

Doctoral Thesis ETH No. 18350

**Protein NMR – Cell-Free Expression of Human Proteins and
Structure Determination in Solution**

A dissertation submitted to the

ETH ZÜRICH

for the degree of
Doctor of Sciences

presented by

ERICH MICHEL

Dipl. Natw. ETH
born July 28, 1978
citizen of Lahr, Germany

accepted on the recommendation of

Prof. Dr. Kurt Wüthrich, examiner
Prof. Dr. Andreas Plückthun, co-examiner

2009

Summary

The production of sufficient amounts of human proteins has always been a major bottleneck in structural and functional studies and often leads to an early cessation of many biomedically important research projects, such as the structure determination at atomic resolution of proteins involved in diseases. Based on the information of the molecular shape of a protein, conclusions on the function of the protein can be deduced which will help us to better understand the molecular basis of both diseases and life in general. An efficient method for the production of functional human proteins would therefore provide new perspectives.

In the main part of this thesis, I have investigated the potential of an *E. coli*-based cell-free expression system for producing milligram quantities of soluble, active human proteins. In this context I generated an *E. coli* BL21 (DE3) RIPL-Star strain for the preparation of a cell extract optimized for cell-free expression of eukaryotic proteins. This strain contains an inactive RNase E and additional genes encoding rare-codon tRNAs, thereby combining beneficial features from two different *E. coli* strains. This strain was then used in an optimized protocol for the preparation of S30 extract, which allowed the production of 15 mL quantities of highly active extract within 24 h. To further enhance cell-free protein production, a high-yield cell-free expression vector including an N-terminal GB1-solubility domain was constructed, which increased the yields of soluble human proteins at least two-fold, in particular also for target proteins where the expression level was not limited by protein solubility. An estimate of the expense per mg protein by either batch mode or continuous-exchange mode cell-free synthesis indicated that batch mode expression is about two-fold more economical, and I therefore carried out the cell-free expression in batch mode.

Ten human target proteins with molecular masses between 15 and 25 kDa were selected for production in this cell-free system. I was able to express all ten proteins in soluble form in batch-mode reactions, with yields between 0.18 to 0.86 mg of purified protein per mL reaction mixture. This result is documented with 2D [¹⁵N,¹H]-HSQC spectra of nine of the targets.

Based on this cell-free platform, an approach for mg-production of soluble, active proteins containing disulfide-bonds was then developed, based on redox potential adjustment and

supplementation of the reaction mixture with disulfide bond isomerases. This approach does not require any other modification of the extract, and its redox potential closely resembles the conditions existing in the endoplasmic reticulum, the cellular compartment where disulfide bonds are formed in eukaryotes. The system was systematically optimized by maximizing the soluble yields of the three “test proteins” human doppel, mouse doppel and mouse interleukin-22 containing disulfide bonds, resulting in the production of up to 0.33 milligram of purified protein per milliliter reaction mixture. 2D [¹⁵N,¹H]-HSQC spectra of these proteins confirmed the formation of correctly folded globular domains. These optimized conditions for disulfide bond formation were then applied for soluble expression of a selection of ten human proteins containing up to eight disulfide bonds.

Overall, in this part of the thesis I have generated an *E. coli* cell-free expression protocol for the production of milligram amounts of soluble, active human proteins, including disulfide bond formation under optimized conditions.

In a second part of the thesis, which was pursued during a first phase of my graduate studies in collaboration with Dr. Fred Damberger in the Wüthrich group, I studied the *Bombyx mori* pheromone-binding protein (BmorPBP), which is a carrier protein in silk moths that delivers the hydrophobic pheromone bombykol across the sensillary lymph to the membrane-standing receptor. BmorPBP undergoes a conformational transition from a form containing a cavity suitable for pheromone uptake at neutral pH to a form where the cavity is occupied by a C-terminal α -helix formed by the protein at pH 4.5. This transition suggests that the ligand is ejected to the receptor upon exposure to the reduced pH value near the membrane. We determined the solution structure at pH 6.5 and characterized the pH-dependent conformational changes of a truncated variant, BmorPBP(1–128), which is devoid of the C-terminal helix-forming tetradecapeptide. The structure contains seven well-defined helices 2–12 (α 1a), 16–23 (α 1b), 28–35 (α 2), 46–58 (α 3), 70–80 (α 4), 84–100 (α 5) and 107–124 (α 6) surrounding a large cavity, which provides a rationale for the observation that BmorPBP(1–128) can bind bombykol at neutral pH. Moreover, 2D [¹⁵N,¹H]-HSQC and 3D ¹⁵N-resolved [¹H,¹H]-NOESY spectra indicated that the cavity is also present at pH 4.5, although the residues 8, 11–17, 57–59, 68–70, 72, and 80–82 exhibit conformational exchange. Most of these residues surround a cluster of four histidine

residues (His69, His70, His80, His95), suggesting that protonation of one or more of these residues results in a local destabilization of the polypeptide segment comprised of residues 68–82. In the wild-type protein, this segment undergoes a large conformational change when the pH is reduced to 4.5, so that it can accommodate the C-terminal helix. This suggests that destabilization caused by protonation of histidine residues is the triggering step in the conformational change, which is then followed by formation and insertion of the C-terminal helix.

Zusammenfassung

Die Herstellung ausreichender Mengen an humanen Proteinen hat schon immer einen grossen Engpass in strukturellen und funktionellen Studien dargestellt, welcher oft zu einem frühen Abbruch vieler biomedizinisch wichtiger Forschungsprojekte führt, unter anderem Untersuchungen zur Bestimmung der räumlichen Struktur von Proteinen die mit Krankheiten in Zusammenhang gebracht werden. Anhand der Information über die räumliche Struktur eines Proteins können Rückschlüsse auf dessen Funktion gezogen werden und dieses Wissen wird uns dabei helfen, die molekularen Grundlagen von Krankheiten und allgemeine Prinzipien des Lebens besser zu verstehen. Eine effiziente Methode für die Herstellung ausreichender Mengen an funktionellen menschlichen Proteinen würde daher neue Perspektiven eröffnen.

In dem Hauptteil dieser Doktorarbeit wurde das Potenzial eines *E. coli*-basierten zellfreien Proteinsynthese-Systems für die Produktion von Milligramm-Mengen an löslichen und aktiven humanen Proteinen untersucht. In diesem Zusammenhang wurde ein *E. coli* BL21 (DE3) RIPL-Star Stamm für die Herstellung eines Zellextraktes generiert, der für die Expression eukaryotischer Proteine optimiert ist. Dieser Bakterienstamm enthält eine inaktive Variante der RNase E und zusätzliche Gene, welche tRNA-Moleküle mit seltenen Kodons kodieren, und kombiniert somit vorteilhafte Eigenschaften zweier *E. coli*-Stämme. Dieser Bakterienstamm wurde dann in einem optimierten Protokoll zur Herstellung von Zellextrakt verwendet, welches die Produktion von 15 Milliliter hochaktivem Extrakt innerhalb von 24 Stunden ermöglichte. Weiterhin haben wir einen effizienten Expressionsvektor für zellfreie Proteinsynthese konstruiert, welcher eine N-terminale GB1-Löslichkeitsdomäne enthält. Dieses Konstrukt hat die Ausbeuten an löslichem Protein um mindestens den Faktor 2 erhöht, sogar bei den Proteinen, deren Herstellung nicht durch eine verminderte Löslichkeit des Proteins selbst eingeschränkt war. Eine Kostenschätzung für die Herstellung von einem Milligramm Protein entweder im Batch-Betrieb oder im Dialyse-Betrieb hat angedeutet, dass der Batch-Betrieb ca. zweifach günstiger ist und daher wurde die zellfreie Proteinsynthese im Batch-Betrieb ausgeführt.

Zehn menschliche Proteine mit einer Molekularmasse zwischen 15 und 25 kDa wurden für die zellfreie Herstellung ausgewählt. Alle zehn Proteine konnten in löslicher Form mit Ausbeuten zwischen 0.18 und 0.86 Milligramm aufgereinigtem Protein pro Milliliter

Reaktionsansatz im Batch-Betrieb hergestellt werden. Dieses Ergebnis wurde mit 2D [¹⁵N,¹H]-HSQC Spektren von neun der zehn Proteinen dokumentiert.

Basierend auf diesem zellfreien System wurde dann eine Methode für die Milligramm-Produktion von löslichen und aktiven Proteinen mit Disulfidbrücken entwickelt, welche auf einer Regulierung des Redoxpotenzials und der Zugabe von Disulfidbrücken-Isomerasen zum Reaktionsansatz beruht. Dieses System erfordert keine weiteren Behandlungen des Zell-Extraktes und das Redox-Potenzial gleicht nahezu den Bedingungen des endoplasmatischen Retikulums, dem Ort der Disulfidbrücken-Bildung in Eukaryoten. Das System wurde durch systematische Maximierung der löslichen Ausbeute der drei Disulfidbrücken-enthaltenden Proteine Mensch-Doppel, Maus-Doppel und Maus-Interleukin-22 optimiert, was in der Produktion von bis zu 0.33 Milligramm löslichem Protein pro Milliliter Reaktionsansatz resultierte. 2D [¹⁵N,¹H]-HSQC Spektren dieser Proteine bestätigten die Ausbildung richtig gefalteter globulärer Proteindomänen. Diese optimierten Bedingungen wurden dann zur löslichen Herstellung von zehn ausgewählten menschlichen Proteinen mit bis zu acht Disulfidbrücken angewandt.

Zusammenfassend wurde in diesem Teil der Doktorarbeit ein *E. coli*-basiertes, zellfreies Proteinsynthese-Protokoll für die Produktion von Milligramm-Mengen an löslichen, aktiven menschlichen Proteinen entwickelt, welches auch die Ausbildung von Disulfidbrücken unter optimierten Bedingungen ermöglicht.

In einem zweiten Teil, welcher am Anfang der Doktorarbeit in Zusammenarbeit mit Dr. Fred Damberger aus der Arbeitsgruppe Wüthrich durchgeführt wurde, habe ich das Pheromon-bindende Protein des Seidenspinners *Bombyx mori* (BmorPBP) untersucht, ein Transportprotein, welches ganz gezielt das hydrophobe Pheromon Bombykol durch die Antennenlymphe zum Rezeptorprotein in der Membran befördert. BmorPBP durchläuft eine konformationelle Veränderung von einer Form bei neutralem pH welche einen Hohlraum für die Aufnahme des Pheromons enthält, zu einer Form bei pH 4.5 in welcher dieser Hohlraum von einer C-terminalen α -Helix ausgefüllt ist. Dieser Übergang legt nahe, dass das Pheromon in Gegenwart des reduzierten pH-Wertes in der Nähe der Membran zum Rezeptor herausgestossen wird. Wir haben die Struktur einer Variante von BmorPBP ohne C-terminale α -Helix, BmorPBP(1–128), in Lösung bei pH 6.5 bestimmt und die pH-abhängigen konformationellen Veränderungen charakterisiert. Die Struktur enthält sieben

gut definierte Helices 2–12 (α 1a), 16–23 (α 1b), 28–35 (α 2), 46–58 (α 3), 70–80 (α 4), 84–100 (α 5) und 107–124 (α 6), welche einen grossen Hohlraum umgeben, was in Einklang mit der Beobachtung steht, dass BmorPBP(1–128) Bombykol bei neutralem pH aufnehmen kann. 2D [$^{15}\text{N},^1\text{H}$]-HSQC und 3D ^{15}N -aufgelöste [$^1\text{H},^1\text{H}$]-NOESY Spektren weisen darauf hin, dass der Hohlraum auch bei pH 4.5 ausgebildet ist, obwohl die Aminosäuren 8, 11–17, 57–59, 68–70, 72 und 80–82 konformationellen Austausch aufweisen. Die meisten dieser Aminosäuren umgeben eine Ansammlung von vier Histidinen (His69, His70, His80, His95), was nahe legt, dass die Protonierung von einem oder mehreren Histidinen eine lokale Destabilisierung des Polypeptid-Segmentes bestehend aus Aminosäuren 68–82 hervorruft. Im Wildtyp-Protein durchläuft dieses Segment bei pH 4.5 eine grosse konformationelle Veränderung um die C-terminale α -Helix aufnehmen zu können. Dies deutet darauf hin, dass die Destabilisierung durch Protonierung der Histidine den auslösenden Schritt der konformationellen Veränderung darstellt, dem sich die Bildung und das Einfügen der C-terminalen α -Helix anschliessen.