

Eine bessere Zukunft für Proteindesigner**

Anna Tramontano*

Stichwörter:

Computerbiologie · Proteindesign · Proteinfaltung · Strukturaufklärung

Die Entschlüsselung des Codes, der die Proteinsequenz mit der Struktur verknüpft, ist eine der wichtigsten, aber auch schwierigsten Aufgaben der Computerbiologie. Wäre dieser Code bekannt, könnten wir aus den Sequenzen der Proteine auf ihre Struktur schließen („direktes Faltungsproblem“, Strukturvorhersage) und Sequenzen entwerfen, die vorher festgelegte Faltungen annehmen („inverses Faltungsproblem“, Proteindesign).

Durch Lösen des direkten Faltungsproblems könnten wir Strukturinformationen über zahlreiche Proteine mit bekannter Sequenz erhalten und so auf ihre molekulare Wirkungsweise schließen und diese beeinflussen. Das erfolgreiche Design neuartiger Proteine hätte nicht nur wichtige praktische Auswirkungen für biotechnologische, nanotechnologische und therapeutische Anwendungen, sondern eine ebenso große theoretische Bedeutung für die Formulierung und Prüfung von Hypothesen zu Sequenz-Struktur-Beziehungen von Proteinen.

Historisch gesehen gibt es drei Arten von Proteindesign:^[1] 1) das Redesign der Sequenz eines bekannten Proteins, 2) den Einbau einer neuartigen Funktion in ein gegebenes Proteingerüst

und 3) das Design einer Sequenz, die eine neuartige, bisher noch nicht beobachtete Topologie oder Faltung annimmt. Jeder dieser Ansätze zielt auf unterschiedliche, aber verwandte Fragestellungen ab.

Natürlich vorkommende Proteinsequenzen werden meist aufgrund ihrer Stabilität, Löslichkeit, Wirkungsweise, Kompartimentbildung, ihres kinetischen Verhaltens, der Vermeidung einer nicht erwünschten Konformation oder aus anderen Gründen selektiert. Durch das Redesign der Sequenz eines bekannten Proteins lässt sich besser verstehen, welche relative Bedeutung diese Faktoren haben, welche Rolle sie bei der Formgebung natürlicher Sequenzen spielen und in welchem Umfang der Proteinsequenzraum während der Evolution ausgetestet wurde.^[2] Das zweite Verfahren ermöglicht die Herstellung nützlicher Reagentien, kann aber auch dazu beitragen, die Eigenschaften von aktiven Zentren aufzudecken und die Frage der molekularen Spezifität zu untersuchen.^[3] Das Design einer völlig neuartigen Faltung kann schließlich Aufschluss über die Eigenschaften geben, die eine Faltung physikalisch realisierbar machen, denn während der Evolution bestand sicher nicht die Möglichkeit, den gesamten verfügbaren Faltungsraum auszutesten. Im Idealfall könnten wir auf diese Weise ein Verzeichnis ausführbarer und nichtausführbarer Topologien zusammenstellen, diese testen und unterscheiden und so die Grundregeln der „Faltbarkeit“ verstehen.

Während die beiden ersten Methoden in einigen Fällen gute Ergebnisse lieferten, war die dritte Strategie – bis vor kurzem – weniger erfolgreich. In dieser Hinsicht setzt eine neuere Arbeit von Baker et al.^[4] einen Meilenstein.

Berichtet wird darin über das automatisierte Design eines neuartigen Proteins, was die Hoffnung auf weitere Erkenntnisse zum schwer zugänglichen Problem der Proteinfaltung weckt.

Mit dem Design von Proteinen hatte man in den 80er Jahren begonnen – ausgehend von einer recht optimistischen Betrachtung des Problems: Es war bekannt, dass eine gegebene Sequenz zwar eine charakteristische dreidimensionale Struktur einnimmt, es aber mehrere Sequenzen mit der gleichen Faltung gibt.^[5] Aufgrund dieser Degeneriertheit der Sequenz-Struktur-Beziehung wurde geschlossen, dass das Design einer Sequenz, die sich zu einer bestimmten Struktur faltet, einfacher sein könnte als die Vorhersage der charakteristischen Struktur, die eine bestimmte Aminosäuresequenz annimmt. Dennoch war der Weg zum Erfolg schwieriger als erwartet; in einigen Fällen zeigten die Design-Proteine zwar gewisse Protein-ähnliche Eigenschaften, z.B. kooperative Entfaltung und die erwarteten CD-Spektren, glichen aber sonst eher einem geschmolzenen Kügelchen („molten globule“) als einem natürlich gefalteten Protein.^[6–8]

Tatsächlich zeigten diese Versuche einen Hauptunterschied zwischen dem direkten und dem inversen Faltungsproblem auf: Es ist relativ einfach, die Richtigkeit einer Strukturvorhersage zu beurteilen – sodass Methoden verglichen, kombiniert und schrittweise verbessert werden können –, hingegen ist es angesichts der schwierigen Charakterisierung annähernd nativer Strukturen problematisch zu prüfen, wie erfolgreich eine Design-Faltung ist. Selbst wenn es für das Design eines Proteins mehr akzeptable Lösungen gibt als für die Strukturvorhersage, lässt sich der Erfolg eines Proteindesigns nur bewerten,

[*] Prof. A. Tramontano
Department of Biochemical Sciences
„A. Rossi Fanelli“
Università „La Sapienza“
P. le A. Moro, 00185 Roma (Italien)
Fax: (+39) 06-4440-062
E-mail: anna.tramontano@uniroma1.it

[**] Ich danke David Baker und Gautam Dantas, die uns freundlicherweise die Abbildung 1 zur Verfügung gestellt haben, sowie dem Consiglio Nazionale delle Ricerche („Progetto Finalizzato Genetica Molecolare“) und dem Ministero della Sanità für finanzielle Unterstützung.

wenn eine der Lösungen korrekt und exakt nachgewiesen ist.

Die Arbeit von Baker et al.^[4] bildet einen Durchbruch aus mehreren Gründen: Das Design lieferte ein Protein, das zum nativen Protein ausreichend ähnlich war, um für eine Röntgenkristallstrukturanalyse geeignet zu sein. Das verwendete Verfahren ist automatisiert, offenbar allgemein anwendbar und beruht im Wesentlichen auf dem gleichen Proteinkräftemodell wie die von derselben Arbeitsgruppe entwickelte Rosetta-Methode zur Strukturvorhersage.^[9]

Der erste Schritt der Methode besteht in der Wahl eines Satzes aus Abstands-Restraints, die die gewünschte Faltung definieren. Mit diesen Parametern werden Proteinfragmente bekannter Struktur zu mehreren Anfangskonformationen zusammengesetzt. Die Sequenzoptimierung für jede dieser Ausgangsstrukturen erfolgte mit Monte-Carlo-Simulationen und im Wesentlichen der gleichen Energiefunktion wie bei der Rosetta-Methode. Aufeinanderfolgende Zyklen aus Sequenz- und Strukturoptimierung führten schlussendlich zu dem als Top7 bezeichneten Sequenz-Struktur-Paar. Die biophysikalische Charakterisierung des synthetisierten Proteins ergab, dass Top7 löslich, monomer und stabil ist und eine kooperative Zweizustands-Entfaltung eingeht. Seine hochaufgelöste Struktur im Kristall weist eine eindrucksvolle Übereinstimmung mit dem Design-Modell auf und weicht nur in zwei kurzen exponierten Windungen von diesem ab (Abbildung 1).

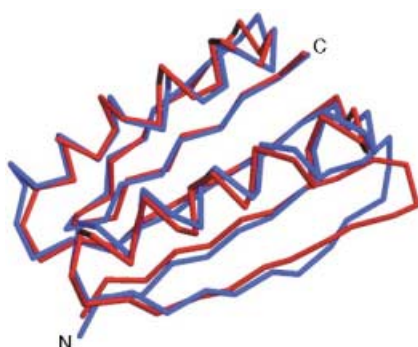


Abbildung 1. α -Überlagerung der durch Design erhaltenen (blau) und der kristallographisch bestimmten Struktur (rot) von Top7. (Mit freundlicher Genehmigung von David Baker und Gautam Dantas, Department of Biochemistry, University of Washington, USA.)

Wie bereits erwähnt wurde, ergeben sich aus einem erfolgreichen Design mehrere Schlussfolgerungen. Zunächst bestätigt es die Wirksamkeit der Strategie, durch Verwendung bekannter Unterfragmente mehrere Modelle auf atomarer Ebene herzustellen, ein Ansatz, der auch zur Vorhersage der Proteinstruktur bemerkenswert erfolgreich ist.^[10] Des Weiteren unterstreicht es die Zweckmäßigkeit einer simultanen Optimierung von Sequenz und Struktur. Schließlich wird eine der Schlussfolgerungen aus Experimenten zur Strukturvorhersage von Proteinen bestätigt: Ein korrektes Modell auf atomarer Ebene lässt sich nach umfassender Optimierung der Seitenkettenpackung leichter identifizieren als auf einer der Zwischenstufen des Verfahrens.^[10]

Interessanterweise berücksichtigte die Vorgehensweise von Baker et al. keinen der Punkte, die als Ursache für frühere Fehlschläge angesehen wurden; weder wurde versucht, vermutete alternative Faltungen zu destabilisieren, noch wurde die Kinetik der Faltung beim Design berücksichtigt. Die Autoren führen die große Übereinstimmung von Design und Vorhersage darauf zurück, dass die Sequenz im Verlauf des Designprozesses geändert werden kann und keine Bedingungen an eine Funktionalität des Design-Proteins vorgegeben wurden.

Einerseits kann die Möglichkeit, die Sequenz und damit die Energielandschaft zu ändern, zumindest teilweise die begrenzte Genauigkeit der verfügbaren Kraftfelder bei der Durchsuchung des komplexen, ausgedehnten und zerklüfteten Konformationsraums von Proteinen kompensieren. Andererseits sind die aktiven Zentren von Proteinen in struktureller Hinsicht höchstwahrscheinlich suboptimal, da sie funktionalen und dynamischen Vorgaben genügen müssen; eine Vorhersage ihrer Konformation auf der Basis einer Energieoptimierung kann daher irreführend sein. Es gibt allerdings mehrere Beispiele für den erfolgreichen Einbau von aktiven Zentren in bekannte Proteine. Die Gibbs-Energie für die Entfaltung von Top7 ist mit $13.2 \text{ kcal mol}^{-1}$ höher als für die meisten Proteine ähnlicher Größe, sodass es nicht überraschen würde, wenn tatsächlich ein funktionelles Zentrum in diesem Des-

ign-Gerüst untergebracht werden könnte. Ein solcher Ansatz würde den Weg zur Konstruktion von maßgeschneiderten molekularen Maschinen ebnen.

Online veröffentlicht am 14. Mai 2004

- [1] C. Sander, *Trends Biotechnol.* **1994**, *12*, 163–167.
- [2] D. T. Jones, *Protein Sci.* **1994**, *3*, 567–574; B. I. Dahiyat, C. A. Sarisky, S. L. Mayo, *J. Mol. Biol.* **1997**, *273*, 789–796; R. B. Hill, D. P. Raleigh, A. Lombardi, W. F. DeGrado, *Acc. Chem. Res.* **2000**, *33*, 745–754; M. Mutter, G. Tuchscherer, *Cell. Mol. Life Sci.* **1997**, *53*, 851–863; S. Dalal, L. Regan, *Protein Sci.* **2000**, *9*, 1651–1659; A. E. Keating, V. N. Malashkevich, B. Tidor, P. S. Kim, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14825–14830; L. L. Looger, H. W. Hellinga, *J. Mol. Biol.* **2001**, *307*, 429–445; S. A. Marshall, S. L. Mayo, *J. Mol. Biol.* **2001**, *305*, 619–631; J. Mendes, R. Guerois, L. Serrano, *Curr. Opin. Struct. Biol.* **2002**, *12*, 441–446; A. Jaramillo, L. Wernisch, S. Hery, S. J. Wodak, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 13554–13559; G. Dantas, B. Kuhlman, D. Callender, M. Wong, D. Baker, *J. Mol. Biol.* **2003**, *332*, 449–460.
- [3] M. Szardenings, B. Vasel, H. J. Hecht, J. Collins, D. Schomburg, *Protein Eng.* **1995**, *8*, 45–52; A. L. Pinto, H. W. Hellinga, J. P. Caradonna, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 5562–5567; A. Skerra, *Biochim. Biophys. Acta* **2000**, *1482*, 337–350; J. Reina, E. Lacroix, S. D. Hobson, G. Fernandez-Ballester, V. Rybin, M. S. Schwab, L. Serrano, C. Gonzalez, *Nat. Struct. Biol.* **2002**, *9*, 621–627; O. Maglio, F. Natri, V. Pavone, A. Lombardi, W. F. DeGrado, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3772–3777.
- [4] B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard, D. Baker, *Science* **2003**, *302*, 1364–1368.
- [5] C. Sander, *Biochem. Soc. Symp.* **1990**, *57*, 25–33.
- [6] A. N. Fedorov, D. A. Dolgikh, V. V. Chemeris, B. K. Chernov, A. V. Finkelstein, A. A. Schulga, Yu. B. Alakhov, M. P. Kirpichnikov, O. B. Ptitsyn, *J. Mol. Biol.* **1992**, *225*, 927–931.
- [7] T. Kortemme, M. Ramirez-Alvarado, L. Serrano, *Science* **1998**, *281*, 253.
- [8] A. Pessi, E. Bianchi, A. Cramer, S. Venturini, A. Tramontano, M. Sollazzo, *Nature* **1993**, *362*, 367–369.
- [9] K. T. Simons, R. Bonneau, I. Ruczinski, D. Baker, *Proteins* **1999**, *Suppl. 3*, 171–176.
- [10] A. Tramontano, V. Morea, *Proteins* **2003**, *Suppl. 6*, 352–368.