellern sind dafür verantwortlich, dass die Wiederholungseinheit auf Sequenzebene gegen die strukturelle Einheit, das Propellerblatt verschoben zu sein scheint. Ein Beispiel ist die WD40-Einheit, die ursprünglich so beschrieben wurde, dass sie zum N-Terminus durch eine Gly-His (GH)-Abfolge und am C-Terminus durch das namensgebende Trp-Asp (WD) begrenzt wird, wobei die Wiederholungseinheit eine Frequenz von ca. 40 Aminosäuren Länge aufweist [72]. In Abb. 1.7 ist jedoch ersichtlich, dass das Propellerblatt um genau einen β -Strang gegen die Sequenzwiederholung verschoben ist.

Damit ein Protein zirkulär permutiert werden kann, muss eine Reihe von Schritten stattfinden, wie in Abb. 1.4 dargestellt. Der codierende Abschnitt eines Gens muss in unmittelbarer Nachbarschaft zu sich selber dupliziert werden und die beiden Kopien müssen unter Erhalt des Leserasters fusionieren, so dass das Ursprungprotein als Homodimer in einem Strang vorliegt. Wenn jetzt der N-Terminus des Dimerproteins deletiert wird, kann seine Funktion von dem N-Terminus der zweiten Kopie übernommen werden. Findet nun eine zweite Deletion am N-Terminus statt, so resultiert wieder das ursprüngliche Protein. Verliert jedoch das verkürzte Homodimerprotein an seinem C-Terminus etwas, so entsteht ein zirkulär permutiertes Protein.

Zirkuläre Permutationen werden dadurch begünstigt, dass in globulären Proteinen die N- und C-Termini häufig nahe beisammen liegen [95]. So werden durch zirkuläre Permutationen die räumlichen Anordnungen und Seitenketteninteraktionen im Protein nicht verändert, jedoch ergeben sich in der Topologie des Proteins neue Verbindungen. Eine systematische Analyse der Proteine in der PDB hat ergeben, dass mindestens 400 von 3000 unsymmetrischen Domänen (bei der angewendeten Methode erscheinen alle symmetrischen Proteine als potentiell zirkulär permutiert) im Laufe der Evolution zirkulär permutiert wurden [50]. Diese Beobachtung unterstreicht noch einmal die Bedeutung der zirkulären Permutation und die Häufigkeit ihres Auftretens.

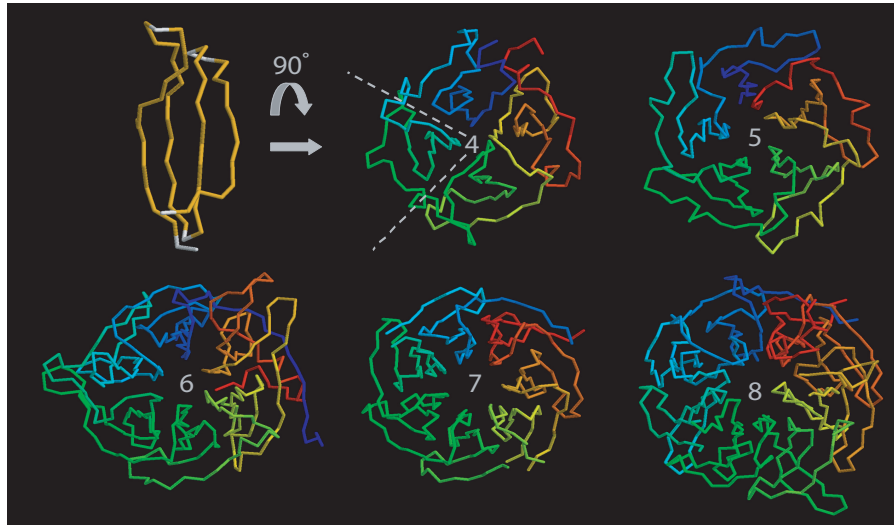


Abbildung 1.5: β -Propeller Proteine sind vielfältig in ihrer Struktur, in der Strukturdatenbank sind Beispiele mit vier bis acht Propellerblättern hinterlegt. Die Proteine sind in allen Fällen aus Wiederholungen eines einzelnen Blattes aufgebaut.

1.2 β -Propeller Proteine

Diese Arbeit behandelt Proteine, die zur Gruppe der β -Propeller gehören. Diese Gruppe ist ausschließlich über strukturelle Gemeinsamkeiten definiert: alle β -Propeller werden aus repetitiven Einheiten aufgebaut, den Propellerblättern, die sich zu einer zirkulären kompakten Proteinstruktur zusammenlagern (s. Abb. 1.1). In der PDB finden sich Proteine, in denen vier bis acht Propellerblätter in einer durchgehenden Aminosäurekette auftreten und zu einer Domäne falten. Die SCOP-Datenbank (Structural Classification of Proteins) [5] teilt die β -Propeller entsprechend der Anzahl ihrer Blätter in fünf *Folds* ein, dies entspricht der Einteilung in fünf Superfamilien in der alternativen Datenbank CATH (Class (C), Architecture (A), Topology (T) und Homologous superfamily (H)) [37]. Anfang 2007 waren in diesen Datenbanken rund 100 Strukturen als Propeller klassifiziert⁷.

In Abb. 1.5 sind Beispiele für Propeller aus den verschiedenen SCOP-Folds gezeigt. Deutlich zu sehen ist der repetitive Aufbau aus vier bis acht Propellerblättern. Die Gesamtstruktur der Propeller ist dabei jedoch gleich: Die Proteine sind diskusförmig mit einer zentralen Pore, eine Geometrie, die als Torus bezeichnet wird. Die Pore bildet eine trichterförmige Öffnung, in die auf der engen Seite häufig Ionen oder Substrate gebunden sind.

⁷Dabei sind in der SCOP-Datenbank immer etwas weniger Einträge als in CATH, da die SCOP-Einteilung manuell für jedes Protein einzeln vorgenommen wird, während CATH automatisch klassifiziert. Die 100 Einträge beziehen sich auf die SCOP Version 1.71, die auf 95% Sequenzidentität gefiltert wurde, um Proteine, die unter verschiedenen Bedingungen kristallisiert wurden, nicht mehrfach zu zählen (die gesamte SCOP 1.71 Datenbank umfasst etwas mehr als 13.000 Strukturen in der gefilterten Version).

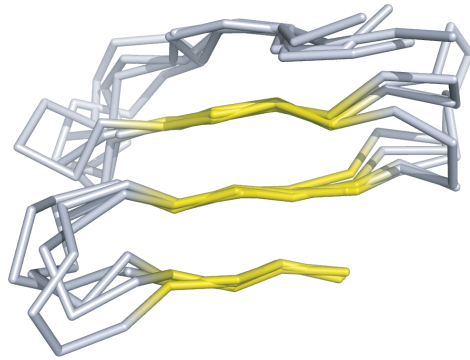


Abbildung 1.6: Überlagerung von Propellerblättern von einem 4-, 5-, 6-, 7- und einem 8-blättrigen β -Propeller. Die inneren drei Stränge A bis C überlagern mit einem RMSD von weniger als 1 Å, unabhängig davon, welche Größe der ursprüngliche Propeller hatte.

Jedes Propellerblatt ist aus einem viersträngigen, antiparallelen β -Faltblatt aufgebaut. Dabei ist jedes Blatt so beschrieben, dass die N-terminale Aminosäure des ersten β -Strangs die Oberseite des Proteins definiert. Von hieraus zieht das Proteinrückgrat mäanderförmig - entsprechend dem antiparallelen Verlauf des Faltblattes - nach außen. Die β -Stränge werden von innen nach außen mit A bis D bezeichnet. β -Faltblätter sind nie flach, sondern weisen alle eine rechtsgängige Drehung auf, wodurch der äußere Strang D nahezu rechtwinklig zu Strang A steht.

Die Überlagerung der Propellerblätter von Propellern unterschiedlicher Größe zeigt, dass die inneren drei β -Stränge aus allen Proteinen nahezu identisch zueinander angeordnet sind, unabhängig davon aus wie vielen Blättern der ursprüngliche Propeller aufgebaut ist (Abb. 1.6). Die strukturellen Einheiten, aus denen die unterschiedlichen Propeller aufgebaut sind, sind also nahezu gleich, sie unterscheiden sich nur im Drehwinkel zwischen Strang A und D.

In vielen β -Propellern ist ein Strang des letzten Blattes zirkulär permutiert und befindet sich am N-Terminus des Proteins. Somit liegen die N- und C-terminalen β -Stränge des Proteins in einem Blatt in direkter räumlicher Nachbarschaft zueinander und sind über Wasserstoffbrückenbindungen miteinander verbunden. Ohne die zirkuläre Permutation befänden sie sich in zwei getrennten Propellerblättern und die Termini wären nicht weiter miteinander verknüpft. Dieser molekulare 'Klettverschluss' (*Molecular Velcro*) dient vermutlich der zusätzlichen Stabilisierung der Proteine. Wenige Propeller haben zusätzlich eine N-terminale Erweiterung, die einen fünften Strang an das C-terminale Blatt anfügt [41]. Dadurch sind die N- und C-Termini über eine Länge von zwei β -Strängen mit Wasserstoffbrückenbindungen verknüpft, ein doppelter Klettverschluss.

Die Permutation von einem C-terminalen β -Strang an den N-Terminus ist die häufigste Variante. Es gibt jedoch auch die Möglichkeit, dass der Verschluss zur Hälfte aus N- und zur anderen Hälfte aus C-terminalen β -Strängen gebildet wird (Bsp. RCC1) [79]. Im Fall von Tachylectin sind sogar drei Stränge

permutiert, dies ist aber eher die Ausnahme [9]. Ebenso treten Proteine ohne *Molecular Velcro* auf, so zum Beispiel Hemopexin, ein vierblättriger Propeller, bei dem der Ringschluss kovalent über eine Disulfidbrücke zwischen N- und C-Terminus erfolgt [36]. Ein weiterer Sonderfall ist die Prolyl Oligopeptidase, die keinerlei zusätzliche Verknüpfung zwischen N- und C-Terminus hat [31]. Dieser Propeller dient als Selektivitätsfilter für eine Proteasedomäne, die an die β -Propellerdomäne angeschlossen ist. In diesem Fall wird angenommen, dass der Propeller durch eine kurzfristige Öffnung seiner zentralen Pore Substraten den Zugang zum aktiven Zentrum der Protease ermöglicht.

1.2.1 Sequenzdiversität in der Gruppe der β -Propeller

Im Gegensatz zu den strukturellen Ähnlichkeiten zwischen den verschiedenen Propellern ist der Sequenzraum der β -Propeller erstaunlich groß. In der Literatur treten Propeller mehrfach auf, dabei wurden meist Familien mit spezifischen Sequenzmustern beschrieben (s. Abb. 1.7):

WD40. Diese Gruppe von 7-blättrigen Propellern wurde als erstes beschrieben und stellt heute die größte beschriebene Gruppe dar, sie wurde zuerst in der β -Untereinheit der heterotrimeren G-Proteine entdeckt [72, 87]. Der Name leitet sich von dem markanten Trp-Asp (WD) Motiv ab, welches mit einer Periode von 40 Aminosäuren auftritt. Der Name sollte Assoziationen zu einem bekannten amerikanischen Öl (WD-40) herstellen, da eine funktionelle Analogie vermutet wurde. In den ersten Veröffentlichungen wurde vermutet, dass das WD Motiv am C-Terminus der Wiederholungseinheit auftritt. In der ersten Kristallstruktur eines WD40 Proteins [101] wurde jedoch deutlich, dass sich das WD Motiv am Ende von β -Strang C befindet. Diese Verschiebung von Sequenzeinheit gegen Struktureinheit lässt sich durch eine zirkuläre Permutation erklären (s. auch 1.1.4 und 1.2).

WD40 Proteine sind an einer Vielzahl von zellulären Prozessen beteiligt, wie z. B. Signaltransduktion, Regulation der Transkription und Apoptose [59]. Für diese Proteinfamilie wurde bisher keine allumfassende Funktion gefunden, jedoch scheinen die meisten Mitglieder an Protein-Protein-Interaktionen beteiligt zu sein.

RCC1. Der Regulator of Chromosome Condensation 1 (RCC1) und verwandte Proteine sind an Transportvorgängen zwischen Nukleus und Zytoplasma und an der Zellzykluskontrolle beteiligt [79]. Wie die WD40 Proteine bilden sie 7-blättrige Propeller mit einer zirkulären Permutation, die zu einem *Velcro* führt. Jedoch sind die Verbindungsschleifen (*Loops*) zwischen den β -Strängen vergrößert, so dass die Wiederholungseinheit 11 Reste länger ist. Am auffallendsten ist ein Asp-Gly (DG) Motiv am Anfang von β -Strang C.

KELCH. Diese Sequenzeinheit wurde nach dem Kelch-Protein aus *Drosophila* benannt, in dem es zum ersten Mal gefunden wurde. Proteine dieser Gruppe falten in 6- und 7-blättrige Propeller. In einigen Fällen wurde beschrieben, dass KELCH-Proteine G_{β} -Untereinheiten von ihrer Bindung an G_{α} verdrängen konnten [34]. Wie die WD40 Proteine enthalten die KELCH-Proteine entsprechendes Tyr-Asp (YD) Motiv am Ende von β -Strang C.

YWTD. Diese Wiederholungseinheit wurde in Proteinen der extrazellulären Matrix und in extrazellulären Domänen von Rezeptoren gefunden [12, 92],

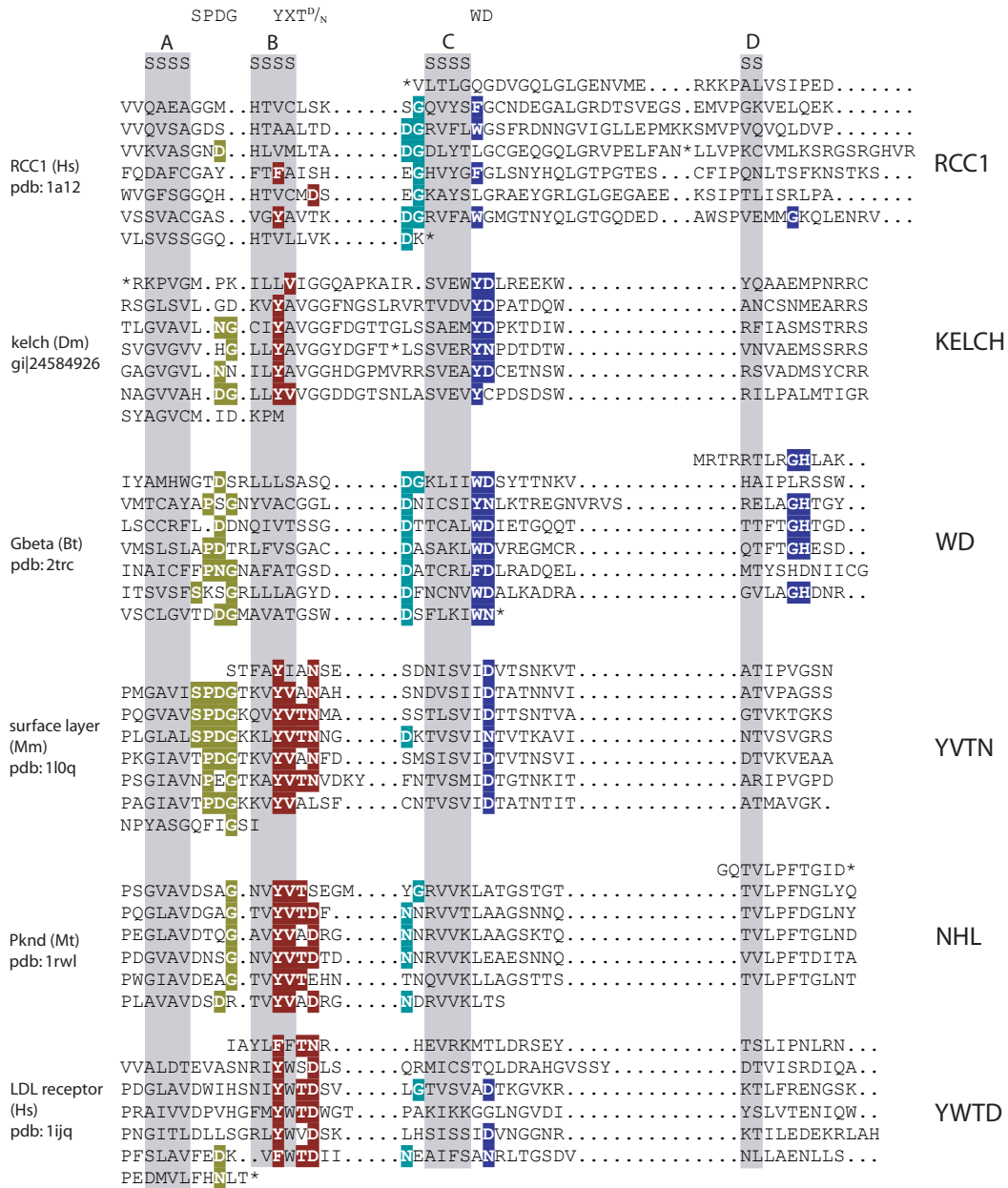


Abbildung 1.7: Strukturbasiertes Sequenzalignment von β -Propeller Proteinen. (Hs: *Homo sapiens*; Dm: *Drosophila melanogaster*; Bt: *Bos taurus*; Mm: *Methanosarcina mazei*; Mt: *Mycobacterium tuberculosis*)

ein Beispiel ist der LDL Rezeptor [46]. Der Name stammt von dem charakteristischen Sequenzmuster Tyr-Trp-Thr-Asp (YWTD), welches das Ende von β -Strang B bildet.

NHL. Diese Proteinfamilie wurde nach ihren ersten drei Mitgliedern benannt, in denen Sequenzhomologie gefunden wurde: NCL-1, HT2A und LIN-41 [86]. Das beschreibende YVTD-Motiv der NHL-Gruppe ist eine Variante des YWTD, welches auch am Ende von Strang B liegt und in etwa zur selben Zeit beschrieben wurde.

YVTN. Diese Wiederholungseinheit wurde als Variante zu YWTD beschrieben, die in Oberflächenproteinen von Archaeen und Metazoen auftritt, allerdings steht die YVTN-Gruppe in näherer Verwandtschaft zu NHL [26]. Die Struktur von einem Oberflächenprotein aus *Methanosarcina*, die stellvertretend für die YVTN-Gruppe ist, zeigt ein ungewöhnlich hohes Maß an struktureller Symmetrie [47]. Zusätzlich zu dem namensgebenden Motiv beinhaltet die Wiederholungseinheit ein Ser-Pro-Asp-Gly (SPDG) Motiv in dem *Loop* zwischen Strang A und B und ein Ile-Asp (ID) Motiv am Ende von Strand C an der äquivalenten Position wie die WD40 Proteine das WD Motiv und die KELCH-Proteine das YD Motiv haben. Keines dieser Motive tritt in NHL Proteinen auf.

Die meisten der oben aufgelisteten Familien wurden unabhängig voneinander beschrieben. Als nach und nach eine strukturelle Ähnlichkeit zu anderen β -Propellern festgestellt wurde, wurde auf Grund der geringen Sequenzübereinstimmungen dies als Ergebnis von konvergenter Evolution angesehen [34, 4]. Bemerkenswert ist, dass für YVTN, NHL und YWTD ein homologer Ursprung vermutet wurde [47]. Uns ist aufgefallen, dass in iterativen Sequenzprofilsuchen Proteine aus verschiedenen Propellerfamilien miteinander in Verbindung gebracht werden können und dass einige Motive, die charakteristisch für eine Familie sind, in anderen Familien -zum Teil in leicht veränderter Form- auftreten. Beide Beobachtungen sind in der Regel Indikatoren für homologe Abstammung. Wir haben daher den Aspekt der homologen im Gegensatz zur analogen Abstammung aller β -Propeller genauer untersucht. Dabei haben wir mit hochentwickelten bioinformatischen Methoden herausgefunden, dass viele, wenn nicht alle β -Propeller höchst wahrscheinlich von einem gemeinsamen Vorfahren abstammen. Der zugrundeliegende evolutionäre Mechanismus scheint die Amplifikation eines einzelnen β -Propellerblattes und anschließende Differenzierung zu sein.

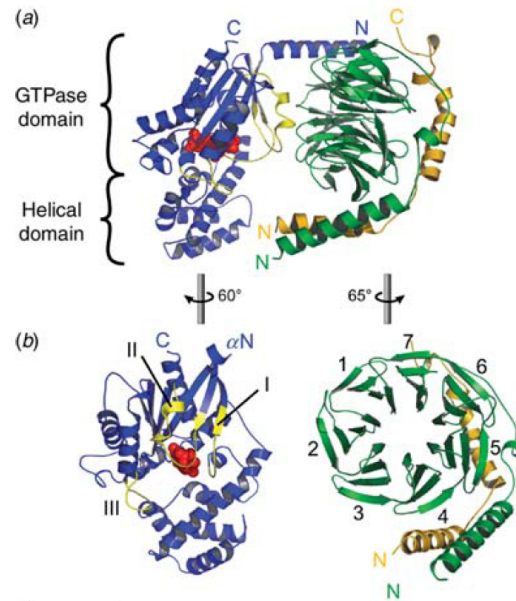


Abbildung 1.8: Kristallstruktur des heterotrimeren G-Proteins (1got). (a) Darstellung des $G_\alpha(\text{GDP})G_{\beta\gamma}$ -Heterotrimeren, die α -Untereinheit ist in blau, β in grün und γ in gold gezeigt. Die Switch Regionen in G_α sind gelb hervorgehoben, das GDP ist in rot gezeigt. (b) Die Untereinheiten wie in (a) jedoch rotiert, um die Interaktionsfläche zwischen den Proteinen zu zeigen. Die Interaktion findet über die αN -Helix und Switch II auf der Seite von G_α und die Blätter 1 bis 3 auf der Seite von G_β statt. nach [73]

1.2.2 Biologische Funktionen von β -Propellern

Ähnlich ihrem großen Sequenzraum können β -Propeller eine ganze Reihe von Funktionen erfüllen. Die bisher beschriebenen Funktionen sind so divers, dass sich keine Gemeinsamkeiten feststellen lassen.

Die am weitesten verbreitete Funktion von β -Propellern ist die Bindung von Substraten, wobei die Bandbreite an Molekülen, die gebunden werden können von Proteinen über Kofaktoren wie Pyrrolochinolin-Chinon (PQQ) und Hämgruppen bis zu Zuckerresten reicht [101, 67, 76].

Am Beispiel der heterotrimeren G-Proteine (s. Abb. 1.8) lässt sich die Funktion der β -Propeller als Vermittler von Protein-Protein-Interaktionen beschreiben. Zudem ist die β -Untereinheit der G-Proteine von zentraler Bedeutung für diese Arbeit.

Proteine aus der Gruppe der heterotrimeren Guaninnukleotid-bindenden Proteine (oder GTP-bindende Proteine), kurz G-Proteine, fungieren als molekulare Schalter im Signaltransduktionsweg. Sie verbinden die aktivierten Rezeptoren⁸ in der Zellmembran/auf der Zelloberfläche mit den intrazellulären Effektoren, den Second Messengern. Nur das Heterotrimer aus α -, β - und γ -

⁸Die Rezeptoren haben charakteristischer Weise sieben Transmembranhelices und werden auf Englisch als G-protein-coupled receptors, kurz GPCR bezeichnet.

Untereinheit bindet an die G-Protein gekoppelten Rezeptoren (GPCR), die für die Signalaufnahme verantwortlich sind. Dabei bilden die G_β - und G_γ Untereinheiten eine feste funktionale Einheit, die sich nur trennen lässt, wenn man die Proteine denaturiert [73].

Die beiden publizierten Kristallstrukturen mit allen drei Untereinheiten [55, 101] zeigen Interaktionen zwischen den G_α - und den $G_{\beta\gamma}$ -Proteinen, wobei keine direkten Kontakte zwischen der α - und der γ -Untereinheit festzustellen sind. Der Vergleich mit den Strukturen der einzelnen Proteine macht deutlich, dass durch die Komplexbildung die α -Untereinheit nur in den Schalterabschnitten Switch I und II verändert wird, während in $G_{\beta\gamma}$ keine signifikanten Strukturveränderungen sichtbar sind [73].

Die Betrachtung der Genanzahl im Humangenom, die für die Proteine der G-Protein vermittelten Signalweiterleitung verantwortlich sind, lässt interessante Schlüsse zu. So werden 865 Genen im Humangenom eine Funktion als G-Protein gekoppelter Rezeptor zugeschrieben, es ist also eine Vielzahl an unterschiedlichen Proteinen für die Signalaufnahme nötig [66]. Im Gegensatz dazu wurden nur 16 G_α -codierende Gene, 5 Gene für G_β und 12 bzw. 14 Gene für G_γ gefunden⁹. Obwohl die meisten β -Einheiten mit den meisten γ -Einheiten interagieren können, bilden sich nicht alle denkbaren Kombinationen. Da zusätzlich verschiedene $G_{\beta\gamma}$ -Proteine mit demselben G_α -Protein einen Komplex bilden können, scheint die Komplexität der Signalvermittlung im Cytosol nicht über verschiedene Proteine, wohl aber durch Variationen in den Komplexzusammensetzungen erreicht zu werden.

Nach der Aktivierung des GPCR dissoziiert das heterotrimere G-Protein in die α -Untereinheit, die ins Cytosol diffundiert und den $\beta\gamma$ -Komplex, der weiterhin membrangebunden bleibt. Beide Komponenten interagieren unabhängig voneinander mit Effektoren in der Signalkaskade. Der Mechanismus, mit dem die $G_{\beta\gamma}$ -Untereinheit mit den nachgeschalteten Proteinen wechselwirkt, ist nicht vollständig geklärt. In vielen aber nicht allen Fällen enthalten die Effektorproteine eine sogenannte PH Domäne¹⁰. Allerdings interagieren nicht alle Proteine mit einer PH Domäne mit $G_{\beta\gamma}$, was eine funktionelle Vorhersage schwierig macht [64].

Einige Propellerdomänen sind nur ein Teil von Multidomänenproteinen. Dabei werden der Propeller und die weitere(n) Domäne(n) als Teil einer durchgehenden Polypeptidkette exprimiert. Teilweise übernimmt der Propeller nur unterstützende Aufgaben. Das aktive Zentrum ist jedoch in der zweiten Domäne lokalisiert, wie im Beispiel der Prolyl Oligopeptidase [31]. In diesem Protein bildet der C-terminale Teil die Proteasedomäne, der β -Propeller fungiert nur als Filter und ist für die Substratspezifität verantwortlich.

Wieder andere Propeller fungieren in der Zelle als Enzyme. Dabei werden häufig kofaktorabhängige Redoxreaktionen oder hydrolytische Spaltungen katalysiert; als Beispiele für Redoxenzyme seien Cytochrom cd_1 Nitrit Reduktase [32] oder die Methanoldehydrogenase [103] genannt, während Phytase [84, 41]

⁹In der Anzahl der α - und β -Untereinheiten herrscht Übereinstimmung in der Literatur, lediglich in der Anzahl der Gene die für die γ -Untereinheit codieren unterscheiden sich [64] und [66].

¹⁰Die PH Domäne (der Name leitet sich von Pleckstrin Homology ab) lässt sich in vielen eukaryotischen Signalkaskadepoteinen finden. Ihr werden unterschiedliche Funktionen zugeschrieben, unter anderem die Fähigkeit Inositolphosphat und phosphorylierte Ser/Thr-Reste zu binden.

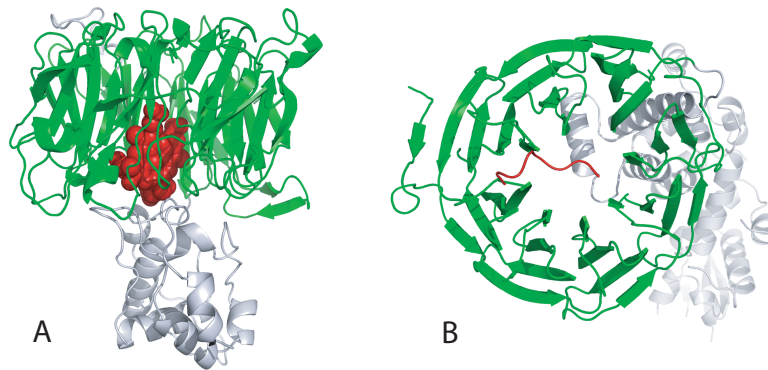
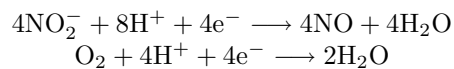


Abbildung 1.9: (A) Kristallstruktur der Cytochrom cd_1 Nitrit Reduktase (1qks), die α -helikale c Domäne ist in grau gezeigt, das katalytische Häm in rot. (B) Kristallstruktur des Skp1-Cdc4-Komplexes (1nex). In grün ist die achtblättrige Propellerdomäne von Cdc4 gezeigt, das in der zentralen Pore gebundene Phosphopeptid in rot.

oder Phosphotriesterase [83] Beispiele für Hydrolasen sind.

Die Cytochrom cd_1 Nitrit Reduktase ist in der Reihe der β -Propeller mit enzymatischer Funktion hervorzuheben, da sie als eine evolutionäre Zwischenstufe angesehen werden kann. So ist die sauerstoffabhängige Atmung für viele Eukaryoten von essentieller Bedeutung, während Prokaryoten häufig alternative Elektronenakzeptoren verwenden können. Castresana und Kollegen haben postuliert, dass die Cytochrom cd_1 Nitrit Reduktase eine Zwischenstellung zwischen den beiden Atmungsformen darstellen könnte [15]. Die Argumentation ist, dass dieses Enzym Elektronen sowohl auf Nitrit (einen der möglichen ancestralen Elektronenakzeptoren) als auch auf molekularen Sauerstoff übertragen kann.



Die Reduktion sowohl von Nitrit als auch von O_2 finden an einer Hämgruppe statt, die in der zentralen Pore eines 8-blättrigen β -Propellers gebunden ist [32] (s. Abb. 1.9).

1.2.3 Biologische Verbreitung der β -Propeller Proteine

β -Propeller haben nicht nur ein breites Spektrum an biologischen Funktionen und einen großen Sequenzraum, sie treten auch in allen drei Ästen der lebenden Organismen auf: Bacteria (z. B. TolB in *Escherichia coli*), Archaea (z. B. Tricorn Protease in *Thermoplasma acidophilum*) und Eukarya (z. B. G_β im *Homo sapiens*). Dabei ist ungeklärt, ob die β -Propeller älter sind als der letzte gemeinsame Vorfahre der drei Hauptdomänen der lebenden Organismen, denn nur bei den Eukaryoten finden sich in der überwiegenden Mehrzahl der Spezies homologe β -Propeller (das heterotrimere G-Protein). In den Bakterien dagegen gibt es kein β -Propellerprotein oder ein verwandtes Protein, das in allen Spezies auftritt. Daher scheint ein lateraler Gentransfer von den Eukaryoten in die

Bakterien eine plausible Erklärung zu sein. Derselbe Mechanismus wird auch für den bekanntesten β -Propeller in den Archaea (Tricorn Protease) angeführt, das Protein scheint einen bakteriellen Ursprung zu haben [80].

Kapitel 2

Homologie der β -Propeller

2.1 Sequenzdiversität bei β -Propellern

In der Einleitung wurde im Abschnitt 1.2.1 die Sequenzdiversität in der Gruppe der β -Propeller beschrieben. Dabei wurde betont, dass trotz ihrer ähnlichen Strukturen die β -Propeller eine erstaunliche Vielfalt an Sequenzen aufweisen, für die sich auch kein übergreifender Konsensus finden lässt. Nicht eine Sequenzposition ist über alle Propellerproteine konserviert. Dennoch lassen sich in einzelnen Familien der β -Propeller spezifische Sequenzwiederholungen feststellen. Eine Reihe solcher Familien ist in der Literatur nach und nach beschrieben worden, dabei wurde zumindest implizit von konvergenter Evolution ausgegangen. Uns ist jedoch aufgefallen, dass in iterativen Sequenzprofilsuchen Proteine aus verschiedenen Propellerfamilien miteinander in Verbindung gebracht werden können und dass einige Motive, die charakteristisch für eine Familie sind auch in anderen Familien -zum Teil in leicht veränderter Form- auftreten. Beide Beobachtungen sind in der Regel Indikatoren für homologe Abstammung. Wir haben daher den Aspekt der homologen im Gegensatz zur analogen Abstammung aller β -Propeller genauer untersucht. Wir sind dabei auf eine Reihe von sehr aussagekräftigen Hinweisen gekommen, die für einen homologen Ursprung aller β -Propeller sprechen. Das Vorgehen und das Ergebnis dieser Arbeit soll in diesem Kapitel beschrieben werden.

2.2 Bioinformatische Methoden

2.2.1 Strukturelle Überlagerung von β -Propellerblättern

Um strukturelle Überlagerungen zu konstruieren, wurden die Proteine in ihre einzelnen Blätter zerteilt und diese interaktiv mit dem Swiss PDB Viewer [40] überlagert. Die für die Überlagerung berücksichtigten Reste entsprechen den β -Strängen A, B und C, die in Abb. 2.2 grau unterlegt sind. Für die Überlagerung in der linken Hälfte von Abb. 2.1 wurde jeweils ein Blatt von einem Propeller mit 4, 5, 6, 7 und 8 Blättern (genaue Bezeichnung der PDBs siehe Abbildungsbeschriftung) so gewählt, dass die *Loops* nicht zu groß sind und die Sicht auf den zentralen β -Mäander versperren. Für den rechten Teil der Abb. 2.1 wurden die beiden Blätter so ausgewählt, dass sie möglichst viele der Sequenzmotive

aus Abb. 2.2 enthalten. Für Abb. 2.2 haben wir die Proteine verwendet, von denen ausgehend die größten Propellerfamilien ursprünglich definiert wurden. Für fünf dieser Proteine ist die Struktur bekannt und eins (Drosophila Kelch) hat ein nah verwandtes Homologes mit bekannter Struktur. Wir überlagerten jeweils alle Blätter eines Proteins, um ein strukturelles *Alignment* für das jeweilige Protein zu erzeugen. Danach haben wir ein Blatt von jedem Propeller verwendet, um ein zweites strukturelles *Alignment* zu generieren, was dann verwendet wurde, um die *Alignments* der einzelnen Proteine auf einander abzustimmen und so ein Gesamtalignment für alle Blätter der ausgewählten Strukturen zu erzeugen. Dabei wurde darauf geachtet, dass die β -Stränge der Propeller sauber übereinstimmen.

2.2.2 Clusterkarte der All- β -Proteine

Wir haben die SCOP-Datenbank [5] in der Version 1.71 verwendet und auf 25% Sequenzidentität gefiltert (SCOP25)¹. Aus diesem Datensatz haben wir die Proteine extrahiert, die als All- β^2 klassifiziert wurden (diese sind in der SCOP-Nomenklatur mit b.xx bezeichnet). Für die HHpred *Alignments* und Hidden Markov Modell (HMM) Datenbanken wurde eine modifizierte Version des PSI-BLAST Algorithmus [2] verwendet³. Buildali.pl verhindert das Hinzufügen von nicht homologen Sequenzabschnitten an den Enden der PSI-BLAST High-Scoring Segment Pairs. BLAST versucht Sequenzabschnitte in der Such- und Zielsequenz zu alignieren und diese zu bewerten. Die am besten übereinstimmenden Abschnitte werden dann erweitert - diese werden als High-Scoring Segment Pairs bezeichnet. Für die All- β -Domänen haben wir mit HHsearch [88] (Version 1.5.0) einen Vergleich aller Proteine gegen alle anderen durchgeführt. HHsearch ist ein sehr sensibles Suchprogramm um entfernte Homologien aufzuspüren, es arbeitet auf der Basis von HMM-HMM-Vergleichen. Wichtig ist, dass in der Version 1.5.0 eine Korrekturmethode für Verzerrungen der Aminosäurezusammensetzung angewandt wird, die es verhindert, dass Proteinalignments nur deshalb gut bewertet werden, weil zwei Proteine eine gleiche Aminosäurekomposition aufweisen.

Wir haben die Proteine auf Grund ihrer paarweisen P-Werte mit CLANS [29] geclustert. CLANS ist eine Anwendung des Fruchterman-Reigold-Algorithmus, welche die log-P-Werte in Anziehungskräfte umwandelt und in einem Kraftfeld darstellt. Für die Darstellung im Kraftfeld wurden paarweise Verbindungen bis zu einem P-Wert von e^{-3} verwendet, die CLANS-Parameter waren folgende: Anziehung und Abstoßung = 10 und der Anziehungsexponent = 2. Die Karte wurde dreidimensional bis zur Konvergenz geclustert. Die Reproduzierbarkeit wurde durch mehrere unabhängige Läufe von zufälligen Ausgangspositionen beginnend sichergestellt.

¹Diese Datenbank kann vom HHpred Webserver [89] am Max Planck Institut für Entwicklungsbiologie heruntergeladen werden.

²All- β wird in der SCOP-Datenbank verwendet um die Klasse an Proteinen zu beschreiben, die ausschließlich β -Stränge als Sekundärstrukturelemente besitzen. All- β wird im Folgenden als Eigenname verwendet.

³Das Programm buildali.pl wurde von J. Söding entwickelt und freundlicher Weise zur Verfügung gestellt.

2.2.3 Matrix mit β -Propellerverbindungen, basierend auf HHsearch

Wir haben alle β -Propeller aus der SCOP25 extrahiert (SCOP-Nomenklatur: *Folds* b.66 bis b.70, insgesamt 60 Propellerproteine) und mit jedem auf der HHpred Webservice Seite (<http://toolkit.tuebingen.mpg.de/hhpred>) gegen die HMM-Datenbank SCOP70 1.71 gesucht. Die Struktur 1n7v wurde nicht berücksichtigt, da sie nur einen partiellen Propeller enthält und keinerlei Sequenzähnlichkeit zu einem anderen Propeller aufweist und zudem noch in SCOP als separater *Fold* (b.126) eingestuft ist. Um Zufallstreffer zu vermeiden, die nur auf Grund von übereinstimmenden Sekundärstrukturelementen zustande kommen könnten, haben wir die HHpred-Option 'Sekundärstruktur bewerten' ausgeschaltet. Die Ergebnisse dieser Untersuchung werden im Text weiter unten diskutiert und sind in Abb. 2.4 zusammengefasst. Wie für paarweise Vergleiche erwartet werden kann, ist die Matrix nahezu symmetrisch, die leichten Abweichungen bei den Wahrscheinlichkeiten (im Gegensatz zu den reinen Punktwertungen) werden hauptsächlich durch die Kalibrierung einzelner HMMs verursacht.

2.2.4 Phylogenie von β -Propellern

Wir haben einen phylogenetisch repräsentativen Satz an β -Untereinheiten von heterotrimeren G-Proteinen und fünf Propeller von *Nostoc* PCC 73102⁴ verwendet. Diese Proteine wurden in ihre einzelnen Blätter zerteilt und diese wiederum einzeln in ClustalX [94] aligniert. Dieses *Alignment* wurde manuell nachbearbeitet, anschließend wurde mit den Standardeinstellungen von ClustalX ein distanzbasierter *Neighbour-joining* Stammbaum errechnet und mit SplitsTree [44] visualisiert - die Ergebnisse sind in Abb. 2.5 gezeigt.

2.3 Ergebnisse/Diskussion Homologie-Untersuchungen

Da es nur eine geringe Zahl an strukturellen Lösungen für eine gefaltete Polypeptidkette gibt, tendieren auch nicht verwandte Sequenzen dazu lokal ähnliche Faltungen anzunehmen. Diese werden als Supersekundärstrukturen bezeichnet, wobei die wichtigsten $\alpha\alpha$ - und $\beta\beta$ -Haarnadeln sowie $\beta\alpha\beta$ -Elemente sind [75] - diese sind auch Abb. 1.2 gezeigt und dort näher beschrieben. Im Gegensatz dazu ist der kombinatorische Sequenzraum gigantisch und eine bestimmte lokale Struktur kann durch verschiedene Sequenzen realisiert werden. Daher ist konvergente Evolution auf der Sequenzebene sehr unwahrscheinlich. Aus diesem Grund wird Sequenzähnlichkeit als primäres Kennzeichen für Homologie angesehen. Um einen homologen Ursprung für die β -Propeller zu evaluieren, haben wir uns daher primär auf Sequenzanalysen konzentriert.

2.3.1 Ein Kontinuum von Sequenzmotiven

Unser Ausgangspunkt war ein Vergleich der Sequenzmotive in verschiedenen β -Propellerfamilien, deren räumliche Anordnung im Propellerblatt in Abb. 2.1 B

⁴Dieses Cyanobakterium ist uns aufgefallen, da sein Genom für sehr viele hochrepetitive WD40 Proteine codiert. So ist auch der kristallisierte Propeller (s. Kapitel 4) aus diesem Organismus.

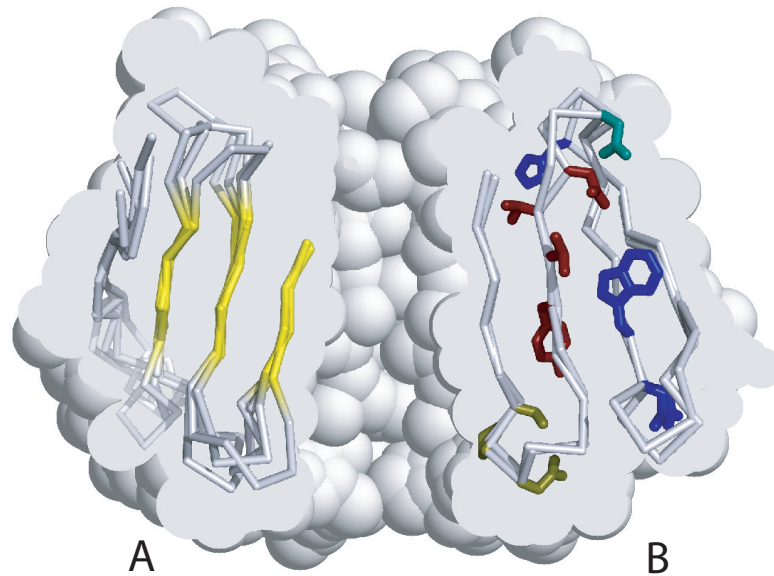


Abbildung 2.1: Strukturelle Eigenschaften von β -Propellern. (A) Überlagerung einzelner Blätter von 4-, 5-, 6-, 7- und 8-blättrigen β -Propellern (Blatt 1 von 1hxn, Blatt 2 von 1gyd, Blatt 4 von 1ijq, Blatt 4 von 2trc und Blatt 7 von 1n90). Die drei inneren β -Stränge überlagern sehr gut, unabhängig von der Größe des Ursprungspropellers. Die aufgeschnittene Kugeldarstellung eines Propellers im Hintergrund soll die Lage der Blätter in der Gesamtstruktur verdeutlichen. (B) Die räumliche Darstellung von charakteristischen Sequenzmotiven: SPDG am Ende von β -Strang A (oliv), YVTN ist der hydrophobe Kern von Strang B (rot), DG ist im Loop zwischen den Strängen B und C (grün), WD am Ende von Strang C und GH am Ende von Strang D (blau). Eine Einordnung der Motive auf Sequenzebene ist in Abb. 2.2 möglich. Da keine Sequenz alle Motive vereinigt, haben wir Blätter von zwei Proteinen überlagert (Blatt 4 von 2trc und Blatt 3 von 1l0q).

gezeigt ist. Dazu haben wir repräsentative Stellvertreter jeder Gruppe ausgewählt und diese Blatt für Blatt in einem multiplen *Alignment* zusammengeführt (s. Abb. 2.2), wo immer möglich wurde das *Alignment* auf Grund von Strukturdaten erstellt. In Abb. 2.2 entsprechen die Stellvertreter den Proteinen, an Hand derer die Familien ursprünglich definiert wurden. Zusätzlich zu den Familien, die in Kapitel 1.2.1 bereits beschrieben wurden, haben wir eine weitere Familie identifiziert, die durch die Proteinkinase Pkwa aus dem Actinobacterium *Thermospora curvata* repräsentiert wird. Wir benennen die neue Gruppe nach den hervorstechenden Resten Ser-Pro-Asp-Gly (SPDG), die die Stränge A und B verbinden. Die SPDG-Proteine sind eine Untergruppe der WD40 Proteine, die allgemein durch gut konservierte Sequenzwiederholungen auffällt. In profilgestützten Sequenzsuchen mit β -Propellern sind die SPDG-Proteine häufig die ersten Proteine, die außerhalb der Familie der Suchsequenz gefunden werden.

Wie in Abb. 2.2 ersichtlich, erstrecken sich die Hauptmotive einer Familie

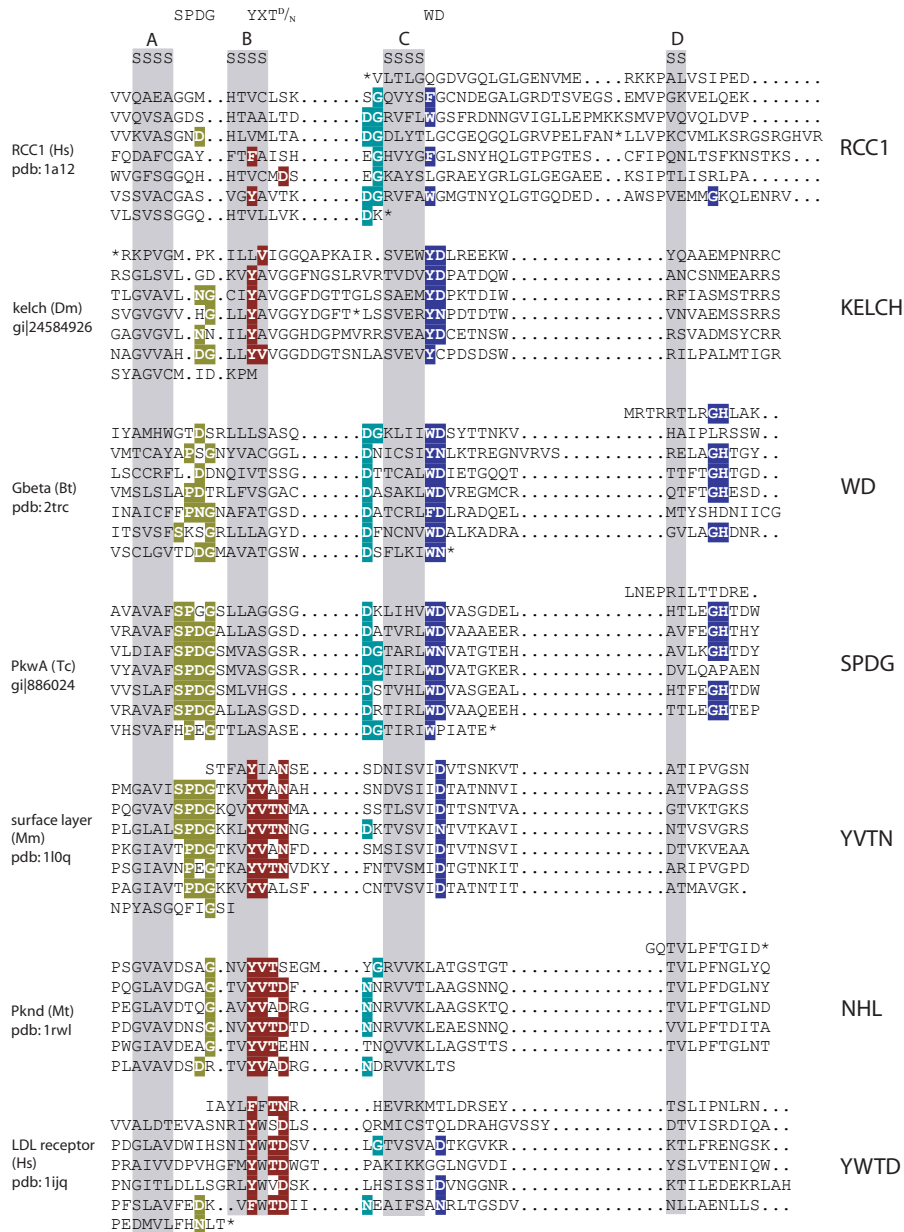


Abbildung 2.2: Strukturbasiertes Sequenzalignment von β -Propeller Proteinen. (Hs: Homo sapiens; Dm: Drosophila melanogaster; Bt: Bos taurus; Tc: Thermomonospora curvata; Mm: Methanosarcina mazei; Mt: Mycobacterium tuberculosis)

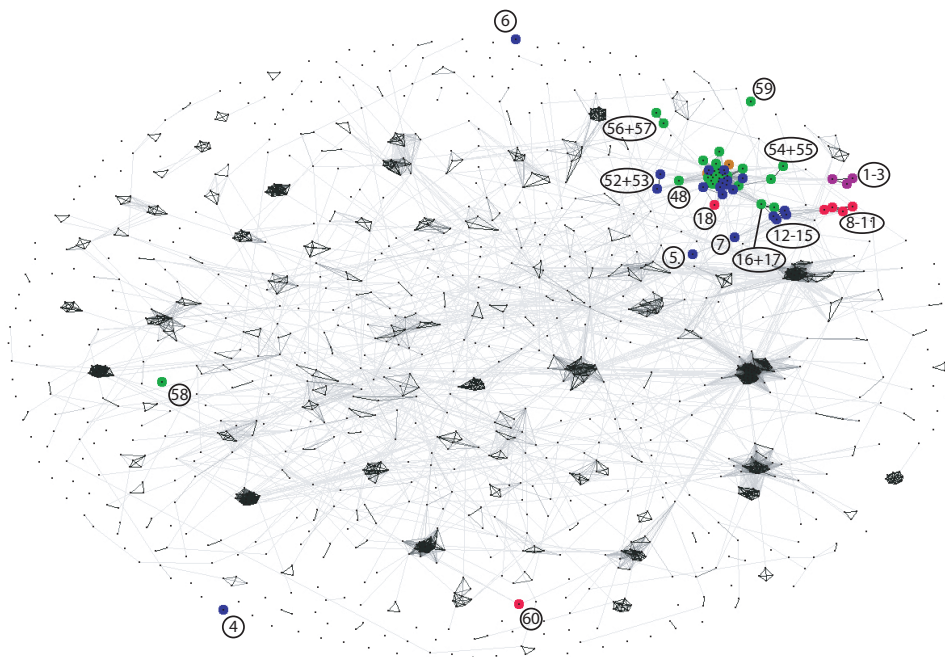


Abbildung 2.3: Clusterkarte von Proteinen, die in der SCOP25 als All- β klassifiziert sind. Die farbigen Punkte markieren die Propeller (violett = 4-blättriger-, rot = 5-blättriger-, blau = 6-blättriger-, grün = 7-blättriger- und braun = 8-blättriger Propeller). Die Farben und die Sequenznummern werden im Rahmen dieser Arbeit beibehalten. Aus Gründen der Übersichtlichkeit sind nur Sequenzen außerhalb des zentralen Clusters beschriftet - Legende zu den Nummern s. Abb. 2.4.

in nahezu allen Fällen über mehrere Proteinfamilien und bilden so einen kontinuierlichen Motivraum innerhalb der β -Propeller. Nur das GH-Motiv am Ende von Strang D scheint spezifisch für die erweiterte WD40 Familie zu sein. Das auffälligste und durchgängigste Motiv besteht aus großen hydrophoben Resten, gefolgt von einem Asp- oder Asn-Rest am Ende von Strang C - dieses Motiv ist zumindest teilweise in den meisten β -Propellern enthalten.

2.3.2 β -Propeller bilden eine Gruppe in einer Clusterkarte der All- β -Proteine

Wir wollten der Frage nachgehen, ob diese Sequenzmotive konvergent entstanden sind und nur die Folge von strukturellen Zwängen von β -Haarnadeln im Allgemeinen und von β -Mäandern (antiparallelen β -Faltblättern) im Besonderen darstellen. Dazu haben wir eine neu entwickelte sensitive Methode um HMMs miteinander zu vergleichen (HHsearch [88]) mit einem Clusteralgorithmus gekoppelt. So haben wir eine Clusterkarte aller All- β -Proteine aus der SCOP25 Datenbank erstellt (s. Abb. 2.3). Diese Clusterkarte wurde ausschließlich auf Grund von Sequenzinformationen erstellt. Die Argumentation bei dieser

Karte ist, dass die All- β -Folds, die zum großen Teil aus $\beta\beta$ -Haarnadeln aufgebaut sind und zahlreiche β -Mäander enthalten, einen ausreichenden Hintergrund an konvergenten Sequenzmotiven enthalten, um Gruppen von konvergenten Propellern auseinanderzuziehen.

Das Ergebnis der Karte ist jedoch, dass alle bis auf vier Propeller in der Karte in einer Region gruppiert sind - unabhängig von ihrer SCOP-Fold-Klassifizierung. Sechs der sieben Proteine aus dem *Alignment* in Abb. 2.2 haben entweder bekannte Strukturen oder wie im Fall von Kelch ein Orthologes mit bekannter Struktur. Von diesen Proteinen clustern fünf in der zentralen Gruppe, während RCC1 (Nr. 54) in der direkten Nachbarschaft erscheint und mit der zentralen Gruppe verbunden ist.

2.3.3 HMM-HMM Vergleiche von β -Propellern

Um die Sequenzbeziehungen zwischen β -Propellern systematisch zu untersuchen, haben wir die 60 Propellerproteine der SCOP25 (Version 1.71, *Folds* b.66 bis b.70) verwendet. Wir haben die HMM-Profile ihrer Sequenzen mit den HMMs der anderen β -Propeller der SCOP70 Datenbank mit Hilfe des HHpred Webservers verglichen. Die Auswertung der einzelnen Suchergebnisse ergab, dass alle Propeller als beste Zielsequenz einen anderen Propeller gefunden haben und dass von den 2597 Sequenzverknüpfungen mit einem E-Wert kleiner als e^{-5} alle mit anderen Propellern waren.

Die besten Verbindungen zu Proteinen mit einem anderen *Fold* bestanden zwischen der *Vibrio cholerae* Sialidase (1w0p) und der Chitobiase aus *Serratia marcescens* (1qba) mit einem E-Wert von $3,9e^{-3}$ und zwischen den Blutegelproteinen Sialidase (2sli) und Chitobiase bei einem E-Wert von $4,4e^{-3}$. Diese Sequenzbeziehung ist bereits beschrieben worden [21, 81] und wurde als mögliches Beispiel für entfernte Homologie zwischen Proteinen aus unterschiedlichen *Folds* diskutiert [21, 61]. Diese Sequenzbeziehung fußt in einer konservierten β -Haarnadel mit einem SxDxGxxW Turn-Motiv, der sogenannten Asp-Box [21], die in den β -Propellern zwischen den Strängen C und D auftritt. Asp-Boxen treten außerdem in immunoglobulin-ähnlichen β -Sandwiches (SCOP b.1.18), mikrobiellen Ribonucleasen (d.1.1), Sialidasen (6-blättrige β -Propeller, b.68) und in Cellobiohydrolasen (7-blättrige β -Propeller, b.69.13) sowie in einer um einen Rest verkürzten Form (SxDxxxW) in der Arabinase/Levansucrase/Invertase-Gruppe der 5-blättrigen β -Propeller (b.67.2) und in kohlenhydratbindenden Domänen (b.18.1) auf. Es ist interessant zu sehen, dass alle Proteine, die eine Asp-Box enthalten, im Zusammenhang mit Kohlenhydraten stehen - eine direkte Partizipation dieses Motivs bei der Kohlenhydratbindung bleibt jedoch unklar.

Die gegenseitigen Beziehungen der Propeller untereinander sind in Abb. 2.4 zusammengefasst. Die gezeigten HHsearch-Wahrscheinlichkeiten entsprechen der Wahrscheinlichkeit, dass zwei Proteine mit demselben E-Wert in SCOP als homolog eingestuft werden (dass sie also in derselben Familie oder Superfamily eingeordnet sind) [88]. Mehr als die Hälfte aller Propeller bildet eine eng verbundene Gruppe (Proteine mit den Nummern 18 bis 51), die durch zahlreiche statistisch hochsignifikante paarweise Verbindungen zusammengehalten wird. Diese zentrale Gruppe ist strukturell und funktionell sehr divers, sie enthält β -Propeller mit sechs, sieben und acht Blättern, deren Funktionen ein breites Spektrum abdecken, peripher ist ein 5-blättriger Propeller, Tachylectin (Nr. 18,

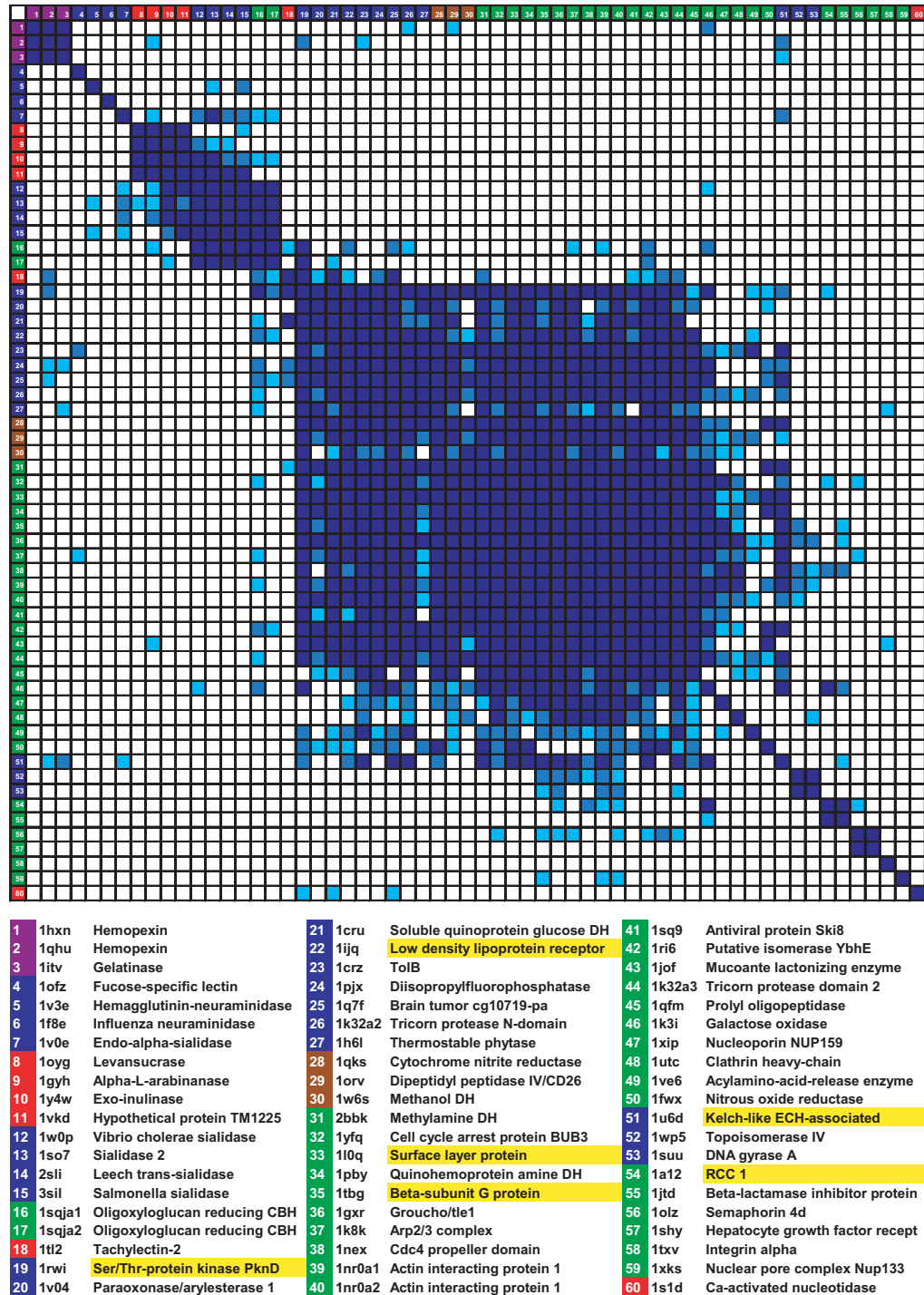


Abbildung 2.4: Paarweise Sequenzvergleiche von β -Propellern in der SCOP25. Die horizontale Achse enthält die Sequenzen, mit denen gesucht wurde. Jede Spalte listet auf, welche anderen Propeller aus dem Datensatz mit HMM-HMM-Vergleichen gefunden werden, dabei entspricht dunkel blau einer Wahrscheinlichkeit von $>90\%$, blau $50-90\%$ und hellblau $20-50\%$ Wahrscheinlichkeit. Dies entspricht den Wahrscheinlichkeiten, dass zwei Proteine mit demselben E-Wert in SCOP als homolog eingestuft sind. Die Farben auf den Achsen entsprechen den Größen der Propeller wie in Abb. 2.3 eingeführt. Die fünf Proteine mit bekannter Struktur aus Abb. 2.2 und das Kelch Protein, dessen humanes Orthologes gelöst wurde, sind in gelb hervorgehoben.

1tl2) der Gruppe zugehörig. Von den sieben Proteinen aus dem *Alignment* in Abb. 2.2 sind die fünf Propeller, die in Abb. 2.3 bereits eng geclustert waren (Nr. 19, 22, 33, 35 und 51), alle in der zentralen Gruppe vertreten.

Die zweitgrößte Gruppe, bestehend aus 5-, 6- und 7-blättrigen Propellern, enthält die oben beschriebenen kohlenhydratmodifizierenden Propeller, bei denen die Asp-Box das dominante Sequenzmerkmal ist (Nr. 7-17). Die Verbindungen in dieser Gruppe werden durch intermediäre Strukturen verstärkt, die sowohl die kanonische als auch die verkürzte Asp-Box enthalten (Salmonella Sialidase, 3sil und Aspergillus inulinase, 1y4w).

Die beiden Hauptgruppen weisen viele Querverbindungen mit $>50\%$ und einige mit $>90\%$ Wahrscheinlichkeit auf und verknüpfen so 45 der 60 β -Propeller. Die Verknüpfung der beiden Gruppen erfolgt in erster Linie über die beiden Propellerdomänen der Cellobiohydrolase (1sqj, Nr. 16+17), die zusätzlich zu einigen Asp-Boxen Motive aufweisen, die vor allem für andere Propellergruppen charakteristisch sind, z. B. ein WD Motiv im ersten Blatt der ersten Propellerdomäne. Das HMM der Cellobiohydrolase enthält noch weitere Verknüpfungssequenzen wie beispielsweise ein Oberflächenprotein aus *Methanoculleus marisnigri* (gi|126179850), welches durch eine hohe interne Sequenzkonservierung hervortritt und kanonische Asp-Boxen mit einem xPDG-Motiv verknüpft, welches genau an der Stelle der SPDG-Box in der erweiterten WD40-Familie auftritt.

Die restlichen 15 Propeller in Abb. 2.4, unter ihnen alle 4-blättrigen Propeller, haben weniger Verbindungen zu den Kernpropellern. Einige jedoch, wie z. B. RCC1 (Nr. 54) können mit einzelnen Verbindungen mit über 90% Wahrscheinlichkeit an die Hauptgruppen angebunden werden. Lediglich drei Proteine haben bei einem Schwellenwert von 10% Wahrscheinlichkeit keine Verbindungen. Es sind dies ein Protein des Kernporenkomplexes und eine Ca-aktivierte Nukleotidase (Nr. 59 bzw. 60) und die Influenza Neuraminidase (1f8e, Nr. 6), von der angenommen werden kann, dass sie ein divergentes Mitglied der Sialidase-Familie ist (SCOP b.68.1) und zu der Gruppe der kohlenhydratmodifizierenden Propeller gehört, auch wenn der Sequenz eine Asp-Box fehlt⁵.

Die Anzahl und die Stärke der Verbindungen zwischen den meisten Propellern in Abb. 2.4 ist ein deutlicher Indikator für Homologie, vergleicht man die Wahrscheinlichkeiten mit der Kalibrierung an Hand der SCOP-Datenbank. Daher ist die Gruppierung der β -Propeller in fünf verschiedene SCOP-Folds (nämlich nach Anzahl der Blätter) nicht konsistent mit dem Grad an Sequenzübereinstimmung, mit dem andere Proteine typischer Weise in einer Superfamily klassifiziert werden. Anscheinend hat im Rahmen der Klassifikation die Anzahl der Blätter gegenüber Sequenzargumenten überwogen und dazu geführt, dass die β -Propeller in verschiedene *Folds* unterteilt wurden. Bei anderen Toroidstrukturen (TIM-Barrels, Glykosidasen) hat eine variable Anzahl an Wiederholungseinheiten einer Gruppierung in einen einzelnen *Fold* nicht entgegengestanden - bei diesen ist die Variabilität aber auch nicht so groß wie bei den Propellern. Wir schlagen daher vor, dass zumindest die 45 gut verknüpften Propeller in Abb. 2.4 als Mitglieder einer Superfamily angesehen werden sollten.

Da das Nichtvorhandensein von Hinweisen kein Hinweis auf Nichtvorhanden-

⁵Virale Proteine sind allgemein für ihre außerordentlich hohen Mutationsraten bekannt. So können Virusproteine einen viel größeren Teil des Sequenzraums, der mit einer bestimmten Faltung kompatibel ist, ausnutzen als Proteine aus zellulären Organismen. Das macht es jedoch sehr schwer, phylogenetische Beziehungen nachzuvollziehen oder Homologien zu entdecken [35].

sein ist, argumentieren wir weiter, dass die schwachen oder fehlenden Verbindungen zu den restlichen 15 Propellern kein direkter Hinweis für einen analogen Ursprung sind. Betrachtet man die wichtige Rolle von verknüpfenden Sequenzen, um Gruppen von Propellern auf der Sequenzebene zu verbinden (wie z. B. Cellobiohydrolase oder Tachylectin), so ist es gut möglich, dass kommende Erweiterungen der Datenbanken solche Verbindungen enthalten werden.

2.4 Evolutionäre Szenarios

Viele globuläre Proteinfamilien sind strukturell repetitiv. In den meisten Fällen sind die Vorgänge, die zu den Wiederholungen führten sehr lange her und fanden noch vor der Verbreitung der Proteinfamilien statt. Daher ist jede Wiederholungseinheit auf der Sequenzebene der entsprechenden Einheit in einem anderen Protein der Familie ähnlicher als den weiteren Einheiten in seinem eigenen Protein. Ein solches Szenario kann auch für einige β -Propeller beobachtet werden. So gruppieren die äquivalenten Blätter der β -Untereinheit der heterotrimeren G-Proteine unter Ausschluss anderer Blätter in phylogenetischen Bäumen zusammen (s. Abb. 2.5A). Allgemein haben die Sequenzen von β -Propellern Eigenschaften, die sie von anderen repetitiven *Folds* unterscheiden:

(I) Ihre Sequenzen umfassen die gesamte Breite an internen Symmetrien, von hoch repetitiv bis vollständig ausdifferenziert. Dies suggeriert, dass die Propellerproteine mehrfach im Laufe der Evolution durch Amplifikation eines einzelnen Blattes entstanden sind. Für die strukturellen Wiederholungseinheiten erwartet man, dass sie sich durch Prozesse wie adaptive Differenzierung oder neutrale Evolution verändern, wenn sie nicht durch ein symmetrisches Substrat oder die Notwendigkeit für eine hohe strukturelle Beständigkeit eingeschränkt sind. Daher kann man allgemein davon ausgehen, dass Propeller mit stark symmetrischen Sequenzen neueren evolutionären Ursprungs sind als voll ausdifferenzierte Propeller. Die Existenz von Propellern mit nahezu identischen Blättern, wie z. B. ORF Npun0227829 aus *Nostoc punctiforme* (gi|23124439), bei dem alle 14 Blätter, die sich auf Sequenzebene vorhersagen lassen, in allen bis auf eine Position identisch sind, zeigt, dass die Evolution von β -Propellern ein fortlaufender Prozess ist. In der Tat ist *Nostoc* ein Organismus mit einer auffallend hohen Anzahl an unlängst amplifizierten Propellern (s. Abb. 2.5B).

(II) Propellerproteine weisen keine anderen Symmetrien auf als die Wiederholung einzelner Blätter. So finden wir z. B. keinen sechsblättrigen Propeller mit einer zwei- oder dreifachen Symmetrie. Wenn ein Propeller Sequenzsymmetrie aufweist, so ist die Wiederholungseinheit immer das einzelne Blatt.

Uns sind keine Fälle bekannt, die einer Situation ähnlich HisA und HisF bei den TIM-Barrels entsprechen, wo eine zweifache Sequenzsymmetrie die grundsätzliche achtfache strukturelle Symmetrie überlagert [42, 56]. Wir schließen daraus, dass Propeller ausschließlich durch die Amplifikation einzelner Blätter und nicht durch Duplikation von Einheiten mit mehr als einem Blatt entstehen.

(III) Die Blätter in einem Propeller haben in der Regel alle dieselben Sequenzmotive. Proteine der WD40-Familie beinhalten den WD40-Konsensus, während YWTD-Propeller den YWTD-Konsensus enthalten, obwohl wir eine geringe Anzahl an Propellerblättern in der Strukturdatenbank beobachtet haben, die näher an einer anderen Proteinfamilie zu sein scheinen (das beste Beispiel ist das bereits erwähnte erste Blatt der Cellobiohydrolase, welches ein WD-Motiv

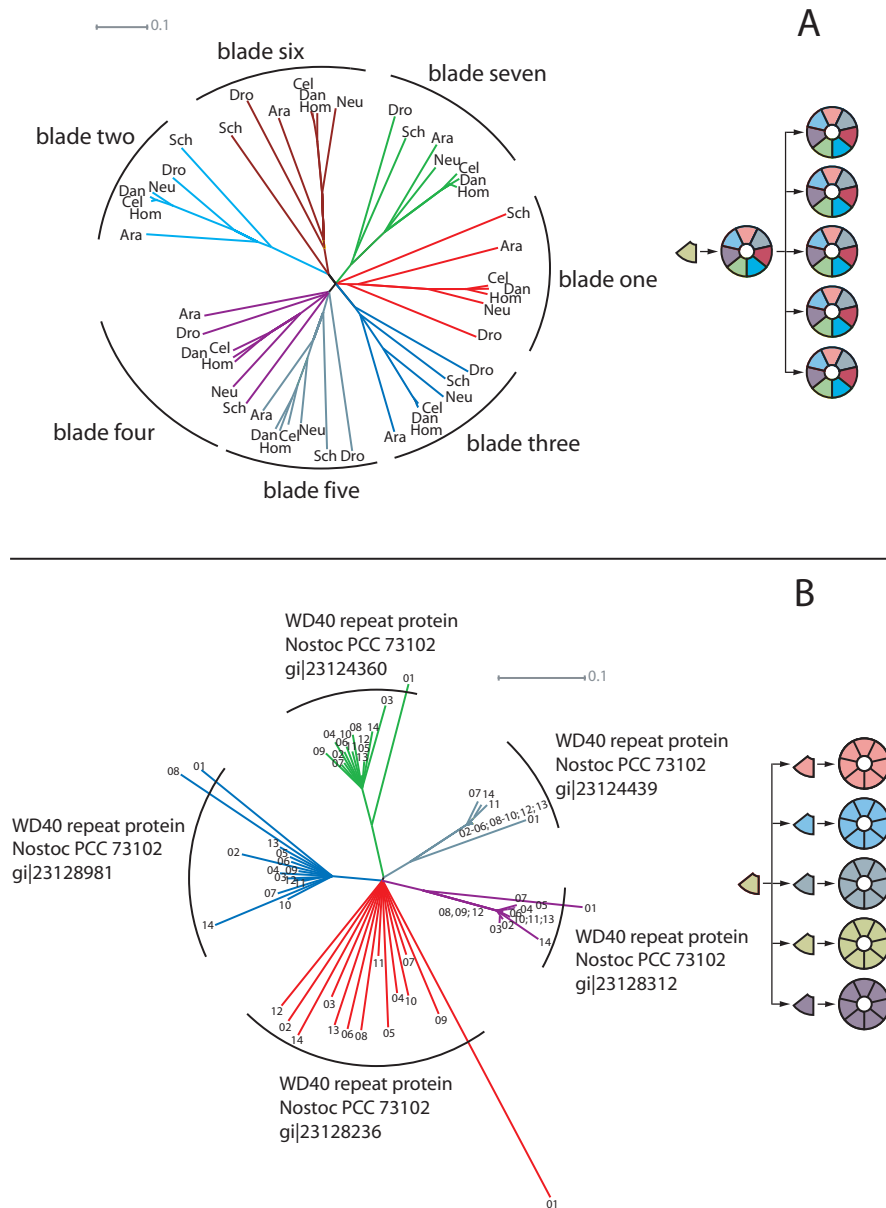


Abbildung 2.5: Zwei evolutionäre Szenarios für β -Propeller als Stammbäume visualisiert (Neighbor-joining Methode). (A) Divergente Evolution eines vollständig differenzierten Propellers, am Beispiel der β -Untereinheit der heterotrimeren G-Proteine (Ara: Arabidopsis thaliana; Cel: Caenorhabditis elegans; Dan: Danio rerio; Dro: Drosophila melanogaster; Hom: Homo sapiens; Neu: Neurospora crassa; Sch: Schizosaccharomyces pombe). (B) Amplifikation eines einzelnen Blattes, am Beispiel der 14-blättrigen Propeller in Nostoc.

an Stelle der Asp-Box besitzt). Wir schließen daraus, dass die Rekombination zwischen Blättern verschiedener Familien selten ist und dass der gemeinsame Vorfahre an Verzweigungspunkten zwischen bedeutenden Propellerfamilien ein einzelnes Blatt war, nicht ein vollständig ausgebildeter Propeller.

Diese Analyse suggeriert, dass die einzelnen ancestralen Propellerblätter möglicherweise in der Lage waren als Oligomere zu falten. Es ist daher auffallend, dass es nur einen oligomeren Propeller in der gesamten Strukturdatenbank gibt. Dies ist das Trimer einer zweiblättrigen Untereinheit, welches noch nicht in SCOP eingeordnet ist (*Ralstonia* Fucose-Specific Lectin, 2bt9, es ist jedoch über HHpred dem Pilzprotein Fucose-Specific Lectin, 1ofz, zu über 90% homolog - 1ofz ist ein sechsblättriger Propeller, der aus einer durchgehenden Polypeptidkette besteht) [54]. Eine mögliche Erklärung für das Nicht-Auftreten könnte sein, dass oligomere Propeller weniger effektiv sind als ihre fusionierten Entsprechungen, beispielsweise auf Grund geringerer Stabilitäten, und daher nur als kurzfristige evolutionäre Zwischenstufen auftreten, die sehr schnell durch die vollständig amplifizierte Propeller ersetzt werden.

2.5 Zusammenfassung

Wir haben die Sequenzen von β -Propellern analysiert und herausgefunden, dass mindestens eine Kerngruppe mit rund 80% der Propeller mit bekannter Struktur einander so ähnlich sind, dass dies auf Homologie schließen lässt. Trotz ihrer großen Sequenzdiversität gibt es mehrere Argumente die auf einen monophyletischen Ursprung hindeuten: (I) gemeinsame Sequenzmotive zwischen Familien, (II) gemeinsame Gruppierung in der Clusterkarte der All- β -Folds und (III) statistisch signifikante Gemeinsamkeiten zwischen ihren Profil-HMMs. Auf Grund der Datenlage können wir nicht unterscheiden, ob die Propeller, die nicht an die Kerngruppe gebunden sind, analoge Entwicklungen sind oder ob es sich um homologe Proteine handelt, bei denen die Sequenzinformation für gemeinsame Abstammung durch Sequenzdivergenz verloren wurde. Wir sehen jedenfalls, dass viele von diesen Proteinen nur wenige Homologe in den momentanen Datenbanken haben. Mit dem rapiden Voranschreiten von Sequenzierungsprojekten ist es gut möglich, dass die HMMs an Tiefe gewinnen und dass Verknüpfungssequenzen ermittelt werden, die die Außenseiter an die zentralen Proteine binden. Außerdem werden die Werkzeuge für die Sequenzanalyse vermutlich an Sensitivität gewinnen, wodurch auch entferntere Verbindungen verlässlich geknüpft werden können.

Diese Situation erinnert an die des TIM-Barrel Folds, ebenfalls eine toroidale Faltung, die eine große Gruppe an Proteinen mit marginaler Sequenzähnlichkeit beschreibt, für die ein monophyletischer Ursprung diskutiert wird [20, 71, 91].

Obwohl wir einen gemeinsamen Ursprung aller β -Propeller vorschlagen, behaupten wir nicht, dass diese alle aus einem ancestralen Propeller entstanden sind. Vielmehr finden wir Hinweise, dass die großen Propellerfamilien unabhängig von einander aus einzelnen Blättern amplifiziert wurden und dass dies ein fortwährender Prozess ist. Unser evolutionäres Szenario baut auf der Hypothese auf, dass gefaltete Proteine aus einem ancestralen Repertoire von Peptiden entstanden sind, die das Potential hatten, Supersekundärstrukturen zu bilden [61, 90]. Eins von diesen war ein β -Mäander mit der Fähigkeit durch Oligome-

risierung toroidale Strukturen zu bilden, ähnlich in Größe und Komplexität zu dem $\beta\alpha\beta$ -Peptid, das die Cradle-Loop Barrels bildete [17, 18] oder dem $\alpha\beta\alpha$ -Element, aus dem sich die Domänen der AAA+ Proteine und der Histone entwickelten [3]. Die Differenzierung der ersten Protopropeller ergab eine Population an neuen β -Mäandern, die als Ausgangspunkte für weitere Amplifikationen dienten. Seit dem hat sich die Welt der β -Propeller durch aufeinanderfolgende Amplifikationen und Diversifikationen kontinuierlich ausgedehnt.

Kapitel 3

Aufbau von β -Propellern aus Fragmenten

3.1 Einleitung

Die grundlegende Fragestellung dieser Arbeit ist, wie große und komplexe Proteindomänen entstehen konnten. Die Analyse der bioinformatischen Daten im vorherigen Kapitel suggeriert, dass einzelne ancestrale Propellerblätter möglicherweise in der Lage waren zu falten und sich als Oligomere zusammenzulagern. Solche Vorstufen zu finden würde die Theorie der ancestralen Peptide auf attraktive Weise untermauern [90, 61].

Die β -Propellerdomäne ist aus strukturellen Wiederholungseinheiten aufgebaut. Übereinstimmend mit den bioinformatischen Daten lässt sich auch auf Grund der Struktur argumentieren, dass einzelne Blätter gute Kandidaten für evolutionäre Vorstufen der Propellerdomäne sind.

In diesem Kapitel sollen die Experimente beschrieben werden, die zeigen, dass sich β -Propellerdomänen tatsächlich aus kleineren Einheiten aufbauen lassen. Dazu werden zunächst die Expressionskonstrukte erläutert, die im Rahmen dieser Arbeit angefertigt wurden um dann exemplarische Reinigungsstrategien aufzulisten. Abschließend werden die biochemischen Untersuchungen der Proteine sowie die Assemblierungsexperimente und deren Ergebnisse zusammengefasst.

3.2 Methoden zur Charakterisierung von Proteinen

Für Methoden, die für diese Arbeit von zentraler Bedeutung sind, werden im Folgenden die biophysikalischen Grundlagen erläutert.

3.2.1 Statische Lichtstreuung

Um die Molmasse von Makromolekülen zu bestimmen ist eine Methode die statische oder elastische Lichtstreuung. Unter bestimmten Voraussetzungen lassen sich weiterhin Aussagen über die Größe und Gestalt sowie intermolekulare

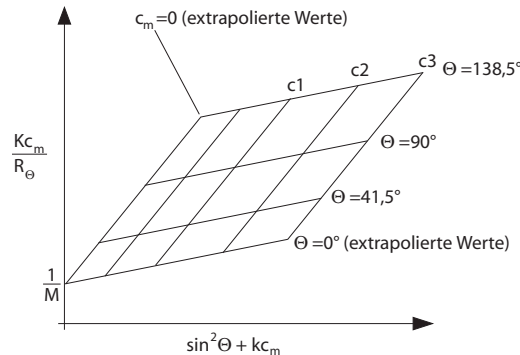


Abbildung 3.1: Beispielhafter Zimm-Plot zur Auswertung von Lichtstreuergebnissen. Die experimentell ermittelten Datenpunkte liegen auf den Schnittpunkten des Gitters. Die Streuwinkel θ entsprechen den Messungen in wässrigen Lösungsmitteln. Zur Auswertung wird auf die Konzentration Null und den Streuwinkel Null Grad extrapoliert. Diese beiden Geraden schneiden sich in einem Punkt, der der reziproken Molmasse des Moleküls entspricht.

Wechselwirkungen der Moleküle machen. Ausführliche Ableitungen der mathematischen Grundlagen können in einschlägigen Lehrbüchern nachgelesen werden [102, 8], hier sollen nur die wichtigsten Abhängigkeiten aufgeführt werden.

Bei der statischen Lichtreuung geht man davon aus, dass die elektromagnetische Strahlung ohne Energieverlust (=elastisch) gestreut wird und dass es keine Absorption durch die Probe gibt. Weiterhin soll polarisiertes Licht einer Wellenlänge eingestrahlt werden, was durch die Verwendung eines Lasers in der Lichtstreuapparatur gegeben ist.

Die Intensität des Streulichts hängt dann von der Intensität des eingestrahlten Lichts, dem Abstand Probe-Detektor, der Richtung des Detektors und der Polarisierbarkeit des Teilchens ab. Weiterhin gibt es eine starke Abhängigkeit zur Wellenlänge ($\sim \lambda^{-4}$), so dass kurzwellige Strahlung viel stärker gestreut wird als langwellige¹. Für die Molmassenbestimmung von Proteinen mit dieser Methode ergibt sich als weitere Komplikation, dass die zu untersuchenden Moleküle größer sind als die Wellenlänge der verwendeten Strahlung, was zur Folge hat, dass das Licht an verschiedenen Punkten im Probenmolekül gestreut wird und die verschiedenen Streustrahlen miteinander interferieren.

Zur Auswertung wird die Probe bei verschiedenen Konzentrationen c_m und bei verschiedenen Streuwinkeln θ vermessen, zur Auswertung werden die Daten auf Konzentration Null und Winkel Null Grad extrapoliert. Diese Auswertung erfolgt im Zimm-Plot, wie in Abb. 3.1 gezeigt.

Die Steigung der Geraden $c_m \rightarrow 0$ ermöglicht eine grobe Abschätzung der Molekülform. Da diese Auswertung für eine weitere Verwendung im Labor jedoch zu ungenau ist, werden diese Daten nicht weiter verwendet und nur die Molmasse als Ergebnis des Experiments ermittelt.

¹Diese Abhängigkeit erklärt auch, warum der Himmel blau erscheint: Der kurzwellige, blaue Anteil des Sonnenlichts wird von den Molekülen in der Atmosphäre stärker gestreut als der längerwellige Anteil.

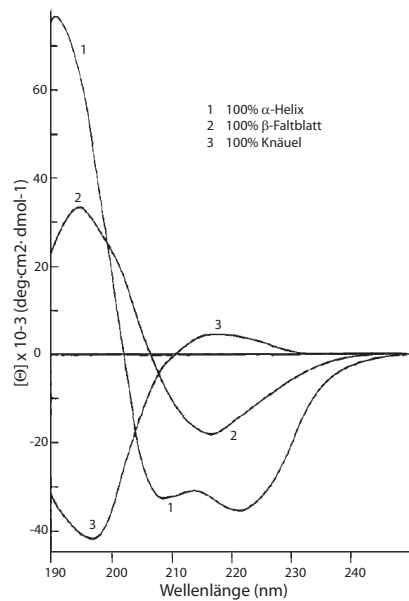


Abbildung 3.2: *Poly-L-Lysin* nimmt unter verschiedenen Bedingungen α -helikale, β -Faltblatt oder eine ungefaltete Konformation an. Diese Spektren eines Peptids, das ausschließlich ein Sekundärstrukturelement aufweist, werden als Referenzspektren verwendet. nach: [38]

3.2.2 Circular dichroismus (CD)

Um Auskunft über das Vorhandensein von Sekundärstrukturelementen in einem Protein zu bekommen, verwendet man CD-Spektroskopie. Dazu wird eine Proteinprobe mit zirkulär polarisiertem Licht bestrahlt und die Elliptizität der Probe bestimmt, die in den unterschiedlichen Absorptionskoeffizienten für links bzw. rechts zirkulär polarisiertem Licht begründet ist. Damit unterscheidet sich der Circular dichroismus von der optischen Rotationsdispersion, bei der nur die Drehung der Schwingungsebene von linear polarisiertem Licht gemessen wird (Bei der optischen Rotationsdispersion werden nur die unterschiedlichen Ausbreitungsgeschwindigkeiten -dargestellt durch die Brechungsindices- von links- und rechtspolarisiertem Licht gemessen).

Als Ursache für die unterschiedlichen Absorptionskoeffizienten wird die Kopplung von im Protein vorhandenen Chromophoren angesehen, die durch die Proteinfaltung in räumliche Nähe zueinander kommen. Besonders wichtig ist dabei der Einfluss der Sekundärstruktur auf die spektrale Charakteristik des Proteins. Um den Sekundärstrukturanteil in einem Protein zu bestimmen, vergleicht man das gemessene Spektrum mit Referenzspektren, in denen Peptide mit ausschließlich einem Sekundärstrukturelement vermessen wurden (s. Abb. 3.2).

Da sich CD-Spektren von ungefalteten Proteinen (Knäuel in Abb. 3.2) stark von denen gefalteter Proteine unterscheiden, lässt sich diese Methode gut anwenden, um Faltungskinetiken zu messen.

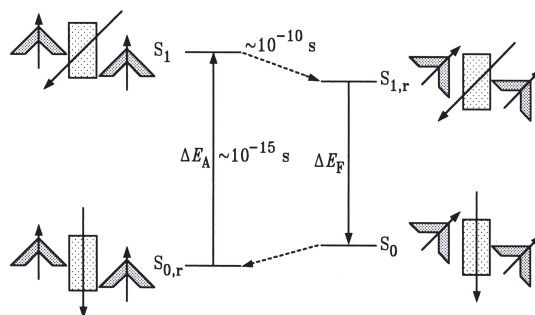


Abbildung 3.3: *Elektromagnetische Strahlung regt in einem Probenmolekül Elektronen an und versetzt sie in den S_1 -Zustand, der ein etwas anderes Dipolmoment aufweist als der Grundzustand. Etwas zeitverzögert lagern sich die umgebenden Wassermoleküle um und stabilisieren dadurch den S_1 -Zustand. Das emittierte Fluoreszenzlicht ΔE_F hat daher eine längere Wellenlänge als das Anregungslicht, die Fluoreszenzstrahlung ist rotverschoben. Das Probenmolekül ist schematisch als Kasten dargestellt, die Wassermoleküle sind als Dreiecke gezeichnet, die Pfeile bezeichnen die Dipolmomente. aus: [102]*

3.2.3 Fluoreszenzspektroskopie

Um über die Polarität der Umgebung von Tryptophanseitenketten² in einem Protein Aussagen treffen zu können, wird häufig die Fluoreszenzspektroskopie eingesetzt. Die Methode beruht darauf, dass elektromagnetische Strahlung in der Lage ist, in der Probe Elektronen in einen angeregten Zustand zu versetzen. Die Rückkehr in den Ausgangszustand erfolgt dann durch eine verzögerte Lichtemission, die Fluoreszenz.

Das angeregte Molekül besitzt ein lokal leicht verändertes Dipolmoment (s. Abb. 3.3). Ist die Probe von Wasser umgeben, so lagern sich die Wassermoleküle im veränderten, angeregten Zustand so um, dass sie diesen stabilisieren, was zu einer energetischen Absenkung führt. Der Übergang vom elektronisch angeregten Zustand zurück in den Grundzustand setzt in diesem Fall weniger Energie frei, als für die Anregung nötig war. Die emittierte Strahlung hat eine längere Wellenlänge. Diese sogenannte Rotverschiebung ist in einem polaren Lösungsmittel stärker als in einer unpolaren Umgebung. Da die meisten Tryptophane im gefalteten Protein im hydrophoben Kern vorliegen, während sie im entfalteten Zustand direkten Kontakt mit dem Lösungsmittel haben, kann man die Trp-Fluoreszenzspektren heranziehen, um Aussagen über den Faltungszustand von Proteinen zu machen.

CD-Spektroskopie und Trp-Fluoreszenz sind dabei komplementäre Methoden: Das CD-Spektrum gibt nur Auskunft darüber, ob Sekundärstrukturelemente vorhanden sind oder nicht. Die Trp-Fluoreszenz erlaubt dagegen die Aussage, ob die Proteinstruktur einen kompakten, hydrophoben Kern besitzt. Nur durch die kombinierte Auswertung beider Spektren lässt sich der sogenannte *Molten*

²Ähnliche Spektren können auch für Tyrosinreste gemessen werden, sollte das Protein keine Tryptophane enthalten. Da dies aber für keines der Proteine in der vorliegenden Arbeit zutrifft, wird die Tyr-Fluoreszenz hier nicht weiter besprochen.



Abbildung 3.4: Konstruktion der DNA-Sequenz für das einzelne Propellerblatt. Die beiden Primer sind in rot geschrieben, die durch die DNA-Polymerase aufgefüllten Nukleotide in grau. In blau die in Aminosäuren übersetzte Sequenz, die beiden Schnittstellen für die Klonierung sind NdeI und HindIII.

Globule Zustand ausschließen, bei dem sich zwar α -Helices und β -Faltblätter ausbilden, diese aber ohne feste Kontakte zueinander bleiben.

3.3 Material und Methoden

3.3.1 Klonierung des einzelnen Propellerblattes

Die in Abb. 3.10 entwickelte Aminosäuresequenz wurde manuell in eine DNA-Sequenz übersetzt, wobei darauf geachtet wurde, dass die von *E. coli* häufig verwendeten Triplets vermehrt eingesetzt wurden. Diese DNA-Sequenz wurde auf zwei revers komplementäre Oligonukleotide (Primer) aufgeteilt, die im mittleren Bereich der Sequenz überlappen, s. auch Abb. 3.4. Die beiden DNA-Primer wurden in stöchiometrischen Mengen gemischt (jeweils 100 pmol) und in einem Metallblock auf ca. 90°C erhitzt und dann langsam abgekühlt, um so eine Zusammenlagerung der beiden Primer zu erreichen. Diese DNA-Lösung wurde für eine Primer-Verlängerungsreaktion eingesetzt³:

30,0	μl	Primerlösung
5,0	μl	Pfu Pufferlösung mit MgSO_4
5,0	μl	dNTP Mix (jew. 10 mM in der Ausgangslösung)
ad 49,0	μl	H_2O
1,0	μl	Pfu DNA Polymerase (2,5 units/ μl) - kommerziell erhältlich bei Fermentas TM

Bei 72°C, der optimalen Elongationstemperatur der Pfu-Polymerase, wurden die in Abb. 3.4 grauen Nukleotide ergänzt. Die nun doppelsträngige DNA⁴ wurde über ein 1,5%iges Agarosegel von einzelsträngigen DNA-Stücken und den restlichen Komponenten der Pfu-Reaktion getrennt. Die aus dem Gel eluierte

³Die hier detailliert beschriebenen Reaktionen wurden -jeweils auf die entsprechende Reaktion angepasst- immer nach den hier aufgelisteten Schemata durchgeführt.

⁴Im Gegensatz zur Taq-Polymerase generiert die Pfu-Polymerase *blunt end* PCR-Produkte, also DNA-Stränge, bei denen am Ende keine Nukleotide überhängen.

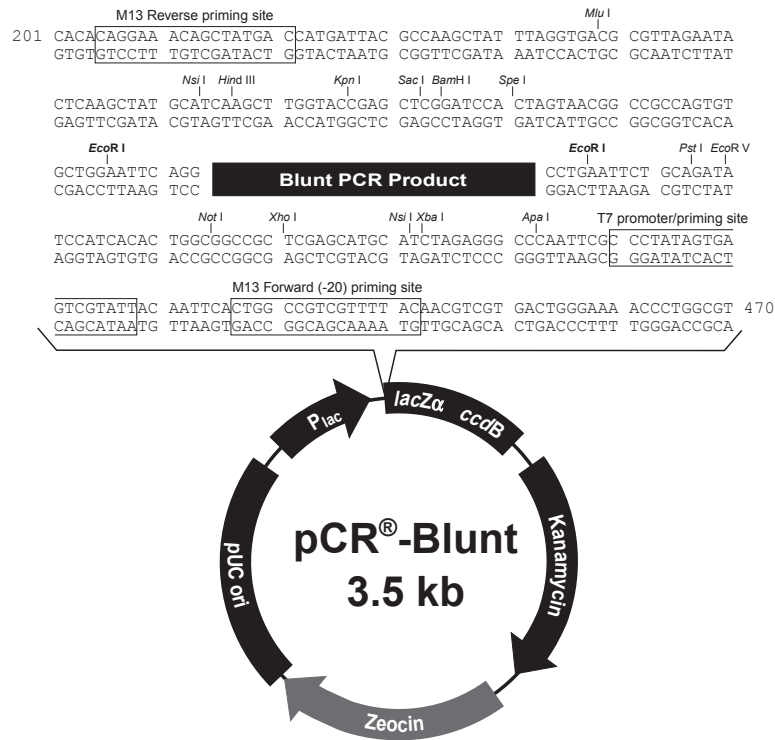


Abbildung 3.5: Vektorkarte des Klonierungsvektors Zeroblunt. Der Vektor ist kommerziell erhältlich bei InvitrogenTM. Die Vektorkarte ist nach dem Handbuch modifiziert wiedergegeben.

Bande wurde in einen Subklonierungsvektor, Zeroblunt, ligiert (Vektorkarte s. Abb. 3.5).

- 7,0 μ l DNA-Lösung mit der doppelsträngigen Bande
- 1,0 μ l Ligasepuffer
- 1,0 μ l T4 DNA Ligase (400 units/ μ l) - kommerziell erhältlich bei New England BioLabsTM (NEB)
- 1,0 μ l Zeroblunt Vektor (25ng/ μ l) - kommerziell erhältlich bei InvitrogenTM
- 1-16 Std Inkubation bei 16°C

Die für ein einzelnes Propellerblatt codierende DNA konnte im Zeroblunt vervielfältigt und sequenziert werden. Die DNA aus einem korrekten Klon wurde isoliert, das codierende DNA-Fragment herausgeschnitten und endgültig in einen Expressionsvektor ligiert. Ein Restriktionsverdau setzte sich wie folgt zusammen:

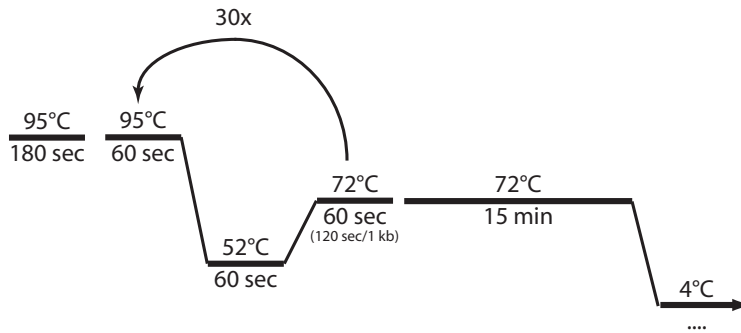


Abbildung 3.6: Zyklus für die Mutations-PCR um die *NdeI*-Schnittstelle im *pET28*-Vektor gegen eine *BamHI*-Schnittstelle auszutauschen. Die Reaktion wurde wie im Text beschrieben zusammengesetzt, die Primergestaltung ist in Abb. 3.7 beschrieben.

16,0	μl	Plasmid-DNA
2,0	μl	NEB Buffer 2 - Restriktionspuffer
1,0	μl	<i>NdeI</i> Restriktionsenzym (20 units/μl) - kommerziell erhältlich bei NEB
1,0	μl	<i>HindIII</i> Restriktionsenzym (10 units/μl) - kommerziell erhältlich bei Fermentas TM
2-16	Std	Inkubation bei 37°C

Die Expressionsvektoren sollten eine effiziente Proteinsynthese eines spezifischen Genprodukts ermöglichen. Um dies zu erreichen und eine einfache Reinigung zu vereinfachen, wurden die Konstrukte so gestaltet, dass die Proteine am N-Terminus mit einem Hexahistidin- (*His₆*-Tag) oder einer Glutathion-S-Transferase (GST) -Fusion exprimiert wurden. Dabei fügt der *pET28*-Vektor einen *His₆*-Tag an das Protein an und *pGEX* eine GST-Fusion (s. Abb. 3.8). Beide Vektoren ermöglichen eine IPTG-induzierbare Expression, die Vektoren werden auch noch im Abschnitt 3.4 beschrieben. Für die Klonierung in *pGEX* musste die 5' *NdeI*-Schnittstelle durch eine *BamHI*-Schnittstelle ersetzt werden, damit das Konstrukt als Fusion mit der Glutathion-S-Transferase exprimiert werden konnte. Dabei war darauf zu achten, dass durch die veränderte Schnittstelle das Leseraster nicht verändert wurde. Dieser Austausch wurde durch eine PCR mit einem geeigneten Oligonukleotid (Mutationsprimer) erreicht (Abb. 3.6). Dieser Primer wurde so erstellt, dass er bis auf den Bereich der zu mutierenden Schnittstelle sowohl auf der 5'- als auch auf der 3'-Seite genau zu dem Vektor passte bzw. der Gensequenz komplementär war. Anhand der Proteinsequenz in Abb. 3.7 wird deutlich, warum die Fusionsproteine nach der Thrombinspaltung alle am N-Terminus mit Gly-Ser beginnen und nicht mit einem Start-Met. Die PCR wurde mit dem Mutationsprimer und einem Rückwärtsprimer (z. B. T7 Terminator oder M13 Reverse) durchgeführt:

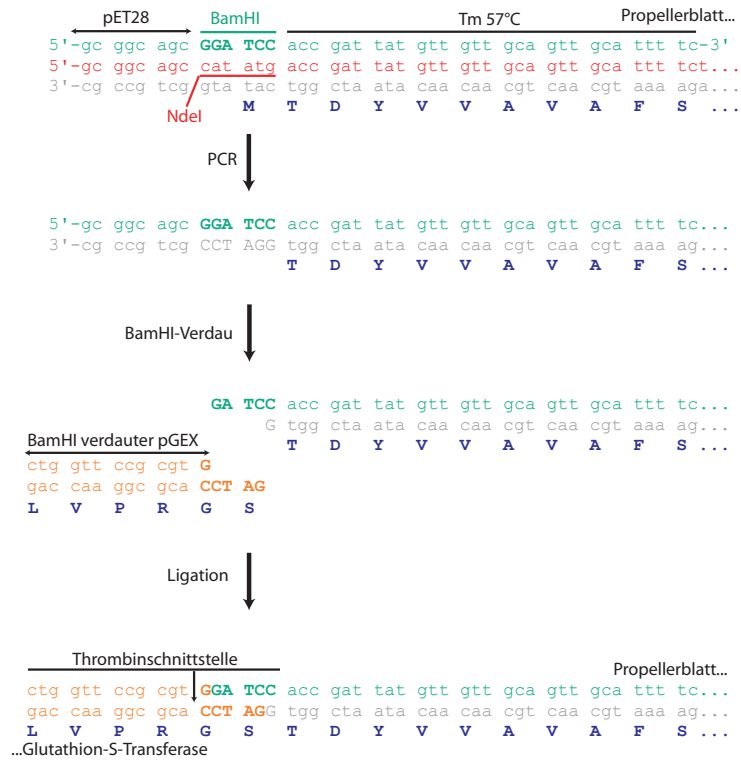


Abbildung 3.7: Schematisches Vorgehen beim Austausch einer Restriktionschnittstelle. Ein PCR-Primer wurde so gestaltet, dass er der DNA-Vorlage komplementär war, nur die auszutauschenden Basenpaare bildeten keine Basenpaarungen aus. Als zweiter Primer wurde ein Standardprimer, z. B. T7 Terminator, verwendet. Das PCR-Produkt wurde verdaut und in einen entsprechend geöffneten Vektor ligiert. Die ursprüngliche Sequenz im pET28-Vektor ist in rot geschrieben, der Primer in grün, der pGEX-Vektor in orange.

0,5	μl	pET28 Vektor mit dem Propellerblatt (ca. 30 ng/μl)
3,0	μl	Mutationsprimer (10 pmol/μl)
3,0	μl	Rückwärtsprimer (10 pmol/μl)
5,0	μl	Pfu Pufferlösung mit MgSO ₄
5,0	μl	dNTP Mix (jew. 10 mM in der Ausgangslösung)
ad 49,0	μl	H ₂ O
1,0	μl	Pfu DNA Polymerase (2,5 units/μl)

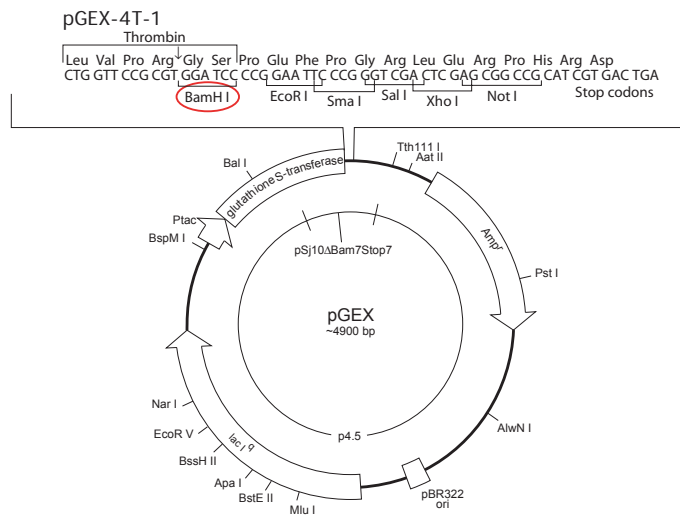
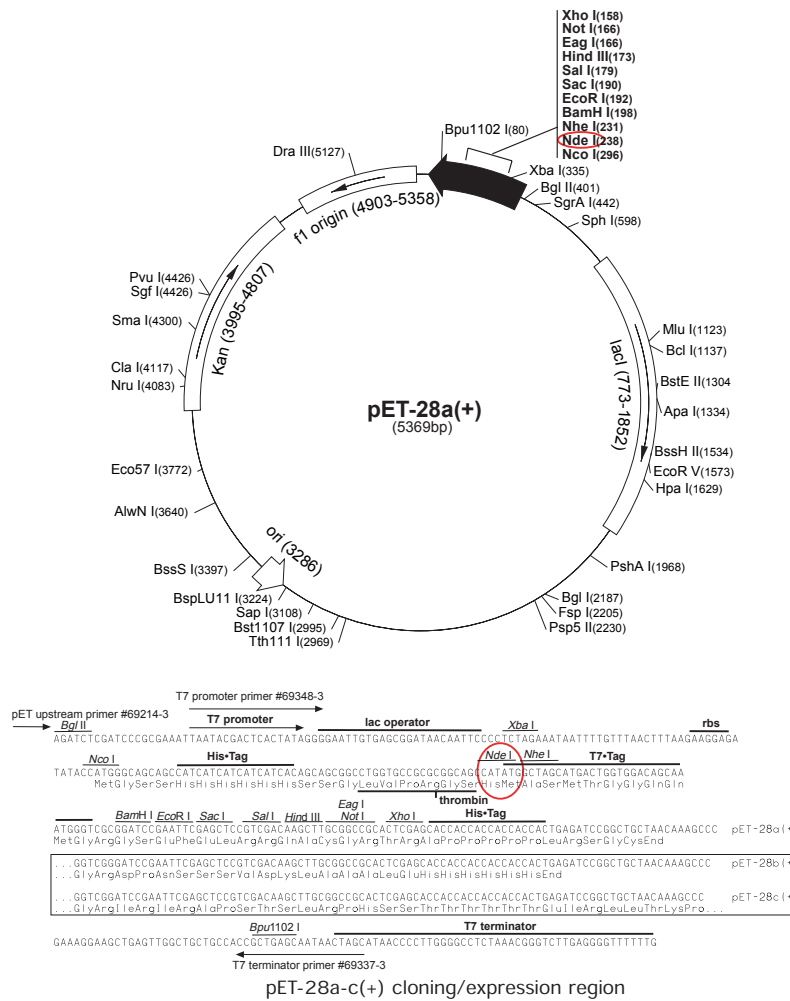


Abbildung 3.8: Vektorkarten der Expressionsvektoren pET28 und pGEX. Die Vektoren sind kommerziell erhältlich bei NovagenTM bzw. GE HealthcareTM. Die Vektorkarten sind aus den jeweiligen Handbüchern modifiziert wiedergegeben.

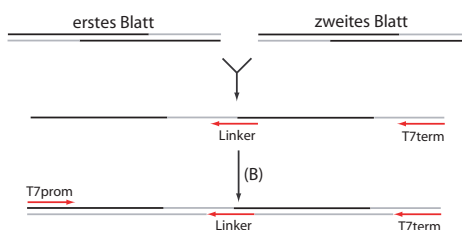


Abbildung 3.9: Verdopplung des einzelnen Propellerblattes. Die Oligonukleotide für die einzelnen Propellerblätter sind in schwarz dargestellt, von der Polymerase aufgefüllte Nukleotide in grau. In der zweiten Reaktion wurden der Linkerprimer und der T7 Terminator zu einem durchgehenden DNA-Strang ergänzt, der dann für zwei Propellerblätter codierte. (B) Zusätzlicher Amplifikationsschritt der bei der Herstellung der längeren Konstrukte angewendet wurde.

3.3.2 Amplifikation der Propellerblätter

Um ein Konstrukt mit zwei Propellerblättern zu bekommen, wurden zwei einzelne Propellerblätter, wie im vorherigen Abschnitt bereits beschrieben, jedoch mit unterschiedlichen Codons aus Primern hergestellt. Die beiden Konstrukte wurden gereinigt und in einer weiteren Primer-Verlängerungsreaktion eingesetzt (s. Abb. 3.9).

17,0	μl	gelgereinigte Bande mit Blatt 1
17,0	μl	gelgereinigte Bande mit Blatt 2
0,3	μl	Linkeroligonukleotid (30 pmol)
3,0	μl	T7 Terminator Primer (30 pmol)
5,0	μl	Pfu Pufferlösung mit MgSO_4
5,0	μl	dNTP Mix (jew. 10 mM in der Ausgangslösung)
ad 49,0	μl	H_2O
0,7	μl	Pfu DNA Polymerase (2,5 units/ μl)

Da diese Methode nur mit einer ausreichenden Effizienz abläuft, wenn sich die DNA-Sequenzen für die einzelnen Blätter unterscheiden, ist dieses Vorgehen nur für die Verdopplung des Propellerblattes tauglich. Unterscheiden sich die Sequenzen nicht, so ergeben sich multiple Bindungsstellen für den Linkerprimer und es kommt nur in sehr geringem Maße zu einer linkervermittelten Verknüpfung von zwei DNA-Konstrukten. Für die Konstrukte mit drei bis fünf Propellerblättern wurde jeweils das längste Konstrukt mit dem vektorspezifischen T7 Promoter Primer und einem Primer, der spezifisch für das Ende des Konstrukts war, amplifiziert. Dieses wurde durch ein einzelnes Blatt ergänzt, das an seinem Ende die Sequenz für den T7 Terminator Primer beinhaltet. Nur wenn diese beiden Teilkonstrukte erfolgreich verbunden wurden, ergab sich ein Konstrukt, welches durch eine PCR mit T7 Promoter- und Terminator Primer amplifiziert werden konnte. Dieser Amplifikationsschritt war entscheidend für den Erfolg dieser Klonierungsstrategie. Auf diese Art und Weise wurden folgende Konstrukte hergestellt:

- das verdoppelte Propellerblatt (in pET30)⁵;

⁵Dieses Konstrukt bzw. das entsprechende Protein wird von der englischen Bezeichnung

- drei Propellerblätter in Folge (in pGEX)⁶;
- vier Propellerblätter in Folge (in pGEX)⁷;
- fünf Propellerblätter in Folge (in pET28)⁸.

3.3.3 Alternative Propellerblätter

Um auszuschließen, dass die gewählte Konsensussequenz nur einen Einzelfall darstellt, wurde eine Reihe verschiedener Einzelblattsequenzen untersucht. Für alle in Abb. 3.12 beschriebenen Konstrukte wurde zunächst die DNA-Sequenz aus Oligonukleotiden künstlich hergestellt und diese dann in den pGEX Expressionsvektor kloniert. Das detaillierte Vorgehen wurde in Abschnitt 3.3.1 bereits erläutert.

3.4 Expression und Reinigung der Konstrukte

Die in dieser Arbeit beschriebenen Proteine sind rekombinant in *Escherichia coli*-Zellen hergestellt worden. Die prinzipiellen Reinigungsschritte werden in den folgenden Abschnitten beschrieben. Vor der Proteinreinigung stand die Expression, die für alle Konstrukte vergleichbar ablief.

Das Expressionsplasmid wurde für die Expression in *E. coli* C41 (DE3) Zellen transformiert⁹. Für Expressionen wurden die Plasmide immer frisch in die Expressionsstämme transformiert. Eine so transformierte Einzelkolonie wird in ca. 100 ml LB-Medium über Nacht bei 37°C geschüttelt. Mit 20 ml Zellsuspension aus dieser Vorkultur wurden 800 ml LB-Medium (entspricht 2,5%) angeimpft und in einer 1 L-Flasche durch ein Glasrohr mit Druckluft begast. Die pET-Vektoren vermitteln eine Kanamycinresistenz, es wurden in der Endkonzentration 50 µg/ml Antibiotikum eingesetzt. Der pGEX-Vektor vermittelt eine Ampicillinresistenz, wobei 100 µg/ml Endkonzentration verwendet wurden. Die Zellen wurden bei einer optischen Dichte von $OD_{600nm} \approx 0,6$ mit 1 mM IPTG induziert. Nach vier bis sechs Stunden wurden die Zellen durch Zentrifugation geerntet (20 min bei 6500 g). Der Zellklumpen wurde gewogen und bei -20°C gelagert.

3.4.1 GST-Fusionsproteine

Die Konstrukte der einzelnen Blätter sowie die Konstrukte für drei und vier Propellerblätter wurden als Glutathion-S-Transferase (GST) -Fusionen exprimiert. Für den Zellaufschluss wurden jeweils ca. 8 g tiefgefrorene Zellen auf 35 ml Volumen mit PBS (20 mM Natriumphosphat pH 7,2; 150 mM NaCl; 2 mM KCl) aufgefüllt und durch Vortexen und Resuspendieren aufgetaut. Die Suspension wurde auf 5mM MgCl₂ eingestellt und mit Proteaseinhibitorcocktail (cOmpleteTMProteaseInhibitor Cocktail Tablets, kommerziell erhältlich von

Two blades IN A Row als TINAR bezeichnet.

⁶Dieses Konstrukt bzw. das entsprechende Protein wird von der englischen Bezeichnung ThREe blades IN A Row als TriNAR bezeichnet.

⁷Dieses Konstrukt bzw. das entsprechende Protein wird von der englischen Bezeichnung FOuR blades IN A Row als FoNAR bezeichnet.

⁸Dieses Konstrukt bzw. das entsprechende Protein wird von der englischen Bezeichnung FIve blades IN A Row als FiNAR bezeichnet.

⁹In Versuchsreihen wurden auch BL21 Gold Zellen sowie unterschiedliche Temperaturen getestet, die besten Ergebnisse ergaben sich jedoch in jedem Fall in C41 Zellen bei 37°C.

Roche) und DNaseI (AppliChem) versetzt. Die Zellen wurden in fünf Durchläufen durch eine French Pressure Zellpresse (American Instrument Co.) bei ca. 1000 psi aufgeschlossen. Im Anschluß an den Zellaufschluß wurde der Suspension EDTA bis zu einer Endkonzentration von 10 mM zugegeben. Die groben Zellbestandteile wurden durch zweimalige Zentrifugation (37.000 g, 30-45 min) abgetrennt. Das so gewonnene Rohlysate wurde durch einen 0,45 μ M Filter gepresst und dann auf eine GSTPrepTM FF 16/10 Säule von GE Healthcare (früher Amersham Pharmacia¹⁰) aufgetragen. Die Säule wurde so lange mit PBS gespült, bis die UV-Absorption_{280nm} wieder einen stabilen Wert annahm. Die Bindung der Glutathion-S-Transferase an die Säule erfolgte auf Grund der Bindung des Proteins an das kovalent an die Säule fixierten GSH-Peptids. Die Elution erfolgte durch kompetitive Verdrängung durch das freie GSH im Elutionspuffer (10 mM reduziertes GSH; 10 mM Tris; pH 8,0). Die eluierte Proteinmenge wurde bestimmt und die Lösung mit 1 *Unit* Thrombin pro mg Protein versetzt. Nach 4 Stunden war ein vollständiger Verdau erfolgt, das Thrombin wurde durch Zugabe von 0,1 M PMSF inhibiert. Durch Dialyse gegen PBS ließ sich das GSH aus der Proteinlösung entfernen bevor die Proteinlösung erneut auf die regenerierte GSTPrep-Säule aufgetragen wurde. Das von der GST abgespaltene Protein wurde im Durchlauf aufgefangen, einkonzentriert und auf einer HiLoad 26/60 Superdex G75 Gelfiltrationssäule endgültig gereinigt.

3.4.2 Das verdoppelte Propellerblatt

Das Konstrukt mit zwei Propellerblättern wurde aus dem pET30 Vektor exprimiert (pET30 entspricht dem bereits vorgestellten pET28 Vektor mit dem Unterschied, dass kein N-terminaler His₆-Tag angefügt wird). Der Zellaufschluss erfolgte genau wie für die GST-Fusionsproteine, nur dass als Aufschluss- und Ladepuffer 50 mM Tris; pH 7,5; 50 mM NaCl verwendet wurde. Die filtrierte Proteinlösung wurde auf eine HiPrep 16/10 Phenyl FF Säule geladen und die Säule mit Ladepuffer gespült. Die Elution erfolgte mit 30% Ethanol. Das eluierte Protein wurde 1:10 mit Ladepuffer verdünnt und an eine 10 ml MonoQ Säule gebunden. Die Elution erfolgte in einem linearen Gradienten zwischen Auftragspuffer und 50 mM Tris; pH 7,5; 1 M NaCl. Das Protein wurde einkonzentriert und auf einer HiLoad 26/60 Superdex G75 Gelfiltrationssäule endgültig gereinigt.

3.4.3 Proteine mit einem His₆-Tag

Das Konstrukt mit fünf Propellerblättern und der 14-blättrige Propeller in Kapitel 4 wurden vom pET28 Vektor exprimiert und wiesen daher einen N-terminalen His₆-Tag auf. Der Zellaufschluss erfolgte wie bei den gerade beschriebenen Reinigungen (Puffer: 50 mM Tris; pH 7,9; 50 mM NaCl). Das Zelllysate wurde auf eine QHP-Anionenaustauschchromatographiesäule gebunden. Die Elution erfolgte in einem linearen Gradienten von 50 bis 1000 mM NaCl. Die stark salzhaltige Lösung mit eluiertem Protein wurde direkt auf eine Ni SepharoseTM High Performance-Säule geladen. Schwach gebundene Verunreinigungen wurden zunächst mit einer Mischung aus 5% Imidazolpuffer (50 mM Tris; pH 7,9; 300 mM

¹⁰ Alle Säulen für die Proteinreinigung sind von diesem Hersteller, daher wird er im Folgenden nicht mehr weiter erwähnt.

NaCl; 600 mM Imidazol) und 95% Auftragspuffer, was einer Imidazolkonzentration von 30 mM entsprach, von der Säule gewaschen. Die endgültige Elution erfolgte mit einem linearen Gradienten von 30 bis 600 mM Imidazol. Wenn eine Abspaltung des His₆-Tags erforderlich war, wurde gegen Thrombinpuffer dialysiert (20 mM Tris; pH 8,5; 100 mM NaCl; 1 mM CaCl₂; 1 mM DTT) und mit 1 *Unit* Thrombin pro mg Protein über Nacht bei 18°C inkubiert. Thrombin wurde durch Zugabe von 0,1 M PMSF inhibiert. In jedem Fall wurde das Protein einkonzentriert und auf einer HiLoad 26/60 Superdex G75 Gelfiltrationssäule endgültig gereinigt, dabei konnte bei Bedarf ein Pufferwechsel durchgeführt werden.

3.4.4 Kristallisationsansätze der Konstrukte

Die Proteine TINAR und FINAR wurden für Kristallisationsansätze verwendet. Die Proteine wurden nur dann verwendet, wenn die Reinheitsüberprüfung ergab, dass das Protein als einzelne, saubere Bande auf einem SDS-Polyacrylamidgel aufgetrennt wird. Zusätzlich wurde vor den Kristallansätzen mit statischer Lichtstreuung sichergestellt, dass die eingesetzte Proteinprobe monodispers war. Für die Kristallansätze wurden Proteinkonzentrationen zwischen 10 und 100 mg/ml eingesetzt. Da kein Ansatz Proteinkristalle mit oligomerisierten Propellern ergab, werden diese Experimente nicht weiter beschrieben.

3.4.5 Vorgehen bei den Mischungsexperimenten

Um die gereinigten Propeller in unterschiedlichen stöchiometrischen Verhältnissen zusammenzubringen, wurden von den Ausgangsproteinen die molaren Konzentrationen bestimmt. Die Mischungen werden entsprechend der Proteinzusammensetzung berechnet. Um die Volumenzunahme durch die Zugabe von Guanidin Hydrochlorid (GuaHCl) zu berücksichtigen, wird von einer Volumenverdopplung ausgegangen. Entsprechend dem verdoppelten Volumen wird GuaHCl eingewogen, um eine Lösung mit 6 M Endkonzentration zu erhalten. Das Salz wird in der Proteinklösung durch Vortexen gelöst und die entfaltete Proteinlösung mindestens 30 min bei Raumtemperatur inkubiert.

Die Rückfaltung in die assemblierten Propeller erfolgt durch Dialyse gegen PBS in Membranschläuchen mit einem Ausschlussvolumen von 1000 Da.

3.5 Hintergründe der verwendeten Konstrukte

Um kleine Proteine künstlich herzustellen, werden im Feld des Proteindesigns häufig Konsensussequenzen verwendet [53, 63]. Dabei wird argumentiert, dass es vorteilhaft ist, statt exemplarischer Proteinsequenzen die häufigste Aminosäure aus einem *Alignment* zu nehmen, da einzelne Sequenzen eventuell destabilisierende (lokale) Mutationen enthalten können oder einzelne lokale Mutationen durch weitere kompensiert werden müssen [60]. Um die β -Propellerdomäne aus Peptidfragmenten aufzubauen, haben wir zunächst versucht, die strukturelle Wiederholungseinheit (das einzelne Propellerblatt) als einzelnes Protein zu exprimieren (für weitere Ansätze siehe Abschnitt 3.5.3).

```

Blatt 7                               LNEPRILTTD
Blatt 1 .REAVAVAFSPGGSLLAGGSGDKLIHVWVDVASGDELHTLEGH
Blatt 2 TDWVRVAVAFSPDGCALLASGSDDATVRLWVAAAEEERAVFECH
Blatt 3 THYVLDIAFSPDCSMVASGSRDGTARLWNVATGTEHAVLKGH
Blatt 4 TDYVYAVAFSPDCSMVASGSRDGTIRLWVDVATGKERDVLQAP
Blatt 5 AENVVSLAFSPDCSMLVHGS.DSTVHLWVDVASGEALHTFECH
Blatt 6 TDWVRVAVAFSPDGCALLASGSDDRTIRLWVDVAAQEEHTTLECH
Blatt 7 TEPVHSAVAFPECTTLASASEDGTIRIWPPIA

Konsensus TDYVVAVAFSPDGSLLASGSDDGTIRLWVDVATGEELAVLECH

```

Abbildung 3.10: Sequenzalignment der C-terminalen Propellerblätter aus dem *PkwA* Protein. Die von uns verwendete Konsensussequenz kommt in exakt dieser Form nicht in dem Protein vor.

3.5.1 Informationen über das Ausgangsprotein für die einzelnen Blätter

Zu Beginn der Arbeit waren die hochrepetitiven Sequenzen aus *Nostoc punctiforme* (s. Kapitel 4) noch nicht bekannt, daher sind wir von einem anderen repetitiven Protein ausgegangen, *PkwA*. Dieses Protein ist in seiner detaillierten Struktur nicht beschrieben, zeichnet sich jedoch durch relativ gute Sequenzkonserviertheit in den einzelnen Propellerblättern aus. Das Sequenzalignment ist in Abb. 3.10 gezeigt. Aus diesen sieben Propellerblättern haben wir den Konsensus gebildet und mit diesem weitergearbeitet.

Der Propeller ist nur der C-terminale Teil eines 79 kDa Proteins. Dies ist als Ser/Thr Protein Kinase beschrieben [45] und stammt aus dem Actinobacterium *Thermomonospora curvata*. Die beiden webbasierten Datenbanken zur Analyse von Proteindomänen, SMART und Pfam [58, 28], stimmen in ihren Vorhersagen für *PkwA* überein:



Abbildung 3.11: Domänenzuordnung für das Vollängenprotein *PkwA* durch den SMART Webserver (<http://smart.embl-heidelberg.de/>).

Die schwarze Domänenrepräsentation am N-Terminus symbolisiert die Proteinkinasedomäne, die rosa Bereiche markierten Abschnitte mit geringem Sequenzinformationsgehalt (low compositional complexity). Die als grüne Dreiecke dargestellten C-terminalen WD40-Sequenzwiederholungen sind in ihrer Sequenz in Abb. 3.10 gezeigt. Die Konsensussequenz haben wir für unsere weiteren Experimente verwendet.

3.5.2 Amplifikation der Propellerblätter

Außer dem einzelnen Propellerblatt sollten auch größere Konstrukte mit sukzessive zwei, drei, vier und fünf Propellerblättern in einem Konstrukt untersucht

werden. Dazu wurde von dem einzelnen Blatt die DNA-Sequenz multipliziert, wie im Abschnitt 3.3.2 im Material- und Methodenteil beschrieben. So wurden folgende Konstrukte hergestellt:

- das verdoppelte Propellerblatt (in pET30)¹¹;
- drei Propellerblätter in Folge (in pGEX)¹²;
- vier Propellerblätter in Folge (in pGEX)¹³;
- fünf Propellerblätter in Folge (in pET28)¹⁴.

3.5.3 Alternative Propellerblätter

Um auszuschließen, dass die gewählte Konsensussequenz nur einen Einzelfall darstellt, wurden eine Reihe verschiedener Einzelblattsequenzen untersucht. Für alle im Folgenden beschriebenen Konstrukte wurde, wie in Abschnitt 3.3.3 beschrieben, zunächst die DNA-Sequenz aus Oligonukleotiden künstlich hergestellt und diese dann in den pGEX Expressionsvektor kloniert.

Um zu überprüfen, ob der Unterschied zwischen dem nicht-faltenden einzelnen Pk_wA-Konsensusblatt und dem faltenden Doppelblatt ein sequenzabhängiger Prozess ist oder ob physikalische Größen eine Rolle spielen, sollten zwei einzelne Blätter kovalent miteinander verbunden werden. Da die Konsensussequenz außer dem N-Terminus keine weitere freie Aminogruppe aufweist und Verbrückungsreagentien, die für Carbonsäuregruppen spezifisch sind, räumlich anspruchsvoll sind, sollten zwei Blätter über eine Disulfidbrücke verbrückt werden. Die Konsensussequenz wurde am N-Terminus durch einen Cysteinrest verlängert und exprimiert.

Als alternative Sequenzen zu der Konsensussequenz, die in Abb. 3.4 abgeleitet wurde, haben wir einen hypothetischen Vorfahren für die Blätter des Pk_wA-Propellers berechnet. Dazu wurden die einzelnen Blätter als Endpunkte in einer Phylogenie angesehen und aus dem zugrundeliegenden Stammbaum auf den wahrscheinlichsten gemeinsamen Vorfahren geschlossen. Dieses Prozedere wurde sowohl für die Protein- als auch für die codierende DNA-Sequenz durchgeführt, wobei zwei leicht unterschiedliche Sequenzen berechnet wurden [14, 30]. Die exprimierten Sequenzen sind in Abb. 3.12 dargestellt.

Weiterhin wurden zwei exemplarische Blätter aus bekannten Proteinen exprimiert. Das eine ist eine strukturelle Wiederholungseinheit aus dem kristallisierten Propeller, der in Kapitel 4 beschrieben wird. In diesem Protein sind alle Blätter nahezu identisch, sie variieren nur in einer Position. Es kann also davon ausgegangen werden, dass sich die Amplifikation dieses Blattes zu einem vollständigen Propeller erst kürzlich vollzogen hat.

Für die zweite natürlich vorkommende Sequenz sollte ein Blatt aus einem thermophilen Organismus verwendet werden. Durch iterative Datenbanksuchen gegen die Datenbank mit archaealen Proteinen¹⁵ wurde ein geeignetes Protein

¹¹Dieses Konstrukt bzw. das entsprechende Protein wird von der englischen Bezeichnung Two blades IN A Row als TINAR bezeichnet.

¹²Dieses Konstrukt bzw. das entsprechende Protein wird von der englischen Bezeichnung ThREe blades iN A Row als TriNAR bezeichnet.

¹³Dieses Konstrukt bzw. das entsprechende Protein wird von der englischen Bezeichnung FOuR blades iN A Row als FoNAR bezeichnet.

¹⁴Dieses Konstrukt bzw. das entsprechende Protein wird von der englischen Bezeichnung FIve blades iN A Row als FiNAR bezeichnet.

¹⁵Das öffentlich zugängliche 'Bioinformatics Toolkit' auf der Homepage des Max Planck Instituts für Entwicklungsbiologie bietet die Möglichkeit Sequenzsuchen mit

PkwA-Konsensus	TDYVVAVAFSPDGSLLASGSDDGTIRLWDVASGEELAVLEGH
Cys-Konsensus	CTDYVVAVAFSPDGSLLASGSDDGTIRLWDVASGEELAVLEGH
Ancestor-Prot	TDYVRAVAFSPDGSMLASGSDDGTIRLWDVATGEERAVLEGH
Ancestor-DNA	TDYVHAVAFSPDGSMVASGSDDGTIRLWDVATGKEHDVLEKGH
HMM-Konsensus	TGWNSVAFSPDGKLLASGSDDKTVRLWDLATGKCLRTLIGH
14-blättriger	SSSVWGVAFSPDGQTIASASDDKTVKLWN-RNGQLLQTLIGH
Sulfolobus	PSGILYDPSNGYIYVTDSKSNTVSVINSSTNOVIANITVKGH

Abbildung 3.12: *Alternative Propellerblätter, die im Rahmen dieser Arbeit exprimiert wurden. Sequenzpositionen, die nicht mit dem Konsensus aus PkwA übereinstimmen, sind in rot hervorgehoben. Die Ableitung der einzelnen Sequenzen ist im Text detailliert beschrieben.*

gefunden: Das hypothetische Protein ST1748 aus dem Archaeon *Sulfolobus tokodaii* (Stamm 7), >gi|15922059|. *Sulfolobus tokodaii* wächst optimal bei 80°C [52], daher kann für das Protein eine entsprechende Thermostabilität angenommen werden. Für das Protein werden sieben Propellerwiederholungen vorhergesagt. Da die Propellerblätter jedoch keine perfekt ausgeprägten Sequenzwiederholungen aufweisen, wurden die einzelnen Blätter in ClustalX [94] aligniert und das *Alignment* manuell nachbearbeitet. Aus diesem Sequenzalignment wurde die Konsensussequenz abgeleitet, dieses ist in Abb. 3.12 gezeigt.

Im Zuge der Argumentation, dass einzelne Sequenzen durch lokale Mutationen destabilisiert sein können und sich solche Effekte durch die Verwendung von Konsensussequenzen vermeiden lassen, haben wir einen weiteren Konsensus berechnet. Dieser sollte möglichst viele Sequenzen umfassen, die alle sicher als Propellerblätter falten. Dazu wurde mit dem Programm zur Sequenzmustererkennung PATTINPROT¹⁶ [19] die SWISSROT+SPTreMBL-Datenbank nach dem Muster S-P-D-G-x(12,16)-W-D durchsucht. Alle 351 gefundenen Sequenzen enthielten die Sequenzmotive SPDG und WD in einem Abstand von 12 bis 16 Resten. Diese Sequenzen wurden mit ClustalX aligniert und das *Alignment* manuell so bearbeitet, dass jene Spalten, in denen Sequenzlücken (*Gaps*) überwogen, gelöscht wurden. Dieses bearbeitete *Alignment* wurde mit dem Programm HMMER [25] in ein Hidden Markov Modell umgewandelt und im Mai 2004 mit diesem HMM gegen die nicht-redundante (nr) Proteindatenbank des NCBI¹⁷ gesucht. Bei dieser Suche wurden 1720 Proteine gefunden, die in ein *Alignment* geschrieben wurden, welches wiederum so modifiziert wurde, dass Positionen mit überwiegender Lücken entfernt wurden. Aus diesem *Alignment* wurde wiederum ein Profil berechnet und für jede Position wurden die drei häufigsten Aminosäuren ausgegeben. Das Ergebnis im Vergleich zu der Konsensussequenz von PkwA ist in Abb. 3.13 gezeigt.

ausgewählte Datenbanken, so z. B. Genome oder Genomgruppen durchzuführen.
<http://toolkit.tuebingen.mpg.de/>

¹⁶ http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_pattinprot.html

¹⁷ <http://www.ncbi.nlm.nih.gov/>

PkwA-Konsensus	TDYVVAVAFSPDGSLLASGSDDGTIRLWDVATGEELAVLEGH
HMM-Konsensus	TGWVNSVAFSPDGKLLASGSDDKTVRLWDLATGKCLRILTGH
	SDS R LS A A G I K VR ELIQ K
	N L S N S R K E
hmm_cons	TGWVNSVAFSPDGKLLASGSDDKTVRLWDLATGKCLRILTGH

Abbildung 3.13: Vergleiche der Konsensussequenzen von *PkwA* und dem Profil aus 1720 Proteinen. In grün die Sequenz, die aus dem Propeller von *PkwA* abgeleitet wurde, in den folgenden Zeilen sind die drei häufigsten Aminosäuren im HMM-Profil aufgelistet, wenn sie mit vergleichbaren Häufigkeiten auftreten. Ist nur eine Aminosäure angegeben, gibt es nur einen dominanten Rest an dieser Position. In der letzten Zeile steht die Sequenz, die exprimiert wurde. In rot sind Aminosäuren hervorgehoben, die sich zwischen den beiden Konsensussequenzen unterscheiden.

3.5.4 Zusammenfassung der Proteinreinigungen

Wir haben eine Propellerdomäne mit hoher Sequenzähnlichkeit zwischen den einzelnen Blättern verwendet, um eine Konsensussequenz zu konstruieren. Diese 42 Aminosäuren lange Sequenz wurde amplifiziert, so dass Konstrukte mit ein bis fünf identischen Propellerblättern exprimiert werden können.

Da sich die meisten Konstrukte nicht ohne weiteres in *E. coli* Zellen exprimieren ließen, wurden die Konstrukte mit einem, drei und vier Blättern als GST-Fusionsproteine hergestellt. Das verdoppelte Blatt ließ sich direkt exprimieren, ebenso das fünfblättrige Konstrukt, welches mit einem N-terminalen His₆-Tag hergestellt wurde. Alle Proteine ließen sich reinigen, die Charakterisierung der einzelnen Proteine wird im Folgenden beschrieben.

3.6 Biochemische Charakterisierung der Konstrukte

3.6.1 Das einzelne Blatt

Das einzelne Blatt wird wie oben beschrieben als GST-Fusion rekombinant hergestellt. Nach der Thrombinspaltung liegt das Protein mit der Sequenz wie in Abb. 3.10 beschrieben vor, von der Thrombinschnittstelle sind am N-Terminus noch zusätzlich ein Gly und ein Ser angefügt.

Das Protein lässt sich mit CD-Spektroskopie und Trp-Fluoreszenzspektroskopie auf seine Faltung hin überprüfen (s. Abb. 3.14). In der Einleitung sind in Abb. 3.2 Beispielspektren für Proteine mit verschiedenen Sekundärstrukturelementen gezeigt. Deutlich unterschiedlich sind dabei Spektren von Proteinen mit einer überwiegend ungeordneten und einer β -Faltblatt Struktur. Das CD-Spektrum des einzelnen Propellerblattes entspricht einer ungefalteten Struktur, das charakteristische Minimum liegt bei 198 nm.

Aus den Trp-Fluoreszenzspektren erhält man eine Aussage über die Hydrophobizität der Umgebung der Trp-Seitenketten des Proteins. Die großen, aromatischen Seitenkette sind meist im hydrophoben Kern des Proteins zu finden, als Referenz wird das andere Extrem, die freie Aminosäure in Lösung verwen-

det. Trp in Lösung hat typischer Weise ein Maximum im Emissionsspektrum bei 357 nm (Abb. 3.14, Teilabbildung E). Je hydrophober die Umgebung der Seitenkette wird, desto weiter verschiebt sich das Maximum zu kleineren Wellenlängen. Hier zeigt das einzelne Blatt ein Maximum bei 350 nm, welches nahezu dem Spektrum der freien Aminosäure entspricht. Diese Indizien deuten darauf hin, dass das einzelne Blatt nicht spontan falten kann.

Um den Einfluss des Lösungsmittels auf die Faltungseigenschaften des Proteins zu überprüfen, wurden zahlreiche unterschiedliche Bedingungen getestet. In den Bildabschnitten C und D wird gezeigt, dass sich die Spektren entscheidend ändern, wenn 1% des Detergenz SB10 zu der Proteinklösung gegeben wird. Das Emissionsmaximum im Trp-Fluoreszenzspektrum verschiebt sich zu kleineren Wellenlängen und das Minimum im CD-Spektrum nimmt die Gestalt an, wie wir sie für alle gefalteten β -Propeller im Rahmen dieser Arbeit sehen. Wird das Detergenz wieder entfernt, dann zeigt sich, dass die Veränderung der Spektren reversibel ist und die Spektren zeigen erneut alle Charakteristiken eines ungefalteten Proteins (Teilbild F). Damit das einzelne Blatt Sekundärstrukturelemente ausbilden kann, ist demnach eine hydrophobe Umgebung nötig. Diese wird in diesem Fall durch das Detergenz bereitgestellt, allerdings schirmen die Detergenzmoleküle die einzelnen Proteinmoleküle auch gegeneinander ab und verhindern so eine Oligomerisierung. In den Experimenten ohne Detergenz müssten die Voraussetzungen für die Faltung durch die anderen Proteinmoleküle geschaffen werden. Anscheinend sind die Kontaktzeiten zwischen zwei Molekülen jedoch zu kurz, um zu einer Faltung und Oligomerisierung zu führen. Versuche, die lokale Proteinkonzentration durch Zugabe von PEG und ähnlichen Substanzen zu erhöhen, waren ebenfalls nicht erfolgreich.

Es wurde weiterhin versucht, das einzelne Blatt gemeinsam mit dem TINAR-Protein in einen Komplex zu inkorporieren. Diese Versuche werden im Anschluss bei der Charakterisierung der Mischungsexperimente beschrieben.

3.6.2 Alternative Propellerblätter

Die in Abschnitt 3.5.3 beschriebenen Varianten des einzelnen Propellerblattes wurden jeweils als GST-Fusionen exprimiert und nach der Abtrennung des GST-Teils charakterisiert. Die spektroskopischen Daten entsprachen dabei durchgehend denen des einzelnen Blattes, keines der Konstrukte gab Hinweise auf einen gefalteten Zustand (Daten nicht gezeigt da identisch mit Abb. 3.14, Teilabbildung A und B).

3.6.3 Das duplizierte Blatt - TINAR

Das Konstrukt, welches für das duplizierte Propellerblatt codiert, und das zugehörige Protein werden wie bereits erwähnt auch als TINAR bezeichnet. TINAR ließ sich als unabhängiges Protein rekombinant herstellen, die Reinigung über die Phenylsepharosesäule, bei der die Elution mit 30% Ethanol in H_2O erfolgt, wurde bereits beschrieben. Es kann spekuliert werden, ob das Protein während der Elution nicht entfaltet vorliegt und sich im Laufe der anschließenden Reinigung erneut faltet. Zumindest liegt am Ende der Reinigung ein gefaltetes Protein vor, welches spontan einen stabilen Propeller bildet. Die biochemische Charakterisierung ist in den Abb. 3.15 bis 3.18 vorgestellt. Die wichtigsten Ergebnisse sind:

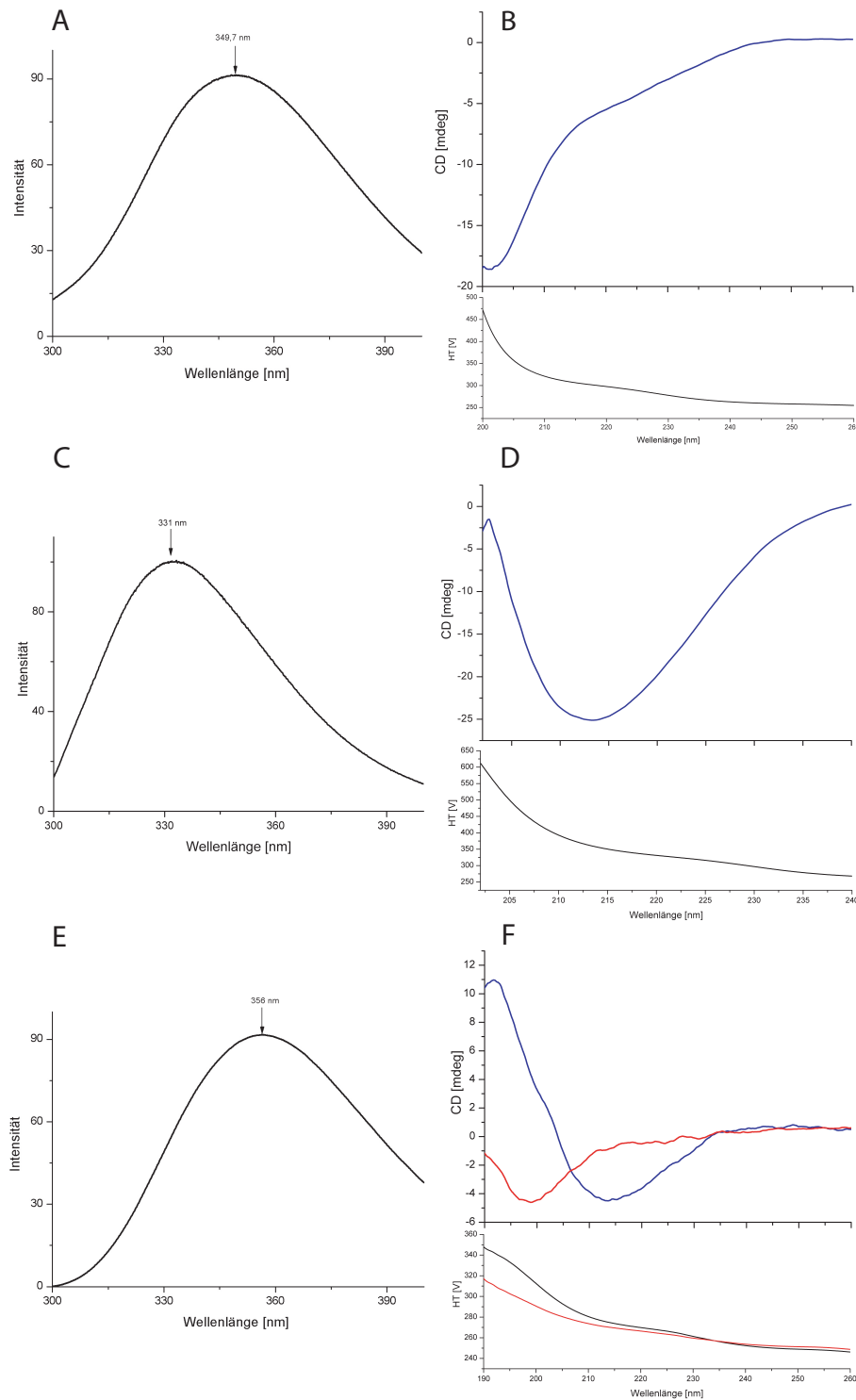


Abbildung 3.14: Spektroskopische Daten des einzelnen Propellerblattes. (A) Trp-Fluoreszenzspektrum des einzelnen Propellerblattes in PBS. (B) CD-Spektrum ebenfalls in PBS. (C) Trp-Fluoreszenzspektrum mit 1% SB10 Detergenz in der Lösung. (D) CD-Spektrum mit 1% SB10 Detergenz. (E) Referenz Trp-Fluoreszenzspektrum für eine vollständig hydrophile Umgebung: die freie Aminosäure Trp in PBS. (F) Vergleich der CD-Spektren vor (blau) und nach (rot) der Entfernung des Detergenz SB10.

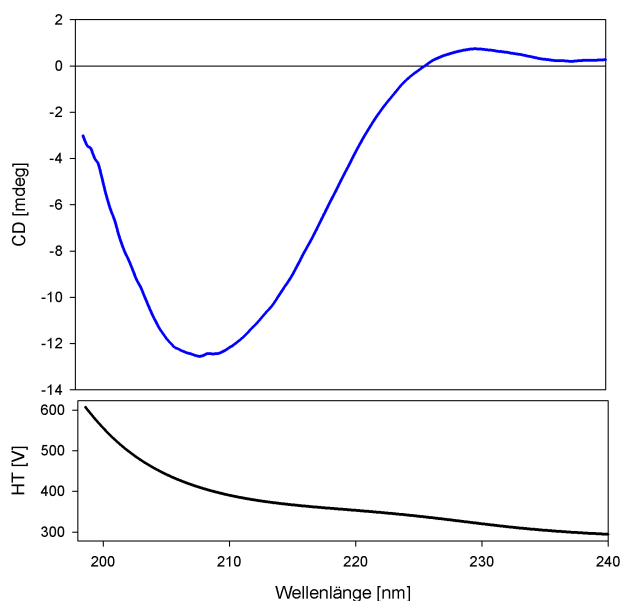


Abbildung 3.15: *CD-Spektrum des oligomerisierten Doppelblattes. Parameter der Messung: Sensitivity Standard; 10 akkumulierte Spektren; 1 mm Küvetten-dicke; 0,24 mg/ml Proteinkonzentration (= 27 μ M); Puffer PBS 1:1 mit H₂O verdünnt.*

- Das CD-Spektrum weist auf einen gefalteten Zustand hin. Das Spektrum zeigt ein charakteristisches Minimum bei < 210 nm und ein kleines Maximum bei 225-230 nm. Diese beiden Extrema entsprechen zwar nicht dem klassischen β -Faltblattspektrum, treten aber konsistent bei allen in dieser Arbeit beschriebenen assemblierten Propellern auf. Da dieses Spektrum unerwartet und untypisch war, haben wir einen TINAR-Propeller zusätzlich mit NMR-spektroskopischen Methoden untersucht. Dabei konnten wir zeigen, dass die Signale (*Peaks*) breit sind und über das Spektrum verteilt auftreten, beides Hinweise auf eine gefaltete Proteinstruktur (M. Coles hat diese Messung durchgeführt und die Datenauswertung dieses Experiments übernommen). Die verbreiterten Signale im gefalteten Protein im Gegensatz zu wenigen, scharfen, klar lokalisierten Signalen im ungefalteten Protein sind dadurch zu begründen, dass im gefalteten Protein jede Aminosäure eine etwas andere chemische Umgebung hat, während im ungefalteten Protein alle Reste mehr oder weniger gleich dem Lösungsmittel ausgesetzt sind.

- Das Emissionsspektrum für die Tryptophanfluoreszenz zeigt eine deutliche Rotverschiebung, wenn das Protein mit 6 M GuaHCl statt mit Puffer verdünnt wird. Ein Emissionsmaximum um 358 nm ist für Tryptophan in wässriger Lösung charakteristisch. Je weiter die Verschiebung des Intensitätsmaximums im gefalteten Protein zu kleineren Wellenlängen ist, desto hydrophober ist die Umgebung der Seitenkette.

- Die Auswertung der Lichtstreudaten zusammen mit den Chromatogrammen der Gelfiltrationsläufe weisen auf eine einheitliche Propellergröße mit acht

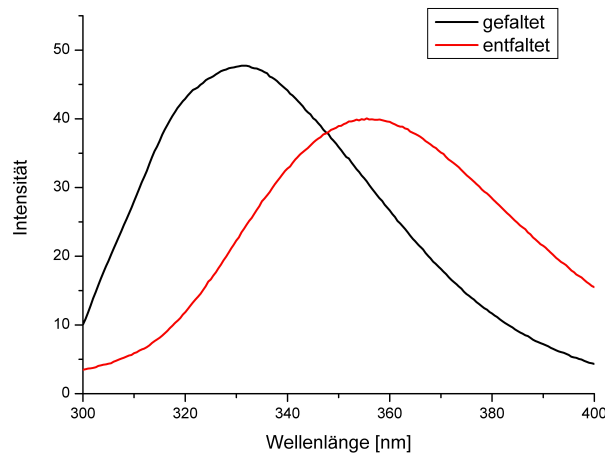


Abbildung 3.16: *Tryptophanfluoreszenzspektren von gefaltetem und oligomisiertem bzw. entfaltetem Doppelblatt. Parameter der Messung: Sensitivity low; 10 akkumulierte Spektren; Anregungswellenlänge 280 nm; 0,34 mg/ml Proteinkonzentration (= 39 μ M); Puffer 20 mM MOPS; 50 mM NaCl; pH 7 bzw. 6 M GuaHCl.*

Propellerblättern hin, s. Abb. 3.17. Dabei lässt sich die mittlere Molmasse von 33,9 kDa bestimmen. Das würde einem Homotetramer des Doppelblattes entsprechen, das einzelne Doppelblatt hat eine theoretische Molmasse von 8769 Da, das Tetramer dementsprechend 35,1 kDa. Bei der Auswertung fällt auf, dass die berechneten Molmassen über das Absorptionsmaximum konstant sind, die Moleküle, die unter diesem Maximum von der Säule eluiert werden, haben alle dieselbe Masse (keine Mischung verschieden großer Moleküle). Eine solche Massenverteilung wird als monodispers bezeichnet.

In Abb. 3.30 sind verschiedene Propellerkonstrukte im Vergleich miteinander aufgetrennt. Im rechten Teil sind als Referenzproteine Propeller aufgetragen, die aus sechs, acht und zehn Propellerblättern bestehen. In der dritten Spur von links ist das Doppelblatt zusammen mit dem vierblättrigen Propeller, der als Referenz für einen achtblättrigen Propeller fungiert, in einer Spur aufgetrennt. Die beiden Proteine zeigen im nativen Acrylamidgel dieselben Laufeigenschaften, woraus wir schließen, dass TINAR ebenfalls einen achtblättrigen Propeller bildet. Diese Interpretation ist konsistent mit den Massenbestimmungen mittels statischer Lichtstreuung.

- Der Proteinkomplex lässt sich thermisch denaturieren, der Schmelzpunkt liegt bei ca. 52°C. Diese Entfaltung ist größtenteils reversibel. Der Übergang zwischen ge- und entfaltetem Protein wird neben der Schmelzkurve auch in den parallel aufgenommenen CD-Spektren deutlich. Das Maximum bei 225-230 nm ist im Ausgangszustand vorhanden. Die Intensität des Maximums nimmt im Bereich der Schmelztemperatur rapide ab, ist bei 90°C nicht mehr vorhanden und bildet sich beim Abkühlen der Probe wieder aus. Dies ist ein Indiz dafür, dass dieses Maximum als Indikator für die Proteinfaltung verwendet werden kann. Die Renaturierungskurve ist nicht identisch mit der Schmelzkurve, da bei hohen Temperaturen geringe Mengen des Proteins präzipitieren.

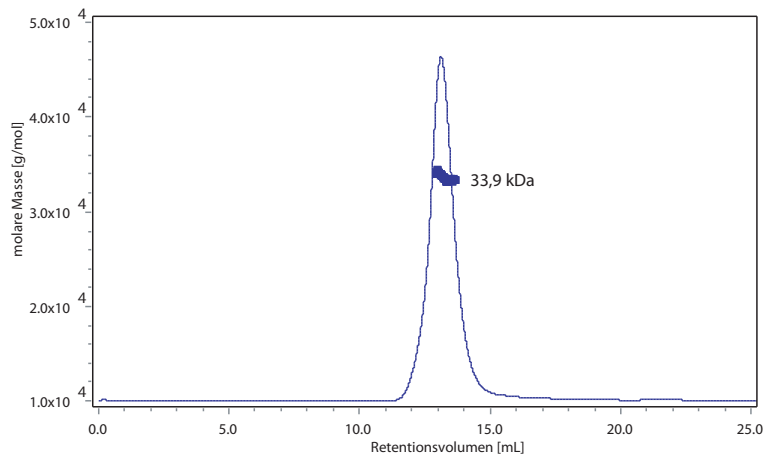


Abbildung 3.17: Bestimmung der Molmasse des TINAR-Proteins. Die mittlere Molmasse im Peakmaximum wurde mit 33,9 kDa bestimmt. Die dünne blaue Line entspricht der UV-Absorption bei 280 nm. Parameter: Flussgeschwindigkeit 0,5 ml/min; Proteinkonz. 1,2 mg/ml; 0,5 ml geladen; Superose 6 Säule.

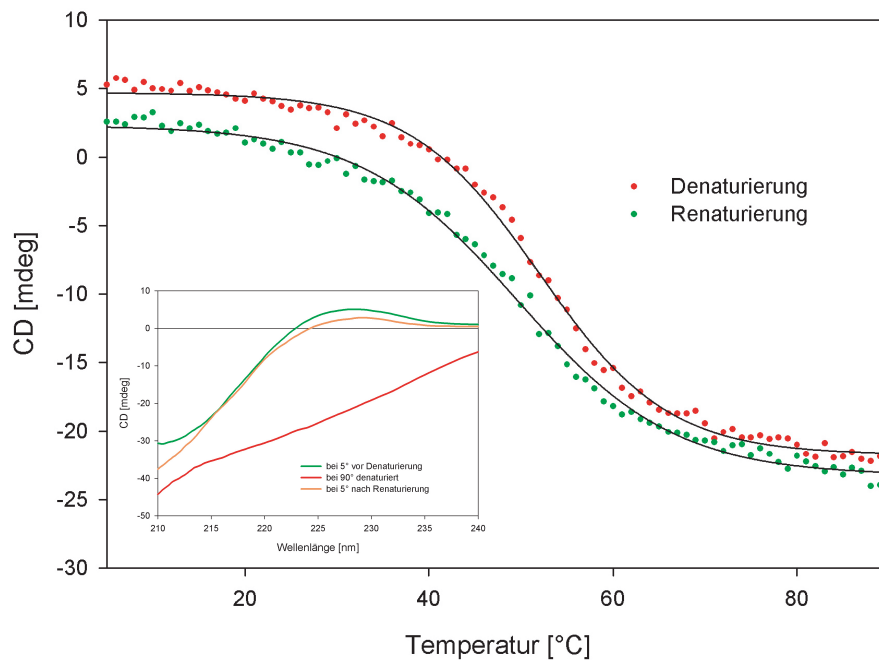


Abbildung 3.18: Schmelz- und Renaturierungskurve des oligomerisierten Doppelblattes. Parameter der Messung: Temperaturgradient $1^\circ\text{C}/\text{min}$; Messung bei 228 nm; 1 mm Küvettdicke; 0,38 mg/ml Proteinkonzentration ($= 44 \mu\text{M}$); Puffer PBS 2:1 mit H_2O verdünnt. Klein in der Abbildung eingefügt sind CD-Spektren, die während der Ent- bzw. Rückfaltung aufgenommen wurden (1 akkumuliertes Spektrum).

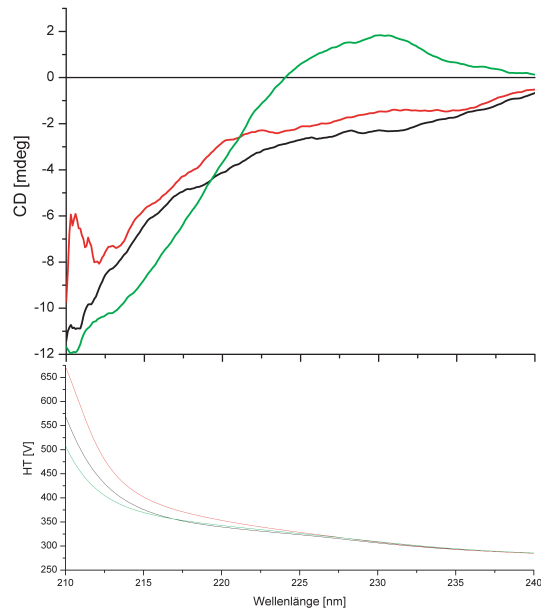


Abbildung 3.19: CD-Spektren des Doppelblattes unter verschiedenen Entfaltungsbedingungen. Die grüne Kurve wurde mit 3 M GuaHCl, die rote mit 6 M GuaHCl und die schwarze mit 6 M Harnstoff aufgenommen. Parameter der Messung: 1 mm Küvettendicke; 0,34 mg/ml Proteinkonzentration (= 39 μ M).

3.6.4 Daten des entfalteten Doppelblattes

Im folgenden Abschnitt 3.7 werden Mischungsexperimente zwischen verschiedenen Konstrukten beschrieben. Eine wichtige Voraussetzung, damit sich heterooligomere Proteinkomplexe bilden können, ist, dass sich die Propeller entfalten, um sich dann in anderen Stöchiometrien neu zusammenzulagern. Um die Entfaltung der Proteine in 6 M Guanidin Hydrochlorid (GuaHCl) zu verifizieren, wurde das Doppelblatt sowohl im CD als auch in Bezug auf sein Trp-Fluoreszenzspektrum vermessen. In Abb. 3.16 sind die Emissionsspektren von ge- und entfaltetem TINAR verglichen. Dabei wird deutlich, dass sich die Trp-Seitenketten in 6 M GuaHCl in einer vergleichbar hydrophilen Umgebung befinden wie freies Trp in Lösung (Maxima der Emissionsspektren in beiden Fällen bei 356 nm). Die CD-Spektren unter verschiedenen Bedingungen sind in Abb. 3.19 gezeigt. Dabei wird deutlich, dass in 3 M GuaHCl die Sekundärstruktur der Doppelblattes noch vorhanden zu sein scheint, wie an dem Maximum um 230 nm zu sehen. Dieses Maximum ist auch bei dem 14-blättrigen Propeller, dessen Kristallstruktur in Kapitel 4 beschrieben wird, charakteristisch und verschwindet, wenn das Protein thermisch denaturiert wird (Abb. 4.5). Löst man das Doppelblatt hingegen in 6 M GuaHCl oder 6 M Harnstoff, so tritt das Maximum bei 230 nm nicht mehr auf und die Spektren gleichen denen von entfaltenen Proteinen (die Spektren sind nur bis 210 nm gezeigt, da bei den eingesetzten Konzentrationen an GuaHCl bzw. Harnstoff eine Messung bis <200 nm nicht möglich ist).

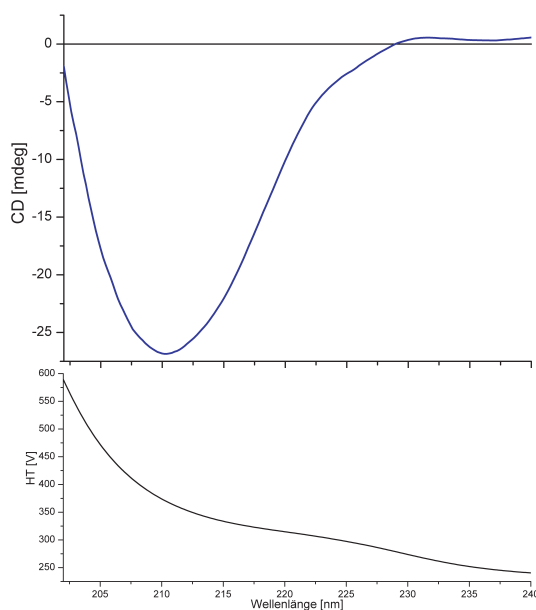


Abbildung 3.20: *CD-Spektrum des oligomerisierten TriNAR-Proteins. Parameter der Messung: Sensitivity Standard; 10 akkumulierte Spektren; 1 mm Küvetten dicke; 0,5 mg/ml Proteinkonzentration (= 37 μ M); Puffer PBS.*

Des Weiteren wurde TINAR in 6 M GuaHCl auf einer analytischen Gel-filtrationssäule aufgetrennt und mit statischer Lichtstreuung ausgewertet. Die Analyse ergab, dass das Protein in zwei Absorptionsmaxima_{280nm} von der Säule eluierte, deren mittlere Massen sich zu 10,5 und 5,3 kDa bestimmen ließen. Aus Referenzmessungen mit Rinderserumalbumin (BSA) ist bekannt, dass die Massenbestimmung mittels statischer Lichtstreuung in 6 M GuaHCl nicht dieselbe Genauigkeit aufweist wie in verdünnten Puffern (es muss mit einer Abweichung von max. 25% gerechnet werden). Dennoch ist die Messgenauigkeit ausreichend, um von einem einzelnen Doppelblatt und nicht von einem oligomeren TINAR-Komplex auszugehen (TINAR einzeln 8,8 kDa / Tetramer: 35,2 kDa). Wir gehen daher davon aus, dass die Propellerkonstrukte in 6 M GuaHCl entfaltet vorliegen.

3.6.5 Das dreiblättrige Propellerkonstrukt - TriNAR

Das Konstrukt, welches für drei konsekutive Propellerblätter codiert, und das zugehörige Protein werden wie bereits erwähnt auch als TriNAR bezeichnet. Der dreiblättrige Proteinbaustein ließ sich ausschließlich als GST-Fusion rekombinant herstellen, eine Expression des Proteins allein war im Gegensatz zum Doppelblatt nicht möglich. Die Reinigung erfolgte über eine Glutathion-Affinitätssäule, anschließend wurden der GST- und der Propellerteil spezifisch durch eine Thrombinspaltung getrennt. Die Glutathion-S-Transferase wurde durch eine weitere Passage über die Affinitätssäule von dem Propeller getrennt und dieser abschließend mittels Gelausschlußchromatographie gereinigt (Reini-

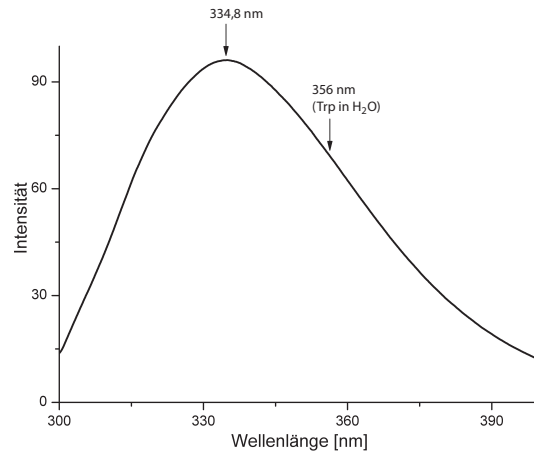


Abbildung 3.21: *Tryptophanfluoreszenzspektren von gefaltetem und oligomerisiertem TriNAR Protein. Parameter der Messung: Sensitivity medium; 10 akkumulierte Spektren; Anregungswellenlänge 293 nm; 0,25 mg/ml Proteinkonzentration (= 18 μ M); Puffer PBS.*

gung siehe Abschnitt 3.4.1.

Die Daten der biochemischen Charakterisierung sind in den Abb. 3.20 bis 3.22 gezeigt. Die wichtigsten Ergebnisse sollen an dieser Stelle diskutiert werden:

- Das CD-Spektrum und das Fluoreszenzemissionsspektrum ähneln denen des Doppelblattes. Das Maximum bei 230 nm ist nur sehr schwach ausgeprägt, aber dennoch vorhanden, das Minimum bei 210 nm gleicht denen der anderen Konstrukte in dieser Arbeit. Ebenso ist das Emissionsspektrum der Trp-Fluoreszenz gegenüber dem freien Trp stark blauverschoben, die Maxima des TriNAR-Proteins (335 nm) und des Doppelblattes (332 nm) gleichen sich annähernd.

- TriNAR faltet und assembliert spontan zu einem stabilen Propeller. Die Molmasse des Komplexes wurde mit statischer Lichtstreuung ermittelt (Abb. 3.22) und ergab, dass sich der dreiblättrige Propellerbaustein ausschließlich zu einem sechsblättrigen Propeller zusammenlagert. Das einzelne Protein hat eine berechnete Masse von 13,6 kDa. Damit wird für den sechsblättrigen Propeller eine Masse von 27,2 kDa erwartet, die Messung ergab 27,8 kDa. Die Massenverteilung über das Absorptionsmaximum ist nahezu konstant, es kann von einer monodispersen Proteinpopulation ausgegangen werden. Auffällig ist jedoch, dass in diesem Fall sowie in einigen der folgenden Messungen die Molmassen mit zunehmendem Retentionsvolumen zunehmen. Nach der Theorie der Gelausschlusschromatographie sollten jedoch größere Moleküle zuerst von der Säule eluieren. Dieses Phänomen ließ sich nicht abschließend erklären, allerdings bleibt festzuhalten, dass die scheinbare Massenzunahme geringer ist als die Messungenauigkeit der angewandten Methode.

Konsistent mit den Daten der statischen Lichtstreuung ist die Auftrennung verschiedener Propellerkomplexe im nicht denaturierenden Acrylamidgel, wie in Abb. 3.30 gezeigt. Dabei sind in den rechten drei Spuren des Gels die Proteinkomplexe aufgetragen, die sich bei spontaner Assemblierung der drei-, vier- und

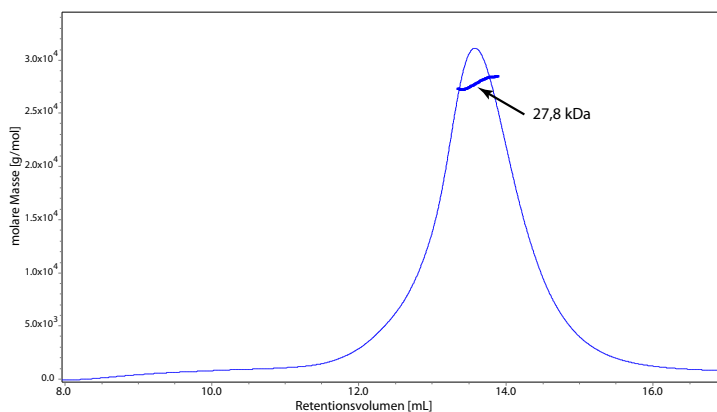


Abbildung 3.22: Bestimmung der Molmasse des TriNAR-Proteins. Die mittlere Molmasse im Peakmaximum wurde mit 27,8 kDa bestimmt. Die dünne blaue Line entspricht der UV-Absorption bei 280 nm. Parameter: Flussgeschwindigkeit 0,5 ml/min; Proteinkonz. 1 mg/ml; 0,5 ml geladen; Superose 6 Säule.

fünfblättrigen Propellerbausteine bilden. Dabei wird deutlich, dass der dreiblättrige Baustein einen kleineren Komplex bildet als die anderen Propeller. Es kann also angenommen werden, dass der dreiblättrige Propellerbaustein dimerisiert und einen sechsblättrigen Propeller bildet.

3.6.6 Das vierblättrige Propellerkonstrukt - FoNAR

Das nächstgrößere Protein ist der Baustein mit vier Propellerblättern in einer durchgehenden Polypeptidkette. Das Konstrukt bzw. das zugehörige Protein werden analog zu der bereits eingeführten Nomenklatur als FoNAR bezeichnet. Das vierblättrige Protein konnte wie die einzelnen Blätter und das dreiblättrige Konstrukt nur als GST-Fusion exprimiert werden. Die Reinigung erfolgte über eine Glutathion-Affinitätssäule mit anschließendem Thrombinverdau, um den GST-Teil von dem Propelleranteil abzuspalten. Die Trennung der beiden Proteine erfolgte über eine weitere Affinitätssäule, um die Glutathion-S-Transferase abzureichern, und eine abschließende Gelausschlußchromatographie.

Die biochemische Charakterisierung ergab, dass sich das FoNAR-Protein sehr ähnlich zu den bereits bekannten Proteinen verhält. Die experimentellen Daten sind in den Abb. 3.23 bis 3.25 zusammengefasst. Die Ergebnisse werden hier diskutiert:

- Das CD-Spektrum und das Fluoreszenzemissionsspektrum ähneln denen des Doppelblattes und des TriNAR-Proteins. Das Maximum bei 230 nm ist wiederum nur schwach ausgeprägt aber dennoch vorhanden, das Minimum bei 210 nm gleicht denen der anderen Konstrukte in dieser Arbeit.

- In der PDB sind Beispiele für Proteine mit vier Propellerblättern hinterlegt, daher wäre es für das FoNAR-Konstrukt theoretisch denkbar, dass es einen vierblättrigen Propeller bildet. Die Bestimmung der Molmasse des FoNAR-Proteins ergibt jedoch, dass der Proteinbaustein dimerisiert und einen achtblättrigen Pro-

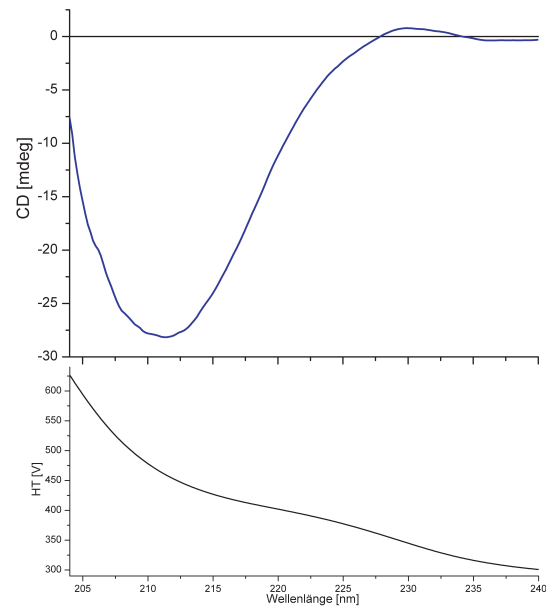


Abbildung 3.23: CD-Spektrum des oligomerisierten FoNAR-Proteins. Parameter der Messung: Sensitivity Standard; 3 akkumulierte Spektren; 1 mm Küvettenstärke; 0,5 mg/ml Proteinkonzentration (= 29 μ M); Puffer PBS 1:1 verdünnt mit H₂O.

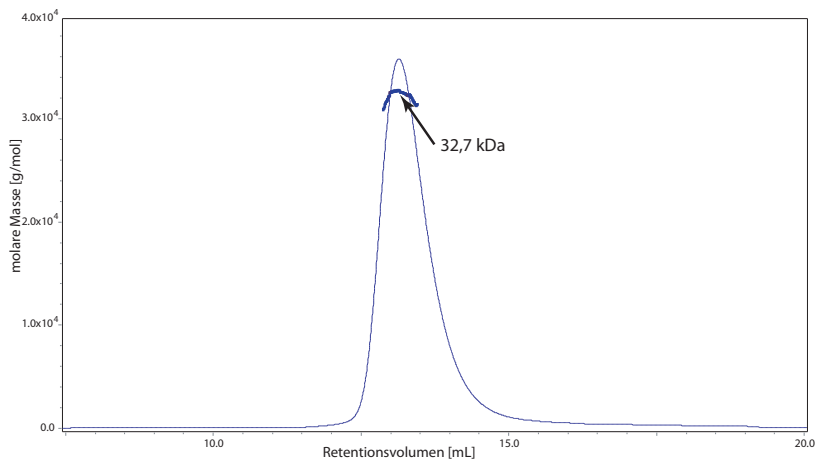


Abbildung 3.24: Bestimmung der Molmasse des FoNAR-Proteins. Die mittlere Molmasse im Peakmaximum wurde mit 32,7 kDa bestimmt. Die dünne blaue Linie entspricht der UV-Absorption bei 280 nm. Parameter: Flussgeschwindigkeit 0,5 ml/min; Proteinkonz. 1,2 mg/ml; 0,5 ml geladen; Superose 6 Säule.

peller bildet. Die bestimmte Molmasse beträgt 32,7 kDa, das einzelne FoNAR-Protein hat eine theoretische Masse von 17,5 kDa. Die Massenverteilung im Absorptionsmaximum ist nahezu monodispers, die Abweichung zwischen erwarteter und bestimmter Masse von 3 kDa bzw. von fast 10% lässt sich durch die Messungenauigkeit erklären (die Molmassenbestimmung wurde reproduziert und ergab in jedem Fall ein Molgewicht von ca. 33 kDa für den Komplex). Man kann jedoch in jedem Fall ausschließen, dass sich ein vierblättriger Propeller oder ein Homotrimer, der einem 12-blättrigen Propeller entspräche, bildet.

Diese Lichtstreuendaten lassen sich auch mit der Auftrennung verschiedener Propellerkomplexe im nicht denaturierenden Acrylamidgel in Einklang bringen. In Abb. 3.30 sind im rechten Teil das TriNAR-Protein, welches einen sechsblättrigen Propeller bildet, das vierblättrige Propellerkonstrukt und das im nächsten Abschnitt diskutierte FiNAR-Konstrukt, welches einen 10-blättrigen Propeller bildet, aufgetrennt. Dabei verhält sich der FoNAR-Komplex in seinen Laufeigenschaften im Gel so, dass er eine intermediäre Laufgeschwindigkeit zwischen dem sechs- und dem 10-blättrigen Propeller aufweist. Zusammengenommen gehen wir auf Grund der Lichtstreuendaten und der Analytik im Nativgel davon aus, dass der vierblättrige Propellerbaustein einen achtblättrigen β -Propeller bildet. Eine genaue Analyse der bekannten vierblättrigen Propeller zeigt auch, dass bei diesen der äußere β -Strang im Vergleich zu den Strängen A bis C deutlich stärker verdreht ist. Weiterhin ist bei diesen Propellern der Strang D entweder durch einen *Loop* im Strang oder eine einfache α -helikale Windung zwischen den Blättern verlängert. Diese besonderen Anforderungen, die an den äußeren Strang eines Propellerblattes eines vierblättrigen β -Propellers gestellt werden, können von der von uns exprimierten Sequenz nicht erfüllt werden. Die von uns verwendete Sequenz wurde von einem ursprünglich siebenblättrigen Propeller abgeleitet. In diesem Zusammenhang ist es erstaunlich, dass die von uns exprimierten und untersuchten Proteine in der Lage sind zu falten und Propeller zu bilden, obwohl die eingesetzten Bausteine in keinem Fall die Rückbildung eines siebenblättrigen Propellers erlauben. Statt dessen bilden die Propellerkonstrukte sowohl sechs- als auch achtblättrige β -Propeller.

- Die Faltung des FoNAR-Komplexes lässt sich reversibel denaturieren. Die Schmelzkurve in Abb. 3.25 zeigt, dass das Protein in einem einstufigen kooperativen Prozess entfaltet. Der Schmelzpunkt für das Protein wird mit 63 °C berechnet. Die Entfaltung des Proteins lässt sich auch durch die Veränderung der zeitgleich zur Erwärmung gemessenen CD-Spektren verfolgen (kleines Bild in der Abbildung). Sowohl das Maximum bei 230 nm als auch das Minimum bei 210 nm gehen mit zunehmender Erwärmung der Probe verloren (Übergang des schwarzen Ausgangsspektrums vor dem Schmelzvorgang in das rote Spektrum bei 85 °C mit dem intermediären grauen Spektrum). Das Ausgangsspektrum gleicht dem Spektrum der anderen β -Propellerproteine in dieser Arbeit, das rote CD-Spektrum entspricht dem eines entfalteten Proteins (eine Messung bis zum eigentlichen Minimum bei 198 nm lässt sich bei den eingesetzten Pufferkonzentrationen und den hohen Temperaturen nicht durchführen).

Es ist wichtig anzumerken, dass die Entfaltung reversibel ist. Durch einfaches Abkühlen der Probe bildet sich die Ausgangsstruktur des Proteins erneut aus (belegt durch ein dem Ausgangsspektrum nahezu identisches grünes CD-Spektrum). Die Rückfaltungskurve ist nicht absolut identisch mit der Schmelzkurve. Das Protein hat bei der Abkühlung eine etwas niedrigere Übergangstemperatur, was für Proteine nicht ungewöhnlich ist und auch schon bei dem Dop-

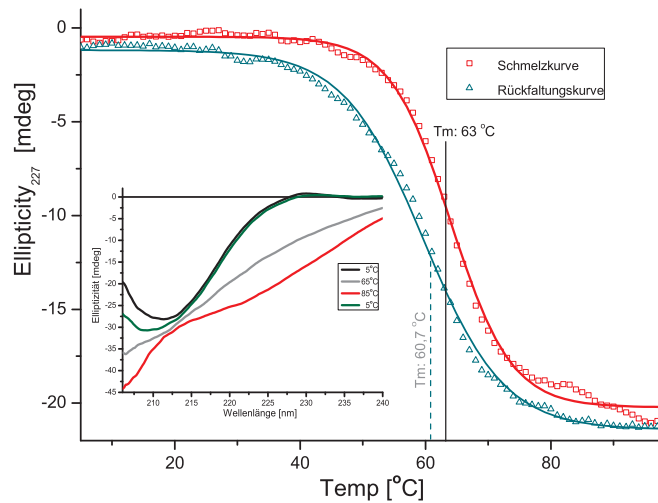


Abbildung 3.25: Schmelz- und Renaturierungskurve des Komplexes, den das FoNAR-Protein spontan bildet. Die Schmelztemperatur T_m während des Schmelzvorgangs lässt sich mit 63°C bestimmen, für die Abkühlung beträgt T_m $60,7^\circ\text{C}$. Parameter der Messung: Temperaturgradient $1^\circ\text{C}/\text{min}$; Messung bei 227 nm ; 1 mm Küvettendicke; $0,5\text{ mg/ml}$ Proteinkonzentration ($= 29\ \mu\text{M}$); Puffer PBS 1:1 verdünnt mit H_2O . Klein in der Abbildung eingefügt sind CD-Spektren, die während der Entfaltung aufgenommen wurden (3 akkumulierte Spektren) bzw. die grüne Kurve wurde nach der erneuten Abkühlung aufgenommen.

pelblatt beobachtet wurde. Im Vergleich der beiden achtblättrigen Propeller ist das Homodimer des vierblättrigen Bausteins deutlich stabiler als das Homotrimer des Doppelblattes (63°C im Vergleich mit 52°C).

Die erhöhte Stabilität des Proteins mit weniger Polypeptidketten könnte ein Grund dafür sein, dass in der Natur (und in der PDB) nahezu keine natürlichen assemblierten Propeller vorkommen. Uns ist lediglich ein solches Protein bekannt, ein Fucose-bindendes Lectin [54]. Dieses bakterielle Protein bildet ein Trimer aus identischen Untereinheiten, die jeweils aus zwei Propellerblättern bestehen. Das Lectin bildet also einen sechsblättrigen Propeller. Andere oligomere Strukturen sind nicht bekannt, was den Schluss erlaubt, dass der evolutionäre Schritt nach der Genamplifikation, die Fusion der einzelnen Abschnitte zu einem durchgehenden Protein, sehr schnell abzulaufen scheint. So bestehen auch die repetitiven Propeller, die im Genom von *Nostoc punctiforme* gefunden werden, aus durchgehenden Polypeptidketten (diese Proteine werden im folgenden Kapitel noch genauer diskutiert). Bei diesen Proteinen gehen wir auf Grund der hohen Sequenzidentität zwischen den Blättern davon aus, dass die Amplifikation erst ein kurz zurückliegendes Ereignis ist - die einzelnen Blätter hatten seit der Amplifikation zu einem vollständigen Propeller noch nicht genügend Zeit, um auf der Sequenzebene zu divergieren. Dennoch codieren diese Gene für Propeller, die aus einer durchgehenden Aminosäurekette bestehen, nach der Amplifikation folgt offensichtlich sofort die Fusion.

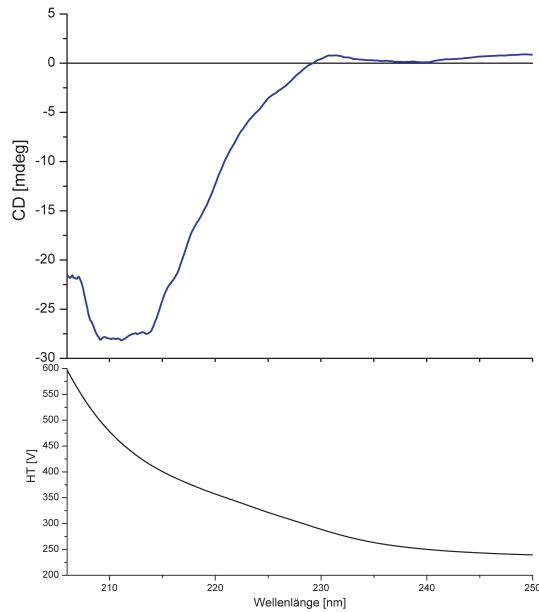


Abbildung 3.26: *CD-Spektrum des oligomerisierten FiNAR-Proteins. Parameter der Messung: Sensitivity Standard, 3 akkumulierte Spektren, 1 mm Küvettendicke; 0,45 mg/ml Proteinkonzentration (= 20 μ M); Puffer 5 mM MOPS; 15 mM Tris; 13 mM NaCl; 38 mM KF; pH = 7,5.*

3.6.7 Das fünfblättrige Propellerkonstrukt - FiNAR

Das Konstrukt, welches für fünf konsekutive Propellerblätter codiert und das zugehörige Protein werden wie bereits erwähnt auch als FiNAR bezeichnet. FiNAR ließ sich sowohl als GST-Fusionsprotein als auch als unabhängiges Protein mit einem N-terminalen His₆-Tag rekombinant herstellen. Die Reinigung über eine Anionenaustauschchromatographie gefolgt von einer Affinitätssäule und einer abschließenden Gelausschlußchromatographie wurde bereits in Abschnitt 3.4.3 beschrieben. Das GST-Fusionsprotein wurde nicht weiter verwendet, da die alternative Expressions- und Reinigungsstrategie höhere Ausbeuten lieferte.

Die biochemische Charakterisierung wird in den Abb. 3.26 bis 3.29 vorgestellt. Die wichtigsten Aussagen sind:

- Das CD-Spektrum und das Fluoreszenzemissionsspektrum ähneln denen der bereits vorgestellten Proteine. So fallen wiederum das breite Minimum bei 210 nm und das schwach ausgeprägte Maximum bei 230 nm auf. Aus dem Fluoreszenzspektrum in Abb. 3.27 lässt sich schließen, dass die Trp-Seitenketten wasserabgeschirmt im hydrophoben Kern des Proteins liegen.

- Die Molmassenbestimmung ergab ein unerwartetes Ergebnis: das fünfblättrige Protein scheint sich zu einem Dimer zusammenzulagern und einen 10-blättrigen Propeller zu bilden. Drei unabhängige Indizien sprechen eindeutig für diese Schlussfolgerung, kein Experiment lieferte gegensätzliche Interpretationsmöglichkeiten. Zunächst wurde die Masse des gereinigten FiNAR-Proteins mit statischer Lichtstreuung bestimmt (Abb. 3.28). Der einfache fünfblättrige

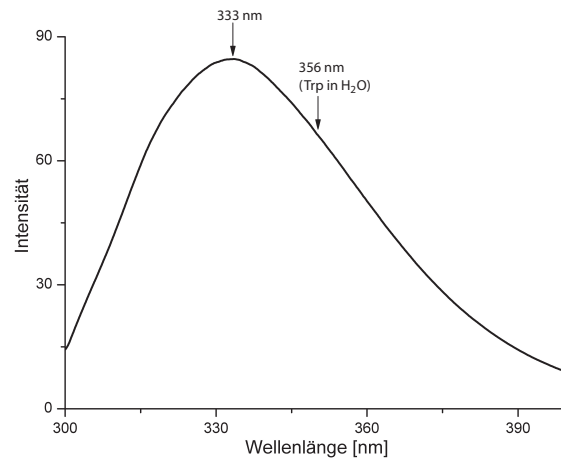


Abbildung 3.27: Tryptophanfluoreszenzspektren von gefaltetem und oligomerisiertem FiNAR Protein. Parameter der Messung: Sensitivity medium; 10 akkumulierte Spektren; Anregungswellenlänge 293 nm; 0,25 mg/ml Proteinkonzentration (= 11 μ M); Puffer PBS.

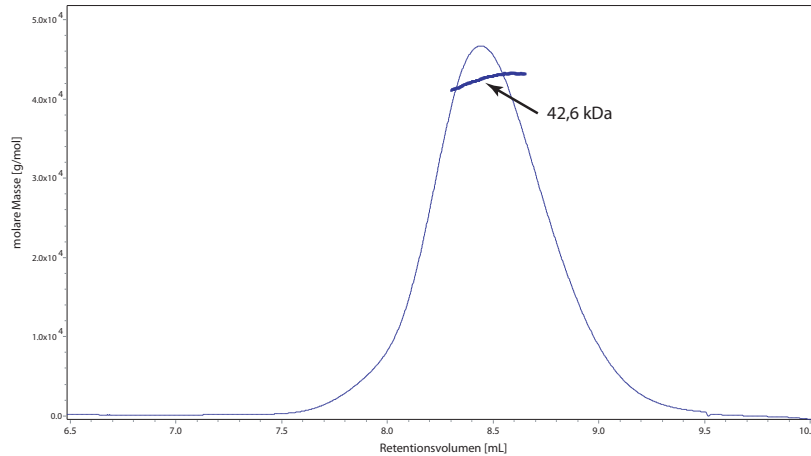


Abbildung 3.28: Bestimmung der Molmasse des FiNAR-Proteins mit statischer Lichtstreuung. Die mittlere Molmasse im Peakmaximum wurde mit 42,6 kDa bestimmt. Die dünne blaue Linie entspricht der UV-Absorption bei 280 nm. Parameter: Flussgeschwindigkeit 0,5 ml/min; Proteinkonz. 11,25 mg/ml; 0,2 ml geladen; Superose 6 Säule.

AUC sedimentation velocity experiment

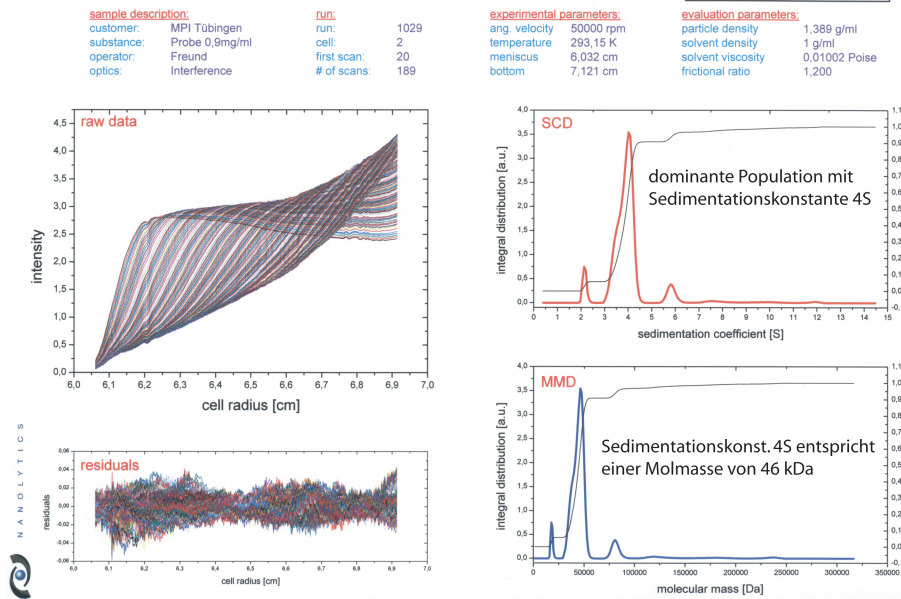


Abbildung 3.29: Molmassenbestimmung des FiNAR Proteins mittels analytischer Ultrazentrifugation. Zunächst wird der Sedimentationskoeffizient bestimmt und aus diesem dann die Molmasse berechnet. Die entsprechenden Profile sind beschriftet.

Baustein ohne His₆-Tag, so wie er für dieses Experiment eingesetzt wurde, hat eine theoretische Molmasse von 22,1 kDa. Die Analytik mittels Gelfiltrationschromatographie ergibt, dass das Protein als homogene Population in einem einzelnen symmetrischen Absorptionsmaximum von der Säule eluiert. Die mittlere molare Masse im Absorptionsmaximum lässt sich zu 42,6 kDa bestimmen, die Massenverteilung ist nahezu monodispers. Dieses Molekulargewicht und das kleine Retentionsvolumen lässt sich nur durch ein FiNAR-Dimer erklären.

In Übereinstimmung mit dieser Beobachtung ist die Auftrennung in einem nicht denaturierenden Acrylamidgel, wie in Abb. 3.30 gezeigt. Die drei rechten Spuren des Gels zeigen, dass der fünfblättrige Propeller als deutlich höhere Bande im Gel läuft als ein sechs- bzw. achtblättriger Propeller. Dieser Vergleich stützt ebenfalls das Postulat des 10-blättrigen Propellers.

Als drittes unabhängiges Experiment wurde das Sedimentationsverhalten des Proteins mittels analytischer Ultrazentrifugation untersucht. Bei der analytischen Ultrazentrifugation wird das Sedimentationsverhalten eines Proteinkomplexes in Lösung bestimmt. Die Probe wird dazu in einer optischen Zelle unter hohen Winkelgeschwindigkeiten in einer Zentrifuge in ein künstliches Schwerfeld gebracht. Das Konzentrationsprofil des Proteins während des Absinkens wird spektroskopisch verfolgt. Aus diesen Daten lässt sich der Sedimentationskoeffizient (angegeben in Svedberg, S) bestimmen und aus diesem die Molmasse des Komplexes.

Die Daten in Abb. 3.29 zeigen, dass die Verteilung von einer Population mit

einer Sedimentationskonstante von 4 S dominiert wird. Dies entspricht einer Molmasse von 46 kDa. In sehr geringen Mengen werden das FINAR-Monomer sowie eine potentielle tetramere Spezies gefunden. Für dieses Experiment wurde der His₆-Tag nicht von dem Protein abgespalten, so dass für das Dimer ein etwas größeres theoretisches Molgewicht von 47,6 kDa berechnet wird.

- Das Protein wurde sowohl mit als auch ohne den N-terminalen His₆-Tag für Kristallisationsansätze eingesetzt, keine Bedingung führte jedoch zu Proteinkristallen.

3.6.8 Zusammenfassung der biochemischen Charakterisierung der Konstrukte

Die proteinbiochemische Untersuchung der exprimierten Konstrukte ergab, dass keines der untersuchten einzelnen Blätter spontan falten konnte. Die weiteren Konstrukte ließen dagegen sehr wohl den Schluss zu, dass sie spontan oligomere, wohlgefaltete Propeller bilden. Dabei bilden das Doppelblatt und der vierblättrige Propellerbaustein achtblättrige Propeller. Es ist interessant, dass der Propeller, der nur aus zwei vierblättrigen Proteinketten besteht, einen deutlich höheren Schmelzpunkt aufweist. Dies deutet auf eine erhöhte Stabilität hin. Das dreiblättrige Konstrukt bildet einen sechsblättrigen Propeller. Verschiedene unabhängige Experimente zeigen, dass sich aus dem fünfblättrigen Konstrukt ein 10-blättriger Propeller bildet.

3.7 Mischungsexperimente

3.7.1 Inkorporation eines einzelnen Propellerblattes in einen bestehenden Propeller

Bei der Charakterisierung der einzelnen Konstrukte, wie sie im vorangegangenen Teil des Kapitels beschrieben wurde, konnten wir zeigen, dass alle Propellerfragmente homooligomere Propeller unterschiedlicher Größe bilden. Lediglich die einzelnen Blätter waren nicht in der Lage zu falten. Da die Bausteine zu meist dimerisierten (das Doppelblatt tetramerisiert), bildeten sich ausschließlich Propeller mit einer geraden Anzahl an Blättern. Eine Überlegung war es, durch Mischung, Entfaltung und gemeinsame Rückfaltung heterooligomere Propeller zu erzeugen. Das Vorgehen für die Entfaltungs- und Mischungsexperimente ist im Material- und Methodenteil beschrieben.

Durch Zusammenführen von einem dreifachen molarem Überschuss des Doppelblattes und dem einzelnen Propellerblatt wollten wir einen siebenblättrigen Propeller herstellen. Damit wollten wir überprüfen, ob das einzelne Blatt im Kontext eines stabilen Propellers in der Lage ist zu falten und sich in einen Komplex zu inkorporieren. Für dieses Experiment wurde das Doppelblatt wie bisher beschrieben gereinigt. Das einzelne Blatt wurde in *E. coli* exprimiert, wobei eine zweifache Isotopenmarkierung über das ¹³C und ¹⁵N enthaltende Wachstumsmedium eingeführt wurde. Dieses ¹³C und ¹⁵N doppelmarkierte Blatt wurde gereinigt und für die NMR-Experimente eingesetzt. Die beiden Proteine wurden zunächst in einem molaren 3:1 Verhältnis gemischt und in einem NMR-Experiment vermessen. (Die Messungen und Auswertungen wurden

am NMR-Zentrum in Garching von M. Coles und V. Truffault in Kooperation mit mir durchgeführt.) Dabei wurde ein HSQC Spektrum (*Heteronuclear Single Quantum Coherence*) des einzelnen Blattes aufgenommen und der Diffusionskoeffizient bestimmt. Auf Grund der Isotopenmarkierung lassen sich das Doppelblatt und das einzelne Blatt im NMR-Experiment unterscheiden. Das Spektrum des einzelnen Blattes zeigt die charakteristischen scharfen und nur wenig aufgelösten Signale eines ungefalteten Proteins.

Die Proteinnischung wurde auf eine Endkonzentration von 6 M GuaHCl gebracht und gegen PBS dialysiert, was zur gemeinsamen Rückfaltung der Proteine führt. Die obigen NMR-Experimente wurden wiederholt. Zwei Dinge fallen bei der Interpretation der Spektren auf:

- Das HSQC Spektrum des einzelnen Blattes zeigt deutlich verbreiterte Peaks und die Signale sind weiter über das Spektrum verteilt. Dies ist ein Anzeichen für ein gefaltetes Protein.

- Der Diffusionskoeffizient des einzelnen Blattes entspricht nicht mehr dem vor der Ent- und Rückfaltung, er ist vielmehr in einem Bereich, den man für einen 30 kDa Propeller erwarten würde.

Die genaue Analyse der Daten ergab jedoch, dass die Spektren keine ausreichende Auflösung besaßen, um eine vollständige 3D-Struktur zu berechnen. Für die Strukturberechnung fehlten die langreichenden Wechselwirkungen zwischen Seitenketten, die in der Primärstruktur nicht nebeneinander liegen. Das Fehlen dieser Informationen interpretieren wir durch eine Anlagerung des einzelnen Blattes an den rückgefalteten TINAR-Propeller. Diese Anlagerung ist aber nicht zu vergleichen mit dem Einbau eines einzelnen Propellerblattes in einen vollständigen Propeller. So lässt sich die Peakverbreiterung und der veränderte Diffusionskoeffizient durch enge Interaktionen des einzelnen Blattes mit dem TINAR-Propeller erklären. Anscheinend findet jedoch kein Einbau in den Propeller statt, der zu einem korrekt gefalteten einzelnen Blatt führen würde.

3.7.2 Kofaltung verschiedener Propellerkonstrukte

Um zu überprüfen, ob Propellerfragmente, die für sich allein falten und oligomerisieren können, in der Lage sind, heterooligomere Propeller zu bilden, haben wir entsprechend dem eben beschriebenen Mischungsexperiment verschiedene Konstrukte miteinander rückgefaltet. Es sei darauf hingewiesen, dass wir in Abschnitt 3.6.4 zeigen konnten, dass das TINAR-Protein in 6 M GuaHCl entfaltet vorliegt und sich wie ein einzelnes entfaltetes Doppelblatt verhält.

Um verschiedene Stöchiometrien zu testen und auch Propeller mit einer ungeraden Anzahl an Propellerblättern zu erhalten, wurden die im Folgenden diskutierten Mischungen hergestellt. Das Vorgehen ist im Material- und Methodenteil beschrieben. Kurz zusammengefasst wurden die Proteinlösungen gemischt, mit festem GuaHCl auf eine Endkonzentration von 6 mol/l gebracht und nach gründlicher Mischung und Inkubation gegen PBS dialysiert, was die Rückfaltung bewirkte.

Auf dem Nativgel in Abb. 3.30 sind in den rechten drei Spuren als Referenzen die Konstrukte aufgetragen, für die festgestellt wurde, dass sie einen sechs-, acht- und 10-blättrigen Propeller bilden.

- Zweifacher Überschuss des TriNAR-Proteins mit dem Doppelblatt: Statt einer homogenen Population an sechs- oder achtblättrigen Propellern erkennt man, dass sich wieder die Ausgangspropeller bilden. Auffällig ist, dass die Bande

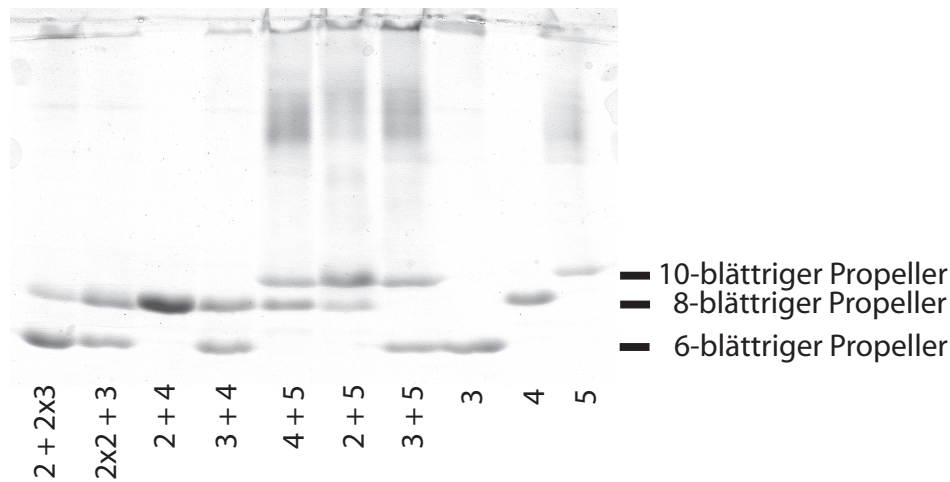


Abbildung 3.30: Auftrennung der verschiedenen β -Propeller in einem nicht denaturierenden Acrylamidgel. Die Propeller wurden durch Assemblierungsreaktionen von unterschiedlichen Fragmenten hergestellt. Unter dem Gel ist angegeben, wie viele Propellerblätter die Ausgangskonstrukte hatten und in welchen stöchiometrischen Mengen sie eingesetzt wurden. Links sind Marker aufgetragen, die einen 6-, 8- und 10-blättrigen Propeller bilden. Die genaue Durchführung der Experimente ist im Text beschrieben (3.7.2 und 3.4.5).

des sechsblättrigen Propellers, der in doppelter Menge eingesetzt wurde, auch stärker angefärbt ist.

- Zweifacher Überschuss des Doppelblattes mit dem TriNAR-Protein: In diesem stöchiometrischen Verhältnis sollten sich siebenblättrige Propeller bilden können. Die Auftrennung auf dem Gel zeigt aber, dass sich wiederum nur die Ausgangspropeller zurückbilden.

- Äquimolare Mischung von Doppelblatt und FoNAR-Protein: Es bilden sich nur achtblättrige Propeller (zu sehen an der deutlich stärker gefärbten Bande), der theoretisch auch mögliche sechsblättrige Propeller ist nicht nachweisbar.

- Äquimolare Mischung von TriNAR und FoNAR: In diesem stöchiometrischen Verhältnis sollten sich ebenfalls siebenblättrige Propeller bilden können. Die Auftrennung auf dem Gel zeigt aber, dass sich nur die Ausgangspropeller zurückbilden.

- Äquimolare Mischung von FoNAR und FiNAR: In diesem Ansatz könnten sich theoretisch neunblättrige Propeller bilden, was wir aber nicht sehen. Im oberen Abschnitt des Gels ist in allen Spuren, in denen FiNAR enthalten ist, eine unspezifische breite Bande zu sehen. Diese Verunreinigung ist nicht erst durch die Entfaltung entstanden, da sie auch in der Referenzspur auftritt.

- Äquimolare Mischung von Doppelblatt und FiNAR: In diesem stöchiometrischen Verhältnis sollten sich ebenfalls siebenblättrige Propeller bilden können. Die Auftrennung auf dem Gel zeigt aber, dass sich nur die Ausgangspropeller zurückbilden.

- Äquimolare Mischung von TriNAR und FiNAR: In diesem Ansatz könn-

ten sich theoretisch auch achtblättrige Propeller bilden, was nach dem Gel zu schließen jedoch nicht der Fall war.

3.7.3 Diskussion

Die Analyse der Mischungsexperimente, so wie gerade beschrieben, führt zu dem Schluss, dass sich keine heterooligomeren Propeller aus den von uns eingesetzten Konstrukten bilden. Dies ist erstaunlich und nur schwer zu erklären. Geht man davon aus, dass die Proteine tatsächlich vollständig entfaltet vorliegen und statistisch durchmischt sind, so muss man folgern, dass die Rückfaltung und Zusammenlagerung mit einem Protein derselben Spezies begünstigt ist. Diese Ergebnisse erlauben keine Aussagen zu energetischen Fragen wie der Stabilität von Propellern mit unterschiedlich vielen Blättern. Es ist verwunderlich, dass sich die unterschiedlichen thermodynamischen Stabilitäten von verschiedenen großen Propellern nicht bei der Rückfaltung auswirken. Man würde erwarten, dass ein Propelleraufbau gegenüber den anderen begünstigt sein sollte. So wurde auf Grund von rein theoretischen Überlegungen postuliert, dass der siebenblättrige Propeller der stabilste sein sollte [69].

Es bieten sich zwei mögliche Erklärungen für das Zustandekommen dieser Ergebnisse an: (a) Die Propeller sind nicht vollständig entfaltet worden und hatten daher gar nicht die Möglichkeit, sich in neuen Stöchiometrien zusammenzufinden. Es wurden jedoch zahlreiche Experimente durchgeführt, die belegen, dass unter den verwendeten Bedingungen das Protein entfaltet vorliegt (Abschnitt 3.6.4). Da die Konzentration des chaotropen Salzes jedoch sicher auf mind. 6 mol/l eingestellt wurde, gehen wir davon aus, dass die mangelnde Entfaltung nicht die Erklärung für die Ergebnisse ist. (b) Die energetischen Unterschiede zwischen den einzelnen Propellern mit verschiedenen vielen Blättern sind so gering, dass sich die thermodynamischen Endzustände unter den gewählten Bedingungen (z. B. Raumtemperatur) nicht auswirken.

3.8 Zusammenfassung

Wie konnte die β -Propellerfaltung in der Evolution entstehen? Ein Szenario ist, dass kleine Peptide, die für sich allein nicht in der Lage sind, eine Funktion auszuüben, sich zusammenlagern und erst der Komplex dieser Fragmente eine ausreichende Stabilität aufweist. Im Fall der β -Propellerdomäne fällt auf, dass die Proteine eine strukturelle Symmetrie aufweisen. Die untereinander sehr ähnlichen einzelnen Blätter wiederholen sich und bilden gemeinsam eine zirkuläre Proteindomäne. Wir haben postuliert, dass die ancestrale Vorstufe der β -Propeller ein Propellerblatt gewesen ist und um diese Hypothese zu überprüfen, haben wir Propellerproteine aus kleineren Fragmenten aufgebaut.

Wir sind dabei von einem bakteriellen Protein PkWA ausgegangen, für das zwar keine Kristallstruktur vorhanden ist, für welches wir jedoch auf Grund von Sequenzvergleichen sicher eine siebenblättrige β -Propellerstruktur annehmen. Dieses Protein zeichnet sich dadurch aus, dass die Blätter ein hohes Maß an Sequenzidentität aufweisen. Diese Propellerblätter haben wir verwendet, um eine Konsensussequenz zu ermitteln, die wir exprimiert, gereinigt und charakterisiert haben. Ebenso haben wir das einzelne Blatt im Labor amplifiziert und

waren in der Lage, Konstrukte mit zwei bis fünf Propellerblättern zu exprimieren.

Die biochemischen Untersuchungen der einzelnen Proteine ergaben, dass lediglich das einzelne Blatt nicht in der Lage ist zu falten und Propeller zu bilden. Das Doppelblatt hingegen assembliert spontan zu einem achtblättrigen Propeller. Das Konstrukt mit drei Propellerblättern dimerisiert und bildet einen Propeller mit sechs Blättern. Das vierblättrige Konstrukt bildet ebenfalls einen achtblättrigen Propeller, zeichnet sich jedoch durch eine erhöhte thermische Stabilität im Vergleich mit dem gleich großen Propeller aus vier Doppelblättern aus. Für das Konstrukt mit fünf konsekutiven Propellerblättern gehen wir von einer 10-blättrigen Struktur aus.

Durch Mischungsexperimente wollten wir Propeller mit einem erweiterten Repertoire an Propellerblättern konstruieren. Die Experimente ergaben jedoch, dass es nicht möglich ist, ein einzelnes Propellerblatt in einen Propeller aus Doppelblättern zu inkorporieren. Ebenso bilden die Mischungen aus unterschiedlichen selbstfaltenden Propellerkonstrukten bei der Rückfaltung wieder die Ausgangspropeller und keine heterooligomeren Komplexe.

Zusammenfassend ist es uns gelungen, β -Propellerdomänen aus kleineren Fragmenten aufzubauen. Im Gegensatz zur bioinformatischen Analyse ist die kleinste Einheit, die spontan einen vollständigen Propeller aufbauen kann, nicht das einzelne Propellerblatt sondern das verdoppelte Blatt. Weiterhin konnten wir zeigen, dass je nach Anzahl der Blätter im Ausgangsbaustein Propeller unterschiedlicher Größe entstehen. Dabei haben alle Blätter dieselbe Sequenz.

Kapitel 4

Struktur eines 14-blättrigen β -Propellers

4.1 Einleitung

Ein Aspekt der vorliegenden Untersuchung zum Thema der β -Propeller war es, die maximale Größe des Propellerfolds zu ermitteln. In der Protein Daten Bank (PDB) sind nur Strukturen mit vier bis acht Propellerblättern in einer zirkulären Domäne bekannt. Wir haben jedoch biochemische Hinweise auf einen β -Propeller mit 10 Blättern in einer Domäne gefunden. Daher stellte sich die Frage, wieviele Blätter maximal in eine zirkuläre Propellerdomäne gefaltet sein können. In diesem Zusammenhang wäre es auch interessant zu bestimmen, was die Größe einer β -Propellerdomäne bestimmt.

Bei der Expression der amplifizierten Propellerblätter in Kapitel 3 fanden wir durch unabhängige Experimente Hinweise darauf, dass das Konstrukt mit fünf konsekutiven Blättern zu einem zehnbältrigen Propeller dimerisiert. Wir konnten auch zeigen, dass dieselbe Sequenz in der Lage ist Propeller verschiedener Größe zu bilden, abhängig davon, wieviele Blätter das eingesetzte Protein aufweist. Diese Experimente unterstreichen die Variabilität und Vielfältigkeit der β -Propellerfaltung.

Weiterhin wurde im Verlauf dieser Arbeit ein Experiment veröffentlicht, bei dem der ursprünglich fünfblättrige β -Propeller Tachylectin fragmentiert und anschließend nach aktiven Komponenten gesucht wurde [104]. Eine solche Proteineinheit, die jedoch nicht weiter charakterisiert wurde, hat in etwa das doppelte Molekulargewicht des fünfblättrigen Ausgangspropellers.

Für ein weiteres Protein (Aip1), das mit zehn Propellerblättern vorhergesagt wurde [74], wurde die Struktur gelöst. Die Kristallstruktur zeigte jedoch, dass das Protein aus zwei Domänen mit jeweils sieben Propellerblättern besteht und dass vier Wiederholungen auf Grund der schlechten Sequenzkonserviertheit bei der Strukturvorhersage übersehen wurden [98]. Damit stellte sich die Frage nach der Stabilität der β -Propellerdomäne und ob es eine Obergrenze für die Anzahl der Blätter dieser Faltung gibt.

Uns ist aufgefallen, dass sich in den Genomen der Cyanobakterien *Nostoc punctiforme* (PCC73120) und *Nostoc anabaena* (PCC7120) zahlreiche Gene finden lassen, die für β -Propellerproteine codieren. Dabei ist es interessant zu

sehen, dass dabei nicht nur Propeller mit vier bis acht Blättern vorhergesagt werden sondern eine ganze Bandbreite an Propellergrößen.

Um einen systematischen Überblick bezüglich der Anzahl der Propeller zu bekommen, haben wir die Genome der beiden *Nostoc* Spezies in allen sechs möglichen Leserastern übersetzt und gegen diese Datenbank BLAST-Suchen durchgeführt. Bei diesen Untersuchungen sind wir auf zwei interessante Proteine gestoßen, die jeweils 14 Propellerblätter aufweisen. Diese Proteine haben wir kloniert und gereinigt, von einem der beiden haben wir die Proteinstruktur mit kristallographischen Methoden gelöst. Die Arbeiten zu diesen Proteinen sollen hier vorgestellt werden.

4.1.1 Cyanobakterien

Im folgenden Kapitel werden Proteine aus dem Cyanobakterium *Nostoc sp.* diskutiert. Die Proteine aus diesen Organismen wurden durch Sequenzvergleiche gefunden und auf Grund von strukturellen Eigenschaften ausgewählt. Um darüber hinaus einen kurzen Einblick in den Ursprungsorganismus dieser Proteine zu geben sollen an dieser Stelle die Cyanobakterien kurz vorgestellt werden.

Cyanobakterien wurden früher auch als Blaualgen bezeichnet und fälschlicherweise den Algen zugeordnet, tatsächlich sind sie aber Prokaryoten und Teil der Bakterien. Cyanobakterien leben als autotrophe Organismen, die in der Lage sind Photosynthese zu betreiben, und kommen mit CO_2 als einziger Kohlenstoffquelle aus. Des Weiteren sind sie in der Lage Stickstoff zu fixieren (N_2 wird dabei zu NH_3 reduziert).

In der Erdgeschichte sind die ältesten Fossilien von lebenden Organismen die von Cyanobakterien. Die ältesten Funde werden auf ein Alter von ca. 3,5 Mrd. Jahre datiert (zum Vergleich: die ältesten Felsformationen sind 3,8 Mrd. Jahre alt).

Nach der Endosymbiontentheorie haben sich die Mitochondrien und Chloroplasten aus frühen Cyanobakterien entwickelt. Es wird auch angenommen, dass der Sauerstoff in der Atmosphäre von Cyanobakterien im späten Archaikum oder frühen Proterozoikum durch Photosynthese gebildet wurde.

Heute finden sich Cyanobakterien in allen lichtzugänglichen Bereichen der Biosphäre. Beispiele für Cyanobakterien aus der täglichen Umgebung sind der grünlich, schleimige Bewuchs auf feuchten Blumentöpfen und Hauswänden oder die schwarzen Streifen in Bereichen mit ablaufendem Wasser auf Kalkfelsen (schwäbische Alb). Die Blaualgenblüte auf eutrophierten Teichen ist ein weiteres Beispiel.

Cyanobakterien treten in der Regel einzellig auf, wachsen aber häufig in großen Kolonien. *Nostoc*-Arten produzieren relativ kurze, bewegliche Filamente, eine Eigenschaft durch die sie sich von der nah verwandten Gattung *Anabaena* unterscheiden. *Anabaena* zeichnet sich durch die Besonderheit aus, dass die Zellen in langen Fäden hintereinander wachsen.

Das Genom von *Nostoc punctiforme* ist mit 9,02 Mb relativ groß und codiert für 7672 Proteine. (Für *Nostoc punctiforme* gibt es mehrere Bezeichnungen, so ist die Nummer ATCC 29133 gleichbedeutend mit PCC 73102.)¹ Im Gegensatz dazu hat das Genom von *Nostoc anabaena* nur eine Größe von 6,41 Mb und codiert für 5366 Proteine (PCC 7120) [51].

¹Da das Genom noch nicht in seiner endgültigen Form publiziert wurde, lassen sich aktuelle Informationen über das NCBI, Abschnitt 'Genomes' abrufen (<http://www.ncbi.nlm.nih.gov/>).

4.2 Kristallstrukturanalyse von Proteinen

Möchte man die Struktur von Proteinen mit atomarer Auflösung bestimmen, so kann man diese Informationen über NMR-Spektroskopie oder Kristallstrukturanalyse gewinnen². Das Vorgehen bei der Röntgenstrukturanalyse gliedert sich dabei in drei Schritte: (a) Herstellung von geeigneten Proteinkristallen, (b) Beugungsexperimente zur Datengewinnung und (c) Datenauswertung.

Proteinkristallisation

In den meisten Fällen ist die Züchtung geeigneter Einkristalle der limitierende Schritt der Proteinstrukturbestimmung. Trotz vieler theoretischer Arbeiten und experimenteller Anstrengungen sind es nach wie vor empirische Ansätze und das Prinzip von Versuch und Irrtum die unternommen werden um Proteine zu kristallisieren.

Proteine werden aus übersättigten Lösungen kristallisiert, diese werden durch Dampfdiffusion erzeugt. Dabei werden einzelne Tropfen der Proteinlösung mit einem Fällungsmittel gemischt und in einer geschlossenen Kammer mit einem großen Volumen an Fällungsflüssigkeit platziert. Im Tropfen vermischt sich die Proteinlösung mit der Fällungsflüssigkeit, wodurch das Fällungsmittel im Tropfen verdünnt wird. Im Bodenreservoir ist ein ca. 100 bis 1000 mal größeres Volumen an reiner Fällungsflüssigkeit als im Tropfen. Durch Verdunstung und Dampfdiffusion findet ein Nettowassertransfer vom verdünnten Tropfen zum konzentrierten Pufferreservoir statt, bis sich die Konzentrationen ausgeglichen haben. Durch den Wasserverlust im Tropfen steigt die Proteinkonzentration an, wird dabei der Zustand der Übersättigung erreicht, kann das Protein kristallin ausfallen. Im allgemeinen muss man sehr viele (mehrere hundert) Bedingungen testen, um für die Röntgenstrukturanalyse geeignete Kristalle zu erhalten. Oft genug passiert es, dass ein Protein gar nicht kristallisiert werden kann.

Im Kristall lagern sich die Proteinmoleküle in einer Art 'geordneten stabilisierten Lösung' zusammen, wobei bis zu 80% des Kristalls aus Wasser bestehen kann. Durch die präzisen Wiederholungen im Kristall werden die Intensitäten der gebeugten Reflexe verstärkt. Ein für die Datengewinnung tauglicher Kristall muss isomorph sein, die Röntgenstrahlung stark beugen und groß genug sein.

Datengewinnung

Zur Lösung der in dieser Arbeit beschriebenen Struktur wurden geeignete Proteinkristalle mit Synchrotronstrahlung bestrahlt. Im Synchrotron werden geladene Teilchen (hier Elektronen) auf Geschwindigkeiten nahe der Lichtgeschwindigkeit beschleunigt. Durch große Magneten werden die Teilchen von einer geraden auf eine Kreisbahn gezwungen und derart beschleunigte Elektronen emittieren Energie (Synchrotronstrahlung).

Am Synchrotron kann die Wellenlänge für das Beugungsexperiment nahezu frei gewählt werden. Neben der frei wähl- und veränderbaren Wellenlänge zeichnet sich Synchrotronstrahlung durch eine viel höhere Intensität aus als Röntgenstrahlung von einer konventionellen Kupferanode.

²Beide Methoden sind inzwischen so etabliert, dass sie in zahlreichen Lehrbüchern detailliert beschrieben werden, ich möchte hier nur die Grundlagen beschreiben und darlegen, welche Anwendungen die Methoden in dieser Arbeit hatten.

Die Proteinkristalle geben ein Beugungsmuster, wenn sie mit Röntgenstrahlung bestrahlt werden. Durch Auswertung des Beugungsmusters kann die Kristall- und Molekülstruktur bestimmt werden.

Es sei angemerkt, dass das Muster der Reflexe nur vom Kristallgitter abhängt und nicht von der Proteinstruktur. Die Strukturinformationen sind in den unterschiedlichen Intensitäten der Reflexe enthalten. Dabei tragen alle Elektronen als Ganzes³ zur Beugung bei - alle Elektronen haben an einen Anteil an jedem Schwärzungspunkt und in jedem Punkt des Beugungsmusters sind Informationen über die gesamte Elektronenstruktur enthalten.

Datenauswertung

Wenn ein fokussierter Röntgenstrahl mit den Elektronen in einem Proteinkristall interagiert, wird der Strahl abgelenkt und bildet abhängig vom Kristallaufbau ein festgelegtes Beugungsmuster. Dieser Vorgang entspricht Licht, welches von einem Objekt kommt und durch eine Linse auf eine virtuelle Darstellung abgebildet wird. Um bei dieser Analogie zu bleiben, müssten die Lichtstrahlen durch eine zweite Linse gebrochen werden um ein reales Abbild des Objekts zu bilden. In der Kristallographie ist das Problem eben dieser zweite Schritt, da Röntgenstrahlen nicht durch elektromagnetische Felder (wie in der Elektronenmikroskopie angewandt) oder durch Linsen (wie in der klassischen Optik) fokussiert werden können. Die Lösung für den Kristallographen ist es, die zweite Linse im Computer zu simulieren. Das Verhältnis zwischen Objekt und Beugungsbild lässt sich präzise durch eine Fouriertransformation beschreiben und diese mathematische Operation ist umkehrbar.

Für die Umkehrung der Fouriertransformation bedarf es einer vollständigen Beschreibung der zugrundeliegenden Welle und dafür sind Frequenz, Amplitude und Phase der Welle nötig. Die Frequenz ist die der verwendeten Röntgenstrahlung, die Amplitude kann aus der Intensität der Reflektionen bestimmt werden, die Phase geht jedoch beim Beugungsexperiment verloren. Um die Phase zu bestimmen (Phasenproblem), macht man sich zunutze, dass die Gesamtheit der Elektronen einen Anteil an jeder einzelnen Reflektion hat. Wenn es gelingt, die Elektronenstruktur der Proteinmoleküle durch Bindung von Schwermetallerivaten zu verändern ohne dabei die Proteinstruktur zu verändern, so kann man einen zweiten leicht veränderten Datensatz aufnehmen⁴. Aus diesen beiden Datensätzen lassen sich ungefähre Phasen bestimmen. Mit diesen Angaben lässt sich ein erstes Arbeitsmodell berechnen, im Folgenden versucht man dieses Arbeitsmodell zu verbessern und sukzessive bessere Phasen zu erhalten. Das Ziel ist ein 'Elektronendichtetunnel' bei dem auch die Seitenketten sichtbar sind und aus dem das endgültige Modell gebaut werden kann.

³Die Röntgenstrahlen werden abhängig von der Elektronendichteverteilung im Proteinkristall gebeugt, die Beschreibung der Struktur durch Atomkoordinaten hat nur den Zweck, dass Modell einfacher zu machen.

⁴Außer der Schwermetallmethode werden noch eine Reihe weiterer Ansätze genutzt um das Phasenproblem zu lösen, die Struktur in dieser Arbeit wurde durch Derivatisierung gelöst.

4.3 Material und Methoden

4.3.1 Systematische bioinformatische Analyse der Nostoc-Genome

Die Genome der beiden Nostoc Spezies *Nostoc punctiforme* und *Nostoc anaebaena* wurden in allen sechs möglichen Leserastern übersetzt und gegen diese Datenbank BLAST-Suchen gestartet⁵. Als Suchsequenz wurde das Protein gi|23130132 verwendet. Dieses Protein ist als "WD40 repeat [Nostoc punctiforme PCC 73102]" annotiert. Die beiden Webserver SMART und Pfam sagen übereinstimmend 14 WD40-Wiederholungen sowie eine Coiled Coil Region im N-terminalen Bereich voraus. Für die Datenbanksuchen wurde nur der C-terminale Propellerbereich des Proteins verwendet.

Die mit der TBLASTN-Suche identifizierten Proteine wurden mit dem SMART Webserver analysiert und manuell aligniert. Zwei Proteine werden wie im Ergebnisabschnitt 4.4 beschrieben ausgewählt und in Expressionsvektoren kloniert, exprimiert und weiter biochemisch untersucht (die Sequenzen der beiden ausgewählten Propeller sind in Abb. 4.12 gezeigt). Die beiden Proteine werden entsprechend der Abb. als 'linker' und 'perfekt' bezeichnet - die Namensgebung bezieht sich darauf, dass das Linkerprotein einen Asn-Rest nach dem siebten Blatt hat und damit einen Verknüpfungs- oder Linkerrest, während das zweite Protein 14 (nahezu) perfekte Sequenzwiederholungen aufweist. Die zugehörigen Proteine sind gi|23130132 und gi|23124439.

4.3.2 Klonierung und Reinigung der β -Propeller aus Nostoc

Die Genabschnitte, die für die Propeller codieren, wurden aus genomischer DNA mittels PCR amplifiziert. Die DNA wurde von LGC Promochem erworben: AT-CC 29133 - dies entspricht der genomischen DNA von *N. punctiforme*, PCC 73102. Die Oligonukleotide für die PCR wurden so entworfen, dass genau die in Abb. 4.12 beschriebenen Sequenzen amplifiziert wurden. Spezifität wurde dadurch erreicht, dass die Primer auch in den flankierenden, nicht repetitiven Genabschnitten binden.

Linkerprotein

Die lyophilisierte genomische DNA wurde in 200 μ l EB-Puffer (10 mM Tris; pH 8,5) gelöst, was einer Endkonzentration von 12 ng/ μ l entspricht.

Es wurden 15 PCR Reaktionen mit 3 unterschiedlichen MgCl₂-Konzentrationen (2 mM; 3 mM und 4mM) und fünf verschiedenen *Annealingtemperaturen* durchgeführt (70°C; 65°C; 60°C; 55°C und 50°C). Der PCR-Zyklus ist in Abb. 4.1 gezeigt.

⁵Die Suche mit einer Proteinsequenz gegen eine Nukleotiddatenbank wurde mit TBLASTN durchgeführt [1]. Dieses Programm kann auf der öffentlich zugänglichen Homepage des Max Planck Instituts für Entwicklungsbiologie unter '*Bioinformatics Toolkit*' abgerufen werden und bietet die Möglichkeit Suchen mit ausgewählten Datenbanken, so z. B. Genomen oder Genomgruppen durchzuführen. <http://toolkit.tuebingen.mpg.de/>

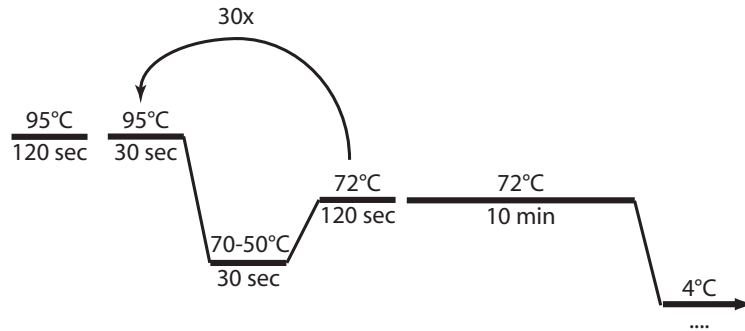


Abbildung 4.1: Zyklus für die PCR um das Linkerpropellergen aus genomischer DNA zu amplifizieren. Die Reaktion wurde wie im Text beschrieben zusammengesetzt.

5,0	μl	genomische DNA (12 ng/ μl)
0,1	μl	Forwardprimer (10 pmol)
0,1	μl	Reverseprimer (10 pmol)
5,0	μl	Pufferlösung
5,0	μl	dNTP Mix (jew. 10 mM in der Ausgangslösung)
ad 49,0	μl	H ₂ O
1,0	μl	PfuTurbo DNA Polymerase (2,5 units/ μl) - kommerziell bei Stragene TM erhältlich

In vier Reaktionen war in der Agarosegeleanalytik eine Bande bei der erwarteten Größe von 1,8 kb zu sehen: 3 mM MgCl₂ bei 70 bis 60°C und bei 70°C mit 4 mM MgCl₂. Diese Banden wurden gemeinsam weiterverwendet, um sie zunächst in den Zeroblunt Vektor und von dort in den pET30 Expressionsvektor zu klonieren (in Abschnitt 3.3.1 beschrieben).

Das Expressionsplasmid wurde in *E. coli* C41 (DE3) Zellen transformiert. Für Expressionen wurden die Plasmide immer frisch transformiert. Eine so transformierte Einzelkolonie wurde in ca. 100 ml LB-Medium über Nacht bei 37°C geschüttelt. Mit 20 ml Zellsuspension aus dieser Vorkultur wurden 800 ml LB-Medium_{Kanamycin} (entspricht 2,5%) angeimpft und in einer 1 L-Flasche durch ein Glasrohr mit Druckluft begast. Die Zellen wurden bei einer optischen Dichte von OD_{600nm} \approx 0,6 mit 1 mM IPTG induziert. Nach vier Stunden wurden die Zellen durch Zentrifugation geerntet (20 min bei 6500 g). Der Zellklumpen wurde gewogen und bei -20°C gelagert.

Für den Zellaufschluss wurden jeweils ca. 8 g tiefgefrorene Zellen auf 35 ml Volumen mit Ladepuffer (50 mM MES pH 6,0; 50 mM NaP_i; 5mM MgCl₂; 1 mM EDTA) aufgefüllt und durch Vortexen und Resuspendieren aufgetaut. Die Suspension wurde mit Proteaseinhibitorcocktail (cOmpleteTM ProteaseInhibitor Cocktail Tablets, kommerziell erhältlich von Roche) und DNaseI (AppliChem) versetzt. Die Zellen ließen sich in fünf Durchläufen durch eine French Pressure Zellpresse (American Instrument Co.) bei ca. 1000 psi aufschließen. Die groben Zellbestandteile wurden durch zweimalige Zentrifugation (37.000 g, 30-45 min) abgetrennt. Das so gewonnene Rohlysate wurde durch einen 0,45 μM Filter gepresst und dann auf eine SP Sepharose Säule (Kationentauscher) von

GE Healthcare (früher Amersham Pharmacia⁶) geladen. Da das Protein einen isoelektrischen Punkt pI von 7,2 hatte, band das Protein nicht und wurde im Durchlauf aufgefangen. Die so vorgereinigte Proteinlösung wurde auf eine MonoQ-Säule (Anionentauscher) geladen und in einem Salzgradienten bis 1 M NaCl eluiert. Das Protein wurde einkonzentriert und auf einer HiLoad 26/60 Superdex G75 Gelfiltrationssäule endgültig gereinigt, wobei ein Pufferwechsel in 50 mM Na₂HPO₄ durchgeführt wurde.

Die gereinigte Proteinlösung wurde auf 6 mg/ml konzentriert und 20 µl dieser Lösung werden mit 10 µl einer 0,02 mg/ml Trypsinlösung versetzt. Der Verdau wurde nach 15 min bei Raumtemperatur durch Zugabe von PMSF gestoppt und die Reaktionsmischung auf einer analytischen Superose 12 Säule aufgetrennt.

Protein mit 14 perfekten Sequenzwiederholungen

Das Gen wurde wie im Falle des Linkerproteins aus der genomischen DNA, die von LGC Promochem erworben wurde, mit Oligonukleotiden amplifiziert, die durch Bindung in den flankierenden, nicht repetitiven Sequenzen ihre Spezifität erhalten. Zunächst wurde eine PCR unter analogen Bedingungen zu den erfolgreichen Reaktionen im Falle des Linkerproteins durchgeführt (Zugabe von MgCl₂), diese führte jedoch zu keinem PCR-Produkt. Eine *Touch Down* PCR, bei der die Anlagerungstemperatur in den ersten 20 Zyklen um jeweils 0,5°C von 68°C auf 58°C gesenkt wurde, gefolgt von 15 Zyklen bei 68°C, war dagegen erfolgreich. Da das PCR-Produkt nur eine schwache Bande im Agarosegel gab, wurde die Bande in einer weiteren PCR (mit Taq Polymerase) nochmals amplifiziert. Da eine erste Reinigung analog zur Reinigung des Linkerproteins kein ausreichend reines Protein für Kristallisationsansätze ergab, wurde das Gen nochmals in pET28 umklontiert (zusätzlicher His₆-Tag).

Die Expression erfolgte in gleicher Weise wie sie für das Linkerprotein beschrieben wurde. Für den Zellaufschluss wurden jeweils ca. 8 g tiefgefrorene Zellen auf 35 ml Volumen mit Ladepuffer (50 mM Tris pH 7,9; 50 mM NaCl; 5mM MgCl₂; 1 mM EDTA) aufgefüllt und durch Vortexen und Resuspendieren aufgetaut. Die Suspension wurde mit Proteaseinhibitorcocktail (cCompleteTM ProteaseInhibitor Cocktail Tablets, kommerziell erhältlich von Roche) und einer Spatelspitze DNaseI (AppliChem) versetzt. Die Zellen wurden in fünf Durchläufen durch eine French Pressure Zellpresse (American Instrument Co.) bei ca. 1000 psi aufgeschlossen. Die groben Zellbestandteile wurden durch zweimalige Zentrifugation (37.000 g, 30-45 min) abgetrennt. Das so gewonnene Rohlysate wurde durch einen 0,45 µm Filter gepresst und dann auf eine QHP Säule (Anionentauscher) geladen. Da das Protein einen isoelektrischen Punkt pI von 9,5 hatte, band das Protein an die Säule und konnte in einem Salzgradienten von 50 mM bis 1 M NaCl eluiert werden. Die stark salzhaltige Lösung mit eluiertem Protein wurde direkt auf eine Ni Sepharose High Performance-Säule geladen. Schwach gebundene Verunreinigungen wurden zunächst mit 5% Imidazolpuffer (50 mM Tris; pH 7,9; 300 mM NaCl; 600 mM Imidazol), was einer Imidazolkonzentration von 30 mM entspricht, von der Säule gewaschen. Die endgültige Elution erfolgte mit einem linearen Gradienten von 30 bis 600 mM Imidazol. Da eine Abspaltung des His₆-Tags erforderlich war, wurde gegen Thrombinpuffer dialysiert (20 mM Tris; pH 8,5; 100 mM NaCl;

⁶Alle Säulen für die Proteinreinigung waren von diesem Hersteller, daher wird er im Folgenden nicht mehr weiter erwähnt.

1 mM CaCl₂; 1 mM DTT) und mit 1 *Unit* Thrombin pro mg Protein über Nacht bei 18°C inkubiert. Thrombin wurde durch Zugabe von 0,1 M PMSF inhibiert. Anschließend wurde das Protein einkonzentriert und auf einer HiLoad 26/60 Superdex G75 Gelfiltrationssäule endgültig gereinigt. Dabei wurde ein Pufferwechsel in 20 mM MOPS; pH 7,0; 50 mM NaCl durchgeführt.

Dieser Puffer wurde für die Kristallansätze verwendet. Ebenso wurden alle weiteren biochemischen Charakterisierungen in diesem Puffer vorgenommen.

Für die Schmelzkurve wurde das Protein auf 4 μ M Protein und 10 mM MOPS, 50 mM NaCl verdünnt.

Für den Proteaseverdau wurde das gereinigte Protein mit Proteinase K (5 μ g/ml Endkonzentration) für 15 min bei Raumtemperatur inkubiert, die Reaktion wurde anschließend mit PMSF gestoppt. Als Positivkontrolle wurde Casein unter denselben Bedingungen verdaut.

4.3.3 Kristallisation des 14-blättrigen β -Propellers

Die Kristallisationsexperimente wurden nach der Dampfdiffusionsmethode mit hängenden Tropfen durchgeführt. Die Tropfen bestanden aus 1 μ l Proteinlösung und 1 μ l Pufferlösung, welche gegen 500 μ l Pufferreservoir equilibriert wurden. Dabei wurden kommerziell erhältliche Platten von Hampton Research mit etablierten Pufferlösungen verwendet. Die für die weiteren Experimente benutzten Proteinkristalle sind bei 18°C unter folgenden Bedingungen gewachsen:

Index Screen (Hampton Research);

Bedingung 24: 2,8 M NaAcetat; pH 7,0;

Proteinkonzentration 14 mg/ml und 5 mg/ml.

Die Schwermetallderivate wurden mit kommerziell erhältlichen Platinsalzen hergestellt. Erfolgreiche Schwermetallbindung erfolgte unter folgenden Bedingungen:

Heavy Atom Screen Pt (Hampton Research);

Bedingung 2: Ammonium Tetrachloroplatinat (II);

Bedingung 4: Kalium Tetranitroplatinat (II);

Bedingung 6: Dichloroethylendiamin Platin (II).

Die Kristalle wurden vor den Röntgenbeugungsexperimenten in flüssigem Stickstoff eingefroren. Die Beugungsexperimente für den nativen Datensatz wurden an der Beamline PXII-X10SA an der Swiss Light Source (SLS - Villigen, Schweiz) durchgeführt. Die derivatisierten Kristalle wurden an der Beamline BM14 der European Synchrotron Radiation Facility (ESRF - Grenoble, Frankreich) gemessen. Die Phasierung wurde mit dem Programm SOLVE/RESOLVE durchgeführt [93].

4.4 Ergebnisse und Diskussion

4.4.1 Repetitive Propeller in *Nostoc* sp.

Uns ist aufgefallen, dass sich in den Genomen der Cyanobakterien *Nostoc punctiforme* (PCC73120) und *Nostoc anabaena* (PCC7120) zahlreiche Gene finden lassen, die für β -Propellerproteine codieren. Durch systematische Suchen gegen die in allen sechs Leserastern übersetzten Genome konnte eine Vielzahl an Genen

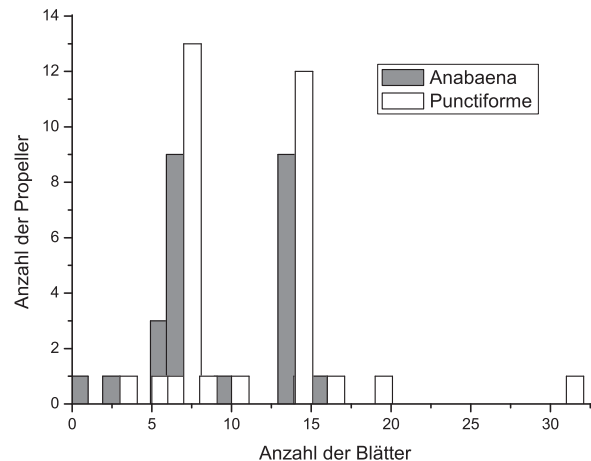


Abbildung 4.2: Die Genome von *N. punctiforme* und *N. anabaena* codieren für zahlreiche β -Propellerproteine. Auffällig ist, dass die Propeller eine weite Spanne an Wiederholungseinheiten aufweisen (die Blätter wurden mit dem SMART Webserver identifiziert).

identifiziert werden, die β -Propellerdomänen enthalten. Von den jeweiligen Genen - 26 in *Anabaena* und 33 in *Punctiforme* - wurden die Vollängensequenzen mit dem SMART Webserver auf ihre Domänenstruktur untersucht und die Propellerbereiche manuell aligniert. Die Verteilung der Propellergrößen gegen die Anzahl der gefundenen Proteine ist in Abb. 4.2 dargestellt. Es ist interessant zu sehen, dass in diesen Genomen die Mehrheit der Propeller aus 7 bzw. 14 Blättern aufgebaut sind, sich aber auch Propeller anderer Größen finden lassen. So werden für ein *N. punctiforme* Protein 31 WD40-Wiederholungen vorhergesagt. In *N. anabaena* findet sich ein potentielles Gen mit nur einem vorhergesagten Propellerblatt. Für dieses Protein ließ sich jedoch nicht abschließend klären, ob es ein tatsächlich exprimiertes Gen ist und ob bei der Sequenzanalyse weitere Propellerblätter nicht vorhergesagt wurden. Aus den Nostoc Propellerproteinen wurden zwei ausgewählt und die Gensequenz für diese beiden Propeller in Expressionsvektoren kloniert.

Biochemische Charakterisierung des Linkerproteins

Das in Abb. 4.12 als 'Linker' beschriebene Protein wurde exprimiert und proteinbiochemisch untersucht. Dafür wurde das gereinigte Protein mit Trypsin verdaut und der Verdau auf einer Gelfiltrationssäule aufgetrennt. Die Analytik ist in Abb. 4.3 gezeigt und macht deutlich, dass unter den gewählten Bedingungen der Verdau nur unvollständig ist, das Protein eluiert in zwei Populationen von der Gelfiltrationssäule. Das erste Absorptionsmaximum enthält noch unverdauten Linkerpropeller (61,8 kDa), während im zweiten Absorptionsmaximum die zwei Proteinspezies mit ca. 30 kDa Molekulargewicht enthalten sind.

Dieses Ergebnis legt nahe, dass es sich bei dem Linkerpropeller um ein Protein mit zwei Domänen handelt, die aus jeweils sieben Propellerblättern bestehen. Nimmt man an, dass Trypsin die Peptidbindung nach dem Lys am Ende

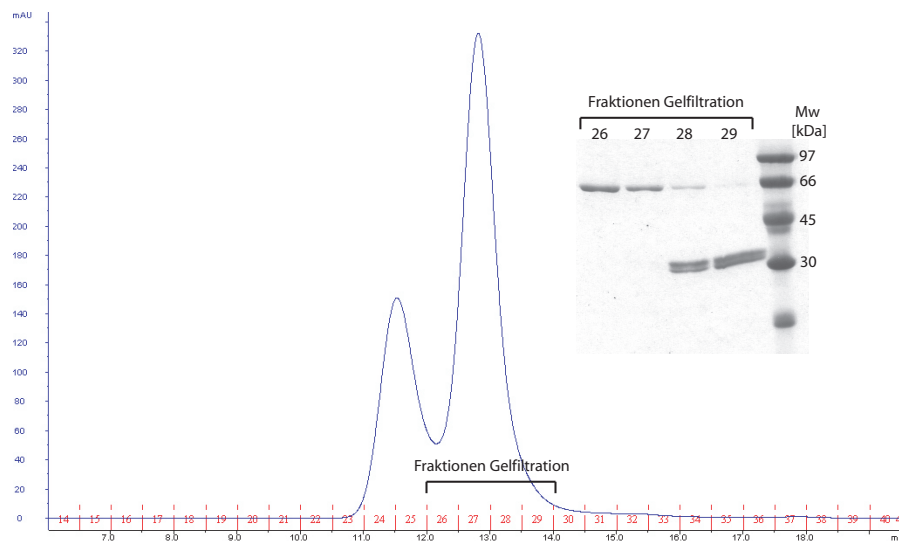


Abbildung 4.3: Auftrennung des Trypsinverdaus des Linkerproteins auf einer Superose6 Gelfiltrationssäule. Das SDS-Gel zeigt die Analytik des zweiten Absorptionsmaximums. Der vollständige Linkerpropeller hat ein Molekulargewicht von 61,8 kDa.

von β -Strang D spaltet, so entstehen zwei Fragmente mit 29,9 bzw. 31,9 kDa. Die hypothetische Schnittstelle ist in Abb. 4.12 markiert, das Lys zwei Aminosäuren früher ist als Schnittstelle unwahrscheinlicher, da es in einem β -Strang vorkommt. In den Fraktionen 28 und 29 in Abb. 4.3 ist eine Doppelbande zu sehen, die diesen Fragmenten entspricht.

Analytik des Proteins mit 14 perfekten Sequenzwiederholungen

Nachdem bei dem Linkerprotein die Evidenzen auf eine Proteinstruktur mit zwei Domänen aus je sieben Wiederholungseinheiten hindeuteten, haben wir bei weiteren Datenbanksuchen einen β -Propeller mit 14 nahezu identischen Blättern identifiziert. Dieses Protein hat keine offensichtliche 'Bruchstelle'. Aus der Sequenz lässt sich eine Faltung in zwei gleich große Domänen nicht ableiten. Dieses Protein, welches in Abb. 4.12 als 'Perfekt' beschrieben ist, wurde ebenfalls kloniert, exprimiert und charakterisiert.

Für die proteinbiochemische Charakterisierung des 14-blättrigen Propellers mit nahezu identischen Blättern wurde ein Proteinase-K-Verdau durchgeführt. Die Daten des Proteinase-K-Verdaus sind in Abb. 4.4 gezeigt und mit dem Trypsinverdau des Linkerproteins verglichen. Dabei wird deutlich, dass im Gegensatz zu dem Linkerprotein der perfekte Propeller nicht in zwei Domänen geschnitten wird. Dies spricht dafür, dass entweder keine Linkersequenz zwischen zwei Domänen vorhanden oder diese für die Protease nicht zugänglich ist.

Einen weiteren Hinweis auf die Stabilität des Proteins ergab die Schmelzkurve im CD-Spektrometer (Abb. 4.5). Zu der Schmelzkurve konnte keine Renaturierungskurve aufgenommen werden, da das Protein bei über 80°C zu prä-

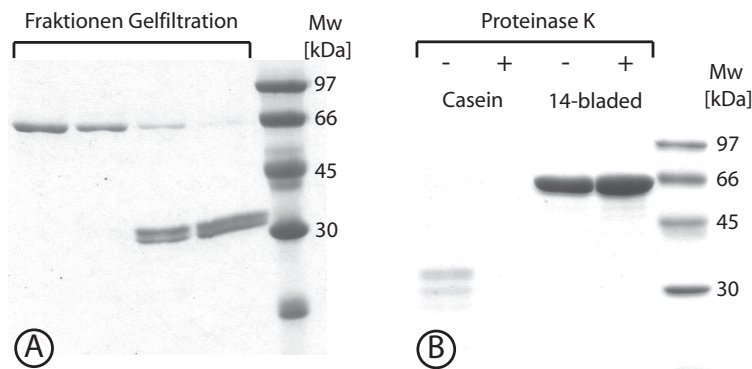


Abbildung 4.4: (A) Trypsinverdau des 14-blättrigen Linkerpropellers. Nach dem (unvollständigen) Proteaseverdau wurde die Reaktionsmischung auf einer Gelfiltration aufgetrennt. Bei 61 kDa sieht man in den frühen Fraktionen unverdaute Propeller. In den späteren Fraktionen erkennt man die Doppelbande der beiden 7-blättrigen Propeller. (B) Proteinase-K-Verdau des 14-blättrigen Propellers mit identischen Blättern. Die Positivkontrolle zeigt, dass Casein vollständig abgebaut wird, während der Propeller eine unveränderte Mobilität im Gel zeigt.

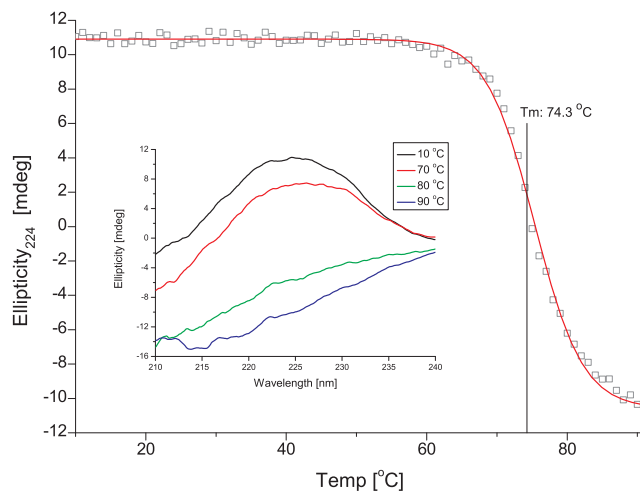


Abbildung 4.5: Schmelzkurve des 14-blättrigen β -Propellers mit 14 perfekten Blättern. Parameter der Messung: Temperaturgradient 1 °C/min; Messung bei 224 nm; 1 mm Küvettendicke; 0,24 mg/ml Proteinkonzentration (= 4 μ M); Puffer 10 mM MOPS; pH 7,0; 50 mM NaCl. Klein in der Abbildung eingefügt sind CD-Spektren die während der Entfaltung aufgenommen wurden (1 akkumuliertes Spektrum).

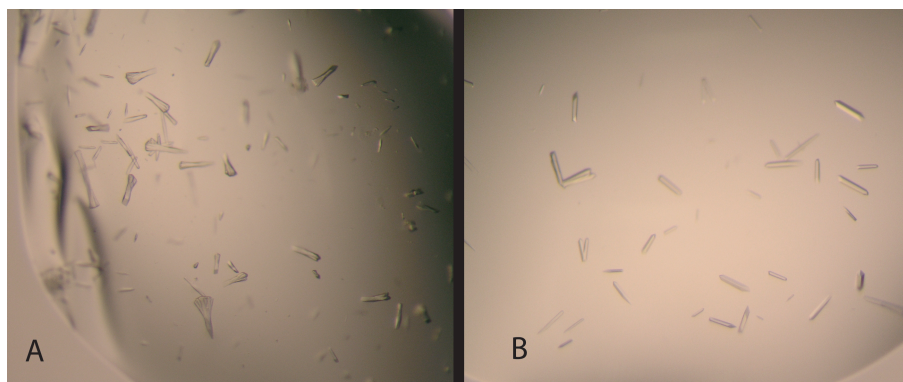


Abbildung 4.6: Kristalle des 14-blättrigen β -Propellers mit 14 gleichen Sequenzwiederholungen. (A) Kristalle aus Proteinelösung mit 14 mg/ml angesetzt. (B) Kristalle aus Proteinelösung mit 5 mg/ml angesetzt. Das experimentelle Vorgehen und die detaillierte Kristallbeschreibung erfolgt im Text.

zipitieren begann.

Die Schmelzkurve kann als eine einstufige, kooperative Entfaltungskurve mit einem Schmelzpunkt bei etwa 74°C interpretiert werden. In der Entfaltungskinetik lässt sich kein Anhaltspunkt für eine Auffaltung in zwei Domänen finden. In diesem Punkt unterscheiden sich die beiden Proteine 'Linker' und 'Perfekt' nicht, die Schmelzkurven beider Proteine sind sehr ähnlich.

4.4.2 Kristallisation und Datengewinnung

Das Vorgehen bei den Kristallansätzen ist im Abschnitt "Kristallisation des 14-blättrigen Propellers" (4.3.3) beschrieben. Von den getesteten Bedingungen kam es nur mit 2,8 M NaAcetat; pH 7,0 zu Kristallwachstum. Unter diesen Bedingungen wuchsen Kristalle sowohl mit der konzentrierten (14 mg/ml) als auch mit der verdünnten (5 mg/ml) Proteinelösung (s. Abb. 4.6). Die Kristalle wuchsen in einzelnen feinen Nadeln. Einige der Kristalle ließen sich in Schwermetallderivate überführen und wurden für die Phasenbestimmung verwendet.

Die Auswertung der Beugungsdaten ergab die orthorhombische Raumgruppe $P2_12_12_1$ mit den Zellkonstanten $a = 88,45 \text{ \AA}$; $b = 107,71 \text{ \AA}$ und $c = 136,24 \text{ \AA}$ und den Winkeln α , β und $\gamma = 90^\circ$. Die Auflösung betrug 1,8 \AA .

4.4.3 Lösung des Phasenproblems und Strukturlösung

Zur Lösung des Phasenproblems wurde das Verfahren des *Multiple Isomorphous Replacement* (MIR) mit schwermetallgetränkten Kristallen eingesetzt. Im Material und Methodenteil wurde beschrieben, dass mit einigen Platinderivaten aus dem nativen Kristall isomorphe Kristalle hergestellt werden konnten, bei denen Schwermetalle an das Protein gebunden haben. Die Phasen wurden mit dem Programm SOLVE/RESOLVE bestimmt und verfeinert [93].

In der generierten initialen Dichte war zu erkennen, dass die asymmetrische Einheit vermutlich aus vier bereits in Ansätzen erkennbaren Propellerdomänen bestehen würde. Später stellte sich heraus, dass sich die Propeller

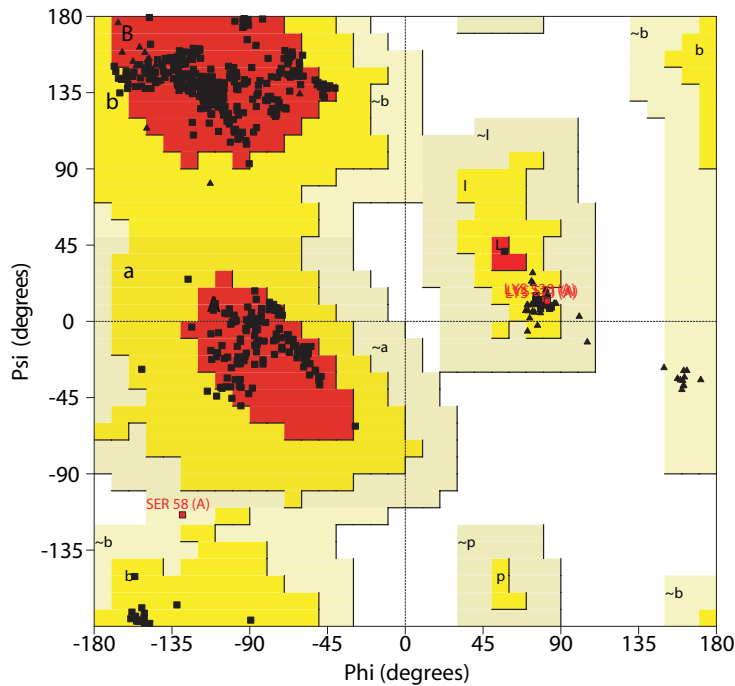


Abbildung 4.7: Ramachandran Plot. Über 90% der Reste befinden sich in erlaubten Bereichen, kein Rest weist verbotene Diederwinkel in seiner Peptidbindung auf.

auf zwei jeweils aus zwei Domänen bestehenden Propellern verteilt. Als Ausgangspunkt für unser Modell wählten wir den Propeller Aip1 (PDB 1nr0). Aus diesem Proteinmodell schnitten wir ein Blatt heraus und ersetzten alle Reste durch Alanine. Von diesem Polyalaninblatt platzierten wir zunächst sieben Kopien in der Elektronendichte und starteten einen ersten Verfeinerungszyklus mit REFMAC5 [78, 68]. In späteren Schritten duplizierten wir das Modell für die erste von uns konstruierten Propellerdomäne und konnten so das vollständige erste Propellermodell erstellen. Dieses Modell wurde verwendet um auch die zweite Proteinkette in der asymmetrischen Einheit zu modellieren. In mehreren Zyklen von manuellem modellieren mit COOT [27] sowie O [48] und Verfeinerung mit REFMAC5 konnten wir den Blättern die korrekte Primärstruktur zuweisen. Nachdem wir in einem späteren Stadium auch die Moleküle der ersten Hydratisierungshülle in das Modell mit aufnahmen, gelangten wir zu einem finalen Modell mit einem R-Faktor von 20,9% und einem R_{free} von 24,2%. (Die Struktur wurde von mir mit Unterstützung von K. Zeth gelöst.)

Um die Qualität der Struktur des 14-blättrigen Propellers zu validieren, wurden die PDB-Koordinaten mit PROCHECK [57] überprüft. Das Ramachandran Diagramm (s. Abb. 4.7) analysiert die Diederwinkel der Peptidbindung der einzelnen Aminosäurereste. Dabei sind verschiedene Kombinationen energetisch ungünstig und einige sogar sterisch verboten.

Die Analyse des 14-blättrigen Propellers ergibt, dass über 90% der Reste

erlaubte Diederwinkel aufweisen, knapp 9% der Peptidbindungen im erweiterten erlaubten Bereich liegen und kein verbotener Winkel auftritt. Für ein Protein mit der von uns erzielten Auflösung sind dies Indikatoren für eine qualitativ hochwertige Struktur.

4.5 Strukturbeschreibung

Die Kristallstruktur des 14-blättrigen Propellers mit nahezu identischen Blättern ergibt, dass dieses Protein genau wie Aip1 und für das Linkerprotein angenommen in zwei Domänen faltet. Die Struktur ist in Abb. 4.8 gezeigt, die Sequenz in Abb. 4.12. Der N-Terminus ist in blau gezeigt, der erste β -Strang bildet den *Molecular Velcro* mit dem letzten β -Strang am C-Terminus (rot). Im Anschluss an den blauen β -Strang faltet das Protein in die blau/grün gefärbte Propellerdomäne. Diese faltet in einen ununterbrochenen β -Propeller mit sieben Propellerblättern. Es ist interessant zu sehen, dass diese Domäne ausschließlich aus Blättern mit vier kovalent verbundenen β -Strängen besteht, also keinen *Velcro* hat. Im Anschluss an diese Domäne wechselt der Proteinfaden wieder in die Domäne, die die N- und C-Termini enthält. Auch diese Domäne ist aus sieben Propellerblättern aufgebaut, jedes von ihnen bis auf das Blatt mit den Termini besteht aus ununterbrochenen Struktureinheiten. Mit den drei C-terminalen β -Strängen schließt sich die Struktur. Durch den *Velcro* sind die antiparallelen β -Stränge von N- und C-Terminus nicht nur durch hydrophobe Wechselwirkungen miteinander verbunden, wie es der Fall wäre, wenn die Enden in unterschiedlichen Propellerblättern wären, sondern zusätzlich durch Wasserstoffbrückenbindungen miteinander verknüpft, was zu einer erhöhten Stabilität des Proteins führt.

Die extreme Sequenzkonserviertheit zwischen den einzelnen Blättern spiegelt sich auch in der Struktur der Blätter wieder. Die Überlagerung der 14 Propellerblätter zeigt, dass sogar die Atome der Seitenketten in allen Blättern gleich angeordnet sind. Die Atome des Proteinerückgrats stimmen bis auf einen RMSD von 0,34 bis 0,38 Å überein (s. Abb. 4.9).

Sogar die Verbindungssequenz nach Blatt sieben, die die Brücke zwischen den beiden Domänen bildet, hat dieselbe Konformation wie in den anderen Blättern, die Proteinkette ist nur um den Thr-Rest vor dem GH-Motiv gedreht (in Abb. 4.9 eingekreist).

Ebenso sind die beiden Domänen des Proteins mit einer RMSD von 0,5 Å überlagerbar, wie in Abb. 4.10 gezeigt. Auch die beiden Proteinmoleküle in der asymmetrischen Einheit überlagern mit einer RMSD von 0,5 Å.

In der PDB sind nur zwei andere Strukturen von β -Propellern mit 14 Blättern hinterlegt. Dies sind das bereits erwähnte Protein Aip1 und ein Oligoxyloglucan reduzierendes Enzym [98, 105]. Auch diese Proteine falten in zwei Domänen mit jeweils sieben Propellerblättern. Das Oligoxyloglucan reduzierende Enzym ist von der Sequenz her relativ weit entfernt von den hier diskutierten Proteinen. Die beiden Domänen dieses Proteins spielen jedoch eine entscheidende Rolle, um bei den bioinformatischen Untersuchungen die beiden Hauptcluster miteinander zu verknüpfen - s. Abb. 2.4, Protein 1sqj (Nr. 16+17). Bei der strukturellen Untersuchung ergibt sich, dass die beiden Domänen von 1sqj einander ebenfalls sehr ähnlich sind, sie stehen jedoch nahezu im rechten Winkel zueinander.

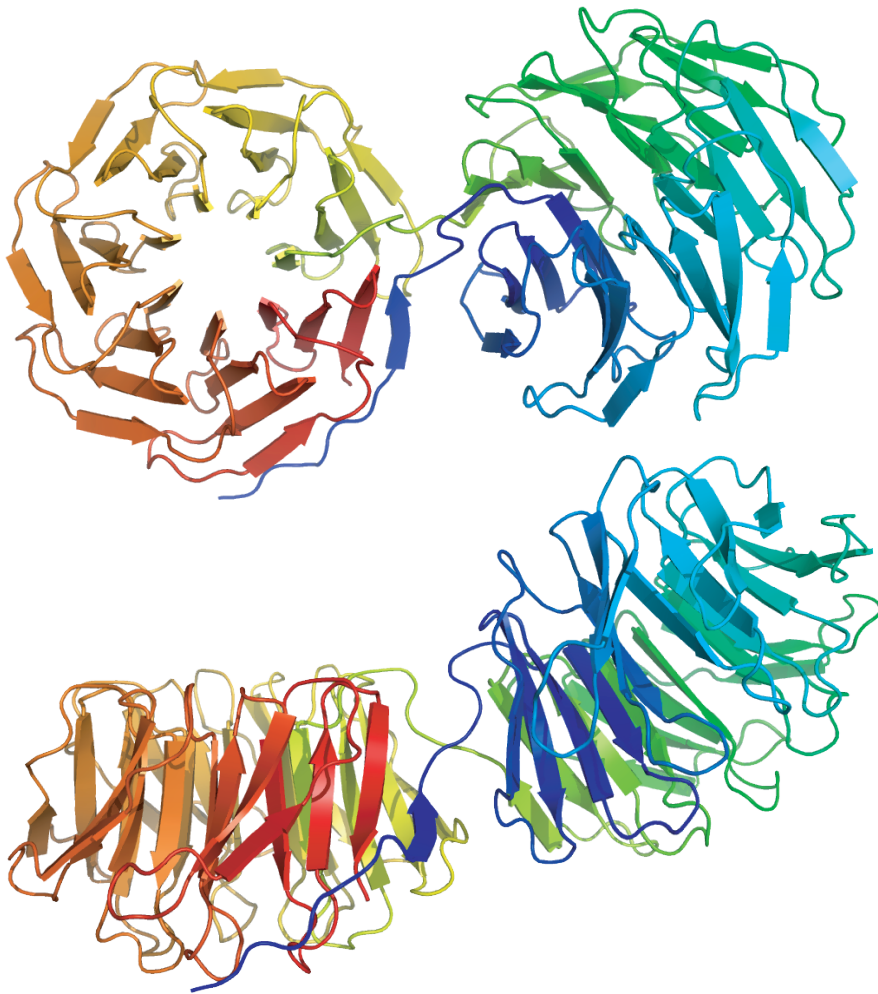


Abbildung 4.8: *Struktur des 14-blättrigen Propellers mit nahezu identischen Blättern. Eine detaillierte Strukturbeschreibung erfolgt im Text.*

```

blade.01..VRGVAFSPDGQTIASASDDKTVKLWNRNGQLLQTLTGHSSS
blade.02..VWGVAFSPDGQTIASASDDKTVKLWNRNGQLLQTLTGHSSS
blade.03..VRGVAFSPDGQTIASASDDKTVKLWNRNGQLLQTLTGHSSS
blade.04..VWGVAFSPDGQTIASASDDKTVKLWNRNGQLLQTLTGHSSS
blade.05..VWGVAFSPDGQTIASASDDKTVKLWNRNGQLLQTLTGHSSS
blade.06..VRGVAFSPDGQTIASASDDKTVKLWNRNGQLLQTLTGHSSS
blade.07..VWGVAFSPDGQTIASASDDKTVKLWNRNGQLLQTLTGHSSS
blade.08..VWGVAFSPDGQTIASASDDKTVKLWNRNGQLLQTLTGHSSS
blade.09..VWGVAFSPDGQTIASASDDKTVKLWNRNGQLLQTLTGHSSS
blade.10..VRGVAFSPDGQTIASASDDKTVKLWNRNGQLLQTLTGHSSS
blade.11..VWGVAFSPDGQTIASASDDKTVKLWNRNGQLLQTLTGHSSS
blade.12..VRGVAFSPDGQTIASASDDKTVKLWNRNGQLLQTLTGHSSS
blade.13..VRGVAFSPDGQTIASASDDKTVKLWNRNGQLLQTLTGHSSS
blade.14..VWGVAFSPDGQTIASASDDKTVKLWN

```

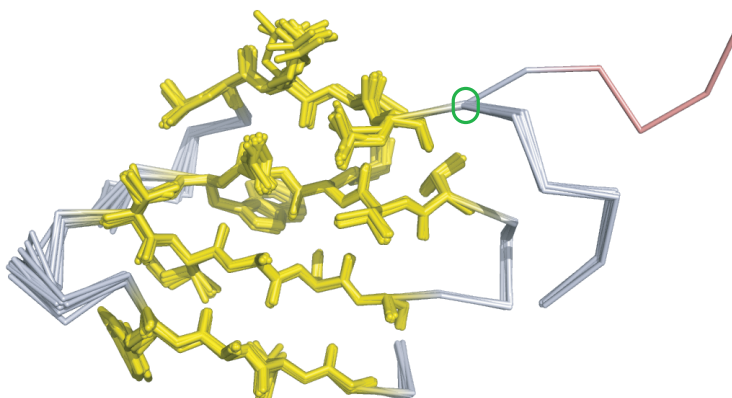


Abbildung 4.9: Sequenzalignment des 14-blättrigen β -Propellers. Im unteren Teil der Abbildung sind die Strukturen der 14 Propellerblätter überlagert. Die zentralen Reste der β -Stränge sind im Sequenzalignment fett hervorgehoben und in der strukturellen Überlagerung in gelb mit den jeweiligen Seitenketten gezeigt. Die Verbindungssequenz zwischen den beiden Domänen ist in rosa gezeigt. Der Thr-Rest um den die Verbindungssequenz rotiert ist, ist eingekreigt.

In Abb. 4.12 sind die Sequenzen von Aip1 und die Sequenz des von uns in seiner Struktur gelösten Proteins gegenübergestellt. Dabei fällt auf, dass die prominenten Sequenzmotive, die in dem von uns kristallisierten Propeller so dominant auftreten, auch in der Sequenz von Aip1 zu finden sind. Auf Grund der Sequenzinhomogenität sind sie jedoch weit weniger stark ausgeprägt. Die Sequenzähnlichkeit spiegelt sich auch in den beiden Strukturen wieder. Der Vergleich der Struktur von Aip1 und dem repetitiven Propeller zeigt, dass während der von uns gelöste Propeller durch eine konstante Abfolge von gleich langen β -Strängen charakterisiert ist, die einzelnen Sekundärstrukturelemente bei Aip1 sehr unterschiedlich ausgeprägt sind. Ein Beispiel ist der stark verlängerte β -Strang D in Blatt aip1.02 mit der Sequenz QTTHILKTT, der in Abb. 4.11 direkt neben dem grauen Balken zur Markierung des Kippwinkels im Vordergrund der grauen Struktur zu sehen ist.

Der markanteste Unterschied zwischen den beiden Strukturen sind die unterschiedlichen Kippwinkel, mit denen die beiden Propellerdomänen zueinander stehen. Bei Aip1 kommen die beiden Domänen mit einem Winkel von 110° zueinander zu liegen, im Falle des repetitiven Propellers ist der Winkel ca. 20° flacher. Bei Aip1 sind die beiden Propellerdomänen auch stärker gegeneinander verdreht.

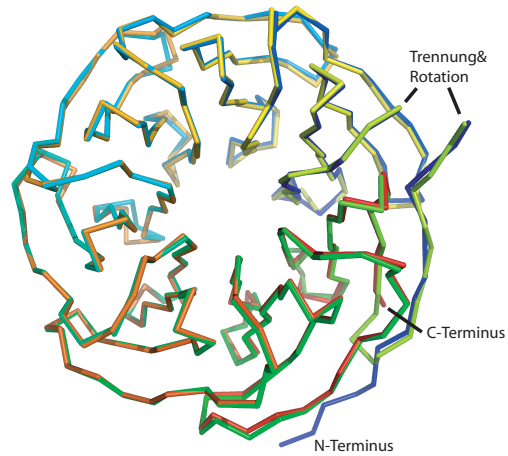


Abbildung 4.10: Überlagerung der N- und der C-terminalen Domäne des Propellers mit identischen Blättern. Die Struktur ist analog zu Abb. 4.8 eingefärbt.

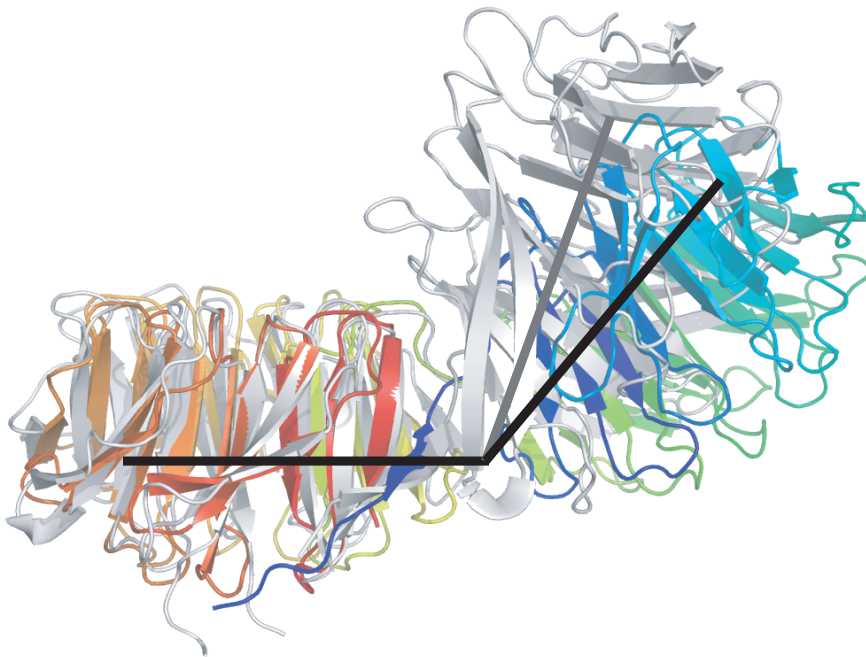


Abbildung 4.11: Überlagerung der N-terminalen Domänen des 14-blättrigen Propellers mit identischen Blättern (farbig, wie in Abb. 4.8) und Aip1 (hellgrau). Die Proteindomänen haben sehr ähnliche Strukturen, die beiden Proteine unterscheiden sich jedoch im Kippwinkel zwischen den Domänen (durch die Balken veranschaulicht).

4.5.1 Überlegungen zur maximalen Anzahl von Propellerblättern in einer Domäne

In Abb. 4.12 sind drei Proteine gegenübergestellt, die aus 14 Propellerblättern bestehen und alle in zwei Domänen zu jeweils sieben Blättern falten. Für das erste Protein Aip1 wurden ursprünglich 10 Propellerblätter vorhergesagt [74]. Nachdem die Struktur gelöst war [98], wurde deutlich, dass die Blätter 01; 10; 11 und 12 nicht als solche erkannt wurden. Bei diesen Blättern ist das WD-Motiv nicht sehr deutlich ausgeprägt. Selbst mit dem Wissen, dass das Aip1-Protein in zwei Domänen faltet, ist es nicht möglich, die Bruchstelle genau vorherzusagen. Gerade im Bereich der Verbindungssequenz zwischen Strang D und dem folgenden Blatt ist die Sequenzvariabilität besonders hoch, so dass man auch nach den Blättern 05; 06 oder 08 Argumente für eine Verbindungssequenz finden kann. Dennoch ergab sich aus der Strukturbestimmung, dass die 14 Propellerblätter in zwei je siebenblättrige Domänen falten. Eine analoge Struktur nehmen wir für das Linkerprotein an. In diesem Protein ist die Sequenzkonserviertheit im Bereich von β -Strang D und den anschließenden Resten auffallend, nur Blatt sieben fällt aus der Reihe. Der extra eingefügte Asn-Rest scheint die Funktion eines Abstandhalters zwischen den beiden Propellern zu haben. Die potentielle Trypsinschnittstelle, die in Abb. 4.12 rot eingekreist ist, an der das Protein geschnitten wird, ist offensichtlich für die Protease zugänglich. Bei diesen beiden Proteinen lassen sich zumindest im Nachhinein Argumente finden, dass das Protein in zwei Domänen bricht.

Die Sequenz des 14-blättrigen Propellers mit identischen Wiederholungen bietet zumindest von der Sequenz her keine Anhaltspunkte für eine offensichtliche Bruchstelle. Die Argumentation bei diesem Protein ist daher: Wenn es einen Propeller mit 14 Blättern in einer Domäne gibt, dann sollte er eine repetitive Sequenz aufweisen, entsprechend der Sequenz des kristallisierten Propellerproteins. Sollte dieser Propeller nicht in eine große Domäne falten, so würde die Sequenz auch eine Faltung in einen sechs- und einen achtblättrigen Propeller nicht ausschließen. Der 14-blättrige Propeller kristallisierte, was dafür spricht, dass die Proteinlösung homogen aus nur einer Proteinspezies besteht. Da dieses Protein in zwei siebenblättrige Domänen faltet, gehen wir davon aus, dass β -Propeller mit 14 Strukturwiederholungen in einer Domäne nicht existieren.

Aussagen über die Existenz eines 12-blättrigen Propellers können wir auf Grund unserer Experimente nicht machen. Gegen das Auftreten einer solchen Faltung spricht aber, dass wir bei den Assemblierungsreaktionen sechsblättrige Propeller erhalten haben, was für die Stabilität dieser Anordnung spricht. Alternativ hätten sich aus dem dreiblättrigen Baustein Oligomere mit neun oder zwölf Blättern bilden können, diese wurden aber nicht gefunden.

Die Protease Tricorn ist in jeder Untereinheit aus 13 Propellerblättern aufgebaut, diese falten in einen sechs- und einen siebenblättrigen Propeller [13]. Dabei sind die einzelnen Propellerblätter auf der Sequenzebene ähnlich heterogen wie Aip1 und die beiden Propeller durch eine 16-Aminosäuren-lange Verbindungssequenz getrennt. Die beiden Propeller sind auch räumlich etwas voneinander getrennt, da sie die C-terminalen Domänen C1 und C2 zwischen sich einschließen.

Im vorhergehenden Kapitel wurden experimentelle Evidenzen geliefert, dass im Rahmen der durchgeführten Assemblierungsexperimente 10-blättrige Propeller gefunden wurden (Abschnitt 3.6.7). Ein solcher 10-blättriger Propeller

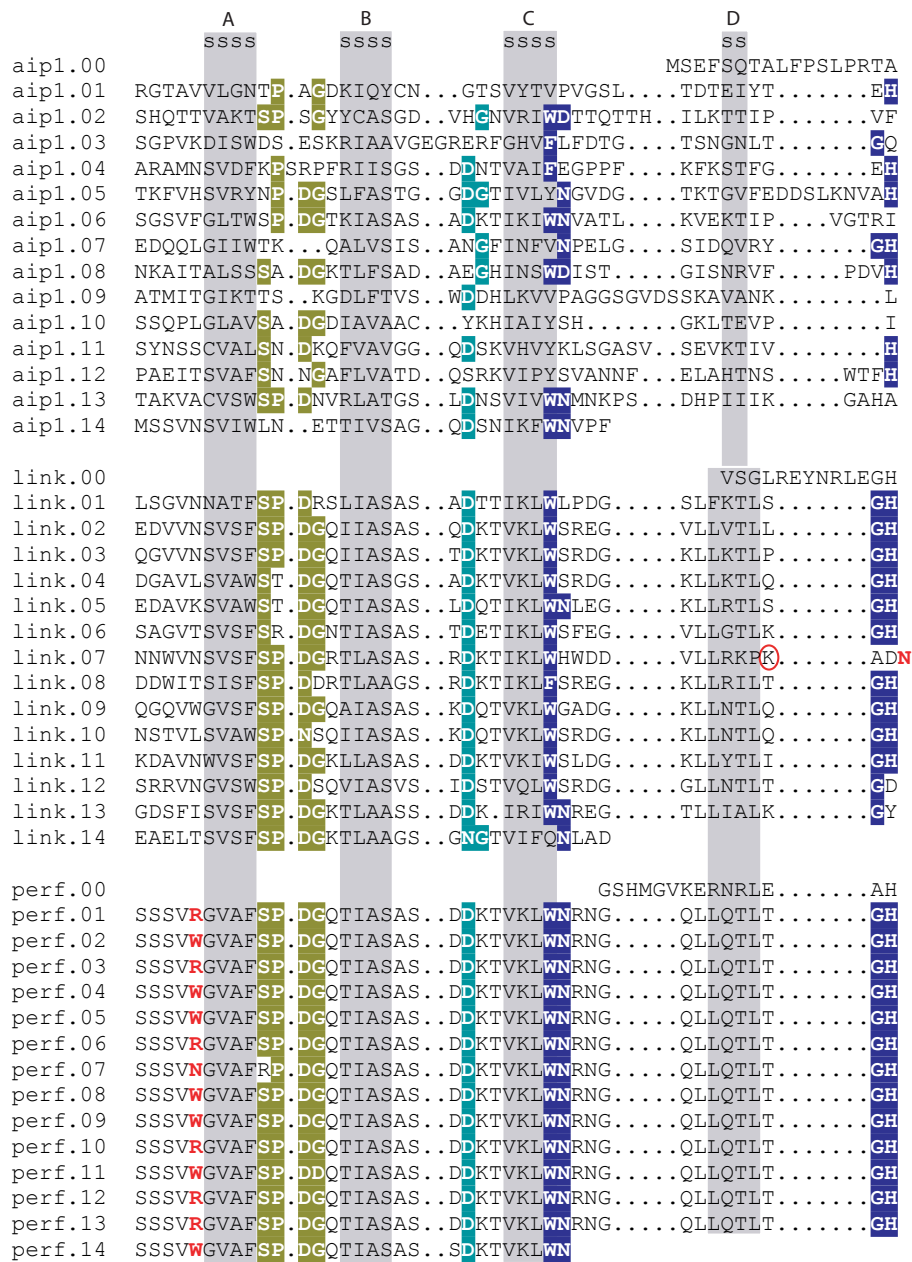


Abbildung 4.12: Strukturelles Alignment der drei β -Propeller Aip1 und der beiden gereinigten Propeller aus Nostoc. Im Linkerprotein ist die potentielle Trypsinschnittstelle am Ende von β -Strang D rot eingekreist.

kann durch Dimerisierung eines fünfblättrigen Propellerbausteins erhalten werden. Die Aussage, dass ein solcher Propeller existiert, beruht dabei auf statischen Lichtstreuexperimenten, relativen Mobilitäten im nicht denaturierenden Acrylamidgel (Nativgel) und der Bestimmung des Sedimentationskoeffizienten mittels analytischer Ultrazentrifugation.

Zusammenfassend gehen wir davon aus, dass die β -Propellerfaltung mit Strukturen bestehend aus 4 bis 10 Propellerblättern in einer Domäne kompatibel ist. Beispiele für vier-, fünf-, sechs-, sieben- und achtblättrige Propeller finden sich bereits in der PDB. Für die Existenz eines 10-blättrigen Propellers haben wir gute biochemische Evidenzen (s. Kapitel 3), und ein 14-blättriger Propeller kann seit dem Wissen um die Struktur des von uns gelösten Propellers mit identischen Propellerblättern ausgeschlossen werden.

4.5.2 Evolutionäre Implikationen der Struktur

Das *Alignment* in Abb. 4.12 zeigt, dass die einzelnen Reste in den 14 Blättern des Propellers identisch sind. Lediglich ein Rest am Anfang von β -Strang A ist variabel. In Strang sieben ist außerdem das Ser des SPDG-Motivs zu einem Arg mutiert. Die extrem hohe Sequenzkonserviertheit zwischen den Blättern lässt den Schluss zu, dass dieser β -Propeller durch Amplifikation aus einem Blatt gebildet wurde. Ein solcher Mechanismus wurde bereits in Kapitel 2 postuliert. In diesem Fall muss die Amplifikation allerdings ein rezentes Ereignis gewesen sein. Das hohe Maß an Sequenzübereinstimmung zwischen den Blättern ist nur zu erklären, wenn man davon ausgeht, dass seit der Bildung des Proteins noch nicht genügend Zeit vergangen ist, so dass die Sequenzen der einzelnen Blätter divergieren konnten.

Die variablen Reste vor β -Strang A bilden den Rand der zentralen Pore, wie in Abb. 4.13 gezeigt. Es ist dabei interessant zu sehen, dass die Mutationen kein regelmäßiges Muster haben. Der Propeller muss daher aus einem einzelnen Blatt entstanden sein, und die Blätter sind später einzeln mutiert. Das *Alignment* der DNA-Sequenz zeigt, dass der Aminosäureaustausch durch eine Veränderung der ersten Base des codierenden Triplets zustande kommt. Da die restliche Sequenz auch auf der DNA-Ebene nahezu mutationsfrei ist, postulieren wir eine funktionale Bedeutung für diese veränderten Reste. Es ist für zahlreiche β -Propeller beschrieben, dass sie Substrate in der zentralen Pore binden (s. Abb. 1.9 in der Einleitung). Vor diesem Hintergrund ist anzunehmen, dass die ringförmig angeordneten Reste um die zentrale Pore eine Bindungsfunktion erfüllen und daher von der strengen Konsensussequenz abweichen.

Es ist extrem selten, dass man Proteine am Anfang ihrer Evolution untersuchen kann, die meisten der heutigen Proteine haben ihre Ausdifferenzierung bereits abgeschlossen. Der hier vorgestellte β -Propeller ist vermutlich eine solche Rarität. Das Protein zeigt in seiner DNA-Sequenz noch wenige weitere Mutationen. Man kann also annehmen, dass das Protein mit seinen Mutationen noch nicht gegen einen Endzustand konvergiert ist. Der 14-blättrige β -Propeller mit nahezu identischen Blättern scheint also ein 'Schnappschuss' eines gerade evolvierenden Proteins zu sein.

4.6 Zusammenfassung

In diesem Kapitel wurde die Kristallstruktur eines 14-blättrigen β -Propellers beschrieben. Die Proteinsequenz des Proteins zeichnet sich dadurch aus, dass die einzelnen Sequenzwiederholungen nahezu identisch sind, lediglich eine Sequenzposition am Rand der zentralen Pore ist variabel. Die Sequenz beinhaltet jedoch keine offensichtliche 'Bruchstelle', wie sie in einem zweiten Propeller auftritt, der sich durch einen Trypsinverdau in zwei gleich große Fragmente zerschneiden lässt. Die von uns gelöste Struktur zeigt, dass das Protein in zwei Domänen mit jeweils sieben Propellerblättern faltet. Dabei sind entsprechend der Sequenzkonserviertheit die einzelnen Blätter auch in ihrer Struktur nahezu identisch.

Diese Struktur ermöglicht zwei Aussagen: Erstens ist mit dieser Struktur nahezu ausgeschlossen, dass β -Propeller mit 14 Blättern in einer Domäne existieren. Zweitens scheint dieses Protein ein 'Schnappschuss' eines gerade evolvierenden Proteins zu sein. Die Sequenzidentität zwischen den Blättern deutet auf eine rezente Amplifikation der Blätter zu einem vollständigen Propeller hin. Die einzige variable Sequenzposition bildet einen Ring um die zentrale Pore des Propellers, was auf eine Funktion als Bindungsstelle hindeutet.

Kapitel 5

Zusammenfassung und Ausblick

In dieser Arbeit wurden verschiedene Aspekte der Evolution der β -Propeller untersucht. Im Mittelpunkt standen dabei die drei thematisch zusammenhängenden Themen: die Frage nach der Entstehung der β -Propellerfaltungsklasse, die Frage wie viele Wiederholungseinheiten maximal zu einer Propellerdomäne zusammengelagert werden können und die Frage, ob die β -Propeller einen homologen Ursprung haben.

Wir konnten für die überwiegende Mehrheit der β -Propeller einen homologen Ursprung nachweisen. Trotz der großen Sequenzdiversität in der Gruppe der β -Propeller konnten wir Ähnlichkeiten zwischen verschiedenen Untergruppen der Propeller finden, die nicht nur struktureller Natur sind. Wir konnten in *Alignments* zeigen, dass Sequenzmotive, die für einzelne Gruppen charakteristisch sind auch in anderen Propellerfamilien auftreten (zum Teil in weniger deutlich ausgeprägter Form). Auch gruppieren die meisten β -Propeller in einem dichten Cluster, wenn man sie mit empfindlichen Sequenzanalysemethoden (HHpred) untereinander und mit allen anderen Proteinen, die ausschließlich aus β -Elementen bestehen vergleicht. Dies ist ein gutes Indiz dafür, dass die Sequenzähnlichkeiten nicht durch Konvergenz auf Grund von Einschränkungen durch die Sekundärstruktur zustande kommen. Weiterhin konnten wir einen homologen Ursprung für 80% der β -Propeller durch statistisch signifikante Gemeinsamkeiten zwischen den Profil-HMMs belegen. Auf Grund der Datenlage können wir nicht unterscheiden, ob die Propeller, die nicht dieser Kerngruppe angehören, analoge Entwicklungen sind, oder ob es sich um homologe Proteine handelt, bei denen die Sequenzinformation für gemeinsame Abstammung durch Sequenzdivergenz verloren gegangen ist.

In der Zukunft wird es interessant sein zu sehen, ob sich durch die voranschreitenden Sequenzierungsprojekte und immer weiter verbesserten Nachweismethoden auch für die wenigen verbleibenden Proteine ein gemeinsamer Ursprung mit der Hauptgruppe der β -Propeller nachweisen lässt.

Uns ist es weiterhin gelungen, β -Propellerdomänen aus kleineren Fragmenten zu assemblieren. Im Fall der Propeller ist in den heute vorkommenden Proteinen

eine strukturelle Symmetrie zu erkennen, in der einzelne Blätter wiederholt aneinandergereiht werden, um die zirkuläre Propellerdomäne zu bilden.

Wir haben verschiedene einzelne Blätter exprimiert und gereinigt und für eines die Amplifikation von einem bis zu fünf Propellerblättern *in vitro* nachvollzogen. Dabei konnten wir zeigen, dass die kleinste selbstfaltende Einheit das duplizierte Blatt ist, welches zu einem achtblättrigen Propeller assembliert, was einem Homotetramer entspricht. Die Propellerbausteine mit drei, vier und fünf Propellerblättern bilden sechs-, acht- und zehnbältrige Propeller. Dies ist ein beachtliches Ergebnis, da die Aminosäuresequenz für alle Blätter konstant gehalten wurde, sich aber dennoch Propeller unterschiedlicher Größe bilden. Das Ergebnis, dass sich Proteindomänen aus kleineren Fragmenten aufbauen lassen, unterstützt die Theorie, dass die heutigen Proteine zum großen Teil aus Supersekundärstrukturelementen gebildet wurden.

Weiterführend wird es interessant sein, die Proteinstruktur eines 10-blättrigen Propellers zu bestimmen. Zur Frage nach der Stabilität der assemblierten Propeller bleibt abzuwarten, ob in der Zukunft in der Literatur über weitere oligomere Propeller berichtet wird. Bisher scheinen die Propeller, die aus einer durchgehenden Polypeptidkette bestehen gegenüber oligomeren Propellern im Vorteil zu sein.

Wir haben die Kristallstruktur eines β -Propellers gelöst, dessen Sequenz aus 14 identischen Wiederholungseinheiten besteht. Das Protein faltet nicht in eine große 14-blättrige Domäne, sondern in zwei strukturell ähnliche Domänen mit jeweils sieben Wiederholungseinheiten. Interessant ist, dass lediglich die Aminosäure in jedem Blatt, die den Ring um die zentrale Pore formt, entweder ein Trp oder ein Arg ist. Dies lässt zwei Schlüsse zu: zum einen scheint der Propeller ein Substrat in der zentralen Öffnung zu binden und zweitens scheint dies Protein einen seltenen Einblick in ein Protein zu geben, das noch im Evolutionsprozess begriffen ist. Seit der Amplifikation des Ausgangsblattes zu einem vollständigen Propeller ist noch nicht genügend Zeit vergangen, damit die einzelnen Blätter divergieren konnten.

Dieses Protein stammt aus dem Cyanobakterium *Nostoc punctiforme*, in dessen Genom sich eine Vielzahl weiterer Propellergene mit unterschiedlich vielen Propellerblättern befindet. Es wäre interessant zu untersuchen, ob sich in dem Genom auch Vorstufen von Propellern finden lassen oder eventuell sogar einzelne Propellerblätter, die sich zu weiteren Proteinen entwickeln könnten.

Die Fragestellungen in dieser Arbeit sind im Zusammenhang der Evolution der β -Propeller bearbeitet worden. Wir postulieren als eine Möglichkeit für die Entstehung dieser Proteinfaltungsklasse, dass einzelne Blätter oligomerisierten und erst später zu einer durchgängigen Polypeptidkette fusionierten. Obwohl wir einen gemeinsamen Ursprung aller β -Propeller vorschlagen, behaupten wir nicht, dass diese alle aus einem ancestralen Propeller entstanden sind. Vielmehr finden wir Hinweise, dass die großen Propellerfamilien unabhängig voneinander aus einzelnen Blättern amplifiziert wurden und dass dies ein fortwährender Prozess ist. Unser evolutionäres Szenario baut auf der Hypothese auf, dass gefaltete Proteine aus einem ancestralen Repertoire von Peptiden entstanden sind, die das Potential hatten, Supersekundärstrukturen zu bilden. Eines von diesen war ein β -Mäander mit der Fähigkeit durch Oligomerisierung toroidale Strukturen zu bilden. Die Differenzierung der ersten Protopropeller ergab eine Population an neuen β -Mäandern, die als Ausgangspunkte für weitere Amplifikationen

dienten. Seitdem hat sich die Welt der β -Propeller durch aufeinanderfolgende Amplifikationen und Diversifikationen kontinuierlich ausgedehnt.

Die Daten in dieser Arbeit deuten jedoch darauf hin, dass die Amplifikation und Fusion eines einzelnen Propellerblattes zu einem vollständigen Propeller vermutlich nicht der limitierende Schritt in der Evolution dieser Faltung sind. Das Auftreten eines einzelnen Blattes in isolierter Form scheint hingegen nur selten zu passieren.

Kapitel 6

Abkürzungen und Erklärungen

Die folgenden Abkürzungen werden in der vorliegenden Arbeit verwendet. Umschreibungen von Begriffen, die sich in der Wissenschaftssprache etabliert haben und sich aus dem Englischen nicht übersetzen lassen, sind ebenfalls aufgelistet und im Text *kursiv* geschrieben..

Alignment	Eine nach bestimmten Kriterien ausgerichtete Schreibweise von Proteinsequenzen.
All- β	Bezeichnung in der SCOP-Datenbank um die Gruppe an Proteinen zu beschreiben, die ausschließlich β -Stränge als Sekundärstrukturelemente besitzt.
blunt end	DNA bei der beide Stränge dieselbe Länge haben und nicht ein Strang mit ungepaarten Nukleotiden den anderen Strang überragt.
Cluster	Anhäufung von Sequenzen, die nahe miteinander verwandt sind.
DNA	Deoxyribonucleic Acid
Erdbeschleunigung	Angegeben in $g = 9,81 \text{ m/s}^2$.
Fold	Kategorisierung in der SCOP-Datenbank. Zwei Proteine mit demselben Fold sind durch eine ähnliche Anordnung von Sekundärstrukturelementen charakterisiert. In SCOP wird davon ausgegangen, dass nicht alle Proteine in einem Fold homolog sind.
GuaHCl	Guanidin Hydrochlorid
HMM	Hidden Markov Modell
IPTG	Isopropyl- β -D-thiogalactopyranosid ist ein künstlicher Induktor des Lactose-Operons in <i>Escherichia coli</i> .
Linker	Verbindungssequenz (entweder DNA oder Protein)

Loop	Ein meist ungeordneter Bereich einer Aminosäurekette, der Sekundärstrukturelemente über größere Entfernung miteinander verbindet.
Molten Globule	Zustand, bei dem sich α -Helices und β -Faltblätter ausbilden, diese aber ohne feste Kontakte zueinander bleiben.
Neighbour-joining	Algorithmus zur Darstellung von Stammbäumen. Dabei werden die Proteinsequenzen paarweise verglichen und die mit der geringsten genetischen Distanz zu einem Ast des Baumes vereinigt. Die genetischen Distanzen der Sequenzen werden erneut berechnet und wieder die nächstverwandten zu einem Ast mit zwei Proteinen zusammengefügt.
Primer	Oligonukleotid. In dieser Arbeit werden Gensequenzen teilweise direkt aus langen Oligonukleotiden aufgebaut. Außerdem werden Primer für PCR-Reaktionen benötigt.
RMSD	Root Mean Square Deviation. Gibt die Abweichung von zwei überlagerten Strukturen zueinander an.
SCOP	Structural Classification of Proteins. Datenbank, die Proteindomänen auf Grund von strukturellen Eigenschaften klassifiziert, aber auch eine Aussage über deren Abstammung macht.
Torus	Geometrisches Gebilde. Die Form ist scheibenförmig mit einem Loch in der Mitte, einem Donut vergleichbar.
Turn	Ein Sekundärstrukturelement, das die Richtung der Aminosäurekette innerhalb weniger Aminosäurereste umkehrt.
Unit	Angabe von Enzymaktivitäten. Eine Unit ist definiert als diejenige Enzymmenge, die unter Standardbedingungen je min ein μmol Substrat umsetzt. Für Restriktionseinheit wird eine Unit bisweilen etwas anders definiert - herstellerabhängig.
Velcro	Klettverschluss. Bei β -Propellern wird so das Ergebnis der zirkulären Permutation beschrieben, durch die ein Propellerblatt aus N- und C-terminalen β -Strängen gemeinsam gebildet wird.

Aminosäuren werden mit dem Ein- oder Dreibuchstabencode angegeben.

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

Literaturverzeichnis

- [1] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410, 1990.
- [2] S. F. Altschul, T. L. Madden, A. A. Schaffer, J. H. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. Gapped blast and psi-blast: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17):3389–3402, 1997.
- [3] V. Alva, M. Ammelburg, J. Soding, and A. N. Lupas. On the origin of the histone fold. *BMC Struct Biol*, 7:17, 2007.
- [4] M. A. Andrade, C. P. Ponting, T. J. Gibson, and P. Bork. Homology-based method for identification of protein repeats using statistical significance estimates. *J Mol Biol*, 298(3):521–37, 2000.
- [5] A. Andreeva, D. Howorth, S. E. Brenner, T. J. P. Hubbard, C. Chothia, and A. G. Murzin. Scop database in 2004: refinements integrate structure and sequence family data. *Nucleic Acids Research*, 32:D226–D229, 2004.
- [6] L. Aravind, L. M. Iyer, and E. V. Koonin. Comparative genomics and structural biology of the molecular innovations of eukaryotes. *Curr Opin Struct Biol*, 16(3):409–19, 2006.
- [7] L. Aravind and E. V. Koonin. Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. *J Mol Biol*, 287(5):1023–40, 1999.
- [8] P. W. Atkins. *Physikalische Chemie*. Wiley-VCH, 2 edition, 2001.
- [9] H. G. Beisel, S. Kawabata, S. Iwanaga, R. Huber, and W. Bode. Tachylectin-2: crystal structure of a specific glcnac/galnac-binding lectin involved in the innate immunity host defense of the japanese horseshoe crab tachypleus tridentatus. *Embo J*, 18(9):2313–22, 1999.
- [10] S. A. Benner, M. A. Cohen, and G. H. Gonnet. Empirical and structural models for insertions and deletions in the divergent evolution of proteins. *J Mol Biol*, 229(4):1065–82, 1993.
- [11] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Res*, 28(1):235–42, 2000.

- [12] P. Bork, A. K. Downing, B. Kieffer, and I. D. Campbell. Structure and distribution of modules in extracellular proteins. *Q Rev Biophys*, 29(2):119–67, 1996.
- [13] H. Brandstetter, J. S. Kim, M. Groll, and R. Huber. Crystal structure of the tricorn protease reveals a protein disassembly line. *Nature*, 414(6862):466–70, 2001.
- [14] W. Cai, J. Pei, and N. V. Grishin. Reconstruction of ancestral protein sequences and its applications. *BMC Evol Biol*, 4:33, 2004.
- [15] J. Castresana, M. Lubben, M. Saraste, and D. G. Higgins. Evolution of cytochrome oxidase, an enzyme older than atmospheric oxygen. *Embo J*, 13(11):2516–25, 1994.
- [16] C. Chothia and A. M. Lesk. The relation between the divergence of sequence and structure in proteins. *Embo J*, 5(4):823–6, 1986.
- [17] M. Coles, T. Diercks, J. Liermann, A. Groger, B. Rockel, W. Baumeister, K. K. Koretke, A. Lupas, J. Peters, and H. Kessler. The solution structure of vat-n reveals a 'missing link' in the evolution of complex enzymes from a simple betaalphabetabeta element. *Curr Biol*, 9(20):1158–68, 1999.
- [18] M. Coles, M. Hulko, S. Djuranovic, V. Truffault, K. Koretke, J. Martin, and A. N. Lupas. Common evolutionary origin of swapped-hairpin and double-psi beta barrels. *Structure*, 14(10):1489–98, 2006.
- [19] C. Combet, C. Blanchet, C. Geourjon, and G. Deleage. Nps@: network protein sequence analysis. *Trends Biochem Sci*, 25(3):147–50, 2000.
- [20] R. R. Copley and P. Bork. Homology among (betaalpha)(8) barrels: implications for the evolution of metabolic pathways. *J Mol Biol*, 303(4):627–41, 2000.
- [21] R. R. Copley, R. B. Russell, and C. P. Ponting. Sialidase-like asp-boxes: sequence-similar structures within different protein folds. *Protein Sci*, 10(2):285–92, 2001.
- [22] A. F. Coulson and J. Moult. A unifold, mesofold, and superfold model of protein fold use. *Proteins*, 46(1):61–71, 2002.
- [23] R. F. Doolittle. Convergent evolution: the need to be explicit. *Trends Biochem Sci*, 19(1):15–8, 1994.
- [24] R. F. Doolittle, D. F. Feng, and G. Cho. Determining divergence times with protein clocks. *Biol Bull*, 196(3):356–7; discussion 357–8, 1999.
- [25] S. R. Eddy. Profile hidden markov models. *Bioinformatics*, 14(9):755–63, 1998.
- [26] T. A. Edwards, B. D. Wilkinson, R. P. Wharton, and A. K. Aggarwal. Model of the brain tumor-pumilio translation repressor complex. *Genes Dev*, 17(20):2508–13, 2003.

- [27] P. Emsley and K. Cowtan. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, 60(Pt 12 Pt 1):2126–32, 2004.
- [28] R. D. Finn, J. Mistry, B. Schuster-Bockler, S. Griffiths-Jones, V. Hollich, T. Lassmann, S. Moxon, M. Marshall, A. Khanna, R. Durbin, S. R. Eddy, E. L. Sonnhammer, and A. Bateman. Pfam: clans, web tools and services. *Nucleic Acids Res*, 34(Database issue):D247–51, 2006.
- [29] T. Frickey and A. Lupas. Clans: a java application for visualizing protein families based on pairwise similarity. *Bioinformatics*, 20(18):3702–3704, 2004.
- [30] T. Frickey and A. N. Lupas. Phylogenie: automated phylome generation and analysis. *Nucleic Acids Res*, 32(17):5231–8, 2004.
- [31] V. Fulop, Z. Bocskei, and L. Polgar. Prolyl oligopeptidase: an unusual beta-propeller domain regulates proteolysis. *Cell*, 94(2):161–70, 1998.
- [32] V. Fulop, J. W. Moir, S. J. Ferguson, and J. Hajdu. The anatomy of a bifunctional enzyme: structural basis for reduction of oxygen to water and synthesis of nitric oxide by cytochrome cd1. *Cell*, 81(3):369–77, 1995.
- [33] D. J. Futuyma. *Evolutionary Biology*. Sinauer Associates, Sunderland, 3 edition, 1998.
- [34] J. Gettemans, K. Meerschaert, J. Vandekerckhove, and V. De Corte. A kelch beta propeller featuring as a g beta structural mimic: reinventing the wheel? *Sci STKE*, 2003(191):PE27, 2003.
- [35] E. Ghedin, N. A. Sengamalay, M. Shumway, J. Zaborsky, T. Feldblyum, V. Subbu, D. J. Spiro, J. Sitz, H. Koo, P. Bolotov, D. Dernovoy, T. Tatusova, Y. Bao, K. St George, J. Taylor, D. J. Lipman, C. M. Fraser, J. K. Taubenberger, and S. L. Salzberg. Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature*, 437(7062):1162–6, 2005.
- [36] F. X. Gomis-Ruth, U. Gohlke, M. Betz, V. Knauper, G. Murphy, C. Lopez-Otin, and W. Bode. The helping hand of collagenase-3 (mmp-13): 2.7 a crystal structure of its c-terminal haemopexin-like domain. *J Mol Biol*, 264(3):556–66, 1996.
- [37] L. H. Greene, T. E. Lewis, S. Addou, A. Cuff, T. Dallman, M. Dibley, O. Redfern, F. Pearl, R. Nambudiry, A. Reid, I. Sillitoe, C. Yeats, J. M. Thornton, and C. A. Orengo. The cath domain structure database: new protocols and classification levels give a more comprehensive resource for exploring evolution. *Nucleic Acids Res*, 35(Database issue):D291–7, 2007.
- [38] N. Greenfield and G. D. Fasman. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry*, 8(10):4108–16, 1969.
- [39] N. V. Grishin. Fold change in evolution of protein structures. *J Struct Biol*, 134(2-3):167–85, 2001.

- [40] N. Guex and M. C. Peitsch. Swiss-model and the swiss-pdbviewer: an environment for comparative protein modeling. *Electrophoresis*, 18(15):2714–23, 1997.
- [41] N. C. Ha, B. C. Oh, S. Shin, H. J. Kim, T. K. Oh, Y. O. Kim, K. Y. Choi, and B. H. Oh. Crystal structures of a novel, thermostable phytase in partially and fully calcium-loaded states. *Nat Struct Biol*, 7(2):147–53, 2000.
- [42] B. Hocker, J. Claren, and R. Sterner. Mimicking enzyme evolution by generating new (beta alpha)(8)-barrels from (beta alpha)(4)-half-barrels. *Proceedings of the National Academy of Sciences of the United States of America*, 101(47):16448–16453, 2004.
- [43] L. Holm and C. Sander. Dali: a network tool for protein structure comparison. *Trends Biochem Sci*, 20(11):478–80, 1995.
- [44] D. H. Huson and D. Bryant. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*, 23(2):254–67, 2006.
- [45] L. Janda, P. Tichy, J. Spizek, and M. Petricek. A deduced thermomonospora curvata protein containing serine/threonine protein kinase and wd-repeat domains. *J Bacteriol*, 178(5):1487–9, 1996.
- [46] H. Jeon, W. Meng, J. Takagi, M. J. Eck, T. A. Springer, and S. C. Blacklow. Implications for familial hypercholesterolemia from the structure of the ldl receptor ywtd-egf domain pair. *Nat Struct Biol*, 8(6):499–504, 2001.
- [47] H. Jing, J. Takagi, J. H. Liu, S. Lindgren, R. G. Zhang, A. Joachimiak, J. H. Wang, and T. A. Springer. Archaeal surface layer proteins contain beta propeller, pkd, and beta helix domains and are related to metazoan cell surface proteins. *Structure*, 10(10):1453–64, 2002.
- [48] T. A. Jones, J. Y. Zou, S. W. Cowan, and M. Kjeldgaard. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A*, 47 (Pt 2):110–9, 1991.
- [49] G. F. Joyce. The antiquity of rna-based evolution. *Nature*, 418(6894):214–21, 2002.
- [50] J. Jung and B. Lee. Circularly permuted proteins in the protein structure database. *Protein Sci*, 10(9):1881–6, 2001.
- [51] T. Kaneko, Y. Nakamura, C. P. Wolk, T. Kuritz, S. Sasamoto, A. Watanabe, M. Iriguchi, A. Ishikawa, K. Kawashima, T. Kimura, Y. Kishida, M. Kohara, M. Matsumoto, A. Matsuno, A. Muraki, N. Nakazaki, S. Shimpo, M. Sugimoto, M. Takazawa, M. Yamada, M. Yasuda, and S. Tabata. Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium anabaena sp. strain pcc 7120. *DNA Res*, 8(5):205–13; 227–53, 2001.

- [52] Y. Kawarabayasi, Y. Hino, H. Horikawa, K. Jin-no, M. Takahashi, M. Sekine, S. Baba, A. Ankai, H. Kosugi, A. Hosoyama, S. Fukui, Y. Nagai, K. Nishijima, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Kato, T. Yoshizawa, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, S. Masuda, M. Yanagii, M. Nishimura, A. Yamagishi, T. Oshima, and H. Kikuchi. Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *sulfolobus tokodaii* strain 7. *DNA Res*, 8(4):123–40, 2001.
- [53] A. Kohl, H. K. Binz, P. Forrer, M. T. Stumpp, A. Pluckthun, and M. G. Grutter. Designed to be stable: crystal structure of a consensus ankyrin repeat protein. *Proc Natl Acad Sci U S A*, 100(4):1700–5, 2003.
- [54] N. Kostlanova, E. P. Mitchell, H. Lortat-Jacob, S. Oscarson, M. Lahmann, N. Gilboa-Garber, G. Chambat, M. Wimmerova, and A. Imberty. The fucose-binding lectin from *ralstonia solanacearum*. a new type of beta-propeller architecture formed by oligomerization and interacting with fucose, fucosyllactose, and plant xyloglucan. *J Biol Chem*, 280(30):27839–49, 2005.
- [55] D. G. Lambright, J. Sondek, A. Bohm, N. P. Skiba, H. E. Hamm, and P. B. Sigler. The 2.0 Å crystal structure of a heterotrimeric G protein. *Nature*, 379(6563):311–9, 1996.
- [56] D. Lang, R. Thoma, M. Henn-Sax, R. Sterner, and M. Wilmanns. Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion. *Science*, 289(5484):1546–50, 2000.
- [57] R. A. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton. Procheck: A program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, (26):283–291, 1993.
- [58] I. Letunic, R. R. Copley, B. Pils, S. Pinkert, J. Schultz, and P. Bork. Smart 5: domains in the context of genomes and networks. *Nucleic Acids Res*, 34(Database issue):D257–60, 2006.
- [59] D. Li and R. Roberts. Wd-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. *Cellular and Molecular Life Sciences*, 58(14):2085–2097, 2001.
- [60] S. W. Lockless and R. Ranganathan. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science*, 286(5438):295–9, 1999.
- [61] A. N. Lupas, C. P. Ponting, and R. B. Russell. On the evolution of protein folds: are similar motifs in different protein folds the result of convergence, insertion, or relics of an ancient peptide world? *J Struct Biol*, 134(2-3):191–203, 2001.
- [62] M. Lynch. Genomics. gene duplication and evolution. *Science*, 297(5583):945–7, 2002.
- [63] E. R. Main, Y. Xiong, M. J. Cocco, L. D’Andrea, and L. Regan. Design of stable alpha-helical arrays from an idealized tpr motif. *Structure*, 11(5):497–508, 2003.

- [64] C. R. McCudden, M. D. Hains, R. J. Kimple, D. P. Siderovski, and F. S. Willard. G-protein signaling: back to the future. *Cell Mol Life Sci*, 62(5):551–77, 2005.
- [65] S. L. Miller. A production of amino acids under possible primitive earth conditions. *Science*, 117(3046):528–9, 1953.
- [66] G. Milligan and E. Kostenis. Heterotrimeric g-proteins: a short history. *Br J Pharmacol*, 147 Suppl 1:S46–55, 2006.
- [67] W. T. Morgan, P. Muster, F. Tatum, S. M. Kao, J. Alam, and A. Smith. Identification of the histidine residues of hemopexin that coordinate with heme-iron and of a receptor-binding region. *J Biol Chem*, 268(9):6256–62, 1993.
- [68] G. N. Murshudov, A. A. Vagin, and E. J. Dodson. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*, 53(Pt 3):240–55, 1997.
- [69] A. G. Murzin. Structural principles for the propeller assembly of beta-sheets: the preference for seven-fold symmetry. *Proteins*, 14(2):191–201, 1992.
- [70] A. G. Murzin. How far divergent evolution goes in proteins. *Curr Opin Struct Biol*, 8(3):380–7, 1998.
- [71] N. Nagano, C. A. Orengo, and J. M. Thornton. One fold with many functions: the evolutionary relationships between tim barrel families based on their sequences, structures and functions. *J Mol Biol*, 321(5):741–65, 2002.
- [72] E. J. Neer, C. J. Schmidt, R. Nambudripad, and T. F. Smith. The ancient regulatory-protein family of wd-repeat proteins. *Nature*, 371(6495):297–300, 1994.
- [73] W. M. Oldham and H. E. Hamm. Structural basis of function in heterotrimeric g proteins. *Q Rev Biophys*, 39(2):117–66, 2006.
- [74] S. Ono. The caenorhabditis elegans unc-78 gene encodes a homologue of actin-interacting protein 1 required for organized assembly of muscle actin filaments. *J Cell Biol*, 152(6):1313–9, 2001.
- [75] C. A. Orengo and J. M. Thornton. Protein families and their evolution—a structural perspective. *Annu Rev Biochem*, 74:867–900, 2005.
- [76] A. Oubrie, H. J. Rozeboom, K. H. Kalk, J. A. Duine, and B. W. Dijkstra. The 1.7 Å crystal structure of the apo form of the soluble quinoprotein glucose dehydrogenase from acinetobacter calcoaceticus reveals a novel internal conserved sequence repeat. *J Mol Biol*, 289(2):319–33, 1999.
- [77] S. Pascarella and P. Argos. Analysis of insertions/deletions in protein structures. *J Mol Biol*, 224(2):461–71, 1992.

- [78] E. Potterton, P. Briggs, M. Turkenburg, and E. Dodson. A graphical user interface to the ccp4 program suite. *Acta Crystallogr D Biol Crystallogr*, 59(Pt 7):1131–7, 2003.
- [79] L. Renault, N. Nassar, I. Vetter, J. Becker, C. Klebe, M. Roth, and A. Wittinghofer. The 1.7 Å crystal structure of the regulator of chromosome condensation (rcc1) reveals a seven-bladed propeller. *Nature*, 392(6671):97–101, 1998.
- [80] A. Ruepp, W. Graml, M. L. Santos-Martinez, K. K. Koretke, C. Volker, H. W. Mewes, D. Frishman, S. Stocker, A. N. Lupas, and W. Baumeister. The genome sequence of the thermoacidophilic scavenger thermoplasma acidophilum. *Nature*, 407(6803):508–13, 2000.
- [81] R. B. Russell. Detection of protein three-dimensional side-chain patterns: new examples of convergent evolution. *J Mol Biol*, 279(5):1211–27, 1998.
- [82] G. M. Salem, E. G. Hutchinson, C. A. Orengo, and J. M. Thornton. Correlation of observed fold frequency with the occurrence of local structural motifs. *J Mol Biol*, 287(5):969–81, 1999.
- [83] E. I. Scharff, J. Koepke, G. Fritsch, C. Lucke, and H. Ruterjans. Crystal structure of diisopropylfluorophosphatase from *loligo vulgaris*. *Structure*, 9(6):493–502, 2001.
- [84] S. Shin, N. C. Ha, B. C. Oh, T. K. Oh, and B. H. Oh. Enzyme mechanism and catalytic property of beta propeller phytase. *Structure*, 9(9):851–8, 2001.
- [85] N. Sinha and R. Nussinov. Point mutations and sequence variability in proteins: redistributions of preexisting populations. *Proc Natl Acad Sci U S A*, 98(6):3139–44, 2001.
- [86] F. J. Slack and G. Ruvkun. A novel repeat domain that is often associated with ring finger and b-box motifs. *Trends Biochem Sci*, 23(12):474–5, 1998.
- [87] T. F. Smith, C. Gaitatzes, K. Saxena, and E. J. Neer. The wd repeat: a common architecture for diverse functions. *Trends Biochem Sci*, 24(5):181–5, 1999.
- [88] J. Soding. Protein homology detection by hmm-hmm comparison. *Bioinformatics*, 21(7):951–60, 2005.
- [89] J. Soding, A. Biegert, and A. N. Lupas. The hhpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research*, 33:W244–W248, 2005.
- [90] J. Soding and A. N. Lupas. More than the sum of their parts: on the evolution of proteins from peptides. *Bioessays*, 25(9):837–46, 2003.
- [91] J. Soding, M. Remmert, and A. Biegert. Hhrep: de novo protein repeat detection and the origin of tim barrels. *Nucleic Acids Res*, 34(Web Server issue):W137–42, 2006.

- [92] T. A. Springer. An extracellular beta-propeller module predicted in lipoprotein and scavenger receptors, tyrosine kinases, epidermal growth factor precursor, and extracellular matrix components. *J Mol Biol*, 283(4):837–62, 1998.
- [93] T. Terwilliger. Solve and resolve: automated structure solution, density modification and model building. *J Synchrotron Radiat*, 11(Pt 1):49–52, 2004.
- [94] J. D. Thompson, T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. The clustal x windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*, 25(24):4876–82, 1997.
- [95] J. M. Thornton and B. L. Sibanda. Amino and carboxy-terminal regions in globular proteins. *J Mol Biol*, 167(2):443–60, 1983.
- [96] S. Tsukiji, S. B. Pattanaik, and H. Suga. An alcohol dehydrogenase ribozyme. *Nat Struct Biol*, 10(9):713–7, 2003.
- [97] B. Venkatesh. Evolution and diversity of fish genomes. *Curr Opin Genet Dev*, 13(6):588–92, 2003.
- [98] W. C. Voegtli, A. Y. Madrona, and D. K. Wilson. The structure of aip1p, a wd repeat protein that regulates cofilin-mediated actin depolymerization. *J Biol Chem*, 278(36):34373–9, 2003.
- [99] D. Voet and J. G. Voet. *Biochemistry*. Wiley, New York, 2 edition.
- [100] C. Vogel and V. Morea. Duplication, divergence and formation of novel protein topologies. *Bioessays*, 28(10):973–8, 2006.
- [101] M. A. Wall, D. E. Coleman, E. Lee, J. A. Iniguez-Lluhi, B. A. Posner, A. G. Gilman, and S. R. Sprang. The structure of the g protein heterotrimer gi alpha 1 beta 1 gamma 2. *Cell*, 83(6):1047–58, 1995.
- [102] R. Winter and F. Noll. *Methoden der Biophysikalischen Chemie*. Teubner Studienbücher, 1998.
- [103] Z. Xia, W. Dai, Y. Zhang, S. A. White, G. D. Boyd, and F. S. Mathews. Determination of the gene sequence and the three-dimensional structure at 2.4 angstroms resolution of methanol dehydrogenase from methylophilus w3a1. *J Mol Biol*, 259(3):480–501, 1996.
- [104] I. Yadid and D. S. Tawfik. Reconstruction of functional beta-propeller lectins via homo-oligomeric assembly of shorter fragments. *J Mol Biol*, 365(1):10–7, 2007.
- [105] K. Yaoi, H. Kondo, N. Noro, M. Suzuki, S. Tsuda, and Y. Mitsuishi. Tandem repeat of a seven-bladed beta-propeller domain in oligoxyloglucan reducing-end-specific cellobiohydrolase. *Structure*, 12(7):1209–1217, 2004.

Mitwirkende und Danksagung

Die vorliegende Arbeit wurde von Januar 2003 bis August 2007 in der Abteilung Proteinevolution am Max-Planck-Institut für Entwicklungsbiologie in Tübingen durchgeführt. An dieser Stelle möchte ich mich bei allen bedanken, die zum Gelingen dieser Arbeit beigetragen haben und die Beiträge von Kooperationspartnern würdigen:

Mein besonderer Dank gilt Herrn Prof. Andrei Lupas für seine stete Unterstützung und sein reges Interesse am Fortgang dieser Arbeit. Ich danke ihm für die interessante Fragestellung und die exzellente wissenschaftliche Betreuung.

Ich danke Herrn Prof. Baumeister für die Betreuung der Arbeit an der Technischen Universität München.

Für die gute Betreuung und praktische Unterstützung während der Laborarbeit danke ich Jörg Martin.

Für die Durchsicht der Arbeit und die konstruktiven Korrekturvorschläge möchte ich mich bei Reinhard Albrecht, Ursula Chaudhuri, Marcus Hartmann, Birte Hernandez, Birte Höcker, Dirk Linke Jörg Martin und Kornelius Zeth bedanken.

Michael Hulko danke ich für die angenehme Zeit an der gemeinsamen Laborbank.

Bei allen Mitarbeitern in der Arbeitsgruppe bedanke ich mich für die angenehme Arbeitsatmosphäre: Alexander Diemand, Astrid Ursinus, Birte Hernandez, Birte Höcker, Carolin Ewers, Cedric Hobel, Christian Mayer, Christian Münch, Dirk Linke, Felipe Figuera, Hedda Ferris, Ines Wanke, Johannes Söding, Julia Fischer, Justyna Serek-Heuberger, Karin Lehmann, Kornelius Zeth, Manjunatha Karpenahalli, Marcus Hartmann, Marina Lamparter, Markus Gruber, Michael Gebert, Michael Habeck, Moritz Ammelburg, Murray Coles, Oswin Ridderbusch, Reinhard Albrecht, Silvia Deiss, Sonja Billerbeck, Stanislaw Dunin-Horkawicz, Tancred Frickey, Thomas Arnold, Verena Schünemann, Vikram Alva, Vincent Truffault.

Besonderen Dank schulde ich meiner Verlobten Ariane Schleich für die Motivation und die Unterstützung - ja, ich will.

Am meisten Danke ich meinen Eltern, die mir meine gesamte Ausbildung ermöglicht haben und mich im Verlauf der Arbeit unentwegt gefördert haben.

An den Projekten waren unterschiedliche Personen beteiligt. Ich möchte an dieser Stelle die Beiträge der einzelnen Wissenschaftler aufzählen und bedanke mich bei allen für die gute Zusammenarbeit:

Die grundlegende Fragestellung wurde mir von Andrei Lupas überlassen.

Im Rahmen der bioinformatischen Analysen:

Vikram Alva hat Informationen zu den Asp-Boxen geliefert.

Christian Mayer und Deepak Suri haben geholfen Abb. 2.4 zu erstellen.

Johannes Soeding hat das Programm buildali.pl programmiert und die Daten für Abb. 2.3 zur Verfügung gestellt.

Im Rahmen der Arbeiten mit den Propellerfragmenten:

Murray Coles und Vincent Truffault haben NMR-Messungen durchgeführt und mir bei der Aufnahme der Spektren im Rahmen der in Abschnitt 3.7 beschriebenen Mischungsexperimente geholfen. Sie haben mich ausserdem bei der Auswertung der Daten sehr kompetent unterstützt.

Im Rahmen der Strukturaufklärung:

Reinhard Albrecht hat die Proteinkristalle mit verschiedenen Schwermetallen getränkt.

Kornelius Zeth hat die Datensätze der Proteinkristalle aufgenommen und hat mich bei der Lösung der Proteinstruktur entscheidend unterstützt.