

Molecular Modelling bei der Analyse von Röntgenstrukturdaten

angefertigt von

Name: **Markus Braner**

E-Mail: XXXXX

Matrikelnummer: **XXXXX**

Studiengang, Semester: Masterstudiengang Chemie, 2. Semester

Abgabedatum: 05. Mai 2011

im Rahmen des Seminars

Biologisch-chemische Datenbanken und Software

Seminarleitung: Wagner Steuer Costa

Andrè Bazzone

im: WS 2010/2011

Inhaltsverzeichnis

1. Einführung in die Röntgenstrukturanalyse und das Molecular Modelling	1
2. Sammlung von Röntgenstrukturdaten	2
3. Die Fouriertransformation und das Phasenproblem	3
4. Die Lösung des Phasenproblems	5
4.1 Die direkte Methode	5
4.2 Die anomale Streuung	6
4.3 Der isomorphe Ersatz	6
4.4 Molecular Replacement	7
4.4.1 Die Verfeinerung des Modells	10
5. Die Software ARP/wARP	11
6. Literatur	18

1. Einführung in die Röntgenstrukturanalyse und das Molecular Modelling

Als Röntgenstrahlung werden elektromagnetische Wellen der Wellenlänge 10 bis 0.1 Å bezeichnet. Für die Röntgenstrukturanalyse von Proteinen wird eine geeignete monochromatische Röntgenstrahlung benötigt, wobei Elektronen aus einem Glühdraht (bspw. Wolfram) emittiert werden und, nachdem sie durch ein elektrisches Feld beschleunigt wurden, auf eine Kupferanode treffen. Die Anode muss ausreichend gekühlt werden, da der größte Anteil der Strahlung durch strahlungslose Absorption unter Wärmeentwicklung verloren geht. Mit Hilfe von Filtern (z.B. Nickelfilter) kann aus der emittierten Röntgenstrahlung monochromatische Kupfer-K_α-Strahlung erhalten werden, welche dann direkt auf den Proteinkristall geschossen wird. Die Röntgenstrukturanalyse beruht bekanntlich auf der Streuung von Röntgenstrahlung an der Materie. Das Reflexmuster hängt dabei von der Elektronendichteverteilung der streuenden Materie im Kristall ab, während die Symmetrie des Reflexmusters durch die Symmetrie des Kristallgitters bestimmt wird. Die Intensität der Reflexe ist dabei abhängig von der Elektronendichteverteilung innerhalb der asymmetrischen Einheit. Die visualisierte Elektronendichteverteilung, welche bspw. am Computer mit Programmen wie COOT¹ dargestellt werden kann, wird dann durch Fouriertransformation erzeugt. Einen nicht zu vernachlässigenden Part nimmt auch das Molecular Modelling ein, wenn es darum geht, die Brücke zwischen den Röntgenstrukturdaten und dem fertigen Strukturmodell zu bauen.

Molecular Modelling ist rational betrachtet eine Ansammlung von computer-gestützten Techniken, um dreidimensionale Strukturen und ihre physikochemischen Eigenschaften zu berechnen, darzustellen oder auch zu bearbeiten. Dabei dienen die Daten aus der Röntgenstrukturanalyse als Basis. Die Werkzeuge des Molecular Modelling sind hingegen quanten- und molekülmechanische Verfahren. Das Ziel des Molecular Modelling ist der Aufbau eines Strukturmodells (sei es die Modellierung von Proteinen oder Protein-Ligand-Wechselwirkungen). Dies alles geschieht durch Kraftfeldmethoden (empirische Verfahren zur Berechnung von Molekülgeometrien und Molekülenergien) und durch quantenmechanische Berechnungen (man versucht die elektrostatische Struktur von Molekülen zu beschreiben; als Grundlage dient die Schrödinger-Gleichung)².

Da sehr viele Genomsequenzen und 3D-Strukturen bereits aufgeklärt sind, steigt die Chance zu einer Aminosäuresequenz ein homologes Protein zu finden dessen Struktur bekannt ist. Mittels Homology Building kann man versuchen für das unbekannte Protein ein Strukturmodell zu berechnen. Dies wäre bspw. durch das Programm SWISS-Model³ von ExPASy möglich. Auch möglich ist Homology Building, wenn man bereits durch

Sequenzvergleiche (z.B. durch die PDB) oder durch Kombination aus Sequenzvergleichen ein passendes Template einer homologen Sequenz gefunden hat. Die Proteinsequenz wird dabei in mehreren Schritten in das Template gefaltet, wobei in jedem Schritt Score-Funktionen lokal (durch Kraftfeldmethoden und quantenmechanische Methoden) optimiert werden. Für hierbei im Modell nicht vorhandene Regionen wird der Zustand angenommen, dass es sich um Loop-Bereiche handelt. Die Vorgänge laufen dabei allgemein so ab, dass die Aminosäurekette in die Elektronendichtekarte eingepasst wird, wobei dieser Vorgang in der Literatur als Model Refinement oder nur Refinement bezeichnet wird. Hierbei werden primär die Kohlenstoffatome des Peptidrückgrats eingepasst (bei Proteinen mit RNA handelt es sich äquivalent um die Phosphatgruppen, da diese die meiste Elektronendichte aufweisen). Hat man die Kette eingepasst, so lassen sich ggf. schon Sekundärstrukturelemente erkennen. Je nach Auflösung der Elektronendichtekarte kann hierbei schon eine Zuordnung der Seitenketten erfolgen, was für große Aminosäuren (Phenylalanin, Tryptophan oder Tyrosin) bedeutend einfacher ist als für kleine Aminosäuren (Glycin, Alanin).

Da der Bereich des Molecular Modelling sehr umfangreich ist und eine ausführliche Betrachtung den Rahmen dieser Arbeit bei Weitem sprengen würde, wird der Schwerpunkt dieser Arbeit auf dem Molecular Replacement (als eine Lösung des Phasenproblems) liegen, welches einen speziellen Fall im Anwendungsgebiet des Molecular Modelling widerspiegelt. Da dieser Bereich eine wichtige Rolle in der Strukturaufklärung spielt, soll hierbei explizit auf eine dafür verwendete Software eingegangen werden. Die Wahl fiel auf ARP/wARP, welches kostenlos verfügbar ist und einen guten Überblick über den Anwendungsbereich des Molecular Replacement gibt.

2. Sammlung von Röntgenstrukturdaten

Die Qualität der gemessenen Daten ist nicht nur enorm ausschlaggebend für die Genauigkeit des Strukturmodells, sondern auch für den Weg der Strukturlösung. Deshalb ist es wichtig, möglichst genaue und vollständige Messungen der Reflexintensitäten zu machen. Nach jeder einzelnen Aufnahme werden die gestreuten und vom Detektor gesammelten Röntgenquanten zusammengefasst und die gemessenen Intensitätswerte zusammen mit den Daten zur Geometrie der Messanordnung gespeichert. Zur Detektion der Reflexe können beispielsweise zweidimensionale, ortsempfindliche Detektoren verwendet werden. Ebenfalls wichtig ist, dass genügend Redundanz vorhanden ist⁴.

3. Die Fouriertransformation und das Phasenproblem

Röntgenstrahlen, die auf den Kristall geschossen werden, verändern ihre Amplitude, da sie eine Beugung erfahren (die Atome jeder Netzebene beugen Röntgenstrahlen scheinbar so, als würden die Wellen optisch reflektiert werden; Braggsches Gesetz s. Abb. 1). Die Amplitude dieser Welle $|F_{hkl}|$ (hkl = Millerschen Indices) ist dabei proportional zur Wurzel der durch den Detektor gemessenen Intensität (oder anders dargestellt in (1)). Dabei hängt die Position der Röntgenbeugung nur von der Zellgeometrie ab, die Amplitude der verschiedenen Beugungsmaxima jedoch wird von der Sorte und der Lage der Atome in der Elementarzelle beeinflusst.

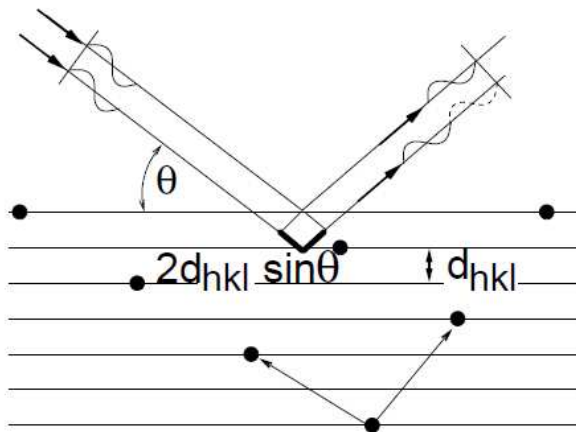


Abb. 1 Darstellung des Braggschen Gesetzes.⁵ Schematisch dargestellt ist eine Netzebenenchar mit dem Abstand d_{hkl} . Die „ausfallenden Wellen“ zeigen dabei eine Verschiebung, resultierend durch den vergleichsweise längeren zurückgelegten Weg im Beugungspunkt.

Die gemessenen Intensitäten werden von dem Strukturfaktor F_{hkl} bestimmt der abhängig von der Art und Verteilung der Atome in der Elementarzelle ist. Die gemessene Intensität ist proportional zum Quadrat der Strukturfaktoramplitude, gleichzeitig ist der Strukturfaktor die Fouriertransformation der Elektronendichte. Dabei stellt I_{hkl} die Intensität des Reflexes und $|F_{hkl}|$ die Strukturfaktoramplitude dar.^{6,7}

$$I_{hkl} \sim |F_{hkl}|^2 \quad (1)$$

Die Fouriertransformation entspricht vereinfacht dargestellt dem Übergang von einem realen Gitter zum reziproken Gitter, ist also die mathematische Beschreibung des

Röntgenbeugungsvorganges. Dies lässt sich durch die sog. Strukturfaktorfunktion $F(\vec{S})$ darstellen.

$$F(\vec{S}) = \int_V \rho(\vec{r}) e^{-2\pi i \vec{r} \cdot \vec{S}} dV \quad (2)$$

Aufgrund von (2) wird ersichtlich, dass es einen direkten Zusammenhang zwischen dem Streuvektor \vec{S} (dem Gittervektor des reziproken Systems) und dem Ortsvektor im realen Gitter \vec{r} gibt. V stellt das Volumen der Elementarzelle und i den Index über alle Einzelmessungen des Röntgenreflexes dar. Betrachtet man das mathematische System ausgehend vom Strukturfaktor, so kann man daraus die Elektronendichte berechnen. Dies wird durch reverse Fouriertransformation möglich. Einhergehend mit der Strukturfaktorfunktion stellt (3) ausgehend von der Elektronendichtefunktion $\rho(\vec{r})$ den rechnerischen Zusammenhang zum reziproken Volumen dar.

$$\rho(\vec{r}) = \int_{V_{\text{reziprok}}} F(\vec{S}) e^{2\pi i \vec{r} \cdot \vec{S}} dV_{\text{reziprok}} \quad (3)$$

Da man jedoch die direkte Beziehung zum Phasenwinkel des Strukturfaktors φ_{hkl} in eine Beziehung zur Strukturfaktorfunktion und zur Elektronendichtefunktion setzen muss, ergibt sich daraus eine Fourierreihe (4), wobei die Atomkoordinaten x , y und z in einer Elementarzelle durch \vec{r} angegeben werden.^{6,7}

$$\begin{aligned} \rho(\vec{r}) &= \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F_{hkl} e^{2\pi i (hx+ky+lz)} = \\ &= \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |F_{hkl}| e^{i\varphi_{hkl}} e^{2\pi i (hx+ky+lz)} \end{aligned} \quad (4)$$

Der Strukturfaktor F_{hkl} ist als Funktion der ortsabhängigen Elektronendichte $\rho(\vec{r})$ dargestellt. Anders ausgedrückt handelt es sich bei der Röntgenstreuung an Kristallen um eine Fouriertransformation der Elektronendichte in den reziproken Raum (Fourieranalyse). Umgekehrt lässt sich die Elektronendichte am Ort xyz durch die entsprechende Rücktransformation (Fouriersynthese) der komplexen Strukturfaktoren berechnen. Dabei kann das Integral (3) durch eine Summation ersetzt werden, da die Strukturfaktorfunktion gemäß den Laue-Bedingungen nur für diskrete Reflexe ungleich Null ist. Die Phasen der Strukturfaktoren sind dagegen experimentell nicht direkt zugänglich und müssen indirekt bestimmt werden, bzw. die Phasenbeziehungen $\varphi(hkl)$ gehen dabei verloren. Dies wird als Phasenproblem der Röntgenstrukturanalyse bezeichnet.

4. Die Lösung des Phasenproblems

Wie alle elektromagnetischen Wellen haben Röntgenstrahlen Amplituden- und Phaseninformationen. Um ein Reflektionsmuster zu erzeugen werden beide Informationen benötigt. Leider können nur die Amplituden mittels Intensitäten gemessen werden, die Phaseninformation geht bei der Aufnahme des Bildes verloren. Diese Problematik wird in der Kristallographie als das „Phasenproblem“ bezeichnet. Es gibt verschiedene Ansätze, das Phasenproblem zu lösen. Beim Vorliegen einer homologen und bereits bekannten Struktur kann diese als Suchmodell zur Phasenbestimmung mittels Molecular Replacement (Molekularer Ersatz) verwendet werden. Als weitere Methode existiert die Phasenbestimmung durch isomorphen Ersatz. Hierbei werden Schweratomverbindungen mit hohen Elektronenzahlen spezifisch an definierten Stellen im Protein gebunden, ohne dass dabei signifikante Änderungen der Kristallpackung oder der Proteinstruktur auftreten. Als häufige Methode der Strukturaufklärung neuer Proteinstrukturen wird die Methode der anomalen Streuung verwendet, bei der ein Schweratomderivat bei mehreren Wellenlängen gemessen wird.

4.1 Die direkte Methode

Die direkte Methode beruht auf der Ausnutzung von Zusammenhängen zwischen den Intensitäten innerhalb von Reflexgruppen und den Phasen. Dadurch wird versucht eine direkte Lösung des Phasenproblems zu finden. Hierbei wird am Anfang ein Startsatz von Reflexen mit bekannten Phasen aufgestellt (nach verschiedenen Konvergenzkriterien). Im nächsten Schritt sucht man nach Reflextripletts (dabei erfolgt auch eine Erweiterung über Phasenbeziehungen in der jeweiligen Raumgruppe) ausgehend von einer Liste von E-Werten. Bei einem E-Wert handelt es sich um eine Normalisierung von Struktur Faktoren. Bei den gesuchten Reflextripletts können neue Phasen aus den bereits bekannten Phasen des Startsatzes (mit hoher Wahrscheinlichkeit) gewonnen werden. Bei genügend bestimmten E-Werten lässt sich eine Fouriersynthese durchführen, sodass man auf diesem Wege zur Struktur gelangt.⁷

4.2 Die anomale Streuung

Die Absorptionskanten der häufigsten Atome in Proteinen (C, N, O, S) liegen sehr weit entfernt von denen in der Proteinkristallographie gebräuchlichen Wellenlängen. In Anwesenheit von schwereren Atomen im Kristall, deren Absorptionskante der einfallenden Strahlung entspricht, unterscheiden sich die Intensität der Friedel-Paare, was als anomale Streuung bezeichnet wird. Diese Anomalie entsteht dadurch, dass zur kohärenten Strahlung die inkohärente Strahlung hinzukommt. Dabei sind die Intensitätsunterschiede abhängig vom Atom und von der Wellenlänge des Röntgenstrahls (wobei das Atom für die anomale Streuung verantwortlich ist). Aus den auftretenden Intensitätsunterschieden zwischen den Friedel-Paaren lässt sich schließlich eine Phaseninformation ableiten. Bei Verwendung von Synchrotron-Strahlung kann die Wellenlänge zwischen 0.6 und 2 Å variiert werden, um die Absorptionskante des Schweratoms exakt zu erfassen und damit den Effekt der anomalen Streuung relativ groß werden zu lassen. Die Differenz zwischen den Friedel-Partnern wird Bijvoet-Differenz genannt^{4,7}. Diese Differenz kann bei den Messungen bestimmt werden und zur Phasenbestimmung herangezogen werden.^{8,9}

4.3 Der isomorphe Ersatz

Bei dieser früher eingesetzten Methode werden mehrere Kristalle vom selben Protein gezüchtet, wobei verschiedene Schweratome in der Struktur eingebaut sind (in der Regel pro Protein ein anderes Schweratom wie beispielsweise Iod, Nickel, Quecksilber). Wichtig ist, dass daraus keine merklichen Änderungen in der Elementarzelle oder den Atomlagen resultieren. Nach der Kristallisation wird für jeden Kristall ein Datensatz gemessen, worauf sich die Schweratomlokalisierung durch Pattersonkarten anschließt, sodass die Möglichkeit zur Phasenberechnung gegeben ist.⁷ Aus diesem Resultat und den Differenzen der Reflexintensitäten aus den verschiedenen Datensätzen der verschiedenen Proteine (mit ihren verschiedenen Schweratomen) lassen sich die Phasen des unsubstituierten Proteins (ohne Schweratom) ermitteln. Dessen 3D-Struktur ist darauf durch Fouriersynthese zugänglich und kann weiter bearbeitet werden.⁷

4.4 Molecular Replacement

Die Methode des Molecular Replacement (MR) kann zur Strukturaufklärung von Proteinen angewendet werden bei denen die Raumstruktur eines ähnlichen Proteins bereits bekannt ist. Proteine mit einer Sequenzidentität von mehr als 30 % besitzen mit hoher Wahrscheinlichkeit auch eine vergleichbare dreidimensionale Faltung.¹⁰ Aufgrund statistischer Überlegungen kann außerdem davon ausgegangen werden, dass für lösliche, globuläre Proteine nur eine begrenzte Anzahl von etwa 500 bis 1000 Strukturfamilien existiert.^{11,12,13} Da die Zahl der veröffentlichten Proteinstrukturen in den letzten Jahren stark anstieg (s. Einträge in der PDB), wird es zunehmend wahrscheinlicher, durch Sequenzvergleiche und Sekundärstrukturvorhersagen ein geeignetes Suchmodell für die Strukturlösung eines Proteins zu finden. Die Methode des MR ermöglicht es, ein atomares Suchmodell korrekt in der asymmetrischen Einheit einer unbekanntem Kristallstruktur zu platzieren. Bei dieser Suche handelt es sich um ein sechsdimensionales Problem, das in eine Rotations- und eine Translationssuche unterteilt werden kann.¹⁴ Das grundlegende Prinzip der Methode ist der Vergleich der Pattersonfunktionen von Suchmodell und Zielmolekül. Aus dem platzierten Modell können anschließend MR-Phasen bestimmt werden, die zusammen mit den experimentell gemessenen Strukturfaktoramplituden zur Berechnung der Elektronendichte herangezogen werden. Für eine erfolgreiche Suche ist der Grad der strukturellen Verwandtschaft zwischen Suchmodell und Zielstruktur von großer Bedeutung. Hochsymmetrische Raumgruppen sowie eine große Anzahl von Molekülen in der Elementarzelle erschweren allerdings die Suche nach der korrekten Lösung. Das MR ist eine leistungsfähige Methode zur Lösung von Kristallstrukturen, wenn eine geeignete Modellstruktur vorhanden ist. Das Modell wird zur Bestimmung der Position und Orientierung des Moleküls in der Elementarzelle benutzt.

Gute Modelle haben in der Regel eine geringe Abweichung von der Zielstruktur; ist das Modell jedoch schlecht, so misslingt das MR. Hierbei spielt die Sequenzidentität eine große Rolle, da geringe Abweichungen sich darin niederschlagen.¹⁵ Deshalb werden Modellstrukturen durch sequenzbasierte Suche identifiziert. Um das Modell jedoch zu verbessern, wendet man folgende Operationen an: Bereiche mit großen Unterschieden werden im Modell entfernt (da sie nur störende Bereiche enthalten) sowie unterschiedliche Seitenketten (wobei diese entweder durch ein C_γ-H oder durch Alanin ersetzt werden).¹⁶ Hat man diese Schritte durchlaufen und eine größere Anzahl an signifikanten Modellen, so kann man dieselben Operationen für jedes Modell noch einmal durchführen, oder man fasst alle Modelle zu einem Modell zusammen. Einzig bei Proteinstrukturen kommt das Problem auf,

dass ähnliche Sequenzen nicht immer ähnliche Tertiärstrukturen bedeuten müssen. Dies tritt jedoch zumeist nur bei Ligandenbindung oder bei unterschiedlicher Packung im Kristall auf. Eine zentrale Bedeutung im MR spielen die sog. Pattersonmethoden. Sie sind sehr nützlich, weil sie direkt aus den gewonnenen Daten berechnet werden können (auch wenn keine Phaseninformationen vorhanden sind; auch können sie aus einem Modell berechnet werden). Für die Zwecke des MR jedoch ist es von großer Bedeutung, dass die Modellstruktur in einen P_1 -Kristall definiert wird, da diese Einheitszelle zum Einen groß ist und zum Anderen viel Platz zwischen den einzelnen Molekülen herrscht. Dies ist dahingehend wichtig, da für die essentielle Pattersonfunktion nur die Vektoren in Betracht gezogen werden, welche nahe dem Ursprung der Vektoren und des dazugehörigen Moleküls liegen, sodass intermolekulare Vektoren (Kreuzvektoren) zwischen Nachbarmolekülen hierbei keine Rolle spielen¹⁵ (Vgl. 5). Daraus ergibt sich auch schon ein weiteres Problem des MR: Größere Moleküle haben mehrere intramolekulare Vektoren und sie enthalten somit mehr Informationen als kleinere Moleküle, die weniger Vektoren besitzen. Deshalb können mit MR auch größere Strukturen einfacher gelöst werden als kleinere. Ebenso problematisch sind längliche Moleküle und Oligomere, da dort eine Überlappung der Vektoren stattfindet. Deshalb wird in solchen Fällen ein Dimer-Modell bevorzugt eingesetzt, da die bisherigen Kreuzvektoren nun zu intramolekularen Vektoren werden.¹⁵

Am Anfang steht also eine generelle Suchstrategie. Es ist bekannt, dass jedes Molekül sechs Parameter besitzt, über die seine Orientierung und Positionierung im Raum definiert wird. Daraus ergeben sich für N Moleküle in einer asymmetrischen Einheit $6N$ Parameter. Dieses Rechenproblem kann jedoch durch zwei separate Suchen gemeistert werden (wobei die Suchoption nach Translationen auf die besten definierten Ergebnisse begrenzt wird). Die Folge daraus ist, dass durch zwei dreidimensionale Suchen die sechs definierenden Molekülparameter an jedem Suchpunkt definiert sind.¹⁵ Daraus erhält man den Vorteil, dass der Strukturfaktor F berechnet werden kann und dabei ein Vergleich zum beobachteten Strukturfaktor erstellt werden kann. Die Berechnung des ersten Strukturfaktors basierend auf den Rotationsparametern kann jedoch nur mit bekannten Translationsparametern durchgeführt werden. Ein möglicher Lösungsansatz basiert auf der Rotationssuche (5).

Die Rotationssuche hat das Ziel die Orientierung des Moleküls im Kristall zu bestimmen. Dazu werden die Pattersonfunktionen des Modells P_{Modell} und der gemessenen Röntgenbeugungsdaten X_{Modell} mit den äquivalenten Daten des Kristalls durch die Kreuzrotationsfunktion $R(C)$ innerhalb eines bestimmten Integrationsvolumens U überlagert (5). Die dabei gefundenen Rotationsparameter zeigen die Überlappung der

Pattersonfunktionen im Ursprung des Vektorsystems. Dort wird die Pattersonfunktion durch Eigenvektoren des Moleküls dominiert, was dahingehend ausschlaggebend ist, da diese keine Translation aufweisen.^{6,7}

$$R(C) = \int_U P_{\text{Kristall}}(X_{\text{Kristall}}) * P_{\text{Modell}}(X_{\text{Modell}}) dX_{\text{Modell}} \quad (5)$$

Rotiert man ein Modell um einen bestimmten Winkel, so rotiert die Pattersonfunktion mit den intermolekularen Vektoren um denselben Winkel. Für die Rotationsfunktion ergibt sich daraus, dass man einen Vergleich zwischen der radiuslimitierten (da somit nur die intramolekularen Vektoren eine Rolle spielen) rotierten Pattersonfunktion des Modells und der radiuslimitierten nicht-rotierenden Pattersonfunktion der gemessenen Daten ziehen kann. So kann formalistisch das Problem der Translation umgangen werden. Die Kristallsymmetrien stellen für die Pattersonfunktion der gemessenen Daten außerdem ein Problem dar, da für jeden Symmetrieoperator nochmals intramolekulare Vektoren um den Ursprung der Pattersonfunktion hinzu kommen, sodass es zu einem schlechteren Signal-zu-Rauschen-Verhältnis (in höher symmetrischen Raumgruppen) kommt.¹⁵ Damit zusammenhängend ergibt sich die Patterson-Produktfunktion, wobei die Produkte der gemessenen Kristallstruktur (bzw. Pattersonfunktion) und die der rotierenden Modellstruktur integriert werden (5). Die zweite Lösungsmöglichkeit ist die „Maximum-Likelihood“-Methode, die statistische Ansätze im reziproken Raum ausnutzt und daraus eine Mittelung aller möglichen Werte für Translationen berechnet (diese wird hier jedoch nicht weiter behandelt).

Mit der Translationsfunktion wird die genaue Lage des Moleküls im Kristall bestimmt. Falls es sich nicht um die bereits erwähnte Raumgruppe P_1 handelt, so enthält die Pattersonfunktion Kreuzvektoren zwischen Atomen, die durch Symmetrieoperatoren verbunden sind. Dies lässt sich dadurch feststellen, indem man das Molekül relativ zu den Kreuzvektoren translatiert, sodass das Symmetrie-verwandte Molekül sich ebenfalls in eine andere Richtung bewegt und die Kreuzvektoren eine Veränderung zeigen. Deshalb ergibt sich, dass die Kreuzvektoren auf eine Translation ansprechen, während die Eigenvektoren dies nicht tun. Dabei beschreibt die Translationsfunktion $T(t)$ die Überlagerung der Pattersonfunktion des Modelles $P(u,t)$ mit derjenigen des Beugungsmusters $P(u)$ (6). Wie gut die Berechnungen sind, lässt sich durch den kristallographischen R-Faktor berechnen. Der R-Faktor ($R = \text{residual}$) ist dahingehend ein Maß für die Übereinstimmung zwischen den berechneten und den beobachteten Strukturmodell, d.h. er gibt im Endeffekt die mittlere relative Abweichung der beiden Strukturamplituden (beobachtet und berechnet) an.

$$T(t) = \int P(u)P(u, t)du \quad (6)$$

$$R = \frac{\sum_{hkl} ||F_{beobachtet}| - k|F_{berechnet}||}{\sum_{hkl} |F_{beobachtet}|} \quad (7)$$

4.4.1 Die Verfeinerung des Modells

Die Verfeinerung des Modells hat zum Ziel, die Minimierung der Differenz zwischen $F_{berechnet}$ und $F_{beobachtet}$ durchzuführen. Daraus ergibt sich die beste Anpassung des Modells an die gemessenen Daten. Als Qualitätsindex kann dafür der bereits eingeführte der kristallographische R-Faktor (7) herangezogen werden. Verändert sich dieser Wert bei weiteren Verfeinerungszyklen nicht mehr, so hat die Verfeinerung einen maximalen Grenzwert erreicht, bzw. ist optimal für das Modell. Die Einführung des freien R-Wertes R_{free} ¹⁷ (8) stellt ein weitaus besseres Konvergenzmerkmal dar. Eine zufällige Auswahl von Reflexen (in der Regel 5-10 % des Datensatzes) mit ähnlichem Beugungswinkel¹⁸ wird dabei nicht in die eigentliche Verfeinerung mit einbezogen. Somit stellt R_{free} ein Kriterium dar, um die Überfittung der Struktur zu verhindern.

$$R_{free} = \frac{\sum_{hkl \in T} ||F_{beobachtet}| - k|F_{berechnet}||}{\sum_{hkl \in T} |F_{beobachtet}|} \quad (8)$$

Geht es um die Verfeinerung, so muss man auch die atomaren Auslenkungsfaktoren (auch B-Faktoren; manchmal auch als Temperaturfaktoren bezeichnet) mit berücksichtigen und diese iterativ verfeinern. Sie sind ein Maß für die mittlere quadratische Auslenkung eines Atoms innerhalb eines sphärischen Volumens aus der Ruhelage.^{6,7}

$$B = 8\pi^2 \overline{u^2} \quad (9)$$

$$B = 8\pi^2 \begin{pmatrix} \langle u_x u_x \rangle & \langle u_y u_y \rangle & \langle u_z u_z \rangle \\ \langle u_x u_y \rangle & \langle u_y u_y \rangle & \langle u_z u_z \rangle \\ \langle u_x u_z \rangle & \langle u_y u_z \rangle & \langle u_z u_z \rangle \end{pmatrix} \quad (10)$$

Problematisch an (9) ist jedoch, dass sie nur für mittlere Auflösungen gut definiert ist. Handelt es sich um Strukturen mit hoher Auflösung (besser als 1,4 Å, da das Daten-Parameter-Verhältnis für eine Verfeinerung besser als zwei sein sollte), so kann (9) nicht mehr angewendet werden, da die Auslenkung von in der Nachbarschaft befindlichen Atomen beeinflusst wird. Da die Beweglichkeit gebundener Atome immer anisotrop ist,^{19,20} lässt sich

das B-Faktormodell anisotrop beschreiben, indem in **(10)** eine Matrix eingeführt wird. Dadurch erhöht sich die Anzahl der zu verfeinernden Parameter pro Atom von Vier auf Neun. Um ein gutes Daten-Parameter-Verhältnis zu erhalten, werden für eine Verfeinerung weitere Informationen ausgenutzt. Diese "restraints" sind Standardgeometrien entnommen, die aus einer großen Menge von Kristallstrukturen resultieren.²¹ Die Beschränkungen der Geometrie lassen sich gegenüber den experimentellen Daten abhängig von ihrer Qualität wichten. Die $2F_{\text{beobachtet}} - F_{\text{berechnet}}$ Dichte gibt ein Abbild der Elektronendichte im Kristall wieder. Die $F_{\text{beobachtet}} - F_{\text{berechnet}}$ Dichte spiegelt dagegen direkt die Differenzen zwischen beobachteter und berechneter Dichte wider. Idealerweise sollte sie am Ende der Verfeinerung gegen Null gehen. Positive Bereiche deuten auf fehlende Teile in der Modellstruktur hin. Negative Dichte heißt, dass die Dichte überinterpretiert ist, d.h. sich zu viele Elektronen in diesem Volumensegment befinden.²² Dies kann durch einen falsch positionierten Rest oder durch eine Fehlinterpretation von Wasserlagen, die eigentlich Ionen darstellen, bedingt sein. Die Elektronendichtekarten lassen sich durch Wichtung mit dem experimentellen Fehler verbessern. Bei solchen σ_A -gewichteten Dichtekarten wird $F_{\text{beobachtet}}$ wie auch $F_{\text{berechnet}}$ mit dem abgeschätzten experimentellen Fehler gewichtet.

Um jedoch einen Einblick zu bekommen, wie Programme mit bestimmten Problemen umgehen und welche Möglichkeiten sie bieten, wurde exemplarisch die Software ARP/wARP ausgewählt, um auf wichtige Aspekte des Molecular Replacement im angewandten Sinne einzugehen.

5. Die Software ARP/wARP

Die kostenlose Software ARP/wARP (Automated refinement program)²³ hat das Ziel, auf Basis einer gegebenen Elektronendichtekarte, eine makromolekulare Modellstruktur abzugeben. Bis zu einer Auflösung der Daten von ca. 2,5 Å ist das Programm in der Lage eine ziemlich genaue und vollständige Proteinmodellierung durchzuführen. Dabei spielen die bei Lamzin & Wilson (1997),²⁴ Perrakis et al. (1999)²⁵ und Morris et al. (2002)²⁶ beschriebenen Algorithmen eine entscheidende Rolle. Das grundlegende Konzept von ARP/wARP ist die einheitliche Sicht auf iterative Bildungs- und Verfeinerungsschritte. Der Modellbau ist hierbei ein integraler Prozess, der zwischen dem Realraum und der Modellanpassung im reziproken Raum und dessen Refinement abwechselt. Ein Vorteil von ARP/wARP ist, dass die Software dem Modell folgt, d.h. es erfolgt kontinuierlich ein Wechsel von Atomen und deren

Positionen auf Basis der vorliegenden Elektronendichte. Der Schlüssel zum Erfolg liegt dabei in den Hybridmodellen. Am Anfang des automatisierten Modellbildungsprozesses steht die Verdichtung der Informationen zu einem Satz an freien Atomen,²⁷ welche keine chemische Identität haben. Sie sollen so gut wie möglich die Elektronendichte ausfüllen, wobei sie in ihrer Anordnung einem Proteinmodell ähneln sollen. Daraufhin „erlangen“ einige Atome eine chemische Identität, d.h. sie werden als Teil einer Proteinkette zugeordnet. Verallgemeinert werden zwei Informationen im sog. Hybridmodell kombiniert.

Wenn man ein Proteinmodell betrachtet, so besteht dieses hauptsächlich aus langen einzelnen Polypeptidketten (mit entsprechenden Resten). Dabei stellt sich heraus, dass es von zentraler Bedeutung ist, eine korrekte Position für die C_{α} -Atome zu finden. Hat man diese lokalisiert, so lässt sich die Kette bilden und ebenfalls die Aminosäurereste. Deshalb muss aus der Elektronendichtekarte zuerst die Polypeptidkette interpretiert werden. Hierbei stellt Abb. 2 einen kurzen Überblick über eine ARP/wARP-Prozedur (warpNtrace genannt) dar.²⁶

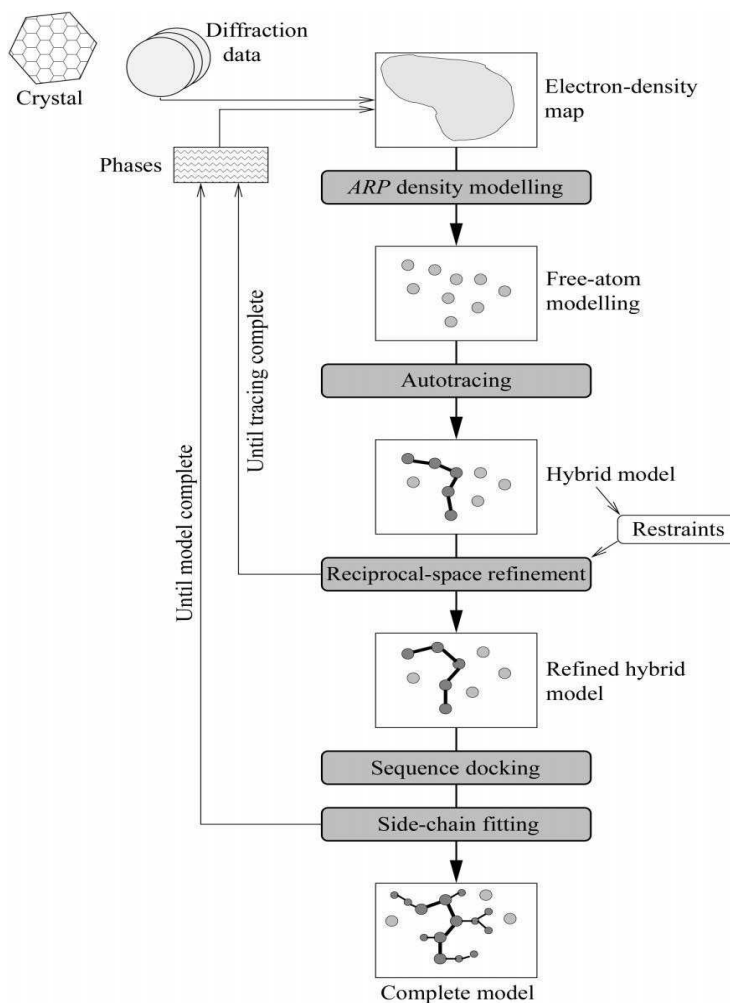


Abb. 2 Ablauf der Modellbildung durch ARP/wARP (warpNtrace).²⁶

Die Algorithmen sowie die Beschreibungen bzgl. Refinement und Überfittung wurden bereits im bisherigen Verlauf der Arbeit erläutert, weshalb an dieser Stelle darauf verwiesen wird. Wie bereits oben beschrieben, ist das durch ARP/wARP erstellte Modell kein konventionelles chemisches Modell, sondern eher eine Ansammlung von freien Atomen, die die Elektronendichte gut reproduzieren. Sind in diesem Schritt keine weiteren Informationen mehr verfügbar, so kann die Annahme getroffen werden, dass der momentane Stand an freien Atomen eine gültige Interpretation der Elektronendichte darstellt. Sollte jedoch das wirkliche Atom unbekannt sein, so können keine stereochemischen Beschränkungen in diesem Schritt modelliert werden. So erfolgt die Verfeinerung ohne Beschränkung, weshalb man auch von freien Atomen spricht. Eine Modellierung diesbezüglich benötigt eine an CPU-Leistung, weshalb große Rechner benötigt werden.

Wie gelangt man nun vom Modell der freien Atome zu einem Proteinmodell?

Die durch ARP/wARP platzierten Atome befinden sich schon ziemlich nahe an der wirklichen finalen Struktur des Modells (Abweichungen in der Regel ca. $0,5 \text{ \AA}$),²⁴ wobei hier immer die Qualität der Daten wichtig ist. Um jedoch die Atome zu identifizieren und zuzuordnen, wird ein sog. Modellbildungsalgorithmus benötigt. Wurde daraus die Atomart und die Bindungen ermittelt, so können stereochemische Eigenschaften abgeleitet werden und in das Modell miteinbezogen werden.²² Dadurch kommt es zu einem besseren Refinement, zu weniger Überfittung und zu besseren Phasenabschätzungen. Aus Abb. 2 ergibt sich, dass nach dem Modell der freien Atome das Hybridmodell steht. Dieses definiert sich durch die Menge der eingebauten Proteinfofragmente mit den verbleibenden freien Atomen. Durch weitere Interpretation der Karte und durch Modellbildung einschließlich erneutem Refinement werden folglich die besten Phasenabschätzungen erzielt. Somit lässt sich auch die eigentlich Kunst von ARP/wARP beschreiben: Sie besteht aus der Verfeinerung des Modells über mehrere Phasenzyklen und durch Iteration bei den Zyklen, was makroskopisch aus Abb. 2 hervorgeht. Bei einem Satz S an N freien Atomen mit den Koordinaten $x_i = (x_i, y_i, z_i)$ für $S = \{x_i/i=1, \dots, N\}$ wird die Teilmenge M von S gesucht,²⁶ bei welcher die stereochemischen Parameter aus den Atomvektoren berechnet werden, welche am besten in die Stereochemie des Proteins passen **(11)**.

$$M = \operatorname{argmax}_{s \in S} P[G(s)] \quad (11)$$

Hierbei wird durch **(11)** das Maximum der Funktion gesucht, wobei s die geometrischen Parameter des ausgewählten Sets S , $G(s)$ die Liste der Parameter und P wie

Wahrscheinlichkeit dieses Sets an beobachteten Parametern beschreibt. Zwar können nun die Parameter für ein entsprechendes Set beschrieben werden, aber es müssen noch die C_α -Atome einschließlich ihrer Positionen identifiziert werden. Logischerweise muss jedes C_α zwei Bindungspartner haben, d.h. es muss zu zwei weiteren C_α verknüpft sein (außer man befindet sich am Kettenende). Bekannt ist, dass die Distanz zwischen zwei C_α ungefähr bei 3.8 Å liegt. Hierbei wird durch Lamzin & Wilson (1997) in zwei Richtungen unterteilt: $-N-C(=O)-C_\alpha$ wird als „reinkommend“ und $-C(=O)-N-C_\alpha$ als „rauskommend“ beschrieben. Um nun das Set an Atomen zu finden, das die Kette beschreibt, kann durch (12) eine mögliche Verbindung der Atome c_{ij} berechnet werden. C_{ij} beschreibt hierbei jedoch nur die „rauskommende“ Richtung zwischen einem freien Atom i und einem weiteren j ($c_{ij} = 1$, falls eine Verbindung besteht, andernfalls wird c_{ij} als Null definiert).²⁶

$$\text{maximize } P[G(s)] = P[G(\{c_{ij} | i, j \in S\})] \quad (12)$$

Um jedoch die Liste an möglichen Verbindungen zu erhalten, ist es essentiell, die Distanzen zwischen allen Atomen abzugleichen, wobei nur die Atome relevant sind, die sich in der Elektronendichte, die passend für eine mögliche Kette ist, befinden. Finden sich darauf zwei C_α , die die passende Entfernung voneinander haben und wo sich die Kette in die Elektronendichte interpretieren lässt, so werden diese durch das Programm verbunden. Hat man mehrere dieser C_α -Paare identifiziert, so erhält man ein komplexes Set von Variablen c_{ij} , die jedoch noch optimiert werden müssen.

Um die besten Näherungen zu bekommen, ist es durchaus von Nutzen, die sog. „Holzhammermethode“ anzuwenden und so die beste Lösung zu finden. Zuverlässige und gute Ergebnisse werden erzielt, wenn es sich um hoch aufgelöste Elektronendichte handelt und jedes C_α einen „reinkommenden“ und einen „rauskommenden“ Part hat.²⁶ Jedoch können mehrere Lösungen auftreten, wenn man die Betrachtung nur auf die Positionsinformation richtet (durchaus mehrere Hundert $C_\alpha-C_\alpha$ -Verbindungen). Folglich müssen die Grenzen und Kriterien, die ARP/wARP zu Grunde liegen, verändert werden. Somit dürfen diese nicht zu eng gesteckt werden, da es sonst (unter Annahme, dass die Elektronendichtekarte der primären Abschätzung verwendet wird) zu falsch platzierten Atomen kommen kann.

Wie hoch ist jedoch nun die Wahrscheinlichkeit die Polypeptidkette der Länge L zu finden, die die Dichte repräsentiert?

Die Wahrscheinlichkeit wird durch $(L-1)$ mögliche Produkte gegeben, wobei n die Anzahl der „rein- und rauskommenden“ Verbindungen und $P(n)$ die Massefunktion (abhängig von der

Phasenqualität und Auflösung) ist. Steigt n an, so nimmt die Qualität der Elektronendichtekarte ab. So ergibt sich:²⁶

$$P(L) = P(n \neq 0)^{L-1} P(n = 0) \quad (13)$$

Da die Wahrscheinlichkeit, L zu finden durch $(L-1)$ gegeben ist und n größer als Null mit der Wahrscheinlichkeit von n gleich Null multipliziert wird, so folgt, dass die Kette nach L -Verbindungen stoppt, sofern keine weiteren mehr möglich sind. Um jedoch die beste Kette zu finden, muss festgelegt werden, wie viele Ketten überprüft werden müssen um ein gutes Modell zu bekommen.²⁶ Dies lässt sich durch die Gleichung (14) ausdrücken:

$$E(L) = \sum_L P(L) \prod_{k \leq L} \sum_n n P(n) = \sum_L P(L) \langle n \rangle^L \quad (14)$$

Hierbei ist $\langle n \rangle$ die durchschnittliche Anzahl der Verzweigungspunkte pro Knoten. Bei einer Aufsummation über L von $P(L)$ ist es nur noch nötig, die maximale Anzahl der Reste (der Kette) zu berücksichtigen. Die Wahrscheinlichkeit der Kettenlänge L muss genauso gegen Null gehen, wie $\langle n \rangle^L$ gegen unendlich tendiert. In der Praxis zeigt sich jedoch, dass die durchschnittliche Anzahl der Verzweigungspunkte die Zahl zwei bei Weitem übersteigt. Dies führt zu einem enormen Problem an möglichen Kombinationen. Um dieses Problem zu lösen, wird die Kette in einzelne kleine strukturelle Einheiten unterteilt.

$$\text{maximize } P[G(s)] = P[G(\{c_{ij} | i, j \in S\})] \approx \prod_u \varphi_u(\{c_{ij} | i, j \in u \subset S\}) \quad (15)$$

Der in (15) eingeführte Index u gilt für alle Einheiten einer Kette bzw. aller Ketten; φ_u ist eine einheitenbasierende Größe. Aus (15) geht hervor, dass man versucht, die Wahrscheinlichkeit P jeder Kette zu erhöhen. Die Struktureinheiten sind jedoch unabhängig, was auch die Vollständigkeit von Helices oder Domänen massiv beeinflusst. Das in (15) beschriebene Produkt kann jedoch durch Summenbildung von φ_u (16) ersetzt werden, was das Problem auf ein endliches Programmproblem reduziert.²⁶

$$\text{maximize } \sum_u \varphi_u(\{c_{ij} | i, j \in u \subset S\}) \quad (16)$$

Die strukturellen Einheiten sollten groß genug sein, um genügend Informationen zu beinhalten um die Struktur gut zu modellieren. Hierbei liegt die Länge der Einheiten bei rund vier C_α -Atomen die eine gute Lösung erlauben. Diese Methoden funktionieren bei Proteinen ohne weitere Probleme, wenn es sich ausnahmslos um all-trans-Peptide handelt. Befinden sich jedoch cis-Peptide in der Struktur, so produziert ARP/wARP an dieser Stelle einen Strangbruch.

Hat man alle bisherigen Probleme überwunden, so kann die Polypeptidkette zusammengefügt werden. Dies geschieht durch Überlappen der letzten drei Atome eines jeden Fragments mit den ersten drei Atomen des nächst folgenden Fragments. Dabei werden die einzelnen Fragmente ihrer Wahrscheinlichkeit nach kontrolliert (abhängig von der Wahrscheinlichkeit der jeweiligen verwendeten Parameter und Winkel).

Um das Modell am Minimum in einen Graph zu passen, wird der Deep-first-Suchalgorithmus²⁸ verwendet. Man sucht hierbei einen Startpunkt (Startknoten), von welchem aus alle weiteren möglichen Knoten gescannt werden und in Form eines Stapels gespeichert werden (der zuletzt gefundene Punkt wird, äquivalent zu einem Kartenspiel, als erstes verarbeitet usw.). Durch Abarbeiten dieses Stapels gelangt der Algorithmus von Knoten zu Knoten. Dabei entscheiden die Länge der Peptidkette und die Verzweigungen über die Komplexität der Rechnung. Ein großes Problem ist jedoch, dass der Algorithmus maximal einen Knoten zurück gehen kann und dass er keine Unterscheidung zwischen verschiedenen Wegen für die Struktur trifft. Daraus folgt, dass frühe Fehler später nicht mehr korrigiert werden können und somit die absolute Strukturlösung gefährdet ist. Aber auch hier wird, wie bei allen anderen Berechnungen, immer die Kette mit dem höchsten möglichen Score verwendet, wobei die Ketten mit guten Scores nacheinander durchgerechnet werden, sodass im Endeffekt die Bewertung des Gesamtmodells zählt.

Eine der problematischsten Regionen beim Modelling stellen die Loops dar. Jones et al.²⁹ haben gezeigt, dass eine Datenbank von Pentapeptiden zur Ergänzung und Verbesserung des Rückgrats durch Skelettierung der Elektronendichtekarte verwendet werden kann. Hierbei kann deren Konformation bspw. durch Diederwinkel zwischen den einzelnen C_{α} -Atomen bestimmt werden.^{30,31} Ausgehend von bekannten Pentapeptiden lässt sich die mögliche Position des fünften C_{α} -Atoms, angesetzt an das letzte Tetrapeptid der Hauptkette, bestimmen. Durch Iteration weiterer terminaler Peptide kann ein Baum bzgl. der möglichen Backbones für die Schleife erstellt werden. Die Schleife, welche am Wahrscheinlichsten in die Struktur passt, wird daraufhin ausgewählt. Für solche Zwecke kann die Zusatzfunktion Loopy in ARP/wARP integriert werden.³² Diese geht nach zwei Möglichkeiten vor: Zum Einen erkennt es die Lücken (Loops) automatisch und wählt die beste Schleife pro Lücke aus oder aber der Benutzer gibt einen zu definierenden Bereich an und das Programm bietet die besten Loops, passend zur Elektronendichte, an. Die Positionierung der Seitenketten der Aminosäuren erfolgt dabei analog zum Algorithmus in ARP/wARP. Nachdem alle Seitenketten angebracht wurden, wird ein Real-space Refinement durchgeführt.

Die Benutzung der Software ARP/wARP ist für alle Vorgänge auf eine Auflösung der Elektronendichte von 1.6 – 2.5 Å optimiert. Sind die Daten besser als 2.0 Å, so scheitert ARP/wARP nur gelegentlich (wenn das Startmodell sehr schlecht ist). Bei einer Auflösung der Daten zwischen 2.0 – 2.5 Å ist das Modell nicht komplett, aber die Software läuft ohne große Hindernisse. Hat man hingegen Daten mit einer Auflösung schlechter als 2.5 Å, so hat man ab und zu die Chance ein zu 80% komplettes Modell zu erhalten. Bei Auflösungen schlechter als 3.0 Å sind keine erfolgreichen Modelle verfügbar.³³

6. Literatur

- [1] P. Emsley, K. Cowtan »Coot: model-building tools for molecular graphics« *Acta Cryst.*, **2004**, D60, 2126-2132.
- [2] H.-D. Höltje, W. Sippl, D. Rognan, G. Folkers »Molecular modelling: basic principles and applications« 3.Auflage, Wiley-VCH, **2008**.
- [3] SWISS-MODEL, www.expasy.ch/swissmod/SWISS-MODEL.html, Stand April **2011**.
- [4] H. Krischner »Einführung in die Röntgenstrukturanalyse« 4.Auflage, Verlag Vieweg, **1990**.
- [5] Leibniz-Institut für Festkörper- und Werkstoffforschung, <http://www.ifw-dresden.de/institutes/imw/lectures/lectures/pwe/c2-beugung.pdf>, Stand April **2011**.
- [6] J. Drenth »Principles of Protein X-Ray Crystallography« 2. Auflage, Springer Verlag, **1999**.
- [7] W. Massa »Kristallstrukturbestimmung« 2. Auflage, Teubner Studienbücher, **1996**.
- [8] A.W. Hendrickson (1985). »Analysis of protein structure from diffraction measurement at multiple wavelength« *Transactions ACA*, **1985**, 21, 11-21.
- [9] A.W. Hendrickson, J.R. Horton, D.M. Le Master »Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): A vehicle for direct determination of three-dimensional structure« *EMBO J.*, **1990**, 9, 1665-1672.
- [10] C. Sander, R. Schneider (1991). »Database of homology-derived protein structures and the structural meaning of sequence alignment« *Proteins: Struct. Funct. Genet.*, **1991**, 9, 56-68.
- [11] G.E. Schulz »Protein Differenzierung: Entwicklung neuartiger Proteine im Laufe der Evolution« *Angew. Chem.* **1981**, 93, 143-151.
- [12] C. Chothia »One thousand families for the molecular biologist« *Nature*, **1992**, 357, 543-544.

- [13] S.E. Benner, C. Chothia, T.J. Hubbard »Population statistics of protein structures: lessons from structural classifications« *Curr. Opin. Struct. Biol.*, **1997**, 7, 369-376.
- [14] M.G. Rossmann, D.M. Blow »The detection of sub-units within the crystallographic asymmetric unit« *Acta Crystallog.*, **1962**, 15, 24-31.
- [15] P. Evans, A. McCoy »An introduction to molecular replacement« *Acta Cryst.*, **2008**, D64, 1-10.
- [16] M.G. Rossmann »The Molecular Replacement Method« *Acta Cryst.*, **1990**, A46, 73-82.
- [17] A.T. Brünger »The free R value: a novel statistical quantity for assessing the accuracy of crystal structures« *Nature*, **1992**, 355, 472-475.
- [18] G.J. Kleywegt, T.A. Jones »Where freedom is given, liberties are taken« *Structure*, **1992**, 3, 535-540.
- [19] C.S. Pratt, B.A. Coyle, J.A. Ibers »Redetermination of the Structure of Nitrosylpentaamminecobalt(III) Dichloride« *J. Chem. Soc.*, **1971**, 2146-2151.
- [20] G.B. Jameson »On structure refinement using data from a twinned crystal« *Acta Cryst.*, **1982**, A38, 817-820.
- [21] A. Engh, R. Huber »Accurate bond and angle parameters for X-ray protein structure refinement« *Acta Cryst.*, **1991**, A47, 392-400.
- [22] A. Perrakis, M. Harkiolaki, K.S. Wilson, V.S. Lamzin »ARP/wARP and molecular replacement« *Acta Cryst.*, **2001**, D57, 1445-1450.
- [23] V.S. Lamzin, A. Perrakis, K.S. Wilson »The ARP/WARP suite for automated construction and refinement of protein models« *Int. Tables for Crystallography*, **2001**.
- [24] V.S. Lamzin, K.S. Wilson »Automated refinement for protein crystallography« *Methods Enzymol.*, **1997**, 277, 269-305.
- [25] A. Perrakis, R. Morris, V.S. Lamzin »Automated protein model building combined with iterative structure refinement« *Nature Struct. Biol.*, **1999**, 5, 458-463.
- [26] R.J. Morris, A. Perrakis, V.S. Lamzin »ARP/wARP's model-building algorithms. I. The main chain« *Acta Cryst.*, **2002**, D58, 968-975.

- [27] R.C. Agarwal, G. Isaacs »Method for obtaining a high resolution protein map starting from a low resolution map« Proceedings of the National Academy of Sciences, **1977**, 74, 2835-2839.
- [28] C.H. Papadimitriou, K. Steiglitz (1998) »Combinatorial Optimization: Algorithms and Complexity« 1.Auflage Dover Publ Inc., New York: Dover, **1998**.
- [29] T.A. Jones, J.-Y. Zou, S.W. Cowan, K. Kjeldgaard »Improved methods for building protein models in electron density maps and the location of errors in these models« Acta Cryst., **1991**, A47, 110-119.
- [30] G.J. Kleywegt »Validation of protein models from C^α coordinates alone« J.Mol.Biol., **1997**, 273, 371-376
- [31] R.M. Esnouf »Polyalanine Reconstruction from Ca Positions Using the Program CALPHA« Acta Cryst., **1997**, D53, 665-672.
- [32] K. Joosten, S.X. Cohen, P. Emsley, W. Mooij, V.S. Lamzin, A. Perrakis »A knowledge-driven approach for crystallographic model protein completion« Acta Cryst., **2008**, D64, 416-424
- [33] S.X. Cohen, M.B. Jelloul, F. Long, A. Vagin, P. Knipscheer, J. Lebbink, T.K. Sixma, V.S. Lamzin, G.N. Murshudov, A. Perrakis »ARP/wARP and molecular replacement: the next generation« Acta Cryst., **2008**, D64, 49-60.