reprinted from:

A. Plückthun and J. Köhler, Protein-Struktur und Funktion: Die Möglichkeiten und Probleme der Computersimulation aus der Sicht des Experimentators. In: Forum '90 Wissenschaft und Technik. Informatik Fachberichte, H.-J. Friemel, G. Müller-Schönberger and A. Schütt, Eds. (Springer-Verlag, Trier, 1990, Proceedings, 1990), vol. 259, pp. 258-274.

Protein-Struktur und Funktion: Die Möglichkeiten und Probleme der Computersimulation aus der Sicht des Experimentators

Andreas Plückthun

Jutta Köhler

Genzentrum, Universität München
c/o Max-Planck-Institut für Biochemie

Martinsried

Zusammenfassung

Die drei-dimensionale Struktur eines Proteins ist die Grundlage seiner spezifischen Funktion. Obwohl die Gentechnologie heute erlaubt, Proteine beliebiger Aminosäure-Sequenz herzustellen, ist die Vorhersage von Struktur und Eigenschaften eines Proteins heute noch nicht möglich, und das enorme Potential der neuen experimentellen Methoden kann deshalb noch nicht voll technisch ausgenutzt werden. Es ist zunächst notwendig, Methoden zu entwickeln, die in verläßlicher Weise die sehr geringen Beträge von Wechselwirkungsenergien vorhersagen, die bei biologischen Prozessen wie Proteinfaltung und Protein-Liganden-Wechselwirkungen die entscheidende Rolle spielen. Dafür sind, um auch nur zu einer qualitativ korrekten Aussage über freie Energien zu kommen, sehr genaue Berechnungen nötig. Viele solcher Prozesse verlaufen nicht nach rein enthalpischen Prinzipien, sondern der Entropieterm kann die entscheidende Rolle spielen, und die Entropie ist nur über die statistische Thermodynamik in aufwendigen Simulationen zugänglich. Der Artikel diskutiert Fortschritte der Molekulardynamik mit empirischen Kraftfeldern und Fallbeispiele an einem Antikörper-Antigen-Komplex, der mit dieser Methode und verschiedenen experimentellen Ansätzen analysiert wurde.

Einleitung

Proteine entfalten ihre Wirkungen als Enzyme, Genregulatoren oder Rezeptoren erst durch die Faltung ihrer Polypeptidkette in eine bestimmte drei-dimensionale Form. Diese spezifische Struktur macht es möglich, daß ein Enzym nur mit einem bestimmten Substrat reagiert, ein Rezeptor nur eine Substanz erkennt und ein Genregulator nur an einer Stelle des Chromosoms bindet. Andererseits unterscheiden sich Proteine in solchem Ausmaß voneinander, daß ein Inhibitor oder Ligand oft nur an ein bestimmtes Protein bindet. Die Grundlage vieler Bestrebungen der pharmazeutischen Chemie ist es, solche spezifischen Liganden (Inhibitoren oder auch Aktivatoren) zu konstruieren, wogegen sich die neue Wissenschaft des "Protein Engineering" bemüht, Proteine mit verbesserten Eigenschaften, also vorbestimmter Faltung, herzustellen. Beiden gemeinsam ist das Problem der Quantifizierung von spezifischen, nicht-kovalenten Wechselwirkungen und der Frage nach der genauen Faltung der Proteinkette.

Warum ist dieses Problem so schwierig? Offensichtlich ist die Zahl der möglichen Konformationen nen der Proteine astronomisch. Doch selbst, wenn es gelänge, alle möglichen Konformationen der Proteinkette und des Liganden zu generieren und deren Energie zu vergleichen (was sowohl für das reale als auch für das simulierte System länger als das Alter des Universums dauern würde (Levinthal, 1968)), würde das heute kaum weiterhelfen: Die energetische Evaluie-

rung der Konformationen erfordert genauere Potentiale als heute erhältlich und, wie unten ausgeführt, entropische Terme sind ebenfalls von Bedeutung. Es soll nun kurz umrissen werden, warum das so ist, um dann verschiedene Lösungsmöglichkeiten zu diskutieren.

Die Faltung einer Proteinkette ist ein hoch-kooperativer Prozeß, und ein Zwei-Zustandsmodell, das nur einen völlig denaturierten Zustand und den völlig nativen beinhaltet, liefert (zumindest für kleine Proteine) eine konsistente Beschreibung der Faltungsthermodynamik. Das bedeutet, daß in der Regel alle Faltungsintermediate und nicht-nativen Formen eine höhere freie Energie sowohl als der native als auch als der "random coil" Zustand haben. Der Unterschied der freien Energie beider Zustände (also gewissermaßen zwischen "Sein" und "Nicht-Sein") liegt für die meisten Proteine bei lediglich 5 bis 15 kcal/mol. Allerdings setzt sich diese geringe Energiedifferenz beider Zustände aus riesigen sich beinahe aufhebenden Komponenten zusammen (Creighton, 1983).

Deshalb erfordert die den Praktiker interessierende Voraussage, ob eine Struktur "möglich" ist, d. h. ob sie stabiler als der völlig ungefaltete Zustand ist, oder sogar stabiler als eine alternative Struktur, extrem genaue Potentialfelder und Berechnungsmethoden (s. u.). Diese Genauigkeit ist bis heute nicht erreicht.

Ein weiteres Problem ist die große Bedeutung der Entropie. Viele wichtige biologische Prozesse sind entropisch bestimmt. Als ein Beispiel sei der hydrophobe Effekt genannt, der bewirkt, daß sich Gleiches in Gleichem löst und Öltröpfchen in Wasser zusammenlagern. Dieser Effekt ist eine wichtige treibende Kraft bei der Proteinfaltung und bei der Bindung von Liganden an Proteine. Offensichtlich würde eine Beschränkung auf enthalpische Parameter (d.h. eine reine Analyse der potentiellen Energien) das Wesen dieses Prozesses verkennen. Aus diesem Grund sind in der Regel Methoden notwendig, wie z.B. die der statistischen Thermodynamik, um von mikroskopischen Simulationsrechnungen zu makroskopischen Vorhersagen zu kommen.

Eine weitere für den Theoretiker unangenehme Eigenschaft der Proteine ist die Tatsache, daß sie so groß sind. Diese Größe ist eine Folge der Notwendigkeit, eine definierte Struktur auszubilden, ohne die wiederum keine spezifischen Prozesse in der Zelle ablaufen könnten. Die Größe ergibt sich aus dem Zwang, daß das Protein ein definiertes hydrophobes Inneres und ein hydrophiles Äußeres besitzen muß. Nur so können stabile Faltungen mit Bindungstaschen für Liganden auftreten; kleine Peptide besitzen zwar oft eine bevorzugte Konformation (Dyson et al., 1988a, b), aber keine stabile Struktur.

Wasser ist ein entscheidender Bestandteil beinahe aller Prozesse, die mit Proteinen zu tun haben. Sein Entropiegewinn bei der Faltung ist eine entscheidende Triebkraft für die Proteinfaltung. Seine hohe Dielektrizitätskonstante ist allerdings ebenfalls eine bestimmende Eigenschaft vieler Wechselwirkungen zwischen Protein und Ligand. Damit ist die Beschreibung elektrostatischer Phänomene vor ein ernstes Problem gestellt. Die Dielektrizitätskonstante ist ein makroskopischer Begriff und die Übertragung auf mikroskopische Simulationsrechnungen ist ein bisher nicht befriedigend gelöstes Problem (Harvey, 1989). Auch die Einbeziehung expliziter Wassermoleküle in Simulationsrechnungen löst dieses Problem nicht automatisch; ein rechentechnisch einfach zu behandelndes, polarisierbares Wassermodell zur Beschreibung dielektrischer Phänomene wäre wünschenswert (Sauniere et al., 1989).

Offensichtlich müssen Simulationsrechnungen, deren Ziel eine Beschreibung und letztlich eine Vorhersage des Experiments ist, in der Lage sein, diese essentiellen Bestandteile des experimentellen Systems Protein-Ligand richtig zu beschreiben. Dabei genügt allerdings meist nicht einmal eine qualitative theoretische Beschreibung. Da die Unterschiede der freien Energie zwischen Bindung und Nicht-Bindung oder Faltung und Nicht-Faltung sehr klein sind, kann nur eine quantitativ sehr genaue Beschreibung überhaupt qualitativ das Experiment richtig beschreiben.

Wir wollen nun einen kurzen Überblick über verschiedene Ansätze geben und dabei den Rechenaufwand, die Bedeutung der Supercomputertechnologie und die Probleme und Perspektiven diskutieren. Es gibt zwei Arten von konkreten Ansätzen, bei denen theoretische Vorhersagen von großem Interesse sind. Die erste ist die Vorhersage einer Struktur, die ähnlich einer bekannten ist. Wie bereits ausgeführt, gibt es derzeit keinerlei Ansatzpunkte, die Vorhersage völlig neuer Strukturen in Angriff zu nehmen. Die Einflüsse bestimmter Austausche von Aminosäuren oder Schleifen des Proteins auf die Struktur sind aber Fragen, die den pharmazeutischen Chemiker und Biotechnologen beschäftigen und für die Vorhersagen von großem Interessse wären. Zudem können sie durch die kombinierte Anwendung von Gentechnologie und Röntgenstrukturanalyse in manchen Fällen experimentell überprüft werden und erlauben deshalb eine experimentelle Verifizierung der theoretischen Methoden. Die zweite Fragestellung ist die Vorhersage von freien Energien der Wechselwirkungen zwischen Ligand und Protein. Damit soll die Frage des Experimentators beantwortet werden, ob eine Variante des Proteins einen bestimmten Liganden besser bindet oder auch eine Variante des Liganden an ein bestimmtes Protein besser binden kann. Hierzu ist die Kenntnis der experimentellen Struktur des Komplexes notwendig.

Strukturvorhersage

Die Frage nach der Struktur eines veränderten Proteins (z. B. durch den Einbau einer neuen Schleife) hat zu Überlegungen geführt, ob die Kenntnis der Datenbank der empirisch bestimmten Strukturen erlaubt, ein Protein gleichsam aus bekannten Teilen zusammenzusetzen. Derartige Programme sind entwickelt worden (Jones & Thirup, 1986), und zumindest in einem Fall konnte gezeigt werden, daß eine Schleife des Proteins auch in neuer Umgebung ihre Struktur behält (Hynes et al., 1989). Die Grenzen dieses Ansatzes sind allerdings schnell erreicht. Keineswegs ist es sicher, daß diese Konserviertheit von Schleifen (Loops) ein generelles Phänomen ist. Der Ansatz negiert alle Wechselwirkungen zwischen entfernten Teilen der Sequenz und er kann keine Aussage treffen, wenn ein derartiger Loop noch nicht beobachtet worden ist. Außerdem ist der Ansatz auf recht grobe Beschreibungen der Struktur beschränkt; eine Vorhersage von Seitenkettenkonformationen ist meistens nicht möglich, da diese durch Wechselwirkungen mit räumlichen Nachbarn bestimmt sind. Diese starken Einschränkungen werden durch einen recht geringen Rechenaufwand kompensiert, der von einem normalen Mikrocomputer bewältigt werden kann.

Alle anderen Strategien beruhen auf Kraftfeldern. Diese sind eine empirische Beschreibung der potentiellen Energie als Funktion der räumlichen Lage aller Atome. Bevor die genauen Formen der Kraftfelder beschrieben werden, sollen kurz die wichtigsten Verfahren diskutiert werden, die mit diesen Kraftfeldern angewendet werden.

Die Energieminimierung ist ein Verfahren, bei dem nach etablierten rechentechnischen Verfahren (steepest descent, van Gunsteren & Karplus, 1980; conjugate gradient, Fletcher & Reeves, 1964) das der Ausgangsstruktur nächstliegende Minimum gefunden werden kann. Da ein überaus komplexes System wie das eines Proteins unzählige Minima auf der Energiehyperfläche besitzt, findet sich in aller Regel ein Minimum in nächster Nachbarschaft zur Ausgangsstruktur. Das so gefundene Minimum ist mit beinahe absoluter Sicherheit nicht das globale. Durch dieses Verfahren ändert sich also die Ausgangsstruktur kaum. Dennoch ist die Energieminimierung ein sinnvoller erster Schritt, weil er die Struktur, ohne sie viel zu verändern, "entspannt" und damit lokal ungünstige Konformationen verbessert, die bei einem anschließend angewandten Verfahren (wie z.B. Molekulardynamik) zu Problemen führen können. Der Aufwand des Verfahrens hängt sehr stark von der Komplexität des Systems ab; ein Protein in einem Wassertropfen von 10.000 bis 20.000 Wassermolekülen kann einen Mikrocomputer mehrere Tage beschäftigen und dauert auf einer Cray/XMP in der Größenordnung von 1 bis 2 Stunden.

Eine Abwandlung dieser Verfahren besteht darin, eine Suche nach Konformationen vorzunehmen. Bei dem Versuch, die Konformation eines Loops vorauszusagen, wurden alle möglichen Konformationen systematisch generiert, indem die Winkel ϕ und ψ , die die relative Lage der aufeinanderfolgenden Peptidbindungen beschreiben, in bestimmten Schritten variiert wurden und mittels eines bekannten Algorithmus die Schleife geschlossen wurde (Bruccoleri & Karplus, 1987). In einem vergleichbaren Ansatz wurde eine Schar von solchen Konformeren durch zufällige Wahl der Winkel generiert, also eine Art Monte-Carlo Ansatz (Fine et al., 1986; Shenkin et al., 1987). Ebenso wurden Seitenketten-Konformere auf diese Art evaluiert (Shih et al., 1985). Der Erfolg der Methode hängt entscheidend davon ab, wie "eng" der Potentialtopf ist, d. h. ob durch die ungünstige Wahl oder zu geringe Zahl der untersuchten Konformationen die mit der niedrigsten Energie übersehen werden kann. Diese Methode wurde bisher nur ohne explizite Lösungsmittelmoleküle angewandt; damit ist sie neben den oben erwähnten Schwierigkeiten für die sich aus dieser Vereinfachung ergebenden Probleme anfällig und, und wie alle Methoden, zusätzlich für alle Ungenauigkeiten des Potentials. Diese Nachteile werden etwas aufgewogen durch den geringen Rechenaufwand. Dieser erreicht allerdings sofort Supercomputerdimensionen, wenn die Interaktion mehrerer Teile des Proteins (z. B. mehrere flexible Loops) erlaubt wird, weil dann die Zahl der zu untersuchenden Konformationen sehr schnell immens anwächst.

Vorhersage thermodynamischer Eigenschaften (Bindung und Stabilität)

Die thermodynamischen Eigenschaften eines experimentellen biochemischen Systems besitzen zwei wichtige Charakteristika: Erstens werden in aller Regel Zustände verglichen, die sich nur durch recht kleine freie Energiebeträge (wenige kcal/mol) unterscheiden. Normalerweise ist ein qualitatives Ergebnis ausreichend (Substanz A bindet; Protein X ist stabil; Substanz B bindet besser als Substanz C). Da aber die Alles oder Nichts-Entscheidung für eine Bindung durch sehr geringe Energieänderungen zustande kommen kann, muß eine Berechnung sehr genaue quantitative Ergebnisse liefern, um überhaupt von Nutzen zu sein. Zweitens handelt es sich bei den Gesamtbeträgen von Enthalpie und Entropie meistens um die Differenzen sehr großer Zahlen, deren Berechnung deshalb mit großen Fehlern behaftet ist.

Das bedeutet im Detail: Die direkte Analyse der potentiellen Energie ist für sich allein nicht ausreichend. Erstens ist die potentielle Energie sehr empfindlich gegen kleinste Verschiebungen der Koordinaten (meist geringer als deren experimentelle Unsicherheit), zweitens werden zur Zeit große Vereinfachungen bei der Beschreibung von dielektrischen, insbesondere Polarisations-Phänomenen gemacht. Drittens wird dabei der Einfluß der Entropie und damit zusammenhängender Kräfte, wie der hydrophobe Effekt nicht berücksichtigt.

In einem originellen Ausweg wurde ein Ansatz versucht (Novotny et al., 1989), bei dem zusätzlich zu dem üblichen enthalpischen Potential empirische Terme für die Entropie und Hydrophobizität verwendet wurden. Ein Vorteil liegt in einer hohen Rechengeschwindigkeit; allerdings sind, wie bei solch weitgehenden Vereinfachungen in der Beschreibung der Entropie zu erwarten, die Übereinstimmungen mit dem Experiment unbefriedigend (Glockshuber et al., 1990).

Molekulardynamik

Wir werden nun eine Methode beschreiben, die prinzipiell Beiträge sowohl zur Strukturvorhersage als auch zu thermodynamischen Eigenschaften eines Proteins liefern kann: Die Molekulardynamik. Dabei werden wir zunächst das Potentialfeld und das Wesen der Dynamikmethode erläutern. Sodann wird die Problematik des Vergleichs einer experimentellen Struktur mit der aus der Dynamik erhaltenen diskutiert und schließlich die Verbindung zur statistischen Thermodynamik erläutert. Die Dynamik eines Makromoleküls ist dabei durchaus bereits von unmittelbarem Interesse. Die Korrelation zwischen biologischen Funktionen von Makromolekülen und ihrem dynamischen Verhalten (z. B. Flexibilität der Erkennungsregionen und der aktiven Zentren von Enzymen, Koordination mit den Übergangszuständen des Substrates etc.) wurden in einem Übersichtsartikel von Bennett und Huber (1984) diskutiert. Mit der Molekulardynamik-Methode lassen sich die Bewegungen eines Moleküls im Pikosekundenbereich erfassen. Einen Überblick über die theoretischen Perspektiven von Dynamik, Struktur und Thermodynamik von Proteinen geben Brooks et al. (1988).

Die Molekulardynamik Methode

Die Bewegungsgleichungen

Die Trajektorie eines N-Teilchen Systems kann durch Integration der Newtonschen Bewegungsgleichungen berechnet werden (siehe z. B. Berendsen & van Gunsteren, 1986):

$$\frac{d^{2} \overline{r_{i}}(t)}{dt^{2}} = \frac{\overline{F_{i}}(\overline{r_{1}}....,\overline{r_{N}})}{m_{i}}$$
 $i = 1,....,N$

Die Kräfte sind konservativ, d.h. sie sind nur von der Position r_i der Teilchen i mit der Masse m_i abhängig. Die Kräfte F_i sind die Ableitungen des Potentials V, welches die Wechselwirkungen zwischen den Teilchen beschreibt.

$$\overline{F}_i = -\nabla_i V(\overline{r}_1, ..., \overline{r}_N)$$

Das Potential ist ein Paar-Potential, d.h. das Gesamtpotential ist die Summe aller Paarpotentiale.

$$V(\overline{r_1},...,\overline{r_N}) = \sum_{i < j} V_{ij}(\overline{r_{ij}})$$

wobei

$$-\frac{\partial}{\partial r_{ij}} V_{ij} (\bar{r}_{ij}) = \bar{F}_{ij} = -\bar{F}_{ji}$$

und

$$\bar{F}_i = \sum_{j} \bar{F}_{ij}$$

Algorithmen

Es gibt verschiedene Algorithmen für die Integration der Newtonschen Bewegungsgleichungen (van Gunsteren & Berendsen, 1977). Die Gear Methode (Gear, 1971) wird in Molekulardynamik-Simulationen verwendet, in denen hochfrequente Bewegungen der kovalenten Bindungen untersucht werden, also z.B. im Femtosekunden Bereich. In Molekülen, in denen die Bindungslängen konstant gehalten werden können, weil ihre Bewegungen von denen der Bindungswinkel und Torsionswinkel relativ stark entkoppelt sind, ist es günstiger, einen anderen Algorithmus zu verwenden (Verlet, 1967; Hockney & Eastwood, 1981; Beeman, 1976). Damit ist es dann möglich, die Integrationsschrittweite so zu vergrößern, daß man auch über etwas längere Zeiträume (Pikosekunden) simulieren kann (Berendsen & van Gunsteren, 1986; van Gunsteren & Berendsen, 1977). Eine Analyse der Genauigkeit von Langevin und Molekulardynamik Algorithmen wurde kürzlich von Pastor et al. (1988) durchgeführt.

Molekulardynamik Berechnungen können bei definierten Temperaturen T_0 und Drucken p_0 erfolgen,

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau_T}$$

$$\frac{dp}{dt} = \frac{p_0 - p}{\tau_p}$$

wobei τ_T und τ_p die Temperatur- und Druck-Relaxationszeiten sind, mit denen das System an ein externes Bad gekoppelt wird (Berendsen et al., 1984; Hermansson et al., 1988).

Die Potential-Funktion für Biomoleküle.

Die Wahl der Potentialfunktion hat erheblichen Einfluß auf die Genauigkeit der Ergebnisse. Die benutzbaren Funktionen sind auch in starkem Maße von den zur Verfügung stehenden Compu-

tern abhängig. In der heutigen Molekulardynamik-Methode für biologische Moleküle werden die Atome als Massenpunkte mit Punktladungen aufgefaßt, d.h. elektronische Effekte wie Polarisation können nicht im Detail beschrieben werden. Die Wechselwirkungen zwischen den Atomen werden durch sog. 'effektive Potentiale' dargestellt. Diese wurden derart geeicht, daß sowohl makroskopische Ensembleeigenschaften wie z.B. die Verdampfungsenthalpie, als auch charakteristische chemische Eigenschaften, wie z.B. die Unterschiede zwischen Sauerstoffatomen aus Ester-, Carbonyl-, Carboxyl-, Hydroxyl-gruppen, oder Wasser-Sauerstoffatomen berücksichtigt werden (Hermans et al., 1984; Postma, 1985). Der Vorteil des Konzepts eines effektiven Paarpotentials ist, daß die mathematische Form rechentechnisch einfach zu handhaben ist, daß aber trotzdem durch die Parametrisierung viele zusätzliche Ensemble-Eigenschaften implizit berücksichtigt werden.

Die Potentialfunktion für ein Molekül enthält Terme für die Bindungslängen b, die Bindungswinkel θ , die Torsionswinkel ξ (mit einem eingeschränkten Winkelbereich, wie etwa in Ringen), sowie die Torsionswinkel ϕ (für Seitenketten mit Rotationsmöglichkeit um 360°). Die Werte für die Idealgeometrien sind dabei mit den Indices 0 versehen. Die Potentialfunktion enthält weiterhin Terme für die nicht-gebundenen Wechselwirkungen wie van der Waals- und Coulomb-Wechselwirkungen. Dabei werden die elektrostatischen Kräfte durch Ladungen q_i auf den Atomzentren modelliert (van Gunsteren & Karplus, 1982). Die Kraftkonstanten K_i und Lennart-Jones-Konstanten C_6 und C_{12} sind ebenfalls jeweils so gewählt, daß das gesamte Potential die experimentellen Eigenschaften der Moleküle möglichst genau wiedergegeben kann (empirisches Kraftfeld). Das GROMOS Potential (Berendsen et al., 1981; Hermans et al., 1984) hat folgende Form :

$$\begin{split} V\left(\overline{r}_{1},...,\overline{r}_{N}\right) &= \\ \sum_{l=1}^{N_{b}} \frac{1}{2} K_{b_{l}} \left(b_{l} - b_{0_{l}}\right)^{2} + \sum_{l=1}^{N_{\theta}} \frac{1}{2} K_{\theta_{l}} \left(\theta_{l} - \theta_{0_{l}}\right)^{2} + \\ \sum_{l=1}^{N_{\xi}} \frac{1}{2} K_{\xi_{l}} \left(\xi_{l} - \xi_{0_{l}}\right)^{2} + \sum_{l=1}^{N_{\phi}} K_{\phi_{l}} \left(1 + \cos\left(n_{l} \Phi_{l} - \delta_{l}\right)\right) \\ + \sum_{i < j}^{N} \left(\frac{C_{12}(ij)}{r_{ij}}\right)^{12} - \left(\frac{C_{6}(ij)}{r_{ij}}\right)^{6} + \frac{q_{i}q_{j}}{4\pi\varepsilon_{0}\varepsilon_{r}r_{ij}} \end{split}$$

Es wurden auch andere Potentialfunktionen entwickelt, implementiert in verschiedenen Programmpaketen (z. B. AMBER (Weiner et al., 1984; Weiner et al., 1986); CHARMM (Brooks et al., 1987; Nilsson & Karplus, 1986), DISCOVER (Hagler et al., 1974; Maple et al., 1988), EPEN/2 (Momany et al., 1974; Snir et al., 1978), OPLS- Parameter (Jorgensen & Tirado-Rives, 1988), X-PLOR (Bruenger, 1988)), die z. B. zusätzliche Kreuzterme oder weitere explizite Terme, z. B. für Wasserstoffbrückenbindungen, enthalten.

Um alle Wechselwirkungen des Systems zu erfassen, ist es im Prinzip notwendig, über alle 1/2 N(N-1) Paare von Teilchen zu summieren. Trotzdem werden in der Praxis jedoch nur Nachbarteilchen innerhalb eines definierten Radius, des sog. Cut-off Radius, um jedes Teilchen berücksichtigt.

Die Potentialfunktion für Wasser

Weil die meisten Reaktionen biologischer Moleküle in wäßriger Lösung ablaufen, sind Molekulardynamik-Simulationen in Wasser von größter Bedeutung. Für die adequate Beschreibung der Ensemble-Eigenschaften sind einfache Paarpotentiale nicht ausreichend. Beispielsweise können damit nicht gleichzeitig die Eigenschaften von Wasser in der Gasphase, in Lösung und im Festkörper beschrieben werden, da kooperative Beiträge in der Größenordnung von 20-30 % auftreten können (Barnes et al., 1979; Clementi et al., 1980; Owicki et al., 1975; Vernon et al., 1982; Berendsen et al., 1981; Koehler et al., 1987a; Koehler & Saenger, 1987). Dieser große Beitrag von kooperativen Kräften wird hauptsächlich durch Polarisationseffekte der O-H Gruppen verursacht. Das kann zu einer Änderung des Dipolmomentes von ca. 25 % fühn (Berendsen et al., 1981) und wird sehr stark durch die Umgebung des Wassermoleküls beeinflußt (Berendsen & van Gunsteren, 1984; Barnes et al., 1979; Berendsen et al., 1981). Ein verbessertes effektives Paarpotential für flüssiges Wasser wurde von Berendsen et al. (1981) entwickelt.

Dieses Simple Point Charge (SPC) Modell besteht aus einem Wassermolekül mit festgehaltener Geometrie sowie Punktladungen auf den Atomen für die elektrostatischen Wechselwirkungen und van der Waals Wechselwirkungen.

Vergleich einer experimentellen Struktur mit der aus der Trajektorie erhaltenen.

Das Ergebnis einer Röntgenstrukturanalyse ist eine Elektronendichte, die so interpretiert wird, daß man die statische Proteinstruktur erzeugt, deren berechnete Röntgenbeugungseigenschaften am Besten mit den gemessenen übereinstimmen. Flexible und ungeordnete Regionen sind dabei über die sogenannten B-Faktoren zu erkennen. Das Ergebnis einer Molekulardynamik-Simulation ist dagegen der Bewegungsablauf eines Proteins über mehrere Pikosekunden. Würde diese simulierte Struktur in völligem Gleichgewicht stehen und würde sie keine multiplen Minima von ähnlicher Energie aufweisen, wäre eine Durchschnittsstruktur der Trajektorie die relevante Struktur, die es zu vergleichen gilt. Beides ist jedoch meist nicht der Fall. Eine sorgfältige Analyse der Zahl der Minima des Proteins und des Vergleichs mit der Röntgenstruktur wurden Dn Elber und Karplus (1987) durchgeführt. Die Übereinstimmungen und Unterschiede zwischen Neutronendiffraktionsstudien und Molekulardynamik Simulationen bei Raumtemperatur und bei 120 K im Kristall (Koehler et al., 1987b,c) sowie in Lösung (Koehler et al., 1988b) wurden an Cyclodextrinen ausführlich untersucht, wobei auch die Frage der Flip-Flop- und Mehrzentren-Wasserstoffbrückenbindungen untersucht wurde (Koehler et al., 1988a,c).

Berechnungen von Differenzen zwischen freien Energien

Die methodischen Konzepte zur Berechnung von Differenzen von freien Energien mit der Molekulardynamik-Methode sind in Übersichtsartikeln von Kollman & van Gunsteren (1987), Singh et al. (1987), van Gunsteren (1988), Cieplak & Kollman (1988), und van Gunsteren & Weiner (1989) dargestellt. Die Bedeutung und eine etwas andere Methode der Berechnung von freien Reaktionsenergien beschreiben auch Warshel et al. (1988). Die berechneten Differenzen zwischen freien Energien zweier verschiedener Systeme können mit experimentell bestimmten Bindungskonstanten verglichen werden:

Enzym + Substrat
$$_1$$
 Enzym•Substrat $_1$ - Komplex ΔG_3 ΔG_2 Enzym + Substrat $_2$ Enzym•Substrat $_2$ - Komplex

In diesem Beispiel können die Werte für ΔG_1 und ΔG_2 aus dem Experiment bestimmt werden (Bindungskonstanten K ; ΔG = RT In K), und ΔG_3 sowie ΔG_4 werden durch Berechnung aus Molekulardynamik-Simulationen bestimmt. Die G - Funktion ist eine Zustandsfunktion, und deshalb ist die Summe aller oben erwähnten ΔG gleich Null. Die Art und Weise, in der die senkrechten Prozesse durchgeführt werden, ist ohne direkte physikalische Realität: Die Kraftfeldparameter, welche das Substrat 1 beschreiben, werden in diejenigen des Substrates 2 überführt, um ΔG_3 zu erhalten; entsprechend wird für die rechte Seite des Gleichungssystems verfahren. Während des reversiblen Übergangs von einem Zustand zum anderen werden alle Konfigurationen mit ihren enthalpischen und entropischen Beiträgen berechnet und addiert, so daß sich die Differenz $\Delta G = \Delta H$ - $T\Delta S$ zwischen Anfangs- und Endzustand ergibt. Die methodischen Grundlagen sind beschrieben in van Gunsteren & Weiner (1989).

Das wesentliche Konzept dabei ist, daß die Molekulardynamik-Simulation benutzt wird, um ein Ensemble von Konformationen zu generieren. Zuerst wird der Hamilton Operator H (p,r), der das System durch seine potentielle Energie V(r) und seine Impulse p beschreibt, in eine Funktion des Kopplungsparameters λ verwandelt, so daß H (p,r, λ_A) den Ausgangszustand und H (p,r, λ_B) den Endzustand darstellt, für die die Differenz der freien Energie ΔG bestimmt werden soll.

$$H (p,r,\lambda) = \sum_{i=1}^{N} \frac{p_i^2}{2m_i(\lambda)} + V (r_1,...,r_N,\lambda)$$

Die Verbindung zwischen der freien Energie und der Simulation ist über die Zustandssumme Z der statistischen Thermodynamik gegeben.

$$\Delta G_{BA} = \Delta H_{BA} - T \Delta S_{BA}$$

$$\Delta F_{BA} = \Delta U_{BA} - T \Delta S_{BA}$$

$$G(\lambda) = NkT - kT ln Z(\lambda)$$

$$F(\lambda) = -kT \ln Z(\lambda)$$

$$\Delta F_{BA} = F(\lambda_B) - F(\lambda_A) = -kT \ln \frac{Z(\lambda_B)}{Z(\lambda_A)}$$

Differenziert man die Helmholtzsche freie Energie $F(\lambda)$ nach λ bei konstanter Temperatur T, erhält man am Schluß einen Ausdruck, der einen Mittelwert des Ensembles von Konformationen für verschiedene λ darstellt.

$$\left(\frac{\partial F(\lambda)}{\partial \lambda}\right)_{T} = -\frac{kT}{Z(\lambda)} \left(\frac{\partial Z(\lambda)}{\partial \lambda}\right)_{T} = <\frac{\partial H(p,r,\lambda)}{\partial \lambda} > \lambda$$

Wenn λ sehr langsam während der Molekulardynamik Simulation verändert wird, so kann gleichzeitig integriert werden. Die erhaltene freie Energiedifferenz ΔF_{BA} ist dann und nur dann richtig, wenn das System immer im Gleichgewicht bleibt (reversible Änderung von einem Intermediat zum nächsten).

$$\Delta F_{BA} = \int_{\lambda_{\Delta}}^{\lambda_{B}} < \frac{\partial H(p,r,\lambda)}{\partial \lambda} > \lambda d\lambda$$

Für ein isobares Ensemble gilt eine entsprechende Formel (van Gunsteren & Berendsen, 1987).

Die Berechnung der Differenz der Entropien erfolgt ebenfalls über eine Ensemblemittelwertbildung, denn es gilt

so daß letztendlich die Differenz der freien Energien der Zustände A und B aus der Simulation bestimmt werden kann.

Fallstudien an einem Antikörper

Alle angesprochenen Probleme werden in Antikörper-Antigenkomplexen tangiert, deren medizinische und biotechnologische Bedeutung hier nicht erläutert werden muß. Ein besonders geeignetes Modellsystem für derartige Untersuchungen ist das Maus Immunglobulin A mit Namen McPC603, welches das Antigen Phosphorylcholin bindet. Der Grund für die Auswahl dieses Systems liegt in den zahlreichen Vorarbeiten, die für dieses Protein bereits geleistet wurden: Die Röntgenstruktur des Proteins mit und ohne gebundenes Hapten (Segal et al., 1974; Satow et al., 1986) ist ebenso wie zahlreiche Bindungsparameter bekannt. Das System kann gerade noch mit theoretischen Methoden behandelt werden, denn ein Fragment des Antikörpers (das sogenannte Fv- Fragment, bestehend aus 2 Ketten mit 115 und 122 Aminosäuren) zeigt bereits dieselben Bindungseigenschaften wie der gesamte Antikörper (Skerra & Plückthun, 1988). Aus unseren neueren experimentellen Ergebnissen liegt außerdem eine größere Anzahl von Daten darüber vor, welche Antigene beim gentechnologischen Austausch von bestimmten Aminosäuren ein verbessertes oder verschlechtertes Bindungsverhalten aufweisen (Glockshuber et al., 1990).

Antigen-bindende Regionen bestehen aus jeweils drei hypervariablen Loops in jeder der beiden Ketten. Mehrere theoretische Ansätze zur Strukturvorhersage der hypervariablen Loops wurden bereits berichtet (Snow & Amzel, 1986; Fine et al., 1986; de la Paz et al., 1986; Bruccoleri et al., 1988), jedoch wurde noch keine Molekulardynamik Trajektorie für das Molekül berechnet, in der alle Atome frei beweglich sowie Wassermoleküle eingeschlossen waren. Eine solche Studie wurde nun durchgeführt. Sie spiegelt die Flexibilität der Bindungstasche und des Phosphorylcholinmoleküls besser wider und ermöglicht nun quantitative Vergleiche mit der Röntgenstruktur. Außerdem ist sie, wie unten angeführt, die Voraussetzung für Berechnungen von Bindungsenergien des Antigens.

Wir haben von drei verwandten Antikörper-Antigen Komplexen diese Trajektorien über 30 Pikosekunden berechnet. Dabei konnte im Falle des Antikörpers McPC603 von der Röntgenstruktur des Komplexes mit Phosphorylcholin ausgegangen werden. Die Ausgängsstrukturen der verwandten Antikörper mit den Namen M167 und T15 wurden, ausgehend von derselben Kristallstruktur, an einer Graphic-Workstation nach geometrischen Kriterien modelliert. Die Aminosäure-Sequenzen der beiden Antikörper M167 und T15 stimmen zu ca. 80% mit denen von McPC603 überein. Alle Komplexe wurden bei 293K in wäßriger Lösung mit dem GROMOS87 Programmpaket simuliert, wobei das SPCE Wassermodell verwendet wurde.

Die Strukturen, die während der Simulation durchlaufen werden, zeigen nach unserer bisher vorläufigen Analyse, daß McPC603 ein equilibriertes Verhalten während der 30 Pikosekunden

aufweist. Demgegenüber sind in den beiden anderen Systemen stärkere Fluktuationen zu beobachten, die darauf hindeuten, daß diese Konformationen noch einer längeren Simulationsdauer bedürfen, um ins Gleichgewicht zu kommen. Diese hatten als Startkonfiguration ja nicht ihre eigenen Kristallstrukturen zur Verfügung. Genaue Analysen und Vergleiche mit der Kristallstruktur werden zur Zeit durchgeführt, um den Wert von Molekulardynamik zur Strukturvorhersage besser abschätzen zu können.

Die Bildung des Antigen-Antikörper-Komplexes dürfte in diesem Fall zu einem erheblichen Teil auf elektrostatische Wechselwirkungen zurückzuführen sein. Deshalb ist eine detaillierte Studie über alle möglichen Bindungskräfte, d.h. van der Waals Kräfte, H-Brücken sowie in besonderem Maße Elektrostatik, notwendig. Zur detaillierten Analyse des Verhaltens der Kräfte während der Simulation wurde ein Programm entwickelt, das die elektrostatischen (nur Monopolterme) und van der Waals- Wechselwirkungen in statischen Strukturen zwischen allen Atompaaren berechnet.

Moleküle wie das Hapten Phosphorylcholin müssen, da sie nicht in Standardbibliotheken von Aminosäuren vorkommen, neu parametrisiert werden. Auch dafür mußten verbesserte Methoden entwickelt werden. Mit quantenmechanischen Methoden werden solche nicht-parametrisierten Moleküle optimiert, so daß der experimentellen Struktur und auch der theoretischen Methode so weit wie möglich entsprochen wird. In diesem Beispiel wurden Einflüsse wie Basisatzabhängigkeit und Berechnung der Ladungsverteilung nach unterschiedlichen Methoden untersucht: Mullikensche Populationsanalyse, elektrostatische Anpassung nach Chirlian & Francl (1985), sowie Verwendung von natürlichen Orbitalen (NO's) (Reed et al., 1985). Für eine Anzahl von Phosphorylcholin-Derivaten wurden die Ladungsverteilungen nach den oben genannten Methoden bestimmt (Schaumberger, 1989). Aus den Ergebnissen läßt sich ableiten, daß die NO-Methode bezüglich Basissatzunabhängigkeit, qualitativer Übereinstimmung mit Elektronegativitäten und Rechengeschwindigkeit vorteilhaft ist. Damit liegt nun eine konsistente Lagungsverteilung für eine Reihe von Derivaten des Antigens Phosphorylcholin vor, die nach einer Umskalierung (Köhler, 1990) als Ladungsparameter für die Molekulardynamik Simulationen verwendet werden können.

Die Berücksichtigung von höheren elektrostatischen Termen (Dipole) und eventuell Polarisationseffekten, im Kraftfeld während der Simulation müssen in Zukunft getestet werden. Die Verwendung weiterer Terme wird allerdings möglicherweise eine neue Gesamtparametrisierung des Kraftfeldes erforden. In Anbetracht der geringen Bindungsenergieunterschiede verschiedener Komplexe aus Protein und Ligand wird dieser Schritt aber notwendig werden. Es sei in diesem Zusammenhang auf Arbeiten verwiesen, in denen die Dipole in kleinen Molekülen mit der ab initio Methode bestimmt wurden, und die Anhaltspunkte für dieses Vorgehen liefert (Williams, 1988; Spackman, 1989; Klapper et al., 1986).

Probleme der Molekulardynamik-Methode und Perspektiven

Der erste Problemkreis ist das Potentialfeld. Dabei sind besonders die dielektrischen Phänomene der Elektrostatik zu nennen, die in den bisherigen Ansätzen nur implizit berücksichtigt werden. Das führt besonders bei geladenen Molekülen zu falschen Ergebnissen.

Der zweite Problemkreis ist die Qualität und Länge der Simulation. Hierfür sei auf den ausführli-

chen Artikel über die Genauigkeit von Berechnungen der Differenzen freier Energien von Pearlman & Kollman (1989) verwiesen. Es müssen bei dieser Methode viele Kompromisse geschlossen werden, die zum Teil von besseren Supercomputern unnötig gemacht werden könnten. Hierbei sei darauf hingewiesen, daß dafür auch in besonderem Maße vektorisierbare, parallelrechnende Computer oder gar Transputer von erheblicher Bedeutung sind und noch mehr sein werden (Lybrand & Kollman, 1985; Sauniere et al., 1989; Goodfellow & Vovelle, 1989).

- (1) Die derzeitige Beschränkung der Simulationsdauer auf den Pikosekundenbereich führt oft zu der Frage, ob der gewünschte physikalische Prozeß innerhalb dieser kurzen Zeitspanne überhaupt beobachtbar ist. Eine Studie über die Verlängerung der Zeitskala von Molekulardynamik wurde von Cartling (1989) durchgeführt.
- (2) Bei thermodynamischen Fragestellungen (wie der reversibel durchzuführenden Berechnung der Differenzen der freien Energien) ist die Statistik durch die Kürze der Simulation für den zu untersuchenden Konformationsraum nicht ausreichend. Daraus resultiert eine große Ungenauigkeit.
- (3) Weiterhin kann eine nicht ausreichende numerische Rechengenauigkeit dazu führen, daß eine längere Trajektorie einen anderen Verlauf nimmt. Auch aus diesem Grund sind heute nur Simulationen im Pikosekundenbereich durchführbar. Prozesse von chemischem oder biochemischem Interesse, die in größeren Zeiträumen ablaufen, (z. B. Konformationsänderungen im Mikrosekunden- oder Millisekundenbereich) sind daher bisher nicht zu handhaben.
- (4) Es werden approximierende Methoden für die Integration der Newtonschen Bewegungsgleichungen verwendet.
- (5) Die Suche der nicht-gebundenen Nachbarn ist sehr aufwendig, was die benötigte Rechenzeit anbelangt (Berendsen & van Gunsteren, 1986). Rechenzeit wird momentan dadurch gespart, daß die angelegten Nachbarlisten nur alle 10 bis 20 Schritte erneut bestimmt werden.
- (6) Es werden nur die Monopolterme der Elektrostatik explizit berechnet. Einflüsse von Dipolund höheren Termen schlagen sich nur implizit durch die Parametrisierung nieder.
- (7) Die Monopolterme der Elektrostatik haben eine Reichweite von ca. 30 Å, aber die verwendeten Cut-off Radien zur Bestimmung der Nachbarn eines Atoms, für die dieser Term berechnet wird, liegen derzeit in der Größenordnung von ca. 8-12 Å. Dadurch wird die Berechnung der Kräfte durch den Cut-off Radius beeinflußt, so daß artifizielle Fluktuationen und Rauschen entstehen. Es gibt allerdings die Möglichkeit, das Potential durch eine sogenannte 'Switching Funktion' so zu verändern, daß diese Fehler etwas eingeschränkt werden (Berendsen & van Gunsteren, 1986). Wenn keine Korrekturen angebracht werden, erkennt man das Rauschen an dem Drift, den der Algorithmus produziert.
- (8) Um die Beschreibung von langreichweitigen Wechselwirkungen und Polarisationseffekten zu verbessern, werden derzeit neue Wassermodelle entwickelt.
- (9) Der Einschluß von additiven Kräften, die über die Paarwechselwirkung hinausgehen, ist im Prinzip möglich, aber erfordert neue Konzepte und erheblich mehr Rechenzeit.

Ausblick

Die Leistungsfähigkeit der Computer ist sicher ein limitierender Faktor in der Qualität der Molekulardynamik-Simulationen und möglicherweise derzeit der Engpaß. Die wirklichen Grenzen liegen aber auch in der Qualität der physikalischen Beschreibung, doch diese Grenzen sind teilweise wegen der geringen Computerleistungen noch gar nicht erkennbar. Die Molekulardynamik-Methode ist ein Gegenstand der Forschung; als solcher ist sie sicher noch kein etabliertes Werkzeug, das dem Experimentalisten zu Diensten ist. Das wird erst möglich sein, wenn die Genauigkeit energetischer Voraussagen besser als 1 bis 2 kcal/mol ist, denn oft ist dies die Barriere einer biologischen Ja/Nein-Entscheidung. Um diesem wichtigen Ziel näher zu kommen, ist allerdings die enge Zusammenarbeit zwischen theoretisch und experimentell arbeitenden Biochemikern sowie den Computerfachleuten die Voraussetzung. Die Zeit, in der diese Methodik einen direkten Einfluß auf Medizin, Pharmazie und Biotechnologie haben wird, ist sicherlich absehbar, und die Geschwindigkeit dieser Entwicklung wird entscheidend durch die Fortschritte der Computertechnologie mitbestimmt werden.

Literatur

Barnes P, Finney JL, Nicholas JD, Quinn JE (1979). Nature 282: 459

Beeman D (1976). J. Comp. Phys. 20: 130

Bennett WS, Huber R (1984). CRC Crit. Rev. Biochem. 15: 291

Berendsen HJC, Gunsteren WF van (1984). In: Barnes AJ et al. (eds.). NATO- ASI Series C135, Reidel Publishing Company, Dordrecht, pp 475

Berendsen HJC, Postma JPM, Gunsteren WF van, Hermans J (1981). In: Pullman B (ed.). Intermolecular Forces. Reidel Publishing Company, Dordrecht, pp 331

Berendsen HJC, Postma JPM, Gunsteren WF van, DiNola A, Haak JR (1984). J. Chem. Phys. 81: 3684

Berendsen HJC, Gunsteren WF van (1986). In: Ciccotti G, Hoover WG (eds.). Proc. Enrico Fermi School of Physics, Varenna, North Holland Physics Publishing, pp. 43

Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983). J. Comp. Chem. 4: 187

Brooks CL, Karplus M, Pettitt BM (1988). In: Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics. Advances in Chemical Physics, Vol. LXXI; John Wiley & Sons

Bruccoleri RE, Haber E, Novotny J (1988). Nature 335: 564

Bruccoleri RE, Karplus M (1987). Biopolymers 26: 137

Bruenger AT (1988). J. Mol. Biol. 203: 803

Burns A, (1988). Programming in OCCAM 2. Addison Wesley Publishing Company.

Cartling B (1989). J. Chem. Phys. 91: 427

Chirlian LE, Francl MM (1985). Quantum Chemistry Program Exchange, Program 524. Cieplak P, Kollman PA (1988). J. Am. Chem. Soc. 110: 3734

Clementi E, Kolos W, Lie GC, Ranghino G (1980). Int. J. Quant. Chem. 17: 377

Creighton TE (1983). Proteins. Freeman, New York.

Dyson HJ, Rance M, Houghten RA, Lerner RA, Wright PE (1988a). J. Mol. Biol. 201: 161

Dyson HJ, Rance M, Houghten RA, Wright PE, Lerner RA (1988b). J. Mol. Biol. 201: 201

Elber R, Karplus M (1987) Science 235: 318

Fine RM, Wang H, Shenkins PS, Yarmush DL, Levinthal C (1986). Proteins 1: 342

Fletcher R, Reeves CM (1964). The Computer J. 7: 149

Gear CW (1971). Numerical Initial Value Problems in Ordinary Differential Equations. Prentice Hall, Englewood Cliffs, N.J.

Glockshuber R, Stadlmüller J, Plückthun A (1990). Zur Publikation eingereicht.

Goodfellow JM, Vovelle F (1989). Eur. Biophys. J. 17: 167

Gunsteren WF van (1988). Protein Eng. 2:5

Gunsteren WF van, Berendsen HJC (1977). Mol. Phys. 34: 1311

Gunsteren WF van, Berendsen HJC (1987). J. Comput.-Aided Moi. Design 1: 171

Gunsteren WF van, Karplus M (1980). J. Comp. Chem. 1: 266

Gunsteren WF van, Karplus M (1982). Macromolecules 15: 1528

Gunsteren WF van, Weiner PK, eds. (1989). Computer Simulation of Biomolecular Systems. ES-COM Science Publishers B.V., Leiden.

Hagler AT, Huler E, Lifson S (1974). J. Am. Chem. Soc. 96: 5319

Harvey SC (1989). Proteins 5: 78

Hermans J, Berendsen HJC, Gunsteren WF van, Postma JPM (1984). Bioploymers 23: 1513

Hermansson K, Lie GC, Clementi E (1988). J. Comp. Chem. 9: 200