Combinatorial Approaches To Novel Proteins

Tomoaki Matsuura,[a] Andreas Ernst,[b] David L. Zechel,[b] and Andreas Plückthun*[b]

Introduction

The number of naturally occurring protein sequences is merely a tiny fraction of all possible protein sequences, hence there is little doubt that nature has not explored all of protein-sequence space. In general, proteins found in nature are compactly folded, resistant to proteolysis, rich in secondary structure, exhibit cooperative unfolding behavior, and possess a defined oligomerization state. In contrast, non-natural polypeptide sequences very rarely possess these properties. Therefore, it is clear that only a small fraction of all possible polypeptide sequences will exhibit properties similar to natural proteins, but there is currently no estimate, not even of the order of magnitude, of how frequently sequences with native-like properties would occur.

The number of protein folds (the three-dimensional arrangement of secondary structures) found in nature is most likely limited.[1] Even though the exact number is a matter of debate[1–4] and is influenced by how narrow the bins that define a particular fold are chosen, the homology between genomes and the presence of sequence families with many members automatically put a cap on this number. The number of folds may eventually be more accurately estimated when the primary sequence of a protein can be used to predict its structure. This, however, is a challenge that easily equals the discovery and creation of new folds in terms of difficulty.

To shed light on these fundamental questions researchers have begun to explore new regions of sequence space with the intent of creating and/or discovering novel, non-natural proteins. There are two approaches one can consider. The first mimics the evolution of natural proteins in that a particular function (e.g., ligand binding) is used to select stably folded proteins. The second approach, which is not likely to have a parallel in nature, selects for compactly folded and stable proteins without a requirement for function (e.g., stability to proteolysis, denaturants, hydrophobicity, temperature etc.). It does have the great advantage, however, that there are probably far more solutions than in selections based on function. In either approach the following question is posited: is it possible to find proteins with a stable fold in unexplored regions of sequence space and, if so, what fraction of all sequences possess these properties?

One can address these questions also by computational de novo protein design. Computational approaches were devised to find a sequence which is compatible with one predetermined fold, usually a natural one,[5, 6] even though a few non-natural ones have been reported.[7, 8] In such calculations, more variants can be tested than possible in experimental libraries, as long as simple energy and scoring functions are being used. Nevertheless, computational approaches have also specifically been devised to find novel folds in sequence space by stacking small building blocks (α-helices and β-strands).[9] It was shown that naturally occurring four helix bundles could be reproduced and even new four helix bundle folds could be identified which are suggested to have a high "designability" (several unrelated sequences can adopt the same fold).[10]

We will, however, not discuss the algorithms or computational strategies but we will instead concentrate on the experimental approaches to generate stable and folded proteins from a pool of sequences. Starting from the development of highly diverse polypeptide libraries containing potentially meaningful sequences, we will focus also on selection techniques suitable to select proteins on the basis of their biophysical properties.

Random Polypeptide Libraries

Creating a random polypeptide sequence library and characterizing the biophysical properties of the proteins contained within is one method to explore sequence space. For more than a decade researchers have worked with random polypeptide libraries (Figure 1a), with the aim of obtaining meaningful (functional and/or folded) sequences.[11–19]

Recently, Keefe and Szostak[15] achieved a major step forward in this field by using mRNA display to select functional 80-residue proteins from a random polypeptide library with a diversity of 6×10^{12}. Such large libraries are only accessible to in vitro selection systems, such as the puromycin-based selection used by Roberts and Szostak[20] or ribosome display (see below).[21, 22] Of the originally selected four peptides, one bound

[a] *Dr. T. Matsuura*
Department of Bioinformatics Science
Graduate School of Information and Science Technology
Osaka University
and PRESTO, Japan Science and Technology Corporation
2 – 1 Yamadaoka, Suita, Osaka 565 – 0871 (Japan)

[b] *A. Ernst, Dr. D. L. Zechel, Prof. Dr. A. Plückthun*
Biochemisches Institut, Universität Zürich
Winterthurerstrasse 190, 8057 Zürich (Switzerland)
Fax: (+41) 1-635-5712
E-mail: plueckthun@bioc.unizh.ch

 DOI: 10.1002/cbic.200300755

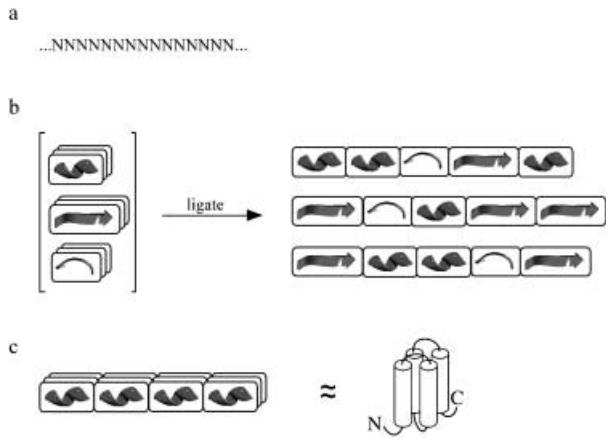

Figure 1. *Schematic representation of the libraries discussed in the text. a) total random library,[11–19] b) library of secondary structure modules, themselves made by synthetic oligonucleotides in which trinucleotide mixtures have been used to achieve the desired secondary structure propensities. The modules are then ligated at the DNA level to achieve open reading frames of about 100 amino acids. For details, see text and ref. [38]. c) Library of potential four-helix bundles made by a synthetic gene based on binary patterning. For details, see text and Kamtekar et al.[26]*

ATP with K_D values in the nanomolar range and required Zn^{2+} to do so, indicating the presence of a defined binding site.[15] Based on the assumption that only 10% of the available 6×10^{12} sequences are potentially functional in the first round of selection, they estimated that roughly 1 in 10^{11} molecules in the original random sequence library have ATP binding activity. Nevertheless, structural characterization has so far been limited because the selected proteins are soluble only as fusions with maltose binding protein, and thus the "foldedness" of the peptide is not yet clear.

In contrast to the work of Keefe and Szostak,[15] which started from a large random sequence library, Yamauchi et al.[23] reported the evolution of esterase activity from merely ten clones that were arbitrarily chosen from a random sequence polypeptide library.[24] Using phage display, the ten clones were assayed for binding to a transition state analogue (TSA) for an esterase reaction. The sequence of the clone with highest affinity to the TSA was randomized and another ten clones were arbitrarily chosen and assessed by phage display as in the previous round. After 13 such rounds of mutation and selection, they obtained a protein (approximately 120 amino acid residues) with a sevenfold higher k_{cat}/K_m value for esterase activity relative to the initial clones. In addition, the TSA was shown to inhibit the esterase activity of the peptide, further suggesting the presence of a specific active site. Oddly, however, CD spectroscopy indicated that there was no significant increase of secondary structure in the evolved peptides.

These studies demonstrate that it is possible to obtain functional proteins from random sequence libraries. The fact that these proteins show some function implies that at least a part of each sequence must be structured, and this structured state must be sufficiently populated, even though it may be in rapid equilibrium with unfolded or less compact states. Nevertheless, the emergence of stable folds from these libraries has yet to be observed.

Combinatorial Protein Libraries

The number of sequences that can be screened experimentally at any one time is $< 10^{14}$,[14] whereas the number of all possible sequences of proteins with a length of 100 amino acids is $20^{100} \approx 10^{130}$. The size of the libraries that can be screened experimentally is unlikely to increase, as 10^{14} genes with a size of 1000 base pairs corresponds to 100 μg DNA, and in order to increase this diversity, one would have to work with grams of DNA, which is very unrealistic. To be able to answer these fundamental questions of stable structure, which may involve exceedingly rare events in sequence space, it may thus be important to rationally design a library with biased sequences to increase the chance of obtaining functional and/or folded proteins. Furthermore, instead of trying to have all conceivable variants present at once, it is a much more effective strategy to create them by an evolutionary process in multiple "generations", that is, a succession of randomization and selection.

The Hecht group has reported the de novo design of protein libraries using binary patterning.[25–34] Previously they have constructed a library of four-helix bundle proteins (Figure 1c) by constraining the pattern of polar and non-polar residues, but not the precise side chains,[26] and have found that a substantial proportion of these proteins exhibit cooperative thermal unfolding[31] and amide protons that are protected from exchange.[30] A large fraction of the four helix bundle library members have been shown to bind heme,[28] possibly because heme generally prefers to be sandwiched between helices. Subsequently, library members have been shown to display peroxidase activity, higher than what had been achieved in other artificial systems.[27] More recently, they have grafted a binary patterned library to the beginning and end of the helices of one of the members[35] arbitrarily chosen from the previously

Andreas Plückthun, born 1956, studied Chemistry at the University of Heidelberg and received his Ph.D. from the University of California at San Diego in 1982. He was a postdoctoral fellow at Harvard University with Prof. Jeremy Knowles, before becoming group leader at the Gene Center and Max-Planck-Institut for Biochemistry in Martinsried, Germany. Since 1993 he is a Full Professor of Biochemistry at the University of Zürich, Switzerland. He is an elected member of EMBO and the German Academy of Natural Scientists, Leopoldina. Among the awards he has received are the Karl-Heinz-Beckurts Prize (Germany), the JP Morgan Chase Health Award (USA), the Wilhelm-Exner Medal (Austria), and the Jury's Prize in the European Grand Prix of Innovation. He is a cofounder of the German Biotech company Morphosys.

designed library.[26] Subsequent biophysical characterization of 5 proteins from this second generation library by CD and NMR-measurements showed a substantially higher stability and improved native-like properties compared to the initial protein, and a structure was recently solved.[36] Binary patterns of alternating polar and nonpolar residues has also been used for designing β-sheet proteins with the main finding that these sequences protect their hydrophobic surface by forming amyloid-like fibrils[33] or assemble as mono-layers on an aqueous/air interface.[34] However, by applying negative design principles[37] it has been shown that these fibril forming β-sheet proteins could be converted into monomeric and soluble proteins by introducing a lysine residue at one of the nonpolar positions in the edge β-strand.[32] Recently, with the long term goal of creating artificial biomaterial, a de novo designed peptide was shown to be capable of assembling into β-sheet fibers on a highly ordered pyrolytic graphite surface that was used to direct the assembly of a de novo designed peptide into β-sheet fibers.[25]

These results demonstrate that a protein library created by using binary patterning to encode a specific given fold (four-helix bundle or β-sheet) will contain a substantial number of proteins with the expected folding properties. This immediately raises the possibility that other simple folds can be obtained from binary patterning.

In contrast to the above studies, we have constructed a combinatorial library that does not restrict the topology to a certain fold (Figure 1b).[38] As chain topology is defined by the arrangement of secondary structure modules, we generated different libraries with an average length of 100 amino acid residues by randomly polymerizing DNA modules encoding secondary structures (α-helix, β-strand and β-turn), based primarily on binary patterning of polar and non-polar residues.[39] The DNA sequences were prepared using codon (trinucleotide) building blocks,[40] which allowed us to tailor the corresponding amino acid mixtures to favor formation of the desired secondary structure elements. We observed that approximately 1 in 6 members in those libraries which only contain α-helical modules and turn modules, while distant from natural proteins in sequence space, possess favorable properties, including the expected α-helical secondary structure, a defined oligomerization state as well as cooperative equilibrium unfolding properties in urea. These proteins, however, far from being fully evolved, also showed properties consistent with a molten-globule state,[38] which may be a stepping stone to a fully stable native structure.

While natural proteins, in general, have a distinct global free energy minimum which allows them to fold into one unique structure, molten globules lack this distinct global minimum, and thus do not have specific or compact tertiary structure,[41] even though they may have a particular topology. How may a protein evolve to have a distinct global minimum?

Nature obviously did not compute and create proteins with native-like properties, but only through successive rounds of mutation and selection were modern proteins obtained. Sequence variation and the resulting evolution of proteins arise from point mutations, deletions, insertions, elongation and recombination. All these mechanisms can be reproduced with current techniques in molecular biology. However, one still requires a means of selection and a clearly defined selection pressure. Undoubtedly all proteins have evolved on the basis of their functionality, and a stable structure is simply the prerequisite to fulfilling a function rather than a goal in itself. In addition, proteins form a compactly folded structure by shielding hydrophobic residues from the solvent, in the form of a well packed hydrophobic core, and by maximizing the number of hydrogen bonds possible. These properties are very difficult, if not impossible to accurately calculate a priori at the current time. Therefore, it is important to examine the evolutionary origins of these properties. It is possible that the *types* of selection pressure available in the laboratory were also operative during natural evolution of proteins, even though the means of implementing them were of course very different.

In the next section we will review the strategies that can potentially be used to select and evolve protein folds.

Selecting Stably Folded Proteins

One selection pressure that has been explored in searching for "foldedness" is protease resistance. It has often been observed that unfolded proteins are digested much more rapidly than folded ones,[42, 43] and that a number of cellular machines even actively unfold proteins for degradation.[44] Partial proteolysis is an established technique for defining domains,[45, 46] for example, for crystallizing proteins. The key to selectivity seems to lie in the fact than many active sites of proteases are at least somewhat recessed, and therefore only accessible to a piece of sequence which can enter this site. Nevertheless, most proteases have some primary sequence preferences, and unstructured polypeptides might evade the pressure by simply avoiding these sequence motifs. Furthermore, perfectly folded proteins can have accessible loops, and all fusion proteins need to be linked in some way and thereby expose a potentially vulnerable linker. Most successes with this selection strategy have therefore been reported for cases in which point mutants of the same fold were challenged, rather than different architectures compared (see below).

It is not clear which role proteolysis plays during natural evolution as proteins and the cellular proteases coevolve to achieve a fine balance of optimized lifetimes in vivo. One method to select proteins with higher protease resistance using phage display[47] is to fuse a library of proteins at the N terminus of the minor coat protein (pIII) of the M13 filamentous phage, followed by selection for variants that are resistant to proteolytic digestion (Figure 2a). This was demonstrated for point mutants of a very stable protein, ubiquitin, and selected clones were shown by CD and NMR to possess native-like properties.[48] Similar approaches were applied to RNase T1,[49] barnase[50] and cold shock protein B from *Bacillus subtilis* (Bs-CspB),[51] with the exception that in these cases the target proteins were inserted between the domains of pIII (Figure 2a), thus linking protease resistance to the infectivity of the phages. Thorough studies on Bs-CspB[51, 52] have shown that protease resistance could successfully enrich the Bs-CspB variant with the highest midpoint of

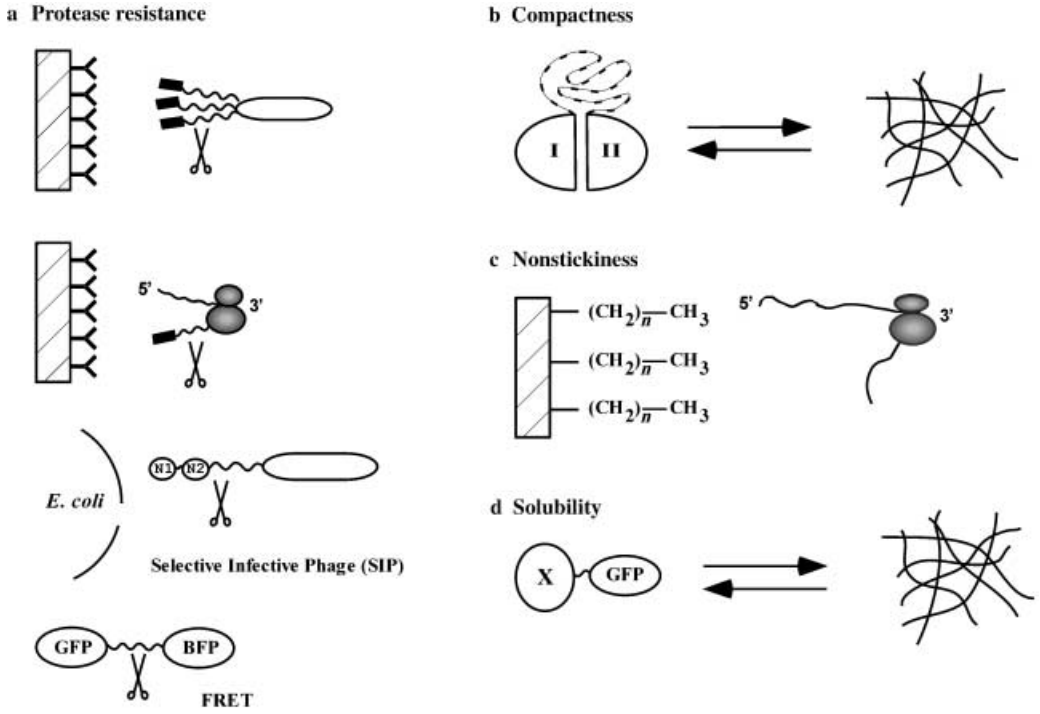

Figure 2. *Schematic representation of the selection methods for folded states discussed in the text in more detail. a) Selection for protease resistance. From top to bottom, phage display with an N-terminal tag and a cognate anti-tag antibody is shown, and the connection of the tag to the phage is challenged by a protease digesting the inserted polypeptide, indicated by wavy lines.[47, 48] The same principle can be applied to ribosome display (second from top), thereby leading to an in vitro selection not requiring any transformation.[64] Instead of an N-terminal tag in phage display, the presence of the N-terminal domains themselves can be used as a selection principle, as they are strictly needed for infecting* E. coli *(third from top).[49–53] Finally, E. coli proteases can be used in vivo, if they disconnect blue fluorescent protein (BFP) from green fluorescent protein (GFP) and thereby stop the possibility of fluorescence energy transfer (FRET).[58] b) Compactness may (in principle) be selected for, if the polypeptide of interest is inserted into a protein, where two domains or subdomains need to be close to allow function. This has been applied for dihydrofolate reductase[60] and an SH2 domain.[18, 19] c) Nonstickiness has been used as a selection criterion using ribosome display by removing unfolded clones, which presumably expose part of their hydrophobic core, and thus bind to a hydrophobic resin, carefully adjusted to provide the right amount of selection pressure.[64] d) Solubility itself can be selected for in fusion proteins to GFP, as an aggregating protein will also lead to the precipitation of GFP, and thus a lack of fluorescence.[55, 56]*

thermal unfolding and Gibbs free energy of unfolding, values which even exceeded that of the naturally thermostable homolog Bc-Csp from *Bacillus caldolyticus*. It has to be noted that in all these cases variants of well-folding, natural proteins were investigated.

The potential of proteolysis selection in combination with phage display has also been used to investigate the possibility of creating novel proteins with stable folds.[53] Riechmann and Winter fused the gene encoding the N-terminal half of the cold shock protein CspA with the fragmented genome of *Escherichia coli*, then selected the corresponding proteins that survived proteolysis. They selected chimeric proteins which could be solubly expressed and exhibited cooperative unfolding. Despite the lack of sequence homology to the C-terminus of CspA, one of the selected chimeras acquired sequence from a fold similar to CspA and another chimera picked up sequences from a protein with a different fold. One of the selected proteins has now been crystallized and shows that the key features of the CspA fold are present, albeit in a multimeric, domain-swapped fashion.[54] These results provide evidence that proteolysis selection has the potential to select domains for compactness and foldedness.

Globular proteins must be soluble to exhibit function. Waldo et al. have developed a method that is capable of screening soluble proteins in *E. coli* based on the idea that soluble proteins will fluoresce when fused with green fluorescent protein (GFP), whereas aggregating proteins will prevent GFP from folding and thus block fluorescence (Figure 2 d).[55] They have recently applied this method to improve the expression properties of hyperthermophilic proteins, which are expressed predominantly as inclusion bodies in *E. coli*. The resulting solubly expressed proteins could then be used for structural studies.[56] It is, however, necessary to match the folding properties of the target with those of GFP. If the target is extremely insoluble, an evolved form of soluble GFP is useful.[57] If the target is not too insoluble, wild-type GFP provides the necessary discrimination. These experiments provide a selection for solubility, but this may be necessary but not sufficient to select artificial proteins for folding.

In a related method proteins have been selected for resistance to *E. coli* proteases by monitoring fluorescence energy transfer (FRET) between blue fluorescent protein (BFP) and green fluorescent protein (GFP) fused to the N and C termini, respectively, of the protein of interest (Figure 2 a).[58] During the time required for the fluorophor to develop (24 h), presumably by air oxidation, these less stable proteins are degraded in the cytoplasm of *E. coli*. Because GFP and BFP are no longer held in

proximity, the FRET signal is lost. In this way an antibody V_L domain was selected that exhibits improved stability. However, this method is probably limited to misfolded proteins which do not form inclusion bodies in the cytoplasm of *E. coli*, and it therefore complements the work of Waldo et al.,[55, 56] which measures absolute GFP fluorescence and thus solubility.

Insertion of an unfolded polypeptide sequence into the loop of a host protein will decrease the stability of the host due to the entropic cost associated with the unfolded sequence. If the inserted sequence is properly folded, however, the entropic cost is minimized (Figure 2b). Based on this theory, Baker and colleagues have attempted to select folded proteins from a random polypeptide sequence library.[18, 19] They inserted the library into the loop of the SH2 protein and selected for binding to the SH2 ligand using phage display.[19] Surprisingly, the selected polypeptide sequence was predominantly random coil, yet had very little effect on the stability of SH2 or its affinity to the ligand. The authors believe that the tolerance of the random coil may stem from the solubility of the protein and not necessarily the presence of a fold. Nevertheless, the insertion may provide an interesting route to evolve multi-domain proteins, and whole domain insertions into another domain are common in multidomain proteins.[59]

Previously, we have attempted to select for folded proteins by inserting secondary structure based libraries[38] between the two domains of murine dihydrofolate reductase (mDHFR; Figure 2b).[60] mDHFR can be divided into two inactive domains that will regain activity when they reassociate with the help of, for instance, two interacting peptides or a cognate antigen-antibody pair.[61–63] Since only the bacterial DHFR is sensitive to trimethoprim, *E. coli* can acquire trimethoprim (TMP) resistance when mDHFR is expressed in an active form, and the growth rate correlates with mDHFR activity. Therefore, when the inserted library protein has a stable fold, the two domains of the DHFR are expected to associate, which will allow *E. coli* cells to grow in the presence of TMP. We thus expected to obtain a protein insert with at least a compact structure, simply by selecting for faster growing clones in the presence of TMP. Indeed, a clear enrichment of a specific sequence was observed. The clones were then further characterized as isolated proteins, in the absence of mDHFR. We observed, however, that the selected sequences had very similar biophysical properties to the proteins arbitrarily selected from the initial library; these proteins were highly soluble but possessed a significant fraction of random coil.[60] This result and the analogous work using the SH2 domain[18, 19] both lead to the conclusion that the solubility of the hybrid protein, rather than a compact, folded structure, is the main determinant of whether the loop insertion is accepted or not.

Recently we reported the development of an in vitro selection system which can be tailored to reward the function, solubility and protease resistance of a protein.[64] We aimed to establish ribosome display as a method where we are able to apply different selection pressures using a common setup. Ribosome display[21] is an in vitro selection system capable of screening very large libraries ($> 10^{12}$), and has been shown to be a powerful tool to select proteins for ligand binding properties.[22, 65] It can be adopted to select for high affinity binding, but it can also be used to select the most stable variant, again by exploiting ligand binding.[66] We have recently further adapted ribosome display to select for proteins based on their folding properties alone, independent of any binding properties. This is based on two properties of misfolded proteins: 1) increased sensitivity to proteolysis (Figure 2a) and 2) greater exposure of hydrophobic residues (Figure 2c). By targeting these properties, we have shown that compactly folded and soluble proteins can be enriched over insoluble and random coil proteins.[64]

Since ribosome display works entirely in vitro, a larger library can be screened is possible with methods where a DNA library has first to be introduced into cells. Moreover, because of the defined nature of the system, possible problems of genetic instability or toxicity of the proteins, which may occur in vivo, can be circumvented. These properties are very likely critical when performing selections on artificial libraries, such as those built from secondary structure modules.[38]

Future Perspectives

There is great interest in generating truly novel proteins, rather than merely varying natural proteins, and the challenge is to establish a system capable of obtaining such molecules. Creating protein libraries, in combination with an appropriate selection method, is one approach that may achieve this goal. Although computational de novo design of proteins has been developing rapidly, the complexity of the physics underlying the protein topology itself, but especially the process of folding, with alternative pathways of misfolding and aggregation, continues to be a major challenge to accurately predict the fate of a given sequence.

The design of a protein library and the method of selecting meaningful sequences from that library are the two fundamental aspects that must be considered. Statistics alone dictate that there are almost certainly novel folds hidden in the unexplored regions of protein sequence space. With current developments in gene-assembly technologies and selection methods, the tools are now here, and the exploration of sequence space can be expected to proceed more rapidly and efficiently.[67]

Keywords: combinatorial chemistry · de novo proteins · molecular evolution · protein design · ribosome display

[1] C. Chothia, *Nature* **1992**, *357*, 543 – 544.
[2] C. A. Orengo, D. T. Jones, J. M. Thornton, *Nature* **1994**, *372*, 631 – 634.
[3] Y. I. Wolf, N. V. Grishin, E. V. Koonin, *J. Mol. Biol.* **2000**, *299*, 897 – 905.
[4] S. Govindarajan, R. Recabarren, R. A. Goldstein, *Proteins* **1999**, *35*, 408 – 414.
[5] C. M. Kraemer-Pecore, A. M. Wollacott, J. R. Desjarlais, *Curr. Opin. Chem. Biol.* **2001**, *5*, 690 – 695.
[6] J. G. Saven, *Curr. Opin. Struct. Biol.* **2002**, *12*, 453 – 458.
[7] T. Kortemme, M. Ramirez-Alvarado, L. Serrano, *Science* **1998**, *281*, 253 – 256.
[8] P. B. Harbury, T. Zhang, P. S. Kim, T. Alber, *Science* **1993**, *262*, 1401 – 1407.
[9] E. G. Emberly, N. S. Wingreen, C. Tang, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11163 – 11168.
[10] H. Li, R. Helling, C. Tang, N. Wingreen, *Science* **1996**, *273*, 666 – 669.
[11] W. Mandecki, *Protein Eng.* **1990**, *3*, 221 – 226.

[12] A. R. Davidson, R. T. Sauer, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 2146 – 2150.
[13] A. R. Davidson, K. J. Lumb, R. T. Sauer, *Nat. Struct. Biol.* **1995**, *2*, 856 – 864.
[14] G. Cho, A. D. Keefe, R. Liu, D. S. Wilson, J. W. Szostak, *J. Mol. Biol.* **2000**, *297*, 309 – 319.
[15] A. D. Keefe, J. W. Szostak, *Nature* **2001**, *410*, 715 – 718.
[16] I. D. Prijambada, T. Yomo, F. Tanaka, T. Kawama, K. Yamamoto, A. Hasegawa, Y. Shima, S. Negoro, I. Urabe, *FEBS Lett.* **1996**, *382*, 21 – 25.
[17] A. Yamauchi, T. Yomo, F. Tanaka, I. D. Prijambada, S. Ohhashi, K. Yamamoto, Y. Shima, K. Ogasahara, K. Yutani, M. Kataoka, I. Urabe, *FEBS Lett.* **1998**, *421*, 147 – 151.
[18] P. Minard, M. Scalley-Kim, A. Watters, D. Baker, *Protein Sci.* **2001**, *10*, 129 – 134.
[19] M. Scalley-Kim, P. Minard, D. Baker, *Protein Sci.* **2003**, *12*, 197 – 206.
[20] R. W. Roberts, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12297 – 12302.
[21] J. Hanes, A. Plückthun, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 4937 – 4942.
[22] A. Plückthun, C. Schaffitzel, J. Hanes, L. Jermutus, *Adv. Protein Chem.* **2000**, *55*, 367 – 403.
[23] A. Yamauchi, T. Nakashima, N. Tokuriki, M. Hosokawa, H. Nogami, S. Arioka, I. Urabe, T. Yomo, *Protein Eng.* **2002**, *15*, 619 – 626.
[24] T. Nakashima, N. Ishiguro, M. Yamaguchi, A. Yamauchi, Y. Shima, C. Nozaki, I. Urabe, T. Yomo, *J Biosci. Bioeng.* **2000**, *90*, 253 – 259.
[25] C. L. Brown, I. A. Aksay, D. A. Saville, M. H. Hecht, *J. Am. Chem. Soc.* **2002**, *124*, 6846 – 6848.
[26] S. Kamtekar, J. M. Schiffer, H. Xiong, J. M. Babik, M. H. Hecht, *Science* **1993**, *262*, 1680 – 1685.
[27] D. A. Moffet, M. A. Case, J. C. House, K. Vogel, R. D. Williams, T. G. Spiro, G. L. McLendon, M. H. Hecht, *J. Am. Chem. Soc.* **2001**, *123*, 2109 – 2115.
[28] D. A. Moffet, L. K. Certain, A. J. Smith, A. J. Kessel, K. A. Beckwith, M. H. Hecht, *J. Am. Chem. Soc.* **2000**, *122*, 7612 – 7613.
[29] N. R. Rojas, S. Kamtekar, C. T. Simons, J. E. McLean, K. M. Vogel, T. G. Spiro, R. S. Farid, M. H. Hecht, *Protein Sci.* **1997**, *6*, 2512 – 2524.
[30] D. M. Rosenbaum, S. Roy, M. H. Hecht, *J. Am. Chem. Soc.* **1999**, *121*, 9509 – 9513.
[31] S. Roy, M. H. Hecht, *Biochemistry* **2000**, *39*, 4603 – 4607.
[32] W. Wang, M. H. Hecht, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2760 – 2765.
[33] M. W. West, W. Wang, J. Patterson, J. D. Mancias, J. R. Beasley, M. H. Hecht, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11211 – 11216.
[34] G. Xu, W. Wang, J. T. Groves, M. H. Hecht, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 3652 – 3657.
[35] Y. Wei, T. Liu, S. L. Sazinsky, D. A. Moffet, I. Pelczer, M. H. Hecht, *Protein Sci.* **2003**, *12*, 92 – 102.
[36] M. Hecht, personal communication.
[37] J. S. Richardson, D. C. Richardson, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2754 – 2759.
[38] T. Matsuura, A. Ernst, A. Plückthun, *Protein Sci.* **2002**, *11*, 2631 – 2643.
[39] M. W. West, M. H. Hecht, *Protein Sci.* **1995**, *4*, 2032 – 2039.
[40] B. Virnekäs, L. Ge, A. Plückthun, K. C. Schneider, G. Wellnhofer, S. E. Moroney, *Nucleic Acids Res.* **1994**, *22*, 5600 – 5607.
[41] M. Arai, K. Kuwajima, *Adv. Protein. Chem.* **2000**, *53*, 209 – 282.
[42] A. Fontana, P. Polverino de Laureto, V. De Filippis, E. Scaramella, M. Zambonin, *Folding Des.* **1997**, *2*, R17 – 26.
[43] S. J. Hubbard, F. Eisenmenger, J. M. Thornton, *Protein Sci.* **1994**, *3*, 757 – 768.
[44] J. R. Hoskins, S. Sharma, B. K. Sathyanarayana, S. Wickner, *Adv. Protein Chem.* **2001**, *59*, 413 – 429.
[45] S. J. Hubbard, *Biochim. Biophys. Acta* **1998**, *1382*, 191 – 206.
[46] C. O. Pabo, R. T. Sauer, J. M. Sturtevant, M. Ptashne, *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 1608 – 1612.
[47] M. D. Finucane, M. Tuna, J. H. Lees, D. N. Woolfson, *Biochemistry* **1999**, *38*, 11604 – 11612.
[48] M. D. Finucane, D. N. Woolfson, *Biochemistry* **1999**, *38*, 11613 – 11623.
[49] V. Sieber, A. Plückthun, F. X. Schmid, *Nat. Biotechnol.* **1998**, *16*, 955 – 960.
[50] P. Kristensen, G. Winter, *Folding Des.* **1998**, *3*, 321 – 328.
[51] A. Martin, V. Sieber, F. X. Schmid, *J. Mol. Biol.* **2001**, *309*, 717 – 726.
[52] A. Martin, I. Kather, F. X. Schmid, *J. Mol. Biol.* **2002**, *318*, 1341 – 1349.
[53] L. Riechmann, G. Winter, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 10068 – 10073.
[54] G. Winter, personal communication.
[55] G. S. Waldo, B. M. Standish, J. Berendzen, T. C. Terwilliger, *Nat. Biotechnol.* **1999**, *17*, 691 – 695.
[56] J. D. Pedelacq, E. Piltch, E. C. Liong, J. Berendzen, C. Y. Kim, B. S. Rho, M. S. Park, T. C. Terwilliger, G. S. Waldo, *Nat. Biotechnol.* **2002**, *20*, 927 – 932.
[57] A. Crameri, E. A. Whitehorn, E. Tate, W. P. Stemmer, *Nat. Biotechnol.* **1996**, *14*, 315 – 319.
[58] B. Philipps, J. Hennecke, R. Glockshuber, *J. Mol. Biol.* **2003**, *327*, 239 – 249.
[59] N. V. Grishin, *J. Struct. Biol.* **2001**, *134*, 167 – 185.
[60] A. Ernst, T. Matsuura, A. Plückthun, unpublished results.
[61] E. Mössner, H. Koch, A. Plückthun, *J. Mol. Biol.* **2001**, *308*, 115 – 122.
[62] J. N. Pelletier, F. X. Campbell-Valois, S. W. Michnick, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12141 – 12146.
[63] J. N. Pelletier, K. M. Arndt, A. Plückthun, S. W. Michnick, *Nat. Biotechnol.* **1999**, *17*, 683 – 690.
[64] T. Matsuura, A. Plückthun, *FEBS Lett.* **2003**, *539*, 24 – 28.
[65] C. Schaffitzel, I. Berger, J. Postberg, J. Hanes, H. J. Lipps, A. Plückthun, *Proc. Natl. Acad. Sci. USA* **2001**, *3*, 3.
[66] L. Jermutus, A. Honegger, F. Schwesinger, J. Hanes, A. Plückthun, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 75 – 80.
[67] Note added in proof: A remarkable success in computational design and experimental verification of a novel protein fold was recently reported by B. Kuhlmann, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard, D. Baker, *Science* **2003**, *302*, 1364 – 1368.

Received: August 26, 2003 [M755]

