

Grundlagen über Proteine

aus: R. Merkl / S. Waack, Bioinformatik interaktiv. Wiley/VCH, Weinheim 2003, S. 5-14.

1.2 Proteine

Proteine sind ebenfalls lineare Makromoleküle, Bausteine sind in diesem Fall die 20 natürlich vorkommenden Aminosäuren. Der Aufbau dieser Molekülfamilie ist einheitlich und besteht aus einem, in allen Aminosäuren identischen, sowie einem variablen Teil, der häufig auch Aminosäurerest genannt wird (siehe **Abb. 1.3**). Form und Art dieses Restes beeinflussen die Wechselwirkungen zwischen den Bausteinen. Die wichtigsten Wechselwirkungen sind Wasserstoffbrückenbindungen zwischen polaren Seitenketten.

Struktur von Aminosäuren

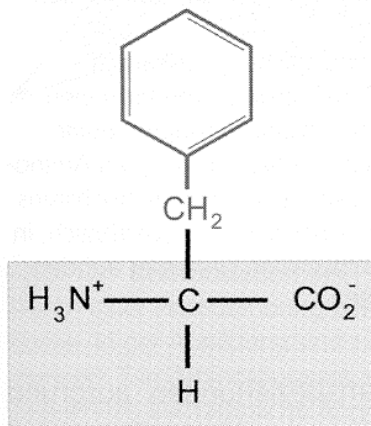


Abb. 1.3: Strukturformel der Aminosäure Phenylalanin.

Der in allen Aminosäuren gleichartige Anteil ist in der Strukturformel grau unterlegt. In jeder Aminosäure ist mit dem zentralen C-Atom ein Wasserstoffatom (unten), eine Aminogruppe (links), eine Carboxylgruppe (rechts) und eine Seitengruppe (hier rot gezeichnet) verknüpft. Das zentrale C-Atom wird wegen seiner Lage im Molekül häufig als C_α-Atom bezeichnet.

Eigenschaften von Aminosäuren

Aufgrund des unterschiedlichen Aufbaus der Seitenkette haben die Aminosäuren von einander abweichende physiko-chemische Eigenschaften. Sie lassen sich z. B. bezüglich der ionischen Ladung in die Gruppen **basisch**, **sauer** und **neutral** einteilen. Unter den neutralen Aminosäuren, die keine elektrische Gesamtladung tragen, finden sich wiederum **polare**, d. h. solche, die innerhalb des Moleküls eine unterschiedliche Ladungsverteilung aufweisen. Apolare, neutrale Aminosäuren sind **hydrophob** (wasserabstoßend). Sie tendieren dazu, untereinander und mit anderen hydrophoben Gruppen zu interagieren. Mit **hydrophil** bezeichnet man Moleküle die gut wasserlöslich sind. Ein Spezialfall ist Prolin, eine zyklische Iminosäure. Nach der Ausbildung der Peptidbindung steht in dieser Aminosäure kein Wasserstoff mehr zur Ausbildung von Wasserstoffbrückenbindungen zur Verfügung. Diese Eigenart hat erheblichen Einfluss auf die Proteinstruktur.

Gruppierung hinsichtlich physikalisch-chemischer Eigenschaften

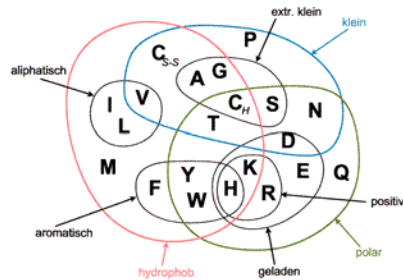


Abb. 1.4: Venn-Diagramm der 20 natürlichen, in Proteinen vorkommenden Aminosäuren. Die Aminosäuren wurden in der Abbildung aufgrund solcher physikalisch-chemischer Eigenschaften geclustert, die für die Tertiärstruktur von Proteinen wichtig sind. Die Aminosäuren sind im Wesentlichen in zwei Gruppen (polar und hydrophob) eingeteilt, eine dritte Gruppe (klein) umfasst die kleineren Aminosäuren. Die Menge extrem klein enthält diejenigen Aminosäuren, die höchstens zwei Seitenkettenatome besitzen. Cys in reduzierter Form ($C_{\beta-S}$) ist Ser ähnlich, in oxidierter Form ($C_{\beta-S-S}$) ähnelt es Val. Aufgrund des speziellen Einflusses auf den Hauptkettenverlauf liegt Pro isoliert, nach [Taylor86].

Diese hier dargestellten Verwandtschaftsbeziehungen aufgrund physikalischer und chemischer Eigenschaften der Aminosäuren sind die Grundlage für praktisch alle Sequenzvergleichs- und Alignment-

Haupt- und Seitenkette tragen zur Stabilisierung der Proteinkonformation bei. Das erste Kohlenstoffatom, das im Rest auf das C_{α} -Atom folgt, wird C_{β} -Atom genannt.

1.4 Hierarchische Beschreibung von Proteinstrukturen

Die Eigenschaften der Seitenketten bestimmen die Wechselwirkungen innerhalb des Proteins und damit dessen dreidimensionale Konformation. Dieser Konformationszustand kann auf verschiedenen Abstraktionsebenen beschrieben werden:

- Als **Primärstruktur** auf der Ebene der **Sequenz** durch die Abfolge der Aminosäuren.
- Auf dem Niveau der **Sekundärstruktur**; aus der Polypeptidkette falten sich Sekundärstrukturelemente, die reguläre Arrangements des Hauptkettenverlaufes ergeben.
- Als **Tertiärstruktur**; sie beschreibt die räumliche Anordnung aller Atome im Raum.

Und auf der Ebene der Proteine:

- Als **Quaternärstruktur**; sie definiert die Anordnung von Proteinen in Proteinkomplexen.

Wir werden Algorithmen kennen lernen, um Primär- Sekundär- und Tertiärstruktur von Proteinen zu analysieren, zu vergleichen oder vorherzusagen.

1.5 Sekundärstrukturelemente

Die Grundbausteine der Proteine sind, wie wir wissen, die Aminosäuren. Diese besitzen zwar jeweils individuelle, physiko-chemische Eigenschaften, der Vergleich evolutionär verwandter Proteine belegt jedoch, dass Aminosäuren sich gegenseitig substituieren können. Aus solchen Vergleichen werden Scores für die Ähnlichkeit von Aminosäureseitenketten abgeleitet. Sie bilden die Grundlage für den Vergleich von Proteinsequenzen. Die nächst höhere Abstraktionsebene, auf der Proteine beschrieben werden können, ist die der **Sekundärstruktur**. Sekundärstrukturelemente sind regelmäßige 3D-Substrukturen des Hauptkettenverlaufes (*backbone*) einer Peptidkette. Bei der Klassifizierung von Sekundärstrukturelementen wird Art und Anordnung der Aminosäurereste (Seitenketten) ignoriert. Die Stabilisierung

verfahren. Hierfür werden Scoring-Matrizen benötigt, die wiederum aus Substitutionshäufigkeiten bestimmt werden. Diese Häufigkeiten werden aus dem Vergleich einer Vielzahl ähnlicher Proteine ermittelt und spiegeln gemeinsame Eigenschaften von Aminosäuren wider.

1.3 Peptidbindung

Proteine sind Polypeptidketten, die aus Aminosäuren synthetisiert werden. Bei der Synthese wird die Carboxylgruppe (COOH) der einen Aminosäure mit der Aminogruppe (NH_2) des Nachbarn durch eine kovalente Bindung (Peptid-Bindung) verknüpft. Jede Polypeptidkette beliebiger Länge hat ein freies Amino-Ende (N-Terminus) und ein freies Carboxyl-Ende (C-Terminus). Die Richtung einer Kette ist definiert als vom N-Terminus zum C-Terminus zeigend. Diese Richtung stimmt überein mit der Syntheserichtung *in vivo*, die mit dem Ablesen der mRNA in 5'-3'-Richtung korrespondiert.

Die an der Peptidbindung beteiligten Atome liegen jeweils starr in einer Ebene. Daher wird der Hauptkettenverlauf einer Polypeptidkette durch die Angabe von zwei Winkeln (Φ , Ψ) pro Residuum beschrieben. Diese Winkel geben die Drehung der beiden, am Hauptkettenverlauf beteiligten Bindungen des zentralen C_{α} -Atoms jeder Aminosäure an. Beide Winkel unterliegen weiteren Einschränkungen, die sich aus der Natur des jeweiligen Aminosäurerestes herleiten. Die Rigidität der Peptidbindung und die sterische Hinderung zwischen

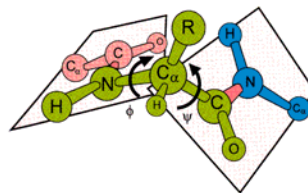


Abb. 1.5: Konformation der Peptidbindung. Die an einer Peptidbindung beteiligten sechs Atome liegen jeweils in einer Ebene. In der Abbildung sind zwei derartige Bindungen gezeigt und rot markiert. Der Aminosäurerest an der betrachteten Position (hier grün) ist mit R bezeichnet. Die räumliche Anordnung des Hauptkettenverlaufes eines Polypeptids $\dots-C_{\alpha}-C-N-C_{\alpha}-C-N-C_{\alpha}\dots$ wird bestimmt durch das, für jede Position (jedes Residuum) anzugebende Paar von Winkeln (Φ , Ψ), mit dem die Lage der durch die Peptidbindung aufgespannten Flächen relativ zum C_{α} -Atom festgelegt ist.

der Sekundärstruktur erfolgt über Wasserstoffbrückenbindungen zwischen den Imino- und Carbonylgruppen **innerhalb der Hauptkette**.

Zusätzlich zu den, hier beschriebenen Bindungskräften, wird die 3D-Struktur eines Proteins im Wesentlichen durch schwache, nicht kovalente Wechselwirkungen der Aminosäureseitenketten, insbesondere durch Wasserstoffbrückenbindungen zwischen polaren Resten bestimmt. Diese Wechselwirkungen spielen bei der Betrachtung der Sekundärstruktur keine Rolle.

Zur Vertiefung des hier dargestellten Stoffes sind die folgenden WWW-basierten Kurse zu empfehlen:

- *Amino Acid and Peptide Structures and Protein Architecture*, beide von W. McClure (Carnegie Mellon University).
- *Principles of Protein Structure Using the Internet* vom Birkbeck College.

Nun wollen wir die beiden wichtigsten Sekundärstrukturelemente etwas genauer betrachten.

1.6 α -Helix

Sind die (Φ , Ψ)-Winkel aufeinander folgender Residuen konstant, so ergeben sich helikale Strukturen. Unter diesen ist die am häufigsten vorkommende die α -Helix. In der α -Helix besteht jeweils eine Wasserstoffbrückenbindung zwischen der CO-Gruppe einer Aminosäure

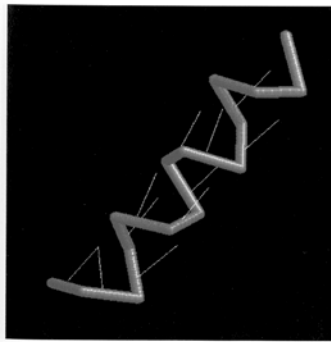


Abb. 1.6: α -Helix aus Karyopherin alpha. Wasserstoffbrücken sind gestrichelt eingezeichnet. Zur Verdeutlichung ist nur der Hauptkettenverlauf dargestellt.

und der NH-Gruppe der viert nächsten. Es machen jeweils 3,6 Aminosäuren eine vollständige Drehung aus. Die Abb. 1.6 zeigt einen typischen Vertreter einer α -Helix.

1.7 β -Faltblätter

Das zweite, wichtige Sekundärstrukturelement ist das β -Faltblatt. Ein β -Faltblatt besteht aus einzelnen β -Strängen, die meist 5–10 Residuen lang sind (siehe Abb. 1.7). Zwischen den Residuen **unterschiedlicher** Stränge werden Wasserstoffbrückenbindungen ausgebildet. Hierbei sind die C=O-Gruppen des einen Stranges mit den NH-Gruppen des nächsten Stranges verknüpft. Auf diese Weise können mehrere Stränge zu einem Blatt verbunden sein. Die C_α-Atome aufeinanderfolgender Residuen kommen abwechselnd über oder unter der Ebene, die durch das Faltblatt aufgespannt wird, zum liegen. Die Stränge können in zwei Richtungen verlaufen:

- **Parallel**, die durch N- und C-Terminus vorgegebene Richtung in nebeneinander liegenden Strängen ist die selbe.
- **Antiparallel**, die Richtung nebeneinander liegender β -Stränge wechselt alternierend.

Im Proteininneren sind die β -Faltblätter meist parallel. An der Proteinoberfläche sind sie meist antiparallel. Dort ragen die Aminosäurereste der einen Seite in die (hydrophile) Umgebung, während die der anderen zum hydrophoben Kern hin ausgerichtet sind. Hieraus ergibt

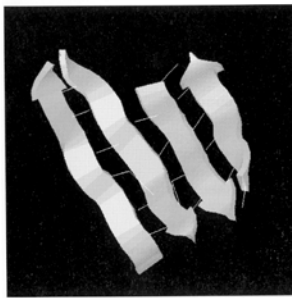


Abb. 1.7: β -Faltblatt aus dem Protein G. Die beiden mittleren β -Stränge verlaufen parallel und werden jeweils von einem antiparallelen Strang flankiert. Faltblätter sind häufig leicht gewölbt. Wasserstoffbrückenbindungen zwischen benachbarten β -Strängen sind als weiße Striche eingetragen.

[Wiene01]. Auf der Begleit-CD ist eine Übung vorbereitet, mit der Sie die Struktur genauer untersuchen können.

1.9 Protein-Domänen

Domänen

Beim Vergleich zweier verwandter Proteinssequenzen fällt häufig auf, dass die Sequenzähnlichkeit nicht über die gesamte Länge hinweg einen konstant hohen Wert aufweist. Häufig wechseln sich Regionen mit signifikant hohen Scores (einem Maß für Sequenzähnlichkeit) ab mit solchen Regionen, die keinerlei Ähnlichkeit zur Vergleichssequenz haben. Ursache für dieses Schwanken des Scores ist der modulare Aufbau von Proteinen aus Domänen. Es sei noch erwähnt, dass in eukaryotischen Genen Domänengrenzen oft mit Intron/Exon-Übergängen zusammenfallen.

Definition

Eine **Proteindomäne** ist die kleinste Einheit mit einer definierten und unabhängig gefalteten Struktur. Proteindomänen bestehen meist aus 50–150 Residuen und führen häufig individuelle Reaktionen aus, deren Zusammenwirken die Gesamtfunktion eines Proteins ausmacht.

CAP-Protein

In Abb. 1.9 sehen Sie die 3D-Struktur eines CAP-Monomers. Dies besteht aus zwei Domänen:

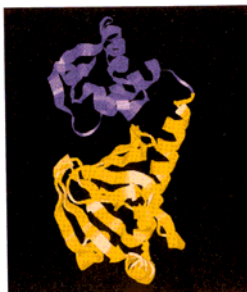


Abb. 1.9: 3D-Struktur eines CAP-Monomers. Die N-terminale Domäne wurde orange, die C-terminale Domäne wurde blau eingefärbt. In vivo lagern sich jeweils zwei CAP-Moleküle zu einem Dimer zusammen, nach [Krip95].

sich im Idealfall in der Sequenz ein charakteristischer Wechsel von hydrophoben und hydrophilen Aminosäuren. Häufig liegen β -Faltblätter sandwichartig aufeinander, ein schönes Beispiel finden Sie auf der Begleit-CD.

Zu den hier angesprochenen Themen finden Sie weiteres Material auf der Begleit-CD.

interaktives Arbeiten



1.8 Supersekundärstrukturelemente

Die regulären Strukturen des **backbones** werden ausgebildet, weil sie energetisch günstig sind. Sie bilden häufig Aggregate, die als Supersekundärstrukturelemente bezeichnet werden. So besteht der klassische **TIM-Barrel-Fold** beispielsweise aus 8 (β)-Einheiten, die rotationssymmetrisch zur Mittelachse angeordnet sind. Die acht β -Stränge bilden eine fassartige Struktur, die außen von den α -Helices bedeckt wird. Das in Abb. 1.8 gezeigte Enzym Fructose-Bisphosphat-Aldolase katalysiert die reversible Spaltung von Fructose-1,6-bisphosphat und Fructose-1-phosphat. Die oben beschriebene, ideale Struktur wird hier durch weitere Sekundärstrukturelemente ergänzt. Die meisten, die in der PDB-Datenbank (einer Sammlung von Protein-3D-Modellen) deponierten Enzymstrukturen weisen eine **TIM-Barrel-Topologie** auf. Sie kommt in vielen Enzymfamilien vor, die völlig unterschiedliche Reaktionen katalysieren. Das Reaktionszentrum (die **active site**) aller **TIM-Barrels** liegen stets an den C-terminalen Enden der β -Stränge. Eine Übersicht zu dieser Faltungstopologie finden Sie in

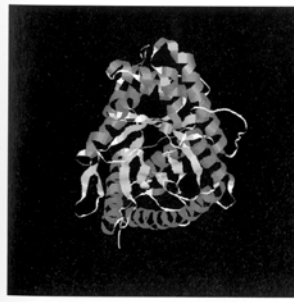


Abb. 1.8: Das TIM-Barrel Protein Fructose-Bisphosphat-Aldolase.

- Die N-terminale Domäne (Residuen 1–135) bindet cAMP und ist an der Dimersierung beteiligt.
- Die C-terminale Domäne (Residuen 136–209) vermittelt die DNA-Bindung des Proteins.

CAP-Dimere, d. h. Aggregate von zwei Monomeren sind in Bakterien an der Aktivierung solcher Gene beteiligt, deren Genprodukte in den Zuckerstoffwechsel eingreifen.

1.10 Vergleich der Domänenstruktur des präsynaptischen Proteins SAP27 und des MAGI-1A Proteins

Auf Domänenebene lassen sich die beiden Proteine wie folgt beschreiben (siehe Abb. 1.10): Beide Proteine enthalten eine GuKc-Domäne und eine unterschiedliche Anzahl von PDZ-Domänen. Die GuKc-Domäne besitzt in aktiven Enzymen Guanylatkinaseaktivität, in membranassoziierten Proteinen zeigt sie nur Proteinbindungsfunktion. Die PDZ-Domänen haben unterschiedliche Bindungsspezifitäten, manche binden C-terminale, andere interne Polypeptide. In MAGI-1A kommt zusätzlich die ww-Domäne zweimal, in SAP27 die SH3-Domäne einmal vor.



Abb. 1.10: Domänenstruktur des präsynaptischen Proteins SAP27 und des MAGI-1A Proteins.

Ein Alignment ist das Ergebnis eines Verfahrens zum paarweisen Sequenzvergleich, das darauf abzielt, in beiden Sequenzen vorkommende Module zu identifizieren und deren Position zu bestimmen. Bei der Darstellung von Alignments in Matrixform machen sich Teilssequenzen mit hoher Sequenzähnlichkeit als diagonalverlaufende Linien bemerkbar. Grundlage für das Quantifizieren der Sequenzähnlichkeit sind **Scores**, die sich aus dem Vergleich von Aminosäuren ableiten. Im Programm **Dotter** [SoDu95] werden Scores für kurze Teilstrings berechnet und als Punkte in eine Matrix eingetragen. Der Score selbst wird als Grauwert codiert. Je dunkler ein Punkt, desto höher der Score. Längere Teilsequenzen mit hohem Score ergeben diagonal verlaufende Linien. Das mehrfache Vorkommen einer Domäne macht sich durch unter- oder nebeneinander liegende Diagonalen bemerkbar. Den Basisalgorithmus lernen Sie im Abschnitt zu Dotplots kennen. In Abb. 1.11 lässt sich schon die oben angegebene

SCOP-Datenbank

Anordnung der in beiden Proteinen gemeinsam vorkommenden Domänen ableiten.

Proteindomänen sind häufig das Kriterium für das Clustern von Proteinen zu Proteinfamilien und Superfamilien. Eine der bekanntesten Datenbanken, die auf diesem Konzept beruht und die wir noch genauer kennen lernen, ist SCOP.

Domänengrenzen und Alignments

Schwache Sequenzähnlichkeiten und daraus resultierende Alignments sind oft nur schwer auf ihre Signifikanz hin zu interpretieren. Sind für eines der an einem Alignment beteiligten Proteine die Domänen bekannt, so können all die Treffer als nicht bedeutsam ausgeschlossen werden, bei denen sich eher unsignifikante Alignments über Domänengrenzen hinwegziehen. Auf diese Problematik werden wir später detailliert eingehen.

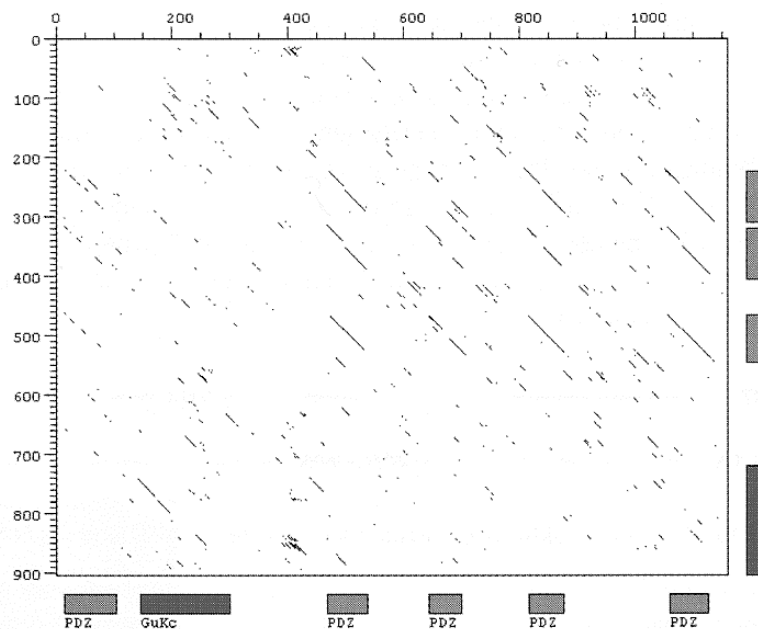


Abb. 1.11: Vergleich der Domänenstruktur von MAGI-1A (horizontal aufgetragen) mit der von SAP97.

In beiden Sequenzen kommen die Domänen PDZ und GuKc vor. Allerdings ist die Anzahl und Reihenfolge der Domänen in den Proteinen unterschiedlich. Die Domänen sind in unterschiedlichen Farben dargestellt. Der Plot wurde unter Verwendung des Programms Dotter [SoDu95] erzeugt. Die erste PDZ-Domäne in MAGI-1A ist weniger stark konserviert als die anderen.