

Die Proteinstrukturfabrik –

Strukturanalyse mit hohem Durchsatz im

Berliner Forschungsverbund

Patrick Umbach¹ und Udo Heinemann^{2,3}

¹Leitprojektverbund Proteinstrukturfabrik, Berlin,

²Max-Delbrück-Centrum für Molekulare Medizin und ³Institut für Chemie – Kristallographie, Freie Universität Berlin

► Nachdem in den letzten Jahren mit großem Aufwand das menschliche Genom entschlüsselt wurde, steht man nun von der Aufgabe, die Sequenzinformation in Wissen über die biologische Funktion der Gene und Genprodukte zu verwandeln. Neben anderen Forschungsansätzen spielt in diesem Zusammenhang die systematische Hochdurchsatzanalyse dreidimensionaler Proteinstrukturen (*structural genomics*) eine wichtige Rolle [1, 2]. Die vom BMBF im Rahmen des Leitprojektwettbewerbs „Molekulare Medizin“ geförderte Berliner „Proteinstrukturfabrik“ [3] ist zur Zeit die wichtigste deutsche Aktivität auf diesem Sektor, der in den letzten zwei Jahren weltweit eine stürmische Entwicklung genommen hat [4-6]. Die „Proteinstrukturfabrik“ hat sich vorgenommen, in enger Kooperation mit dem Deutschen Humangenomprojekt Strukturen menschlicher Proteine mit einem gegenüber konventioneller Strukturanalyse um eine Größenordnung gesteigerten Durchsatz zu bestimmen.

Analog zur Genomsequenzierung stehen wir vor dem Problem der schier Dimension der zu lösenden Aufgabe: War die Proteinstrukturanalyse bislang ein langwieriges und mühsames Geschäft, so wollen wir nun Strukturen am Fließband bestimmen. Prinzipiell sind die Techniken schon lange bekannt: Die Proteingewinnung und -reinigung, spektroskopische und gelelektrophoretische Analyse, Kristallisation und letztlich die Strukturaufklärung durch Röntgendiffraktion beziehungsweise NMR-Spektroskopie. Die prinzipielle Herausforderung besteht daher in einer massiven Automatisierung, Standardisierung und Parallelisierung aller bereits bekannten Vorgänge. Für diese Aufgabe wurde die „Proteinstrukturfabrik“ als ein Verbund aus Berliner Universitäten, außeruniversitären Instituten der Grundlagenforschung, kleinen und mittelständischen Unternehmen (KMU) sowie Pharmaunternehmen gegründet (eine Liste der Projektpartner findet sich auf der Homepage der „Proteinstrukturfabrik“ unter <http://userpage.chemie.fu-berlin.de/~psf/>).

Schon aufgrund der Standardisierung können nicht alle prinzipiell interessieren-

den Strukturen ermittelt werden. Vielmehr startet die „Proteinstrukturfabrik“ mit einer großen Zahl von cDNA-Klonen, die alle Zwischenschritte durchwandern und aus dem Analyseprozess herausfallen, sollten sich Probleme ergeben, die angesichts der verfügbaren Zeit und Ressourcen nicht lösbar sind. Eine wichtige Proteinklasse, die derzeit von der Strukturanalyse im Rahmen der „Proteinstrukturfabrik“ ausgeschlossen ist, ist die Gruppe der integralen Membranproteine. Dies deshalb, weil sie ihre funktionale Struktur nur nach Einbau in eine biologische Membran einnehmen und daher dem nachstehend beschriebenen analytischen Ansatz nicht zugänglich sind.

Prinzipiell ist die „Proteinstrukturfabrik“ wie eine Pipeline mit einigen Bypassen aufgebaut (Abb. 1): Am Anfang stehen die im Ressourcenzentrum des Deutschen Humangenomprojekts (DHGP) gespeicherten humanen cDNAs [7]. Der nächste Schritt, die Bioinformatik, sorgt nun für eine optimale Auswahl und stellt aus den cDNA-Sequenzen eine Kandidatenliste zusammen, die einige experimentelle Schwierigkeiten schon von vornherein vermeidet. So fallen grundsätzlich alle vorhergesagten Transmembranbereiche, unlösliche und nicht globuläre Proteine oder Proteindomänen heraus. Weiterhin gilt eine Größenbeschränkung von 15kD für die NMR-Spektroskopie und von 50kD für die Röntgendiffraktion, wobei letztere nur aus Gründen der einfacheren Handhabung hinsichtlich des hohen Durchsatzes gewählt ist. Prinzipiell kennt die Kristallstrukturanalyse keine Größenlimitierung. Die wichtigste Auswahl wird allerdings auf der Grundlage erwarteter medizinischer Relevanz getroffen. Sind Proteine von besonderem medizinischen Interesse, lohnt sich gegebenenfalls auch ein erhöhter Analyseaufwand. Grundsätzlich eignet sich der experimentelle Absatz der „Proteinstrukturfabrik“ natürlich auch für die Strukturanalyse von Proteinen aus anderen Organismen, beispielsweise Bakterien. Mit der hier getroffenen Beschränkung auf menschliche Proteine konzentriert sich das Projekt auf Moleküle, deren Analyse eher schwierig ist, im Umkehrschluss aber

auch den höchsten Erkenntniszugewinn verspricht.

Es folgt nun die erste Weiche, die gewählten Proteine entweder in *Escherichia coli* oder in Hefe rekombinant herzustellen. Der Hauptteil der Proteine wird in *E. coli* synthetisiert. Entstehen Löslichkeitsprobleme oder lassen sich die Gene nicht oder nur mit geringer Expressionsrate klonieren, wird versucht in Hefen zu exprimieren, wobei *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* und *Pichia pastoris* als Alternativen zur Verfügung stehen. Die Expression kann für beide Strukturlösungsverfahren optimiert werden: In Proteine für die NMR-Spektroskopie werden ¹³C- und ¹⁵N-Isotope, in Proteine für Röntgendiffraktion wird Selen an Stelle von Schwefel eingebaut. Bei letzterem wachsen *E. coli* bzw. Hefekulturen auf einem Minimalmedium, welches den Einbau von Selenomethionin anstelle von Methionin erzwingt. Die Proteinpräparationen werden ständig mit einer leistungsfähigen, automatisierten eindimensionalen Gelelektrophorese auf Homogenität untersucht.

Die getesteten Klone werden nun an KMUs übergeben, die die Fermentation im vollautomatischen 5 l-Maßstab und die Reinigung affinitätsmarkierten Proteins im 10-40mg-Maßstab übernehmen. Nach erfolgreicher Etablierung sollte die Kapazität bei fünf bis zehn Proteinen pro Woche liegen.

Der nächste Schritt in der Pipeline ist die biophysikalische Untersuchung durch Fourier-Transform-Infrarotspektroskopie, Mikrokalorimetrie, Fluoreszenz und dynamische Lichtstreuung sowie die Bestimmung des isoelektrischen Punktes. Hier bekommen wir Hinweise zu korrekter Faltung, pH- und Temperaturstabilität und Homogenität der Proteine. Zum Beispiel möchten wir wissen, ob und wie stark die Probe bereits aggregiert ist.

Für die NMR-Spektroskopie stehen am Standort Berlin-Buch zwei 600MHz-Spektrometer mit einer Tieftemperatureinrichtung für Messungen bei –200 °C und ein 750MHz NMR-Spektrometer zur Verfügung. Ein Aufgabenschwerpunkt in diesem Teilprojekt liegt in der Entwicklung und Etablierung einer automatischen Software, welche die langwierige und schwierige Zuordnung der NMR-Resonanzen erheblich beschleunigen soll. Außerdem wird in Zusammenarbeit mit der Pharmaindustrie eine Methode des NMR-gestützten Ligandendesigns weiterentwickelt. Man verbindet hier einen schwach bindenden, von seiner Wechselwirkung mit dem Zielprotein her bekannten Liganden kovalent mit einem potenziellen, ebenfalls für sich schwach bindenden Liganden. Binden tatsächlich beide Liganden an benachbarten Stellen am selben Protein, ist die Bindung entropisch begünstigt und insgesamt um Größenordnungen stärker als die der einzel-

nen Liganden. Diesen Effekt kann man aufgrund charakteristischer Differenzen im Verschiebungsmuster erkennen, analysieren und zur Gewinnung und Optimierung neuer Leitstrukturen heranziehen.

Die Röntgendiffraktion setzt zunächst die Gewinnung von Proteinkristallen voraus, die in eigens für die *sitting drop*-Methode entwickelten Mikrotiterplatten im 96-Well-Format kristallisiert werden. Das Protein wird zunächst auf der Mikrotiterplatte vorgelegt und anschließend über ein 96-kanaliges Spritzensystem mit dem Fällungsmittel versetzt (*screening*). Für die Ausgangsparameter des Screens werden die Daten der biophysikalischen Untersuchung herangezogen. Die Mikrotiterplatte wird in einem automatischen Lagersystem regelmäßig auto-

matisch auf Kristallbildung untersucht. An einem neu konzipierten Messplatz am Berliner Elektronensynchrotron BESSY II werden die Röntgenbeugungsdiagramme dieser Kristalle aufgezeichnet. Nach erfolgreicher Phasierung der Diffraktionsdaten, Anpassung eines atomaren Proteinmodells an die resultierende Elektronendichte, Strukturverfeinerung und -validierung werden die Strukturdaten zu allen Proteinen in einer internen und schließlich auch in einer öffentlichen Datenbank gespeichert. Entsprechend wird mit den Proteinstrukturen aus der NMR-Analyse verfahren.

Nach einer einleitenden Projektphase, die technischen Entwicklungen und dem Aufbau von Forschungsinfrastruktur gewidmet war, steht die „Proteinstrukturfabrik“ jetzt am

Beginn der Erprobungsphase, an deren Ende die Hochdurchsatz-Strukturanalyse stehen soll. Eine erste Struktur eines doppelt affinitätsmarkierten Proteins wurde kürzlich nach Durchlauf fast aller Stationen der „Fabrik“ bestimmt, weitere befinden sich in verschiedenen Stadien des Analyseprozesses. Wenn das Projekt schließlich mit dem geplanten Durchsatz läuft, sollen sich jeweils fünf Proteinstrukturen parallel im Stadium der Strukturberechnung befinden.

Die Vorhersage funktioneller Parameter auf Grundlage der dreidimensionalen Proteinstrukturen bleibt die abschließende Herausforderung des Unternehmens „Proteinstrukturfabrik“. Sie erfordert eine Kombination aller für ein Proteinmolekül relevanter Daten mit der Kristall- oder NMR-Struktur mit den Mitteln der Bioinformatik.

Literatur

- [1] **Vukmirovic, O.G., and Tilghman, S.M.** 2000. Exploring genome space. *Nature* 405: 820-822
- [2] **Blundell, T.L., and Mizuguchi, K.** 2000. Structural genomics: An overview. *Prog. Biophys. Mol. Biol.* 73: 289-295
- [3] **Heinemann, U., Frevert, J., Hofmann, K.-P., Illing, G., Maurer, C., Oschkinat, H., and Saenger, W.** 2000. An integrated approach to structural genomics. *Prog. Biophys. Mol. Biol.* 73: 347-362
- [4] **Terwilliger, T.C.** 2000. Structural genomics in North America. *Nature Struct. Biol.* 7: 935-939
- [5] **Heinemann, U.** 2000. Structural genomics in Europe: Slow start, strong finish? *Nature Struct. Biol.* 7: 940-942
- [6] **Yokoyama, S., Hirota, H., Kigawa, T., Yabuki, T., Shirouza, M., Terada, T., Iot, Y., Matsuo, Y., Kuroda, Y., Nishimura, Y., Kyogoku, Y., Miki, K., Masui, R., and Kuramitsu, S.** 2000. Structural genomics projects in Japan. *Nature Struct. Biol.* 7: 943-945
- [7] **Wiemann, S., Weil, B., Wellenreuther, R., Gassenhuber, J., Glassl, S., Ansorge, W., Böcher, M., Blöcker, H., Bauersachs, S., Blum, H., Lauber, J., Dusterhöft, A., Beyer, A., Köhrer, K., Strack, N., Mewes, H.-W., Ottenwälder, B., Obermaier, B., Tampe, J., Heubner, D., Wambutt, R., Korn, B., Klein, M., and Poustka, A.** 2001. Towards a catalog of human genes and proteins: Sequencing and analysis of 500 novel complete protein coding human cDNAs. *Genome Res.* 11: 422-435

Korrespondenzadressen

Dr. Patrick Umbach

Leitprojektverbund Proteinstrukturfabrik
Heubnerweg 6
D- 14059 Berlin
Tel.: 030-3263 9194
eMail: PSF@FU-Berlin.de

Prof. Dr. Udo Heinemann

Max-Delbrück-Centrum für Molekulare Medizin
Robert-Rössle-Str. 10
D- 13125 Berlin
Tel.: 030-9406 2270
eMail: Heinemann@MDC-Berlin.de
Zweitanschrift:
Institut für Chemie
Kristallographie
Freie Universität Berlin
Takustr. 6
D- 14195 Berlin

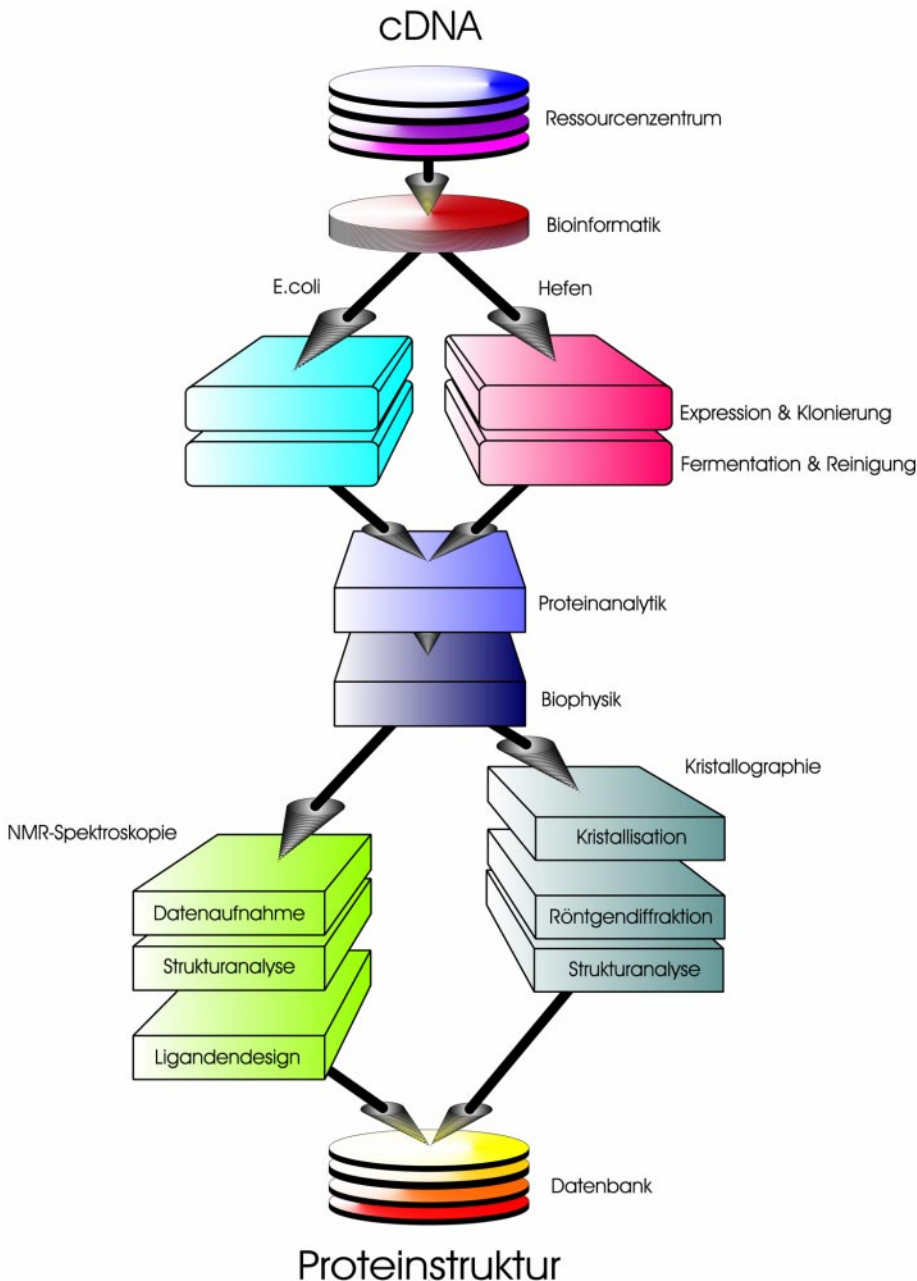


Abb. 1: Weg einer Probe durch die „Proteinstrukturfabrik“ von der cDNA zur dreidimensionalen Proteinstruktur