

## Rechnergestützte Methoden erleichtern die Aufklärung von Proteinfunktionen\*\*

Gerd Folkers\* und Christian D. P. Klein

Die Biochemie steht vor ihrer bisher größten Herausforderung: Sie muss die in immer stärkerem Maße anfallenden Genomdaten in einen funktionellen Zusammenhang stellen. Die Durchsequenzierung eines Genoms stellt in erster Linie logistische Anforderungen an die beteiligten Firmen und Institute. Hier ist durch vermehrten Materialaufwand eine fast beliebige Steigerung des Durchsatzes möglich. Anders ist die Lage, zumindest jetzt noch, beim wichtigsten nachfolgenden Schritt: Der Funktionsaufklärung jener Proteine, welche von den nun bekannten Gensequenzen kodiert werden. Dieser Forschungszweig wird als „Functional Genomics“ bezeichnet.

Von einer DNA-Sequenz kann zwangslässig auf die Sequenz des Proteins geschlossen werden. Proteinsequenzen werden seit Jahrzehnten archiviert, und inzwischen sind viele tausend Sequenzen von Proteinen mit oftmals bekannter Funktion allgemein verfügbar.<sup>[1]</sup> Da Proteine ähnlicher Sequenz in vielen Fällen die gleiche Funktion erfüllen, liegt es nahe, durch Sequenzvergleiche auf die Funktion neuer Proteine zu schließen. Dieser Ansatz wird seit längerer Zeit verfolgt<sup>[2]</sup> und ist durch Verbesserung der zugrunde liegenden Algorithmen zu einem wichtigen Handwerkszeug der Genetiker und Biochemiker geworden. Einer der großen Vorteile dieser Methode ist ihre Schnelligkeit, außerdem benötigt sie keine Daten zur dreidimensionalen Struktur. Sie hat aber auch einen schwerwiegenden Nachteil: Im Laufe der Evolution haben sich Proteine oft konvergent in Richtung auf ihre heutige Funktion entwickelt, d.h., dass nicht entwicklungs geschichtlich verwandte Proteine, deren Aminosäuresequenzen nur wenig ähnlich sind, ähnliche oder identische Funktionen haben.<sup>[3]</sup> So sind sich die Enzyme Trypsin und Subtilisin in großen Teilen sehr unähnlich. Es wäre nicht möglich, eine funktionelle Verwandschaft dieser Enzyme aufgrund ihrer

Primärstruktur festzustellen, da hier die Übereinstimmung nur etwa 10 % beträgt – wenig mehr, als der statistische Zufall erwarten ließe. Ein bestimmter Teil dieser Enzyme ist jedoch sehr ähnlich: Es ist jener Bereich, welcher den eigentlichen katalytischen Apparat ausmacht. Die an der Katalyse beteiligten Aminosäuren sind zwar in der dreidimensionalen Struktur der beiden Enzyme benachbart, befinden sich aber nicht hintereinander auf dem Peptidstrang. Es wäre nicht möglich, diese beiden Enzyme mittels eines paarweisen Sequenzvergleichs als funktionell verwandte Serinproteasen einzuordnen.

Die Konservierung des funktionsbestimmenden Bereichs ist eines der Paradigmen der Proteinchemie. Unabhängig von der Primärstruktur finden sich hier immer wieder ähnliche physikalisch-chemische Eigenschaften, oft sind sogar die beteiligten Aminosäuren identisch.

### *Von der Struktur zur Funktion*

S. Schmitt, M. Hendlich und G. Klebe stellten in dieser Zeitschrift einen neuen Ansatz zur Feststellung funktioneller Ähnlichkeit vor, welcher sich die Konservierung des aktiven Zentrums von Enzymen und des Ligandbindungsbereichs von Rezeptoren zunutze macht.<sup>[4]</sup> Sie beginnen damit, die Bindungstaschen von Proteinen mit bekannter Funktion und bekannter dreidimensionaler Struktur weitgehend zu abstrahieren. Eine Bindungstasche wird dann nicht mehr durch Aminosäuren, sondern durch die räumliche Verteilung bestimmter Pseudofunktionalitäten wie Wasserstoffbrückendorfer oder -acceptor und ähnliche repräsentiert (Abbildung 1). Mit dieser Datenreduktion wird eine einfach zu handhabende, aber immer noch aussagekräftige Datenbank erzeugt. Die

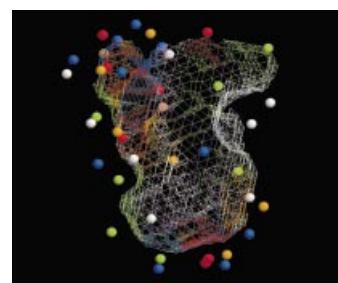


Abbildung 1. Oberflächen und Pseudozentren einer Bindungstasche (aus Lit. [4]).

[\*] Prof. Dr. G. Folkers, Dr. C. D. P. Klein  
Pharmazeutische Chemie  
Departement für Angewandte Biowissenschaften  
ETH Zürich  
Winterthurerstrasse 190, 8057 Zürich (Schweiz)  
Fax: (+41) 1-635-6884  
E-mail: folkers@pharma.ethz.ch

[\*\*] Wir danken Dr. H. Rüther, Köln, Prof. H.-D. Höltje, Universität Düsseldorf, und Prof. Dr. P. A. Schubiger, Paul-Scherrer-Institut, Villingen (Schweiz), für die kritische Durchsicht des Manuskripts.

(vermuteten) Bindungstaschen von Proteinen unbekannter Funktion werden identisch behandelt und können nun mit der Datenbank verglichen werden. Wird dort eine ähnliche räumliche Verteilung der Pseudofunktionalitäten in der Bindungstasche gefunden, so sind Rückschlüsse auf die Funktion des unbekannten Proteins möglich.

Diese Methode ist konzeptionell verwandt mit Techniken, die bisher insbesondere zur Identifizierung von Protein-Wechselwirkungspartnern, also Liganden oder Enzyminhibitoren, verwendet wurden. Sie steht gewissermaßen in der Tradition von heute viel genutzten Programmen wie GRID,<sup>[5]</sup> RELIBASE<sup>[6]</sup> und anderen.

Dies ist nicht der erste Versuch, die Geometrie von Bindungstaschen zur funktionellen Analyse zu verwenden.<sup>[7]</sup> Der wesentliche Vorteil der neuen Methode besteht darin, dass nicht nur die geometrischen Eigenschaften der Bindungstasche (also ihre Form), sondern auch die Anordnung physikalisch-chemischer Eigenschaften berücksichtigt wird.

Für Leser, welche mit den Techniken des Molecular Modeling etwas vertraut sind, ist die Tatsache erstaunlich, dass das von Klebe et al. entwickelte Verfahren mit nur fünf Pseudofunktionalitäten auskommt und dabei sehr gute Ergebnisse erzielt. Programme, welche die intermolekularen Wechselwirkungseigenschaften kleiner Moleküle bestimmen, verwenden sehr viel komplexere Darstellungen. Elektrostatische Eigenschaften der Bindungstasche werden beim Verfahren der Arbeitsgruppe Klebe beispielsweise nicht (oder nur indirekt) betrachtet – dies hat allerdings auch wieder den Vorteil, dass diese Eigenschaften nicht erst berechnet werden müssen.

Um den Ansatz von Schmitt, Hendlich und Klebe nutzen zu können, reicht die Kenntnis der Gensequenz oder der Primärstruktur eines Proteins nicht aus. Es müssen genaue Informationen über die dreidimensionale Struktur des Proteins vorliegen. Solche Daten werden oft durch Röntgenstrukturanalyse erhalten, seit einigen Jahren gewinnen NMR-Verfahren zunehmend an Bedeutung.<sup>[8]</sup> Beide Methoden sind bisher nicht für ihre Schnelligkeit oder Einfachheit bekannt. Wegen des großen kommerziellen Potentials hat dieses Gebiet viel Interesse auf sich gezogen, und es gibt Bestrebungen, die experimentelle Bestimmung von Proteinstrukturen durch zunehmende Automatisierung zum Hochdurchsatzverfahren zu entwickeln.<sup>[9]</sup> Neben den experimentellen Methoden finden theoretische Verfahren breite Verwendung.<sup>[10]</sup> In vielen Fällen ermöglichen sie Vorhersagen, welche für alle praktischen Zwecke ausreichend genau sind. Auch hier gibt es, wie bei der Funktionszuweisung, Ansätze, welche auf der Suche nach ähnlichen Primärstrukturen beruhen. Daneben wurden aber auch Methoden beschrieben, welche ohne solche zusätzlichen, empirisch gewonnenen Informationen auskommen.<sup>[11]</sup>

Die momentan vorliegenden Genomdaten höherer Organismen sind noch mit vielen Fehlern behaftet.<sup>[12]</sup> Wird eine fehlerhafte Proteinsequenz zur Vorhersage einer Proteinstruktur und schließlich zur Funktionsvorhersage verwendet, kann sich die von Schmitt et al. durchgeführte Abstraktion von Vorteil sein, indem der Einfluss einer „falschen“ Aminosäure auf die Repräsentation der Bindungstasche und damit das Ergebnis der Vorhersage verringert wird.

## Gene, Proteine, Arzneimittel

Die Sequenzierung von Genomen wird vor allem von der Hoffnung getrieben, neue therapeutische Ansatzpunkte zu finden. Die Intervention durch Medikamente findet nur sehr selten auf der Ebene der Gene statt, sondern erfolgt meistens an einem Protein. Es macht natürlich keinen Sinn, ein Protein mit unbekannter Funktion als Interventionspunkt auszuwählen. Abbildung 2 zeigt schematisch, worin die Bedeutung von

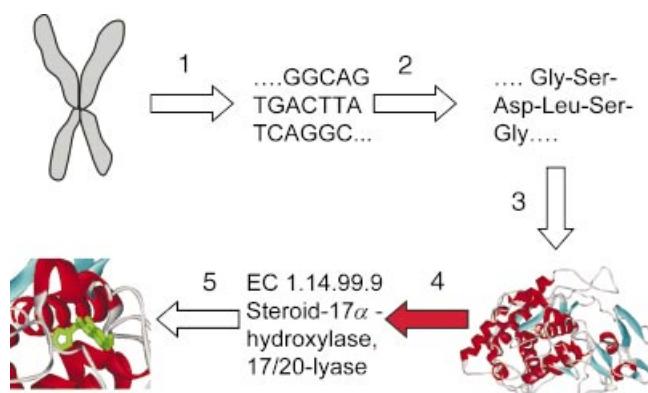


Abbildung 2. 1: Genomsequenzierung; 2: Übersetzung in Primärstruktur; 3: Ableitung oder Bestimmung der räumlichen Struktur; 4: funktionelle Einordnung; 5: Entwurf von Liganden oder Inhibitoren.

Genomdaten bei der Entwicklung neuer Arzneimittel besteht: Die DNA-Sequenz führt unmittelbar zur Proteinsequenz. Daraus kann die dreidimensionale Struktur theoretisch abgeleitet werden, oder es wird eine experimentelle Strukturanalyse durchgeführt. Der nächste Schritt, die Zuweisung einer Funktion, ist kritisch und mit großen Schwierigkeiten verbunden. Eine viel verwendete experimentelle Methode besteht darin, durch gentechnische Methoden Organismen zu erzeugen, denen das betreffende Gen fehlt (z. B. so genannte „Knock-out-Mäuse“), und die Auswirkung der Mutation auf den Organismus zu studieren. Die Generierung entsprechender Organismen ist aufwändig, im Falle des Menschen nicht möglich, und die Interpretation der Resultate fällt oft schwer. Wie bei der Struktaufklärung von Proteinen gibt es Bemühungen, die experimentellen Verfahren, beispielsweise bei der Erzeugung transgener Tiere, weitgehend zu automatisieren.<sup>[13]</sup> Dennoch sind valide theoretische Ansätze ein willkommenes Hilfsmittel.

Auf der Stufe der Funktionszuweisung entscheidet sich, ob ein Protein ein lohnendes Interventionsziel für die Wirkstoffentwicklung, ein „Target“, ist. Falls dem so ist, können z. B. mit den Mitteln des strukturbasierten Wirkstoffdesigns geeignete Inhibitoren oder Antagonisten gesucht werden.

Eine zunächst nicht offensichtliche Anwendung der von Klebe et al. aufgebauten Datenbank besteht darin, dass sie die theoretische Abschätzung von Nebenwirkungen ermöglichen wird. Ein idealer Arzneistoff bindet ausschließlich und mit großer Affinität an das gewünschte Protein. Unerwünschte Wirkungen werden oft dadurch hervorgerufen, dass es auch an andere Proteine bindet. Es gibt bisher kein Verfahren, welches die Abschätzung solcher unerwünschter Bindungen ohne aufwändige biologische Tests erlauben würde. Daher

wird bei neuen Arzneistoffen oft erst in einem späten Entwicklungsstadium, nach der Investition beträchtlicher finanzieller Mittel, festgestellt, dass sie Nebenwirkungen hervorrufen. Die von der Arbeitsgruppe Klebe entwickelte Methode ermöglicht eine „virtuelle“ Suche nach unerwünschten Wechselwirkungspartnern für Arzneistoffe, indem die Datenbank nach solchen Enzymen oder Rezeptoren durchsucht wird, welche den Entwicklungskandidaten binden können. Diese Strategie wäre die Umkehrung einer heute vielfach angewandten Methode zur Suche nach Liganden, dem „virtuellen Screening“ (Abbildung 3A).<sup>[14]</sup> Es würde

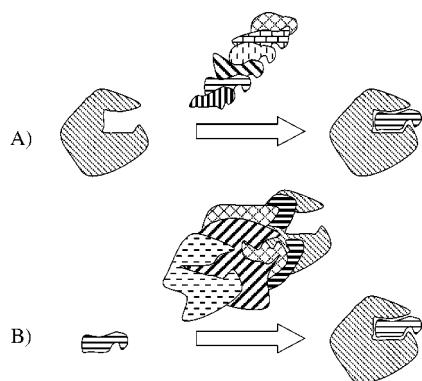


Abbildung 3. Virtuelle Suche nach einem Liganden (A) und nach einem Protein (B).

dabei nicht nach den kleinen Liganden, sondern nach ihren makromolekularen Targets gesucht werden (Abbildung 3B). Solch ein „inverses“ virtuelles Screening wäre nicht nur bei der Suche nach Nebenwirkungen, sondern auch viel allge-

meiner anwendbar. Beispielsweise könnte nach dem Angriffspunkt von biologisch aktiven Molekülen mit unbekanntem Wirkungsmechanismus gesucht werden.

- [1] SWISS-PROT: A. Bairoch, B. Boeckmann, *Nucleic Acids Res.* **1994**, 22, 3578–3580; <http://www.expasy.ch/sprot/>; BLAST: S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipson, *J. Mol. Biol.* **1990**, 215, 403–410; <http://www.ncbi.nlm.nih.gov/BLAST/>, **2001**.
- [2] S. B. Needleman, C. D. Wunsch, *J. Mol. Biol.* **1970**, 48, 443–453; W. R. Pearson, D. J. Lipman, *Methods Enzymol.* **1990**, 183, 63–98.
- [3] A. Fersht, *Structure and Mechanism in Protein Science*, 2. Auflage, W. H. Freeman, New York, **1999**.
- [4] S. Schmitt, M. Hendlich, G. Klebe, *Angew. Chem.* **2001**, 113, 3237–3241; *Angew. Chem. Int. Ed.* **2001**, 40, 3141–3144.
- [5] P. J. Goodford, *J. Med. Chem.* **1985**, 28, 849–857.
- [6] M. Hendlich, *Acta Crystallogr. Sect. D* **1998**, 54, 1178–1182; <http://relibase.ebi.ac.uk/>, **2001**.
- [7] D. Fischer, R. Norel, H. Wolfson, R. Nussinov, *Proteins Struct. Funct. Genet.* **1993**, 16, 278–292; S. L. Moodie, J. B. Mitchell, J. M. Thornton, *J. Mol. Biol.* **1996**, 263, 486–500; N. Kobayashi, N. Go, *Eur. Biophys. J.* **1997**, 26, 135–144; M. Rosen, S. L. Liang, H. Wolfson, R. Nussinov, *J. Mol. Biol.* **1998**, 281, 263–277; M. Stahl, C. Taroni, G. Schneider, *Protein Eng.* **2000**, 13, 83–88.
- [8] A. Saegusa, *Nature* **1998**, 392, 219.
- [9] Protein Structure Factory (Berlin), <http://userpage.chemie.fu-berlin.de/~psf/>, **2001**.
- [10] P. Bork, D. Eisenberg, *Curr. Opin. Struct. Biol.* **1998**, 8, 331–332; J. Skolnick, J. S. Fetrow, *Trends Biotechnol.* **2000**, 18, 34–39; D. Frishman, H. W. Mewes, *Prog. Biophys. Mol. Biol.* **1999**, 72, 1–17.
- [11] D. J. Osguthorpe, *Curr. Opin. Struct. Biol.* **2000**, 10, 146–152.
- [12] Siehe z.B.: S. Karlin, A. Bergman, A. J. Gentles, *Nature* **2001**, 411, 259–260.
- [13] B. P. Zambrowicz, G. A. Friedrich, E. C. Buxton, S. L. Lilleberg, C. Person, A. T. Sands, *Nature* **1998**, 392, 608–611.
- [14] H. J. Böhm, G. Schneider, *Virtual Screening for Bioactive Molecules*, Wiley-VCH, Weinheim, **2000**.