Research News

Protein-fold evolution in the test tube

Christiane Schaffitzel and Andreas Plückthun

Currently, the combination of library selection and directed evolution is the most powerful approach for finding proteins with novel folds or functions. In the past, most studies concentrated either on protein scaffolds with a given fold or on short peptides. With the recent development of potent *in vitro* selection and evolution techniques, the screening of much larger sequence space is possible, allowing for the *de novo* generation of proteins.

Although it is still unclear how frequently proteins with a certain 'fold' and/or affinity to a given target molecule will be identified among peptides with a random sequence, it seems obvious that both frequencies are small and their identification thus requires large libraries as a starting point. For this reason, *in vitro* selection technologies are uniquely suited to find such rare molecules. Anthony Keefe and Jack Szostak recently published an insightful study describing the *in vitro* selection of ATP-binding molecules from a completely randomized peptide library, containing peptides of ~80 amino acids in length[1].

Principles of *in vitro* selection

In vitro selection technologies, by which the proteins are made using cell-free translation, have the great advantage that the size of the library to be screened is not restricted by transformation efficiency. In all other technologies, the proteins are produced within cells, even if the protein is later secreted, such as in phage display[2]. Transformation of the initial library, and of any subsequently created library (e.g. after rounds of *in vitro* mutagenesis), into cells limits the library size achievable in practice to 10^9–10^{10} members. In particular, this makes evolution experiments; that is, repeated cycles of diversification by *in vitro* mutagenesis and *in vivo* selection, very laborious. This limitation is overcome by *in vitro* selection techniques in which the protein molecules of interest are physically linked to their encoding mRNA after *in vitro* translation. Two related technologies (Fig. 1) have been developed to reduce this concept to practice.

In ribosome display[3,4], the link between phenotype and genotype is non-covalent and is achieved during *in vitro* translation by stalling the translating ribosome at the end of the mRNA, which lacks a stop codon (Fig. 1). The encoded protein is not released and can fold into its correct structure while attached to the ribosome. The resulting ternary complexes of mRNA, ribosome and functional protein are directly used for selection against an immobilized target (these ternary complexes are stable for days under appropriate conditions). The mRNA from bound ribosomal complexes is then recovered by dissociation of the complexes with EDTA and amplified by reverse-transcription–polymerase chain reaction (RT–PCR).

The Szostak laboratory developed a technique, mRNA display, that applies a related but somewhat different approach to link phenotype and genotype. In mRNA display[5,6], a DNA linker attached to puromycin is first fused to the 3′ end of mRNA. Subsequently, the protein is translated *in vitro* and the ribosome stalls at the RNA–DNA junction. The puromycin moiety, which mimics aminoacyl-tRNA, then enters the ribosomal A site and accepts the nascent polypeptide. The translated protein is thus covalently linked to its encoding mRNA, and the fused molecules are purified. If desired, the mRNA can be converted into a heteroduplex with cDNA by reverse transcription to minimize the ligand-binding properties of the mRNA itself[7]. The fused molecules can then be used for subsequent selection experiments[5]. After binding to the target, the fused molecules are eluted and the mRNA is then recovered by RT–PCR.

Key advantages: large libraries and built-in evolution

Both mRNA display and ribosome display, although simple in concept, are technically demanding[6,8]. Notwithstanding, both offer significant advantages over conventional *in vivo* selection techniques. First, they allow screening of very large initial libraries ($>10^{12}$ members), the size of which, theoretically, is only limited by the concentration of mRNA and active ribosomes. Increasing the library size; that is, searching a larger sequence space, already leads to a significant improvement in the average affinity of peptides selected against protein targets[3,9,10]. A second significant advantage of these *in vitro* selection techniques is the existence of a powerful built-in evolution process. If low-fidelity polymerases such as Taq polymerase are used to amplify the pool after each selection, mutations are automatically introduced. Thus, diversity can be augmented at will. Although most of the mutations are neutral or detrimental, a small fraction can improve the affinity of the protein for the target, and/or the folding or stability of a member, thereby yielding an advantage for this mutant in the next selection cycle. Thus, proteins can be selected that were not present in the initial library or, in other words, the sequence space sampled is larger than that in the initial pool. In fact, when proteins selected by ribosome display were compared to their respective progenitors, it was found that their target affinities had improved by up to 50-fold[11,12].

In addition to diversification by low-fidelity polymerases, PCR mutagenesis[13] or DNA shuffling[14] can easily be incorporated into the *in vitro* display protocol. In this way, the selection and evolution process has been successfully tailored to improve the stabilities and affinities of antibody single-chain fragments[15]. With ribosome display, single-chain antibody fragments with picomolar affinity have been selected against a variety of targets, including small molecules[15], peptides[11], proteins[12] and unusual DNA conformations[16].

Selecting ATP-binding, structured peptides

In a recent issue of *Nature*, the Szostak laboratory reported the selection of ATP-binding motifs from a library of 6×10^{12} peptides, with a length of 80 amino acids of completely random sequence[1,17], using mRNA display. After nine selection cycles,

 0968-0004/01/$ – see front matter PII: S0968-0004(01)01959-4

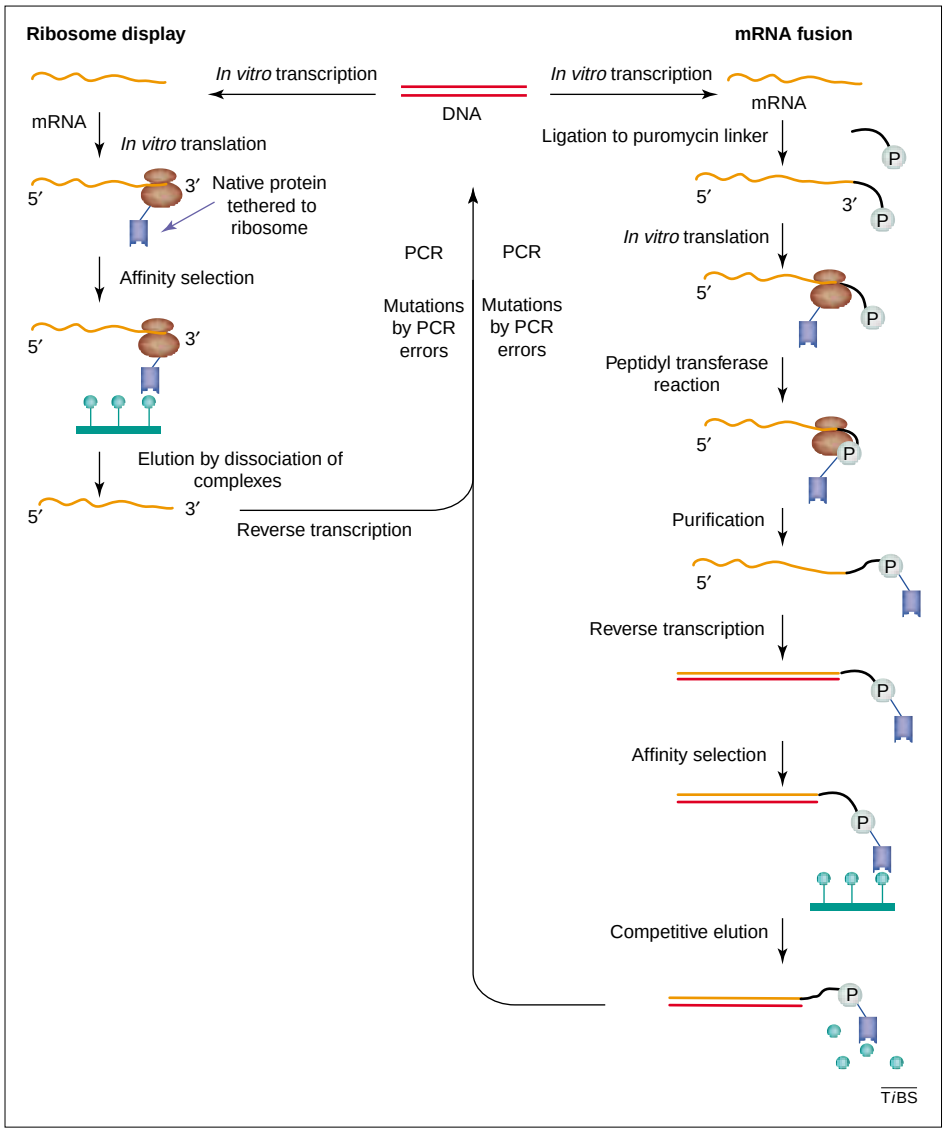

Fig. 1. Two alternative strategies for coupling phenotype and genotype *in vitro*. (Left) Ribosome display. The DNA encoding the library is transcribed *in vitro*. The resulting mRNA lacks a stop codon, giving rise to linked mRNA–ribosome–protein complexes in *in vitro* translation. These can be stabilized and directly used for selection on the immobilized target. The mRNA incorporated in bound complexes is eluted and purified. Reverse transcription–polymerase chain reaction (RT–PCR) can introduce mutations and yields a DNA pool enriched for binders that can be used for the next iteration. (Right) mRNA display. Covalent RNA–protein complexes can be generated by ligation of a DNA-puromycin linker to the *in vitro* transcribed mRNA. The mRNA is translated *in vitro*, and the ribosome stalls at the RNA–DNA junction. Puromycin then binds to the ribosomal A-site. The nascent polypeptide is thereby transferred to puromycin. The resulting covalently linked mRNA–peptide complex is isolated, reverse transcribed and used for selection experiments. Bound complexes are eluted and amplified by PCR.

the pool showed an enrichment of ATP-binding peptides. However, a further enrichment of specific binding peptides could not be achieved in subsequent rounds of selection. The authors suggested that this was a consequence of conformational heterogeneity resulting from inefficient folding of the peptides. At this stage, the pool consisted of four families of ATP binding peptides with unrelated sequence. To overcome the problem of conformational heterogeneity, the authors elegantly combined mRNA display with mutagenesis for three rounds, with the aim of increasing the diversity of the pool again. After a further six rounds of mRNA display, the pool was composed entirely of one family of ATP-binding peptides with related sequence. By this directed evolution approach, further enrichment of ATP-binding peptides could be achieved, and the portion of peptides binding ATP in the selected pool could be increased from 6% (before mutagenesis) to 35% (after mutagenesis).

However, the selected peptides were not themselves sufficiently soluble to be fully characterized and had to be expressed as fusion proteins to maltose-binding protein to increase their solubility. The best peptide selected had improved its affinity for ATP 300-fold, having a K_D value of 100 nM compared with the peptide selected before mutagenesis that bound ATP with only 30 μM affinity. This peptide interacts with several distinct regions of the ATP molecule, indicating the presence of a binding pocket requiring specific folding of the protein. Sequence analysis did not reveal any significant homology of the selected peptide with any known protein. However, two CxxC sequence motifs were identified. One stoichiometric equivalent of Zn^{2+} was also found, and the CxxC motifs probably participate in Zn^{2+} coordination. The presence of Zn^{2+} was essential for ATP binding of the selected protein motif and the authors suggest that metal ion coordination might have been one of the simplest ways of 'developing' a defined protein fold. It is clear that elucidation of the folded structure of this protein motif at atomic resolution will be very important for substantiating decisively the central claims of this study.

The work carried out by Keefe and Szostak was the first to achieve *in vitro* generation of peptides, from a random pool, that are sufficiently structured to bind a small molecule[1]. In previous experiments, linear libraries of peptides only 6–38 amino acids in length were used for selection[18,19]. Such libraries are evidently useful for generating high-affinity ligands to proteins with a defined binding pocket, and the selected peptides were often found to bind the ligand-binding site of the protein target despite there being no obvious selection pressure for this[20,21]. Against protein targets that do not have a defined groove, high-affinity binders have only been selected by introduction of constraints that promote a discrete structure in solution through, for instance, disulfide bond formation[19,22]. However, these approaches, mostly conducted by phage display, have not yet successfully yielded an intrinsically stable protein fold from a random peptide library with the ability to specifically recognize and bind a small molecule with reasonable affinity, a feat now achieved by Keefe and Szostak[1].

In conclusion, directed *in vitro* evolution techniques have clearly proven their suitability as powerful tools for

de novo protein generation through screening sufficiently large sequence spaces and by repeated diversification and selection steps, mimicking the principle of Darwinian evolution *in vitro*.

References

1 Keefe, A.D. and Szostak, J.W. (2001) Functional proteins from a random-sequence library. *Nature* 410, 715–718
2 Winter, G. *et al.* (1994) Making antibodies by phage display technology. *Annu. Rev. Immunol.* 12, 433–455
3 Mattheakis, L.C. *et al.* (1994) An *in vitro* polysome display system for identifying ligands from very large peptide libraries. *Proc. Natl. Acad. Sci. U. S. A.* 91, 9022–9026
4 Hanes, J. and Plückthun, A. (1997) *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. U. S. A.* 94, 4937–4942
5 Roberts, R.W. and Szostak, J.W. (1997) RNA–peptide fusions for the *in vitro* selection of peptides and proteins. *Proc. Natl. Acad. Sci. U. S. A.* 94, 12297–12302
6 Liu, R. *et al.* (2000) Optimized synthesis of RNA–protein fusions for *in vitro* protein selection. *Methods Enzymol.* 318, 268–293
7 Gold, L. *et al.* (1995) Diversity of oligonucleotide functions. *Annu. Rev. Biochem.* 64, 763–797
8 Hanes, J. *et al.* (2000) Selecting and evolving functional proteins *in vitro* by ribosome display. *Methods Enzymol.* 328, 404–430
9 Gersuk, G.M. *et al.* (1997) High-affinity peptide ligands to prostate-specific antigen identified by polysome selection. *Biochem. Biophys. Res. Commun.* 232, 578–582
10 Wilson, D.S. *et al.* (2001) The use of mRNA display to select high-affinity protein-binding peptides. *Proc. Natl. Acad. Sci. U. S. A.* 98, 3750–3755
11 Hanes, J. *et al.* (1998) Ribosome display efficiently selects and evolves high-affinity antibodies *in vitro* from immune libraries. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14130–14135
12 Hanes, J. *et al.* (2000) Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nat. Biotechnol.* 18, 1287–1292
13 Zaccolo, M. *et al.* (1996) An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *J. Mol. Biol.* 255, 589–603
14 Stemmer, W.P. (1994) Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* 370, 389–391
15 Jermutus, L. *et al.* (2001) Tailoring *in vitro* evolution for protein affinity or stability. *Proc. Natl. Acad. Sci. U. S. A.* 98, 75–80
16 Schaffitzel, C. *et al.* (2001) *In vitro* generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylonychia lemnae* macronuclei. *Proc. Natl. Acad. Sci. U. S. A.* 98, 8572–8577
17 Cho, G. *et al.* (2000) Constructing high complexity synthetic libraries of long ORFs using *in vitro* selection. *J. Mol. Biol.* 297, 309–319
18 Scott, J.K. and Smith, G.P. (1990) Searching for peptide ligands with an epitope library. *Science* 249, 386–390
19 Sidhu, S.S. *et al.* (2000) Phage display for selection of novel binding peptides. *Methods Enzymol.* 328, 333–363
20 Kay, B.K. *et al.* (1993) An M13 phage library displaying random 38-amino-acid peptides as a source of novel sequences with affinity to selected targets. *Gene* 128, 59–65
21 Clackson, T. and Wells, J.A. (1994) *In vitro* selection from protein and peptide libraries. *Trends Biotechnol.* 12, 173–184
22 Dennis, M.S. *et al.* (2000) Peptide exosite inhibitors of factor VIIa as anticoagulants. *Nature* 404, 465–470

Christiane Schaffitzel
Andreas Plückthun*
Biochemisches Institut, Winterthurerstr. 190, CH-8057 Zürich, Switzerland.
*e-mail: plueckthun@biocfebs.unizh.ch

Protein Sequence Motif

The following two protein sequence motif articles were submitted simultaneously by independent research groups. As the articles report similar findings, they have been published back-to-back.

The CHASE domain: a predicted ligand-binding module in plant cytokinin receptors and other eukaryotic and bacterial receptors

Vivek Anantharaman and L. Aravind

A novel, 200–230 amino acid extracellular domain was identified in the plant cytokinin receptor Cre1, in the receptor-histidine kinase DhkA and the adenylyl cyclase Acg from the slime mold *Dictyostelium discoideum*, and in a variety of other receptor-like proteins from bacteria and eukaryotes. The domain is predicted to bind diverse low molecular weight ligands, such as the cytokinin-like adenine derivatives or peptides, and mediate signal transduction through the respective receptors.

A diverse set of secreted small molecules function as first messengers in both prokaryotic and eukaryotic signaling systems. Numerous receptors are involved in recognizing and propagating the signals delivered by these first messengers[1]. The most common type of small molecule receptors that have been extensively characterized are the seven-transmembrane receptors found in animals and fungi[1]. Several studies have also revealed another widespread architectural principle in transmembrane receptors for first messengers: a globular extracellular ligand-sensing domain fused to a membrane-spanning segment and intracellular signal-transmitting domains[2–4]. Examples of such extracellular, low molecular weight, ligand-binding domains are the TAR/TAP type α-helical domains of the chemotaxis receptors[3], the periplasmic binding protein domain[4] and the CACHE domain[2]. Here, we report a previously undefined extracellular domain that is predicted to mediate the interactions of a variety of bacterial and eukaryotic receptors with both small-peptide and non-peptide ligands.

Identification of the CHASE domain

To gain a better understanding of the unifying principles of receptor architecture and action, we undertook a systematic computational investigation of receptors whose extracellular regions lacked previously known domains. One such receptor is the recently characterized cytokinin receptor (Cre1), which transduces the signal for cell