

DISS. ETH Nr. 19735

Evolvability of Designed Protein Scaffolds  
for Binding Affinity and Catalytic Activity

A dissertation submitted to the

ETH ZÜRICH

for the degree of

Doctor of Sciences

presented by

MAREN BUTZ

Dipl. Natw. ETHZ

born on January 14, 1980

citizen of Teufen AR

Accepted on the recommendation of

Prof. Dr. D. Hilvert, examiner

Prof. Dr. R. Glockshuber, co-examiner

Prof. Dr. K. Johnsson, co-examiner

Zürich, 2011

## Abstract

Proteins are natural macromolecules that exhibit versatile functions as catalysts, receptors, and structural elements in cells. Darwinian evolution has optimized the functions of these molecules over millions of years by repetitive cycles of mutagenesis, selection and amplification. As a consequence of this optimization process, protein catalysts, also called enzymes, are remarkably efficient, accelerating chemical reactions up to 17 orders of magnitude over background with exacting chemo-, regio- and stereoselectively. Protein receptors have been similarly tailored to bind their partners with high affinity and selectivity.

Although proteins perform an enormous number of different functions, natural biocatalysts may not be suitable for specific applications outside their biological context, for example in research, medicine, or industry. They may not even exist at all. Thus, (re)design of specialized proteins for novel functions is desirable, and much modern research has been devoted to the development of techniques to tailor proteins in the laboratory for new tasks.

One seemingly promising approach to the creation of new biocatalysts involves generating antibodies against transition state analogs. Although many catalysts have been generated by this strategy, the activities of all catalytic antibodies reported to date are modest. Thus, further improvements are required. In Chapter 2 a potentially general strategy for improving catalytic antibodies was investigated with a catalyst for a Diels-Alder reaction. Antibody 1E9 was raised against a hexachloronorbornene derivative and accelerates the Diels-Alder reaction between tetrachlorothiophendioxide and N-ethylmaleimide. A single-chain fragment of this antibody was successfully displayed on the surface of yeast cells in functional form, providing a simple link between genotype and phenotype and thus enabling directed evolution. The genes encoding the antibody were mutagenized and 1E9 variants that bound the norbornene transition state analog with higher affinity were identified using fluorescence activated cell sorting (FACS). It was hoped that increased affinity for the transition state analog should translate into improved catalytic activity. After multiple rounds of mutagenesis and screening, 1E9 variants were identified that exhibited extremely slow dissociation rates for the labeled transition state analog when displayed on yeast, even in the presence of a large excess of unlabeled ligand. Nevertheless, these dramatic improvements were not recapitulated when the antibodies were produced in soluble form, and the largest increases in catalytic activity of 1E9 that were achieved were only about 3-fold. A more active mutant of the wild-type 1E9 catalyst, which harbors a single point mutation at the active site, was not a better starting point for the evolution of improved catalysts. Additional cycles of directed evolution, ideally with a high-throughput

assay for catalytic activity, will presumably be necessary to further increase the activity of this catalytic antibody further.

Another approach to create tailor-made proteins is computational design. A number of novel enzymes and protein interaction partners have already been successfully designed. However, as for catalytic antibodies, only moderate activities and affinities have been achieved. In the laboratory of Prof. Brian Kuhlman (University of North Carolina) an artificial binding protein called “spider roll” was designed to bind the human p21-activated kinase 1; the  $K_D$  value for the resulting complex was reported to be 100  $\mu\text{M}$ . In Chapter 3, yeast cell-surface display was used in combination with FACS to increase the affinity of this design for the target kinase, drawing on experience gained from the experiments described in Chapter 2 to optimize catalytic antibody 1E9. Over three rounds of random mutagenesis and screening, the affinity of the displayed protein for the kinase increased roughly 280-fold. Again, however, when the spider roll variants were produced in soluble form, much weaker binding was observed. The  $K_D$  for the best binder was only 7-fold lower than that of the original computational design. While the evolution of spider roll illustrates the potential of improving computational designs by directed evolution, the gain in *in vitro* affinity is still much too small for possible medical applications and further improvements will be needed if this protein is to be used as an agonist or antagonist of kinases. Moreover, detailed biochemical characterization of the selected protein revealed that it is partially disordered. The selection of a partially disordered protein as an improved binding partner for the kinase was unexpected and shows how directed evolution can sometimes provide surprising solutions for specific tasks. Given the difficulty to fulfill all the simultaneous requirements for protein folding, high affinity, and selectivity, directed evolution is likely to remain an important aid to computational design of tight and selective protein-protein interactions for the foreseeable future.

Directed evolution experiments to improve catalytic activity and affinity depend on the availability of powerful screening or selection systems. In contrast to screening, selection systems directly link the selection and amplification steps of the evolution process. Consequently, much larger libraries of mutagenized proteins can be analyzed, increasing the chances of identifying improved variants. For example, the selection system for chorismate mutase (CM) takes advantage of a CM knock-out strain. This enzyme is essential in the shikimate pathway for phenylalanine and tyrosine production. Elimination of the encoding gene causes auxotrophy. As a consequence, cells can only grow if provided with the gene for a functional enzyme. This complementation system has enabled basic research on natural and unnatural CMs. It has been used, for instance, to study catalytic mechanism, elucidate

the sequence constraints on secondary structural elements, generate topologically novel oligomerization states, and even create an active enzyme composed from only 9 amino acid building blocks. Although powerful, the CM selection system nevertheless has difficulty distinguishing between highly active enzymes, since such catalysts already fully satisfy the metabolic needs of the cell. In Chapter 4, a new strategy for increasing selection stringency by limiting intracellular catalyst concentration is presented. Intracellular protein concentration can be tightly regulated over a very wide range by simultaneously controlling protein production through gene expression from the inducible tetracycline promoter and controlling its lifetime in the cell by appending a suitable peptide tag that directs it to housekeeping proteases. Here, the N-terminal RepA degradation tag (MNQSFISDILYADIES) was investigated and compared to the C-terminal SsrA tag (AANDENYALAA). The RepA tag successfully decreases CM concentrations, although it is less efficient than the SsrA tag. Nevertheless, the RepA tag has a fundamental advantage over the SsrA tag for selection experiments. Specifically, it is not subject to artifacts arising from C-terminal deletions or stop codon mutations that lead to premature truncation and elimination of the tag, restoring high enzyme concentrations and causing loss of selection stringency. This system was then used to evolve a molten globular CM (mMjCM) that could not be evolved before. Three rounds of mutagenesis and selection allowed complementing clones to grow under the most stringent condition (no transcription induction). The best selected variant catalyzed the chorismate-to-prephenate reaction 4-times faster than the starting enzyme. Additionally, it was less molten globular than the template mMjCM as judged by more cooperative thermal denaturation, less ANS binding, and greater peak dispersion in its <sup>1</sup>H-NMR spectrum. It should be possible to extend this approach, which couples a tunable promoter with targeted protein degradation, to adjust selection stringency of many genetic selection systems.

In Chapter 5 alternative strategies for increasing selection stringency by decreasing the intracellular CM concentration were evaluated. These include cytosolic production of anti-sense mRNA to inhibit gene transcription and utilization of impaired ribosome binding sites to slow down protein translation. The anti-sense mRNA strategy took advantage of a T7 promoter site located downstream of the CM gene. In a CM knock-out strain that produces T7 polymerase, basal level transcription of the anti-sense sequence from the T7 promoter resulted in drastic reduction in CM levels in the cell. Although not yet applied in a directed evolution experiment, this approach has the potential to afford extremely high selection stringency. To modulate translation, novel ribosome binding sites (RBSs) were identified. The extensively mutagenized T7 RBS was cloned upstream of the gene for HIV protease, which is

cytotoxic, and variants were selected that enable cell growth even when transcription of the HIV protease is induced. The impaired RBSs were tested for their ability to reduce intracellular CM levels. Although one of the modified RBSs did indeed reduce protein concentration, directed evolution experiments with it did not yield variants of the molten globular CM that were better than those already isolated in the RepA experiments.

Many studies have established the utility of directed evolution for improving the activities of catalysts and the affinities of receptors. By combining this approach with catalytic antibody technology and/or computational design as described in this thesis, it may be possible to create tailored proteins for many interesting applications. Since new selection and screening systems are being continually developed, including systems whose stringencies can be adjusted and fine-tuned, it is probably just a question of time until artificial proteins can be created that perform their functions with the same efficiencies and specificities as their natural counterparts.

## Zusammenfassung

Proteine sind natürliche Makromoleküle, die verschiedenste Aufgaben in der Zelle erfüllen. Sie können als Katalysatoren, Rezeptoren und strukturgebende Elemente fungieren. Über Millionen von Jahren hat die darwinistische Evolution mit ihren repetitiven Zyklen, bestehend aus Mutagenese, Selektion und Vervielfältigung, die Funktion dieser Moleküle optimiert. Dieser Optimierungsprozess liefert bemerkenswert effiziente Proteinkatalysatoren, so genannte Enzyme. Sie sind fähig chemische Reaktionen bis zu  $10^{17}$ -fach relativ zur unkatalysierten Hintergrundreaktion zu beschleunigen und das mit einer exakten Chemo-, Regio- und Stereoselektivität. Ähnlich wurden Proteinrezeptoren massgeschneidert, damit sie ihre Bindungspartner mit hoher Affinität und Selektivität binden.

Trotz der enormen Anzahl verschiedenster Funktionen können natürliche Biokatalysatoren für spezifische Anwendungen ausserhalb ihres natürlichen Kontextes ungeeignet sein, zum Beispiel in Forschung, Medizin oder Industrie. Für manche Anwendungen existieren zudem keine Biokatalysatoren. Es wäre daher wünschenswert, spezialisierte Proteine für neue Aufgaben zu designen oder umzugestalten. Ein Grossteil der modernen Forschung hat sich damit befasst, Techniken zu entwickeln, um im Labor Proteine für neue Aufgaben und Funktionen herzustellen.

Ein scheinbar vielversprechender Ansatz, um neue Enzyme zu gestalten, beruht auf der Generierung von Antikörpern gegen Übergangszustandsanaloge. Obwohl diese Strategie viele Katalysatoren hervorbrachte, ist deren Aktivität bis jetzt moderat. Es ist daher nötig diese Proteine weiter zu verbessern. In Kapitel 2 wurde eine potentiel allgemeingültige Strategie zur Verbesserung von katalytischen Antikörpern an dem Beispiel eines Immunoglobulins, der eine Diels-Alder Reaktion katalysiert, untersucht. Als Antigen für den Antikörper 1E9 wurde ein Hexachloronorboren-derivat verwendet. Dies ermöglichte dem Immunoglobulin die Reaktion zwischen Tetrachlorothiophendioxid und N-Ethylmaleimid zu katalysieren. Für die gerichtete Evolution dieses Katalysators, wurde der Antikörper im Einzelstrangformat auf der Zelloberfläche von Hefe erfolgreich präsentiert, was die nötige Verknüpfung zwischen Genotyp und Phenotyp lieferte. Das Gen des Antikörpers wurde daraufhin mutagenisiert. Varianten von 1E9, die das Norbornen-Übergangszustandsanalog mit hoher Affinität binden konnten, wurden mittels Fluoreszenz-aktivierter-Durchflusszytometrie (FACS) identifiziert. Das Prinzip beruht darauf, dass eine höhere Affinität für das Übergangszustandsanalog zu einer verbesserten katalytischen Aktivität führen sollte. Nach einigen Runden der Mutagenese und Selektion wurden 1E9-Varianten angereichert, die eine stark verlangsamte Dissoziationsrate auf der Zelloberfläche aufwiesen auch in Gegenwart von einem Überschuss an nicht-markierten Liganden. Die aufgereinigten Antikörper konnten diese drastische Verbesserung jedoch nicht reproduzieren. Daher war eine dreifach höhere Aktivität die maximal erreichte Aktivitätssteigerung. Eine Mutante des Wildtypkatalysators mit einer Einfachmutation in der aktiven Tasche wurde als weiterer Startpunkt für ein gerichtetes Evolutionsexperiment verwendet, erzielte aber auch keine

besseren Resultate. Weitere Evolutionszyklen, idealer Weise mit einem auf Aktivität basierenden Test mit hohem Durchsatz, könnten die nötige Aktivitätsverbesserung für diesen katalytischen Antikörper liefern.

Ein anderer Ansatz, um Proteine mit neuen Funktionen auszustatten, beruht auf rechnergestütztem Design. Wie für katalytische Antikörper wurden bisher nur mässige Aktivitäten und Affinitäten erreicht. In dem Labor von Prof. Brian Kuhlman (University of North Carolina) wurde ein künstliches Bindungsprotein entworfen, das „Spider Roll“, das die menschliche p21-aktivierte Kinase1 bindet. Ein  $K_D$  Wert für den resultierenden Komplex wurde mit 100  $\mu\text{M}$  angegeben. In Kapitel 3 wurde die gewonnene Erfahrung aus Kapitel 2 genutzt, um die Affinität des künstlichen Bindungsproteins für die Kinase zu erhöhen, indem das mutagenisierte Proteindesign auf der Zelloberfläche präsentiert wurde. Die besten Varianten wurden mit FACS selektioniert. Nach drei Mutageneserunden und den dazugehörigen Selektionen wurde eine Variante mit einer 280-fach erhöhten Affinität identifiziert. Aber auch bei diesem Experiment hatte das Protein in Lösung eine geringere Affinität als auf der Zelloberfläche, was die 280-fache Verbesserung relativ zum rechengestützten Originaldesign auf das Siebenfache reduzierte. Auch wenn die erreichte *in vitro* Affinität noch weitere Verbesserungen benötigt, um mögliche medizinische Anwendungen als Kinaseagonisten oder –antagonisten zu finden, zeigt das Experiment trotzdem, dass mit Hilfe von gerichteter Evolution Affinitäten von rechengestützten Designs erhöht werden können. Zusätzlich offenbarte die biochemische Untersuchung der selektionierten Variante, dass dieses Protein teilweise unstrukturiert war. Die Selektion eines partiell unstrukturierten Proteins als verbesserter Kinasebindungspartner war unerwartet und zeigt, mit was für erstaunlichen Lösungen gerichtete Evolution aufwarten kann, wenn sie mit speziellen Problemen konfrontiert wird. Angesichts der gleichzeitig gestellten Ansprüche von Proteinfaltung, Affinität und Selektivität, liegt es Nahe das gerichtete Evolution eine wichtige Stütze der rechengestützten Designs sein wird, um starke, selektive Protein-Protein-Interaktionen zu erhalten.

Experimente, die gerichtete Evolution verwenden, um katalytische Aktivitäten und Affinitäten zu verbessern, hängen von der Verfügbarkeit von leistungsfähigen Screening- oder Selektionssystemen ab. Im Gegensatz zu Screeningsystemen verknüpfen Selektionssysteme den Selektions- und Vervielfältigungsschritt des Evolutionsprozesses. Hieraus resultiert, dass grössere Bibliotheken von mutagenisierten Proteinen analysiert werden können, was die Wahrscheinlichkeit erhöht eine verbesserte Variante zu identifizieren. Das Selektionssystem für die Chorismat-Mutase (CM) ist ein typisches Beispiel. CM ist essenziell für die Produktion von Phenylalanin und Tyrosin im Shikimat-biosyntheseweg. In dem Knockout-Stamm, auf welchem das Selektionssystem beruht, wurde das CM-Gen beseitigt, was zu einer Auxothrophie führt. Zellen, die kein Gen für ein funktionelles Enzyme enthalten, können daher nicht wachsen. Dieses Komplementationssystem ermöglichte Grundlagenforschung an natürlichen und

unnatürlichen CM. Zum Beispiel wurde es verwendet, um den katalytischen Mechanismus zu studieren, Sequenzeinschränkungen für sekundäre Strukturelemente zu erforschen, topologisch neue Oligomerisationszustände zu erhalten und sogar um ein aktives Enzyme zu bauen bestehend aus nur neun Aminosäurebausteinen. Auch wenn dieses Selektionssystem sehr leistungsfähig ist, hat es doch Schwierigkeiten Enzyme mit hohen Aktivitäten zu unterscheiden, da diese die metabolischen Bedürfnisse der Zelle hinreichend erfüllen. In Kapitel 4 wird daher ein neue Strategie vorgestellt, um die Selektionsstringenz des Systems zu erhöhen, indem die intrazelluläre Enzymkonzentration reduziert wird und somit für die Zelle limitierend wird. Die intrazelluläre Proteinkonzentration ist über einen grossen Bereich streng regulierbar mittels einer gleichzeitigen Kontrolle der Proteinproduktion sowie des Proteinabbaus. Die Proteinproduktion wird gesteuert über die Genexpression, die mit dem Tetracyclinpromotor reguliert werden kann. Der Proteinabbau wird erreicht, indem das Protein mit einem Abbaumarker versehen wird, die es zu speziellen zellulären Proteasen lenkt. Dies reduziert die Halbwertszeit des Proteins in der Zelle. In dem Projekt wurde der N-terminale Abbaumarker RepA (MNQSFISDILYADIES) erforscht und verglichen mit dem C-terminalen SsrA-Abbaumarker (AANDENYALAA). Der RepA-Marker reduzierte die intrazelluläre Enzymkonzentration erfolgreich, erreichte aber nicht die Effizienz des SsrA-Markers. Trotzdem hat der RepA-Marker verglichen mit dem SsrA-Marker einen fundamentalen Vorteil für Evolutionsexperimente. Mit dem RepA-Marker wurden keine Selektionsartefakte gefunden, die aber mit dem SsrA-Marker auftauchen. Die C-terminale Position SsrA-Markers erlaubt, dass Deletions- und Stopp-Kodonmutationen den Abbaumarker eliminieren. Dies führt zu erhöhten Enzymkonzentrationen, was wiederum eine Reduktion der Selektionsstringenz zur Folge hat. Da der RepA-Marker keine solchen Artefakte hervorbringt, wurde er verwendet, um eine CM mit Eigenschaften eines „molten globules“ (mMjCM) zu verbessern, was mit dem vorherigen Selektionssystem nicht möglich war. Drei Mutagenese- und Selektionsrunden brachten komplementierende Klone hervor, die unter den stringentesten Bedingungen wachsen konnten (keine Transkriptionsinduktion). Die beste, selektierte Variante katalysierte die Umwandlung von Chorismat zu Prephenat viermal schneller als das Ausgangsenzym. Gleichzeitig besass sie weniger ausgeprägte Eigenschaften eines „molten globules“. Dies wurde begründet mit der grösseren Kooperativität bei der Temperatur induzierten Proteinfaltung, der geringeren ANS-Bindung und der stärkeren Peakdispersion in dem <sup>1</sup>H-NMR-Spektrum dieser Variante verglichen mit mMjCM. Es sollte daher möglich sein, diesen Ansatz, bestehend aus regulierbaren Promoter und einem gezielten Proteinabbau, auf andere Selektionssysteme auszuweiten, die eine erhöhte Selektionsstringenz benötigen.

In Kapitel 5 wurden weitere alternative Strategien getestet, um die Stringenz durch eine gezielte Reduktion der intrazellulären CM-Konzentration zu erhöhen. Dies beinhaltete die zytosolische Produktion einer Antisense-mRNA, um die Gentranskription zu behindern, und der Gebrauch einer beeinträchtigten Ribosomenbindungsstelle, die die Translation verlangsamen sollte. Die Strategie der Antisense-mRNA nutzte den T7 Promoter der



stromabwärts des CM-Gens lokalisiert ist. In einem CM-Knockout-Zellstamm der T7 Polymerase produziert, reichte die basale Transkription aus, um genügend Antisense-mRNA zu generieren, dass die CM-Konzentration in der Zelle stark verringert werden konnte. Dieser Ansatz ist vermutlich in der Lage, die Selektionsstringenz stark zu erhöhen. Bisher wurde er aber noch an keinem gerichteten Evolutionsexperiment getestet. Für den zweiten Ansatz wurden neue Ribosomenbindungsstellen (RBS) selektioniert, die die Translationseffizienz beeinflussen. Stromaufwärts eines HIV-Protease-Gens wurde eine stark mutagenisierte T7 RBS eingefügt. Da die HIV-Protease für die Zelle toxisch ist, können bei induzierter Proteintranskription nur Varianten mit schlechter Translation überleben. Die dadurch selektionierten beeinträchtigten RBS wurden daraufhin getestet, ob sie auch die CM-Konzentration verringern können. Obwohl eine der veränderte RBS die Proteinkonzentration tatsächlich verringerte, konnte ein Evolutionsexperiment der besten selektionierten CM der RepA-Selektion keine verbesserte Variante hervorbringen.

Viele Studien haben den Nutzen der gerichteten Evolution zur Verbesserung der Aktivitäten von Enzymen und der Affinitäten von Rezeptoren bestätigt. Wie in dieser Arbeit gezeigt wird, ist die Kombination von gerichteter Evolution mit katalytischen Antikörpern und/oder rechnergestützte Designs vielleicht in der Lage, massgeschneiderte Proteine für vielerlei interessante Anwendungen zu generieren. Da immer neue Screening- und Selektionssysteme entwickelt werden, Systeme eingeschlossen, deren Stringenz verändert und kontrolliert werden können, ist es wohl nur noch eine Frage der Zeit, bis künstliche Proteine erschaffen werden, die ihre Funktion mit gleicher Effizienz wie auch Spezifität ausführen wie ihre natürlichen Vorbilder.