Chemistry & Biology, Vol. 11, 865–874, June, 2004, ©2004 Elsevier Ltd. All rights reserved. DOI 10.1016/j.chembiol.2004.04.006

Evolutionary Optimization of a Nonbiological ATP Binding Protein for Improved Folding Stability

John C. Chaput and Jack W. Szostak*
Howard Hughes Medical Institute and
Department of Molecular Biology
Massachusetts General Hospital
Boston, Massachusetts 02114

Summary

Structural comparison of in vitro evolved proteins with biological proteins will help determine the extent to which biological proteins sample the structural diversity available in protein sequence space. We have previously isolated a family of nonbiological ATP binding proteins from an unconstrained random sequence library. One of these proteins was further optimized for high-affinity binding to ATP, but biophysical characterization proved impossible due to poor solubility. To determine if such nonbiological proteins can be optimized for improved folding stability, we performed multiple rounds of mRNA-display selection under increasingly denaturing conditions. Starting from a pool of protein variants, we evolved a population of proteins capable of binding ATP in 3 M guanidine hydrochloride. One protein was chosen for further characterization. Circular dichroism, tryptophan fluorescence, and ^{1}H-^{15}N correlation NMR studies show that this protein has a unique folded structure.

Introduction

We are interested in the extent to which biological proteins are representative of the total structural diversity available in sequence space. For example, does nature use all possible protein folds or just a subset of possible folds [1]? Our knowledge of protein structure has dramatically improved with recent advances in whole-genome sequencing and the availability of enlarged sets of high-resolution protein structures [2]. From these data, it appears that nature makes use of a relatively small number of distinct protein folds: at most a few thousand and perhaps no more than one thousand [3, 4, 5]. One possibility is that biological protein folds represent all or nearly all of the physically realistic folds [3]. Alternatively, the number of protein folds observed in biology may be only a small fraction of what is actually possible. The biological subset of protein folds could be a result of historical accident (survival of the first) or because biological folds convey certain favorable properties (survival of the fittest). Clearly there is something special about widespread protein folds, such as the $(\alpha/\beta)_8$ TIM-barrel, that makes the evolution of a new catalytic activity more likely to occur by recruitment of a pre-existing TIM-barrel than from other structures. What remains unclear is whether all TIM-barrel proteins derive from one common ancestor or whether this motif has itself evolved independently several times [6]. If contemporary protein structures arose from a small set of primitive protein domains, which over time recombined to form larger motifs of increasing complexity [7], repeated convergence on favorable structures may have been unavoidable [8]. Structures not easily derived from the initial set of protein domains may be under-represented or absent altogether in the proteins we see today [9].

Numerous strategies for the de novo evolution of proteins from combinatorial libraries have been developed in an effort to evolve new protein folds [10, 11]. Many of these rely on the use of rational design to limit the exploration of sequence space to regions thought to be more likely to yield proteins with native-like properties. Binary patterning is one such example, in which protein libraries are biased toward combinations of α helices [12] and β sheets [13] by arranging hydrophobic and hydrophilic residues in a specified repeating order. This approach has recently led to the in vitro evolution of a left-handed four-helix bundle protein with a structure specified by the design of the library [14]. In addition to library patterning, the use of libraries containing reduced sets of amino acids has shown promise [15]. From a random library encoding mainly glutamine, lysine, and arginine, α-helix-rich proteins with thermal melting properties similar to natural proteins were obtained [16]. Computational methods offer a distinctly different route to the design of new protein folds. Structural algorithms that repack existing core residues [17] or find new sequences consistent within pre-existing backbone geometry have been used to identify many independent sequences consistent with the same protein fold [18, 19]. In a dramatic recent advance, a new protein with a folded arrangement of α helices and β strands not before seen in biology was computationally designed, and its three-dimensional structure was verified by X-ray crystallography [20]. Systematic application of this approach has great promise for exploring the range of physically possible protein folds without the constraints of biological evolution. However, these proteins are constructed of α-helical and β strand building blocks, and the design of more divergent structures remains challenging. More challenging still is the de novo design of novel protein folds possessing a desired function.

Our approach to addressing this problem involves determining the three-dimensional structures of functionally active proteins selected from unconstrained libraries of random sequence proteins and comparing these structures to those found in biology. We began with an effort to ascertain the frequency of functionally active proteins in a sampling of all possible sequences of a given length. We used mRNA display to select ATP binding proteins from a random library of 80 contiguous amino acids, unconstrained by any element of design [21]. Starting from a pool of 6×10^{12} unique protein sequences, we determined that roughly 1 in 10^{11} is capable of selectively binding ATP; similar frequencies have been obtained with RNA libraries [22]. Four independent

*Correspondence: szostak@molbio.mgh.harvard.edu

families of ATP binding proteins were isolated. Blast analysis [23] indicated that none of the selected protein sequences bore any significant resemblance to sequences found in nature, although one family (family B) contained two CXXC motifs and indeed turned out to be a zinc binding domain. Unfortunately, biophysical characterization of the selected ATP binding proteins proved impossible due to poor solubility. This observation led to the question of whether protein sequences isolated from unconstrained random sequence libraries could be evolved to adopt a folded state of reasonable stability or whether most such proteins might represent examples of evolutionary dead ends. If sequence space is populated with many protein families of limited functional potential, then of course biology would use the much smaller subset that could be evolved to attain useful properties.

We therefore designed a selection strategy for improved protein folding based on the use of mRNA display to isolate protein variants capable of binding ATP in the presence of increasing concentrations of denaturant. Similar approaches have been described using phage display and ribosome display [24, 25], but neither approach simultaneously achieves both high library complexity ($\geq 10^{14}$ unique sequences) and stable genotype-phenotype linkage under extremely denaturing conditions. In contrast, mRNA display is uniquely suited for this purpose because it is an entirely in vitro selection system based on a covalent linkage between newly translated proteins and their encoding mRNA transcripts [26]. mRNA display has been used to identify novel peptide-drug conjugates [27] and receptors [28] and to deconvolute various protein-protein [29,30] and enzyme-substrate interactions [31]. We reasoned that it might be possible to use mRNA display to evolve a population of ATP binding proteins with improved folding stability if such variants existed and if our starting library contained enough sequence diversity to examine a significant region of sequence space around the ATP binding domain.

Results and Discussion

By selecting for ATP binding proteins in the presence of increasing amounts of denaturant, we hoped to identify functionally active proteins with improved folding stability and better solubility. Initial experiments using clone 18-19 from our previous selection suggested that this protein was quite sensitive to denaturation by urea (50% loss of ATP binding in 1 M urea; J.W.S. and A. Keefe, unpublished data). We then performed a series of ATP binding experiments in the presence of guanidine hydrochloride to determine optimal denaturing conditions for starting our selection. Based on these results, we concluded that our pool would be >95% inactivated in the presence of 1.5 M GuHCl. We therefore carried out the first round of selection for improved folding in the presence of 1.5 M GuHCl.

Since sequencing data from round 18 of our earlier selection suggested that significant sequence diversity remained in the pool, we used the output from round 17 of that selection to directly generate the input pool of mRNA-displayed proteins for the current selection without incorporating additional diversity. Two minor changes were made to the constant regions of the library. At the 5′ end, the TMV translation enhancer was replaced by an AMV enhancer to allow the use of a distinct 5′ PCR primer. At the 3′ end, downstream of the random region, two out-of-frame stop codons were added to minimize fusion formation from out-of-frame mutants, and a psoralen photo-crosslinking domain was introduced to facilitate fusion formation (Figure 1A) [32].

Starting from round 17 of our previous selection, we performed six successive rounds of in vitro selection and amplification. For each round of selection (Figure 1A), the pool DNA was transcribed into RNA, photoligated to a psoralen-DNA-puromycin linker, and translated in the presence of ^{35}S-labeled methionine. Prior to the actual selection step, the mRNA portion of the mRNA-protein fusion was converted to an RNA-DNA heteroduplex by reverse transcription to prevent enrichment of ATP binding RNA aptamers. mRNA-displayed proteins were then incubated with ATP-agarose beads in the presence of GuHCl. The GuHCl concentration was gradually increased from 1.5 to 3 M in an effort to ensure that less than 10% of the input to each round of selection was recovered in the elution step. By inactivating the pool by at least 90% prior to each round of selection, we aimed to enrich for well-folded ATP binding proteins. After washing the column with binding buffer plus GuHCl, functionally active proteins were recovered by eluting with free ATP in GuHCl-supplemented binding buffer. The encoding DNA sequences were PCR amplified and used to generate a new library of mRNA-displayed proteins for input into the next selection cycle. After six rounds of in vitro selection, the concentration of GuHCl had been increased from 1.5 to 3 M, while maintaining the proportion of material recovered during the ATP elution at less than 10% in each round (Figure 1B, blue bars). When the input to each round was reexamined in the presence of 1.5 M GuHCl, the proportion of mRNA-displayed protein eluting with ATP had risen from 0.6% to 47% (Figure 1B, red bars). Material from round 6 was examined in the absence of GuHCl, and 67% bound to the ATP column, compared to ~35% for the input material to round 1.

In order to determine the genetic changes that had led to the improved binding behavior, we cloned and sequenced 24 individual library members from round 6. Sequence alignment showed that a population of ATP binding proteins had emerged with an optimized consensus sequence that differed from the original binding-optimized consensus sequence at 13 out of 80 amino acid positions (Figure 2A). In order to compare the changes in protein sequence over the course of these sequential selections, the consensus sequences from rounds 8 and 18 of the original ATP binding selection were aligned with the consensus sequence obtained from our denaturing ATP binding selection (Figure 2A). We refer to these three distinct consensus sequences as the primordial, binding-optimized, and folding-optimized consensus sequences. The primordial sequence is the parental sequence from which all subsequent variants were derived. The binding-optimized sequence was derived by mutagenesis in rounds 9–11 of the selection followed by 7 additional rounds of selection for improved

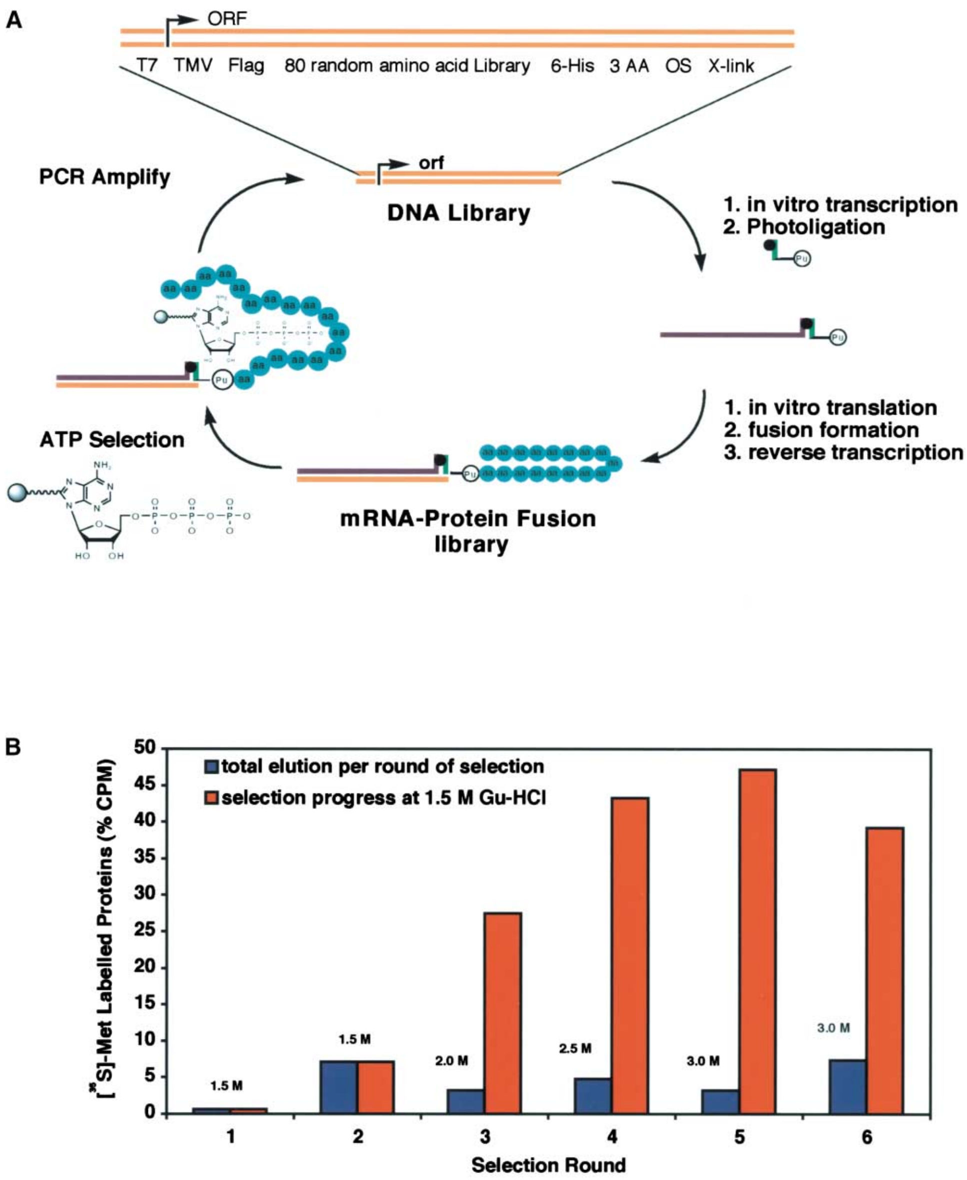

A
ORF
T7 TMV Flag 80 random amino acid Library 6-His 3 AA OS X-link
orf
PCR Amplify
DNA Library
1. in vitro transcription
2. Photoligation
Pu
aa
ATP Selection
1. in vitro translation
2. fusion formation
3. reverse transcription
mRNA-Protein Fusion library
B
total elution per round of selection
selection progress at 1.5 M Gu-HCl
[35S]-Met Labelled Proteins (% CPM)
0
5
10
15
20
25
30
35
40
45
50
1.5 M
1.5 M
2.0 M
2.5 M
3.0 M
3.0 M
1
2
3
4
5
6
Selection Round

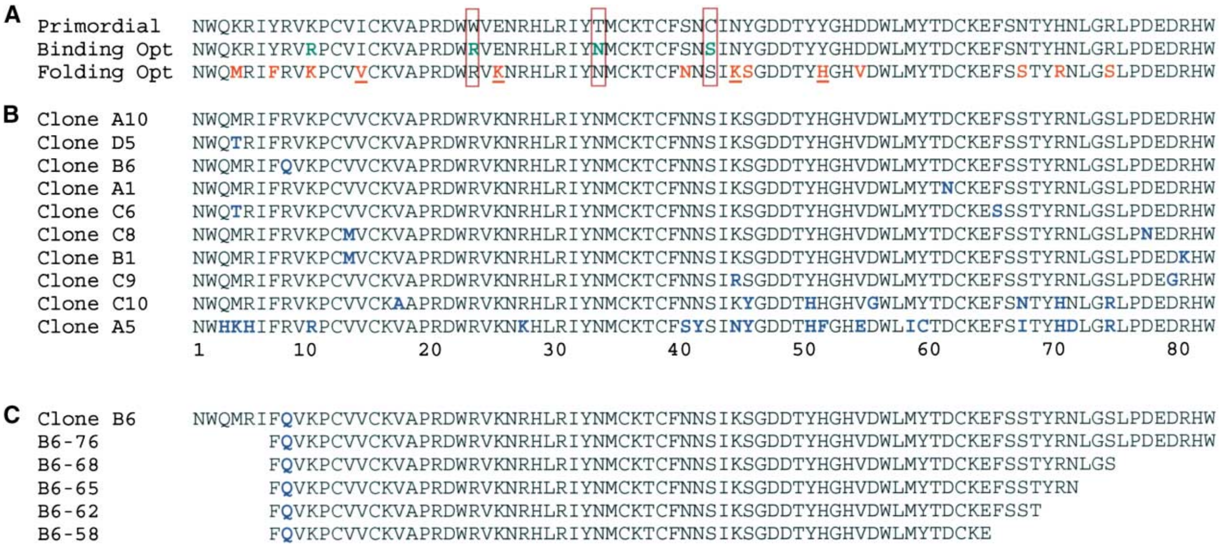
Figure 1. In Vitro Evolution of Nonbiological ATP Binding Proteins

(A) In vitro selection protocol for protein optimization. An 80 amino acid DNA library encoding functionally active but poorly folded ATP binding protein variants was constructed by PCR amplification from the output of our previous selection [21]. The DNA library was transcribed into RNA, photoligated to a DNA-puromycin linker, translated with in situ fusion formation, and reverse transcribed to afford a starting library of mRNA-displayed proteins as input into the first round of selection. mRNA fusion proteins were incubated with ATP agarose under various concentrations of GuHCl, washed with selection buffer, and competitively eluted with free ATP. Eluted fractions were combined and PCR amplified to generate a new library of ATP binding proteins for subsequent rounds of directed evolution.

(B) Selection progress. The total proportion of mRNA-displayed protein library which bound to and eluted from immobilized ATP-agarose per round of selection (blue bars) under increasing concentrations of GuHCl. The relative enrichment per round of selection in 1.5 M GuHCl (red bars).

column binding [21]. The folding-optimized sequence refers to the consensus sequence after selection for ATP binding in the presence of GuHCl. Comparison of the primordial, binding-optimized, and folding-optimized consensus sequences shows that, of the four amino acid substitutions selected for in the transition from the primordial to the binding-optimized sequence (K10R, W23R, T33N, and C42S), three have gone to fixation, while one mutation (K10R) reverted back to a lysine residue in the folding-optimized sequence (Figure 2A). It may be that the arginine residue leads to better ATP binding, but the lysine residue leads to better protein folding. However, while these amino acid changes do improve binding and/or folding, they are not sufficient to allow soluble expression of the free protein in *E. coli*. (J.W.S. and A. Keefe, unpublished data).

Comparison of the sequences obtained from our denaturing selection shows that five of the thirteen consen-



```
A  Primordial     NWQKRIYRVKPCVICKVAPRDWWVENRHLRIYTMCKTCFSNCINYGDDTYYGHDDWLMYTDCKEFSNTYHNLGRLPDEDRHW
   Binding Opt    NWQKRIYRVRPCVICKVAPRDWRVENRHLRIYNMCKTCFSNSINYGDDTYYGHDDWLMYTDCKEFSNTYHNLGRLPDEDRHW
   Folding Opt    NWQMRIFRVKPCVVCKVAPRDWRVKNRHLRIYNMCKTCFNNSIKSGDDTYHGHVDWLMYTDCKEFSSTYRNLGSLPDEDRHW

B  Clone A10      NWQMRIFRVKPCVVCKVAPRDWRVKNRHLRIYNMCKTCFNNSIKSGDDTYHGHVDWLMYTDCKEFSSTYRNLGSLPDEDRHW
   Clone D5       NWQTRIFRVKPCVVCKVAPRDWRVKNRHLRIYNMCKTCFNNSIKSGDDTYHGHVDWLMYTDCKEFSSTYRNLGSLPDEDRHW
   Clone B6       NWQMRIFQVKPCVVCKVAPRDWRVKNRHLRIYNMCKTCFNNSIKSGDDTYHGHVDWLMYTDCKEFSSTYRNLGSLPDEDRHW
   Clone A1       NWQMRIFRVKPCVVCKVAPRDWRVKNRHLRIYNMCKTCFNNSIKSGDDTYHGHVDWLMYTNCKEFSSTYRNLGSLPDEDRHW
   Clone C6       NWQTRIFRVKPCVVCKVAPRDWRVKNRHLRIYNMCKTCFNNSIKSGDDTYHGHVDWLMYTDCKESSSTYRNLGSLPDEDRHW
   Clone C8       NWQMRIFRVKPCMVCKVAPRDWRVKNRHLRIYNMCKTCFNNSIKSGDDTYHGHVDWLMYTDCKEFSSTYRNLGSLPNEDRHW
   Clone B1       NWQMRIFRVKPCMVCKVAPRDWRVKNRHLRIYNMCKTCFNNSIKSGDDTYHGHVDWLMYTDCKEFSSTYRNLGSLPDEDKHW
   Clone C9       NWQMRIFRVKPCVVCKVAPRDWRVKNRHLRIYNMCKTCFNNSIRSGDDTYHGHVDWLMYTDCKEFSSTYRNLGSLPDEGRHW
   Clone C10      NWQMRIFRVKPCVVCKAAPRDWRVKNRHLRIYNMCKTCFNNSIKYGDDTHHGHVGWLMYTDCKEFSNTYHNLGRLPDEDRHW
   Clone A5       NWHKHIFRVRPCVVCKVAPRDWRVKNKHLRIYNMCKTCFSYSINYGDDTHFGHEDWLICTDCKEFSITYHDLGRLPDEDRHW
                  1        10        20        30        40        50        60        70        80

C  Clone B6       NWQMRIFQVKPCVVCKVAPRDWRVKNRHLRIYNMCKTCFNNSIKSGDDTYHGHVDWLMYTDCKEFSSTYRNLGSLPDEDRHW
   B6-76                 FQVKPCVVCKVAPRDWRVKNRHLRIYNMCKTCFNNSIKSGDDTYHGHVDWLMYTDCKEFSSTYRNLGSLPDEDRHW
   B6-68                 FQVKPCVVCKVAPRDWRVKNRHLRIYNMCKTCFNNSIKSGDDTYHGHVDWLMYTDCKEFSSTYRNLGS
   B6-65                 FQVKPCVVCKVAPRDWRVKNRHLRIYNMCKTCFNNSIKSGDDTYHGHVDWLMYTDCKEFSSTYRN
   B6-62                 FQVKPCVVCKVAPRDWRVKNRHLRIYNMCKTCFNNSIKSGDDTYHGHVDWLMYTDCKEFSST
   B6-58                 FQVKPCVVCKVAPRDWRVKNRHLRIYNMCKTCFNNSIKSGDDTYHGHVDWLMYTDCKE
```

Figure 2. Sequence Alignment of Selected ATP Binding Proteins with Improved Folding Stability

(A) Alignment of primordial, binding-optimized, and folding-optimized consensus sequences of family B referring to rounds 8 and 18 of our previous selection [21] and round 6 of our denaturing selection, respectfully. Amino acid residues shown in green indicate strongly selected mutations in the transition from primordial to binding optimized. Amino acid residues displayed in red indicate mutations between the primordial and folding-optimized consensus sequence. Red boxes highlight residues that appear to have gone to fixation following sequence optimization. Underlined characters in the optimized consensus sequence show mutations that were selectively enriched in the binding-optimized sequence.
(B) Amino acid sequence for individual proteins isolated from round six of the denaturing selection. Amino acid mutations shown in blue denote differences between individual protein sequences after sequence optimization relative to the folding-optimized consensus sequence.
(C) Deletion analysis of protein B6 fragments assembled as MBP fusion proteins.

sus mutations (Y7F, I14V, E25K, Y51H, and D54V) were highly selected for (>94%), while eight additional mutations were strongly enriched (>65%). Four of these thirteen mutations (I14V, E25K, N44K, and Y51H) were already present as selectively enriched mutations (≥4/56) in the binding-optimized sequences from round 18 of our previous selection (Figure 2A, underlined residues), suggesting that these positions may be important for protein folding and stability. None of the 13 mutations were mutually exclusive, and clone A10, which contained all 13 mutations and no others, was recovered in 3 out of the 24 sequences.

Based on deletion analysis (see below), the folding-optimized consensus sequence can be divided into three distinct regions: the N-terminal region (residues 1–6), the core ATP binding domain (residues 7–68), and the C-terminal region (residues 69–82). Of the ten amino acid substitutions inside the core ATP binding domain, Y7F, R10K, I14V, S40N, and N67S represent the most conservative changes. The I14V change lies within the first CXXC region of the zinc binding domain and represents a conservative but clearly important substitution that may play a role in repacking the hydrophobic core of the protein. In contrast, amino acid mutations E25K and D54V represent significant changes: charge reversal and a hydrophilic to hydrophobic switch. Such changes could clearly cause large changes in the energy of the folded state and/or in ligand binding. Outside the core ATP binding domain, three other significant amino acid mutations (K4M, H70R, and R74S) were observed. Although these residues are not required for ATP binding, they may nevertheless impart additional folding stability by allowing packing interactions of the N- and C-terminal peptides with the core of the ATP binding protein domain.

To ensure that the selected proteins were functionally active, ten representative sequences (Figure 2B) were translated in vitro as free ATP binding proteins and assayed individually for ATP column binding activity, as previously described (see above). Each of the ^{35}S-labeled proteins bound to and competitively eluted from an ATP agarose column to the extent of 18%–25% in selection buffer supplemented with 2.5 M GuHCl. When the pH was raised from 7.4 to 8.5 while keeping all other conditions the same, the amount of ATP column binding increased to 24%–32% for all ten ATP binding proteins.

In order to determine how selecting for protein folding stability correlated with the evolution of improved solubility, we screened each of the ten ATP binding proteins for solubility at high concentration. All ten ATP binding proteins were cloned as C-terminal fusion proteins of maltose binding protein (MBP) with a thrombin-cleavable linker separating the maltose binding protein from the ATP binding protein. Expression of MBP fusion proteins in *E. coli* and purification by amylose affinity chromatography afforded several milligrams of each fusion protein (>90% purity). As expected from our earlier work with MBP fusion protein (clone 18-19) [21], all ten MBP fusions remained soluble indefinitely at 2 mg/ml. To examine the solubility of the free ATP binding proteins, MBP-fusion proteins were cleaved overnight at room temperature into their respective proteins by thrombin proteolysis. The mixture was centrifuged to separate insoluble aggregates from soluble free protein, and the supernatant was analyzed by SDS gel electrophoresis to detect soluble free ATP binding protein. In the ab-

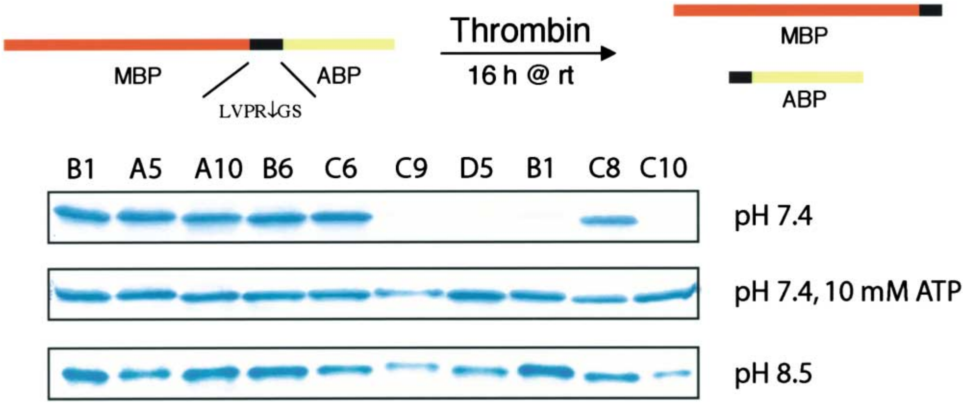


Figure 3. Thrombin Cleavage Assay for Protein Solubility

MBP fusion proteins were cleaved with thrombin to liberate individual ATP binding proteins from the MBP fusion protein. Ten selected ATP binding proteins were assayed for solubility at 2 mg/ml concentration in the presence and absence of free ATP by SDS-PAGE analysis.

sence of ATP, six of the ten proteins remained in solution, while four protein sequences (C9, D5, B1, and C10) formed insoluble precipitates (Figure 3). In the presence of free ATP, three of the four insoluble proteins remained soluble. In light of the fact that our selected proteins appear more active under slightly basic conditions, we performed the thrombin cleavage assay at pH 8.5 in the absence of ATP. Under these conditions, all four insoluble protein clones showed significant improvements in solubility. The fact that 6 out of 10 evolved ATP binding proteins remained soluble even in the absence of ATP demonstrates that selecting for protein folding stability is an effective approach to isolating protein variants with improved solubility properties.

The six MBP fusion proteins with the best overall solubility properties were chosen for further analysis. Apparent dissociation constants ($K_{d\ app}$) for ATP were measured for each MBP fusion protein by spin filtration and were found to range from 180 nM to 1.3 μM [33]. MBP fusion protein B6, which deviated from the consensus sequence only by the additional R8Q mutation, exhibited the best ATP binding affinity, with a $K_{d\ app}$ of 180 nM. FPLC analysis of this protein on a Sephadex-200 gel filtration column indicates that approximately 65% of this protein is monomeric in selection buffer, with the remainder running as higher-order aggregates in the void volume. MBP fusion protein A5, which deviates from the folding-optimized consensus sequence at 18 out of 80 amino acid positions, has an observed K_d of 375 nM and is $>$90% monomeric by gel filtration analysis. Modified expression protocols and alternative buffer conditions did not significantly affect the monomer to aggregate ratios. Overall, this represents a significant improvement in decreased aggregation by members of the folding-optimized pool over members of the binding-optimized pool.

Because previous deletion analysis showed that the entire protein domain is not needed for ATP binding activity, we constructed a series of N- and C-terminal deletion constructs of protein B6 (Figure 2C) in an attempt to eliminate deleterious regions that might nucleate aggregation. Each deletion construct was assayed for solution binding ($K_{d\ app}$) to ATP and for aggregation by size exclusion chromatography. In this manner, we were able to determine the minimal sequence necessary for formation of a stable folded structure capable of ATP recognition. Deletion of six amino acids from the N terminus of the protein had little effect on either protein aggregation or ATP binding. Further N-terminal deletions were not attempted, as previous deletion analysis of clone 18-19 showed that larger deletions led to loss of functional activity [21]. Interestingly, the deletion of eight amino acids (LPDEDRHW, residues 74–82) from the C terminus resulted in a protein that was $>$98% monomeric by gel filtration analysis, suggesting that part or all of this eight amino acid sequence mediates aggregation. This deletion led to a 2-fold improvement in ATP binding ($K_{d\ app} = 100$ nM). Deletion of 6 additional amino acids from the C terminus gave a 62 amino acid (B6-62) core ATP binding domain that was similarly well folded. The solution $K_{d\ app}$ of B6-62 as an MBP fusion protein was approximately 90 nM. Surprisingly, deletion of four additional amino acids from the C terminus of B6-62 reduced ATP binding to levels that were undetectable by this assay ($>$2000-fold increase in $K_{d\ app}$); in contrast, detectable binding was retained following the deletion of 11 additional amino acids from the C terminus of the binding-optimized 18-19 clone [21].

To investigate the possibility that the longer length of the minimum ATP binding domain of our folding-optimized consensus sequence (compared to the minimum domain of our earlier binding-optimized sequence, clone 18-19) allowed additional contacts to be made to the ATP, we examined the specificity of B6-62 as an MBP fusion using a number of ATP analogs. Because the fraction of correctly folded protein can influence apparent solution binding ($K_{d\ app}$) values, true solution binding (K_d) was measured by spin filtration using increasing concentrations of competitor to displace trace amounts of ^{32}P-labeled ATP from protein B6-62 (Figure 4). The true K_d and fraction of folded protein can be derived from this data in combination with the data obtained by titrating the protein concentration. By this approach, the true K_d of the B6-62 protein MBP fusion for ATP was 190 nM, and $>$98% of the protein was correctly folded and active.

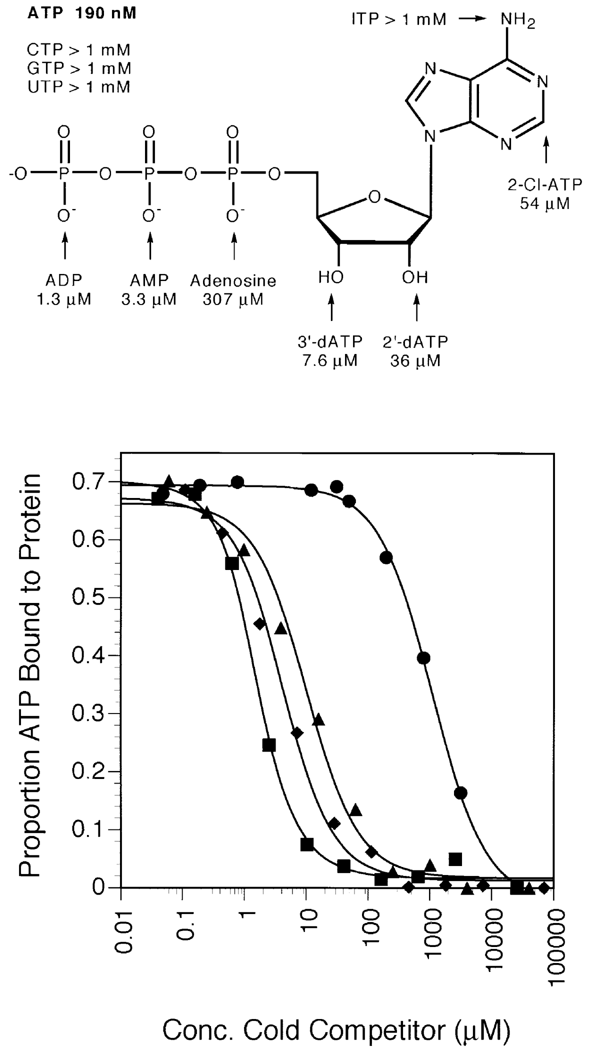


Figure 4. Solution Binding Affinity (K_d) and Specificity of Protein B6-62 for ATP

ATP binding specificity was measured using ATP analogs. IC_{50} plot for B6-62 specificity to ATP (squares), ADP (diamonds), AMP (triangles), and adenosine (circles).

Analysis of ATP analog binding revealed that the MBP: B6-62 fusion protein interacts with several distinct parts of the ATP molecule, as expected, but is surprisingly more specific than the previously studied clone 18-19 MBP fusion protein. For instance, the sequential removal of γ, β, and α-phosphate groups successively raises the K_d (Figure 4), with a large increase associated with removal of the α-phosphate (1600-fold). The loss of each phosphate is associated with an approximately 4- to 7-fold greater effect for B6-62 than for the earlier 18-19 protein. Binding to the 2′- or 3′-deoxy derivatives revealed strong interactions at both positions, with an observed loss in binding of 190- and 40-fold, respectively. Again, these changes are 25- and 9-fold greater for B6-62 than for 18-19. Binding to the nucleobase was probed with the analogs CTP, GTP, UTP, ITP, and 2-chloro-ATP. No binding was observed for the first four analogs (>5000-fold specificity). Substitution of the C-2 hydrogen with chlorine increased the K_d by 284-fold for B6-62 but only by 10-fold for 18-19. These comparison data suggest that the enlarged minimal ATP binding domain of B6-62 might improve contacts to one or more of the above positions, possibly mimicking the way that more complex domains evolve from simpler, less specific ones in nature [8]. However, the increased specificity of our folding-optimized protein (B6-62) compared to our earlier binding-optimized protein (clone 18-19) cannot be explained solely by the evolution of additional or stronger interactions with ATP, as this would lead to higher-affinity ATP binding. Indeed, it has been argued that selecting for higher-affinity binding automatically leads to greater specificity [34]. Here, however, we have selected for greater folding stability and have observed increased specificity with no change in overall affinity for ATP [21]. One reason for this improved selectivity in the absence of better ATP binding may be that protein B6-62 is a more rigid structure that is less able to conform to the geometry of ATP analogs and therefore shows greater discrimination between ATP and its analogs. Alternatively, proteins B6-62 and 18-19 may have similar affinities for ATP because some contacts for B6-62 were optimized at the expense of other untested contacts, with no net effect on binding.

Based on the favorable expression, solubility, and aggregation properties of the B6-62 construct, we proceeded to purify this protein on a larger scale by cleavage from the MBP fusion protein. The MBP: B6-62 fusion protein is expressed at high levels upon induction with IPTG (Figure 5A, lane 2 versus lane 3). Amylose-purified MBP: B6-62 fusion protein was further purified by binding to ATP-agarose beads and washing to remove misfolded proteins. Comparison of the amount of MBP fusion protein present in the flowthrough and wash steps (Figure 5A, lanes 5 and 6) with the total amount of protein loaded onto the resin (lane 4) suggests that less than 10% of the MBP fusion protein was misfolded or otherwise nonfunctional. The MBP protein was separated from the B6-62 domain by overnight cleavage with thrombin, and the free MBP protein was removed from the ATP-agarose column by additional wash steps (lanes 7 and 8). Highly pure, functionally active B6-62 free protein was then competitively eluted from the ATP-agarose column with 10 mM ATP and dialyzed against selection buffer to remove excess ligand. Gel filtration analysis of B6-62 before and after cleavage with thrombin (Figure 5B) shows that both the MBP fusion protein and free ATP binding protein domain give elution profiles consistent with monomeric protein as determined by comparison with known protein standards.

We used the highly purified B6-62 free protein for further biophysical characterization. The stability of the folded state and the nature of the unfolding transition were analyzed by monitoring shifts in tryptophan fluorescence in the presence of increasing concentrations of GuHCl. The denaturing profile of B6-62 in the presence and absence of exogenously added ATP showed a sigmoidal curve consistent with a single cooperative transition between the native and denatured states [35]. As shown in Figure 6A, a shift in transition midpoint ($D_{1/2(F\rightarrow U)}$) from 1.2 to 2.8 M GuHCl was observed for unfolding profiles performed in the absence and presence

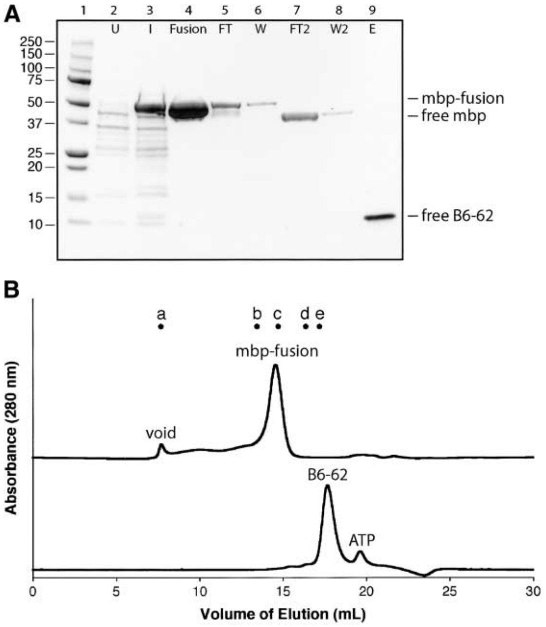


Figure 5. Purification of the Free ATP Binding Protein

(A) SDS-PAGE analysis of the purification of the ATP binding protein B6-62 minimum construct. Lane 1, protein standards; lane 2, uninduced *E. coli* cells; lane 3, induced *E. coli* cells; lane 4, MBP fusion after amylose affinity purification; lane 5, flowthrough after ATP affinity purification; lane 6, wash after ATP affinity purification; lane 7, flowthrough after thrombin cleavage on ATP affinity column; lane 8, wash after thrombin cleavage on ATP affinity column; lane 9, free B6-62 after competitive elution with free ATP.

(B) FPLC analysis of ATP binding protein B6-62 as an MBP-fusion protein (above) and free protein (below) by gel filtration chromatography. MW standards shown (top) correspond to blue dextran, 2000 kDa (a); BSA, 67 kDa (b); ovalbumin, 43 kDa (c); chymotrypsin, 25 kDa (d); and ribonuclease A, 13.7 kDa (e).

of 1 mM ATP, respectively. From the difference in stability, the magnitude of ATP stabilization could be calculated. Using the method of linear extrapolation (Figure 6B) [36, 37], the free energy of the unfolding transition ($\Delta G_{F\rightarrow U}$) was found to be 2.1 Kcal/mol in the absence of ATP and 4.8 Kcal/mol in the presence of 1 mM ATP. This corresponds to an overall stabilization by 1 mM ATP of 2.7 Kcal/mol.

To further examine the folding and structure of free B6-62 protein, we turned to circular dichroism (CD) spectroscopy. The CD spectra of native and denatured B6-62 (Figure 6C) indicate the presence of significant structure in the native form, with the native protein showing a single negative deflection with a minimum at 226 nm. This spectral signature is not consistent with standard α helix (two negative bands near 222 nm and 208 nm and a strong positive band near 190 nm), β sheet (negative band near 217 nm and positive band near 195 nm), or random coil (strong negative band at 200 nm) [38]. One possible explanation for this unexpected CD spectrum is that the structural architecture of this protein is defined by loops collapsed around a zinc-nucleated core. CD spectra taken in the presence of increasing concentrations of ATP did not affect the magnitude or appearance of the 226 nm signal. In contrast, the CD spectrum of B6-62 denatured with 4 M GuHCl changed dramatically, becoming consistent with a random coil peptide, as expected for an unfolded protein.

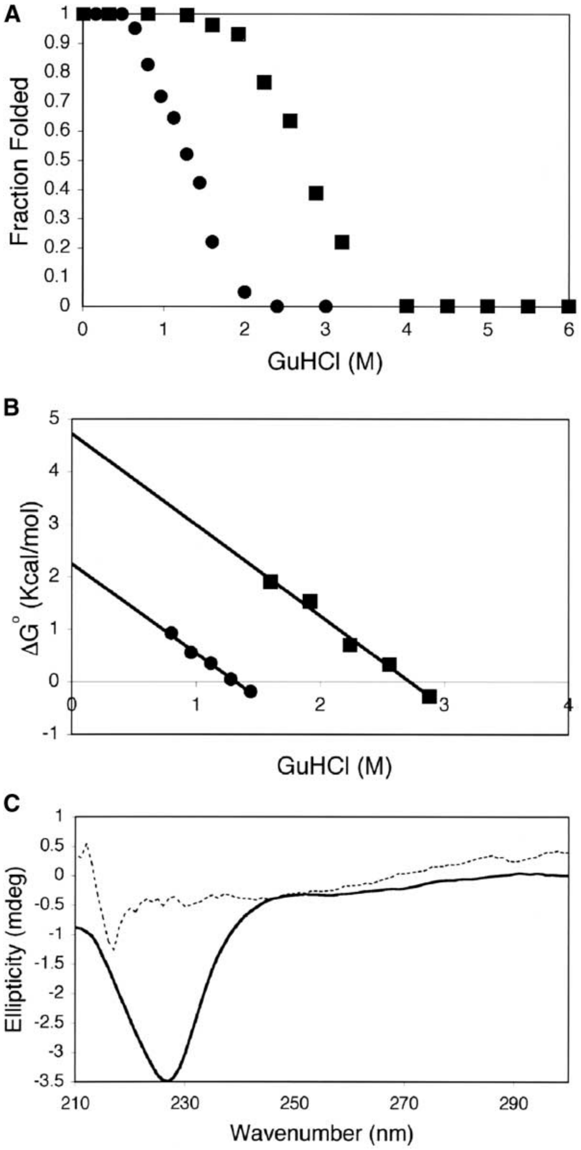


Figure 6. Biophysical Analysis of B6-62 Free Protein

(A) Guanidine hydrochloride denaturing curves for B6-62 free protein in the presence (squares) and absence (circles) of free ATP.

(B) Linear regression analysis of unfolding curves used to determine the Gibbs free energy of stabilization for ATP binding free protein B6-62 in the presence (squares) and absence (circles) of exogenous ligand.

(C) Circular dichroism spectra for B6-62 free protein in the presence (light dashed) and absence (solid) of 4 M GuHCl denaturant. The minimum at 226 nm suggests a folded secondary structure that collapses upon denaturing with GuHCl.

Given the unusual CD spectra indicating the absence of significant α helix or β strand contribution, we decided to investigate the tertiary structure of B6-62 by NMR. The one-dimensional ^{1}H-NMR (data not shown) revealed many sharp peaks distributed across the aliphatic (0–2 ppm), aromatic (6–7 ppm), and amide (7–9 ppm) regions of the spectrum. A few peaks were observed further

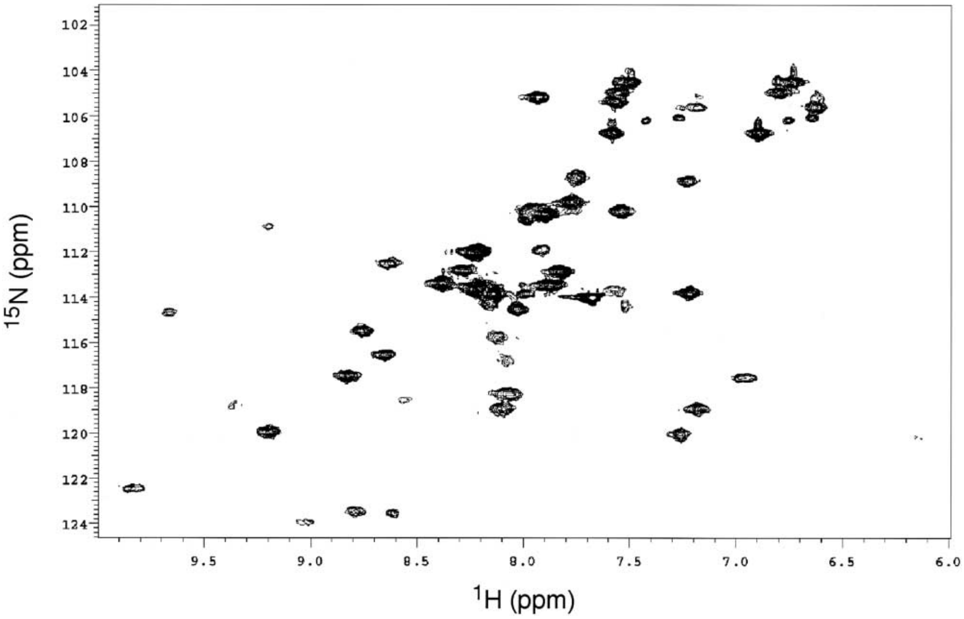


Figure 7. ^{1}H^{15}N HSQC NMR Spectrum of ^{15}N-labeled B6-62 in 10% D_2O

downfield (>9 ppm) of the amide region, indicating possible hydrogen bonding interactions to other parts of the protein. Data from the one-dimensional ^{1}H NMR confirms that B6-62 does not adopt a random coil conformation, since one would expect such a protein to produce few if any distinguishable peaks in the aromatic and amide regions of the ^{1}H-NMR spectrum. To determine if B6-62 is indeed a well-folded structure, a two-dimensional ^{1}H^{15}N-NMR heteronuclear single-quantum coherence (HSQC) experiment was performed (Figure 7) to detect the correlation between the ^{1}H and ^{15}N atoms using ^{15}N-labeled protein prepared from minimal media expression in *E. coli*. The resulting HSQC spectrum indicates the presence of ~45 well-resolved peaks asymmetrically distributed throughout the spectrum, a hallmark of a well-folded protein with a native-like structure. Based on these spectra, it is now evident that B6-62 appears to be a well-ordered protein. Complete structure determination of this protein is currently underway.

In summary, we have applied an in vitro selection strategy for isolating stably folded proteins of a de novo evolved ATP binding protein. We used one of our folding-optimized proteins for biophysical characterization by tryptophan fluorescence, CD spectroscopy, and ^{1}H^{15}N-NMR spectroscopy. Results from these studies indicate that this in vitro evolved protein folds into a well-defined tertiary structure that is stabilized by ATP. However, the results of CD spectroscopy suggest that the folded protein has an unusual structure that may differ from normal biological protein structures.

Significance

Determining whether or not biology uses all or most of the possible protein folds or just a small subset of possible folds will require solving the structures of a number of different proteins that have been evolved in the absence of biological constraints. For this to be possible, in vitro evolved proteins must have a stable folded structure, good solubility, and minimal aggregation. By using mRNA display in combination with in vitro selection, we have demonstrated that proteins selected for function can be optimized for improved folding stability, suggesting that sequence space is not dominated by structures incapable of such optimization. Furthermore, these proteins showed a marked improvement in solubility and ATP binding specificity over the starting population. Our initial data already suggest that this protein has an unusual metal nucleated structure lacking canonical α helices or β strands. We are now poised to obtain the first structure of a protein isolated from a completely unconstrained random sequence library.

Experimental Procedures

Construction of mRNA-Protein Fusion Library

The synthesis of mRNA-display libraries has been described in detail [39, 40, 41]. The initial library used in this work was constructed in a two-step amplification process directly from the output of round 17 of our previous ATP binding selection [21]. No additional diversity, beyond what remained after round 17, was incorporated during the selection. First, PCR was performed using forward and reverse primers designed to modify sequences outside the 80 amino acid random region. The forward primer (5′-d-TTC TAA TAC GAC TCA CTA TAG GGT TTT TAT TTT TAA TTT TCT TTC AAA TAC TTC CAC CAT GGA CTA CAA AGA CGA CGA CGA T-3′) contained the sequences necessary for in vitro transcription and translation, and the reverse primer (5′-d-ATA GCC GGT GCT ACC GCT CAG ACC CTT CGC AGA TCC AGA CAT TCC CAT ATG ATG-3′) incorporated a unique three amino acid signature (AKG), two out-of-frame stop codons, and a psoralen photo-crosslinking site for photochemical attachment to the psoralen-DNA-puromycin linker. PCR amplification (1 ml, 6 cycles: 95°C for 45 s, 58°C for 45 s, and 72°C for 90 s) of round 17 material (1.3 μg yield). A second amplification step (1 ml, 6 cycles, 1.5 μg DNA yield) was then performed to generate the starting library using truncated versions of the forward and reverse

PCR primers. The PCR reaction was phenol/chloroform extracted, ethanol precipitated, and transcribed with T7 RNA polymerase. The mRNA library was purified by denaturing PAGE, treated with RQ1 DNase (Promega), phenol/chloroform extracted, and concentrated by LiCl precipitation. Purified RNA was photochemically ligated to a psoralen-DNA-puromycin linker [5′-psoralen-TAGCCGGTG-$(PEG_9)_2$-dA_{15}CC-puromycin, Glen Research; underlined positions denote 2′-methoxy nucleosides] by irradiating for 15 min at 366 nm in 96-well microtitler plate (50 ul per well) [32]. The crosslinked material was purified by denaturing PAGE and translated in vitro by incubating for 1 hr at 30°C in 1 ml rabbit reticulocyte lysate (Nova Red, Novagen) and [^{35}S]methionine (Amersham Biosciences). In situ fusion formation was promoted by the addition of KCl and $MgCl_2$ to final concentrations of 400 mM and 50 mM, respectively, followed by a second incubation for 15 min at 25°C. mRNA-protein fusions were purified from the crude lysate by oligo (dT) and Ni-NTA chromatography (Qiagen) and quantified by [^{35}S]-scintillation counting. mRNA-protein fusions (100 pmol) were reverse transcribed with Superscript II RT (Gibco) using the RT-primer (5′-d-T_{15}AA CCG CTC AGC TTG GCC TG-3′) to prevent RNA structure formation during the selection. Approximately 25 pmol of the starting library was used in the first round of selection.

Selection of ATP Binding Proteins

Displayed proteins were incubated with C-8 linked ATP agarose beads (Sigma) in selection buffer [200 mM HEPES (pH 7.5), 4 mM $MgCl_2$, 400 mM KCl, 1 mM Na_2HPO_4, 0.1% Trition X-100, and 10 mM BME] supplemented with 1.5–3 M GuHCL for 2 hr at 25°C in a 5 ml disposable column (Biorad). The column was drained, washed with 25 column volumes of selection buffer, and eluted stepwise over 10 min intervals with 0.25 ml aliquots of selection buffer supplemented with 10 mM ATP. Elution fractions were combined, NAP-5 (Pharmacia) exchanged into H_2O, and PCR amplified. Following six rounds of selection, the amplified DNA was cloned into a TOPO vector (invitrogen) and sequenced (MGH Core Facility).

Expression of ATP Binding Proteins as MBP Fusion Proteins

Individual ATP binding protein (ABP) sequences were inserted into the pMal expression vector (pIADL14) [42] as C-terminal fusion proteins between BamHI and HindIII restriction sites with a thrombin cleavage site (LVPRGS) separating the maltose binding protein (MBP) from the ATP binding protein. Plasmid DNA was transformed into *E. coli* BL21(DE3) strain (Invitrogen). Bacteria were grown in LB medium supplemented with 0.1 mM $ZnSO_4$ at 37°C to an A_{600} of 0.6–0.8 and induced with 1 mM IPTG. Following an additional 3 hr of growth, bacteria were harvested by centrifugation and resuspended in 25 ml of 20 mM KH_2PO_4 (pH 8.0), 150 mM KCl, and 5 mM BME (referred to below as phosphate buffer). MBP-fusion proteins were bound to an amylose column (New England Biolabs), washed with phosphate buffer, and eluted with phosphate buffer supplemented with 10 mM maltose. MBP-fusion proteins were concentrated using an Amicon ultrafiltration device (Millipore). The yield of pure MBP-ABP fusion protein obtained from a 1 liter culture of induced *E. coli* was ~30–40 mg.

Dissociation Constants

Equilibrium dissociation constants (K_ds) for purified MBP-ABP fusion proteins were measured by equilibrium ultrafiltration [33]. Apparent K_ds were measured using trace γ-[^{32}P]-ATP (Amersham Biosciences) and a series of concentrations of MBP-ABP fusion protein spanning the $K_{d\ app}$. The data were iteratively fit (through nonlinear regression) using the computer program Deltagraph 4.0 (Red Rock Software) to the equation $y = b + c((x/(x+K_{d\ app}))$, where y is the fraction of ligand bound to protein, x is the protein concentration, b represents nonspecific binding to the protein and/or filter, and c is the maximum fraction of counts that can be bound. True K_ds and fraction of protein folded for ATP and ATP analogs was determined from IC_{50} plots by displacing bound γ-[^{32}P]-ATP from protein with increasing concentrations of competitor as previously described [43, 44].

Thrombin Cleavage of MBP-ABP and ABP Purification

MBP-ABP fusion proteins (2 mg/ml) were bound to an ATP agarose column (Sigma) and cleaved with thrombin (Novagen). Free MBP was removed by washing with phosphate buffer, and free ABP was eluted with phosphate buffer supplemented with 10 mM ATP. The eluted ABP was concentrated using a centricon filter (Millipore) and dialyzed against phosphate buffer. The yield of pure ABP obtained from a 1 liter culture of induced *E. coli* was ~5 mg.

Deletion Analysis of ATP Binding Protein B6

Deletion analysis of the ATP binding protein B6 was performed using sequence-specific PCR primers containing BamHI and HindIII restriction sites. Individual sequences were inserted into the pMal expression vector (pIADL14), transformed into *E. coli* BL21(DE3) strain (Invitrogen), and expressed and purified as described above.

Gel Filtration Chromatography

MBP-ABP fusion proteins and thrombin released free ABP were analyzed by FPLC (BioCAD Sprint Perfusion System) at concentrations of ~2–3 mg/ml using a Sephadex-200 gel filtration column (Pharmacia Biotech) with isocratic elution in phosphate buffer at 4°C. Individual fractions were analyzed by SDS-PAGE. Molecular weights were determined by linear regression analysis using a low molecular weight gel filtration calibration kit (Amersham Biosciences).

Tryptophan Fluorescence

The free energy of folding for free ATP binding protein B6-62 was calculated by using tryptophan fluorescence to monitor protein unfolding [35]. Protein samples were incubated in the presence of increasing concentrations of GuHCl under equilibrium conditions in the presence and absence of free ATP. The fraction of protein folded was determined from the change in the fluorescence emission at 350 nm using a Cary Eclipse fluorescence spectrometer with an excitation wavelength of 295 nm. From this data, the standard free energy of folding (ΔG_{fold}) was determined in the presence and absence of ligand from the relationship $\Delta G^o = -RT \ln K_{fold}$, where *R* is the gas constant and *T* is temperature [36, 37]. The linear extrapolation method then yields the ΔG_{fold} at 0 M GuHCl. Guanidine hydrochloride solution (8 M) was prepared according to the method of Nozaki [45].

Circular Dichroism

CD spectra of free ATP binding protein B6-62 were acquired using an Aviv CD Spectrometer Model 202. Spectra were recorded at 20°C in phosphate buffer under native and denaturing conditions (4 M GuHCl) by monitoring the wavelength dependence of [θ] in 1 nm increments with a sampling time of 10 s.

NMR Spectroscopy

^{1}H and ^{1}H ^{15}N-NMR spectra were acquired using a Varian 500 MHz NMR with ^{15}N-labeled protein (~0.5 mM) in 10% D_2O with 10 mM phosphate buffer and 75 mM KCl at pH 6.5. Protein sample was prepared from minimal media cultures using ^{15}N-labeled NH_4Cl as the sole source of nitrogen.

Acknowledgments

We would like to acknowledge members of the Szostak Lab for technical assistance and helpful discussions. We are particularly grateful to A. Keefe for providing the material necessary to generate our starting pool, for communicating unpublished experiments on the denaturation of the 18-19 protein, and for helpful discussions and comments on the manuscript. We also thank D. Borchardt for assistance with NMR experiments. This work was supported by grants from the NIH and NASA Astrobiology Institute. J.W.S is an Investigator, and J.C.C. is a Research Associate of the Howard Hughes Medical Institute.

Received: February 26, 2004
Revised: April 15, 2004
Accepted: April 21, 2004
Published online: May 20, 2004

References

1. Chothia, C., Hubbard, T., Brenner, S., Barnes, H., and Murzin, A. (1997). Protein folds in the all-α and all-β classes. Annu. Rev. Biophys. Biomol. Struct. *26*, 597–627.
2. Vitkup, D., Melamud, E., Moult, J., and Sander, C. (2001). Completeness in structural genomics. Nat. Struct. Biol. *8*, 559–566.
3. Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., and Thornton, J.M. (1997). CATH—a hierarchic classification of protein domain structures. Structure *5*, 1093–1108.
4. Zhang, C., and DeLisi, C. (1998). Estimating the number of protein folds. J. Mol. Biol. *284*, 1301–1305.
5. Chothia, C. (1992). One thousand families for the molecular biologist. Nature *357*, 543–544.
6. Orengo, C.A., Jones, D.T., and Thornton, J.M. (1994). Protein superfamilies and domain structures. Nature *372*, 631–634.
7. Gilbert, W. (1978). Why genes in pieces? Nature *271*, 501.
8. Doolittle, R.F. (1995). The multiplicity of domains in proteins. Annu. Rev. Biochem. *64*, 287–314.
9. Gerlt, J.A., and Babbitt, P.C. (2001). Divergent evolution of enzymatic function. Annu. Rev. Biochem. *70*, 209–246.
10. DeGrado, W.F., Summa, C.M., Pavone, V., Nastri, F., and Lombardi, A. (1999). De novo design and structural characterization of proteins and metalloproteins. Annu. Rev. Biochem. *68*, 779–819.
11. Moffet, D.A., and Hecht, M.H. (2001). De novo proteins from combinatorial libraries. Chem. Rev. *101*, 3191–3203.
12. Roy, S., Ratnaswamy, G., Boice, J.A., Fairman, R., McLendon, G., and Hecht, M.H. (1997). A protein designed by binary patterning. J. Am. Chem. Soc. *119*, 5302–5306.
13. West, M.W., Wang, W., Patterson, J., Mancias, J.D., Beasley, J.R., and Hecht, M.H. (1999). De novo amyloid proteins from designed combinatorial libraries. Proc. Natl. Acad. Sci. USA *96*, 11211–11216.
14. Wei, Y., Kim, S., Fela, D., Baum, J., and Hecht, M.H. (2003). Solution structure of a de novo protein from a designed combinatorial library. Proc. Natl. Acad. Sci. USA *100*, 13270–13273.
15. Davidson, A.R., and Sauer, R.T. (1994). Folded proteins occur frequently in libraries of random amino acid sequences. Proc. Natl. Acad. Sci. USA *91*, 2146–2150.
16. Davidson, A.R., Lumb, K.J., and Sauer, R.T. (1995). Cooperatively folded proteins in random sequence libraries. Nat. Struct. Biol. *2*, 856–864.
17. Desjarlais, J.R., and Handel, T.M. (1995). De-novo design of the hydrophobic core of proteins. Protein Sci. *4*, 2006–2018.
18. Dahiyat, B.I., and Mayo, S.L. (1997). De novo protein design: fully automated sequence selection. Science *278*, 82–87.
19. Voigt, C.A., Mayo, S.L., Arnold, F.H., and Wang, Z.-G. (2001). Computational method to reduce the search space for directed protein evolution. Proc. Natl. Acad. Sci. USA *98*, 3778–3783.
20. Kuhlman, B., Dantas, G., Ireton, G., Varani, G., Stoddard, B.L., and Baker, D. (2003). Design of a novel globular protein fold with atomic-level accuracy. Science *302*, 1364–1368.
21. Keefe, A.D., and Szostak, J.W. (2001). Functional proteins from a random-sequence library. Nature *410*, 715–718.
22. Sassanfar, M., and Szostak, J.W. (1993). An RNA motif that binds ATP. Nature *364*, 550–553.
23. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.
24. Jung, S., Honegger, A., and Pluckthun, A. (1999). Selection for improved protein stability by phage display. J. Mol. Biol. *294*, 163–180.
25. Matsuura, T., and Pluckthun, A. (2003). Selection based on the folding properties of proteins with ribosome display. FEBS Lett. *539*, 24–28.
26. Roberts, R., and Szostak, J.W. (1997). RNA-peptide fusion for the in vitro selection of peptides and proteins. Proc. Natl. Acad. Sci. USA *94*, 12297–12302.
27. Li, S., and Roberts, R.W. (2003). A novel strategy for in vitro selection of peptide-drug conjugates. Chem. Biol. *10*, 233–239.
28. McPherson, M., Yang, Y., Hammond, P.W., and Kreider, B.L. (2002). Drug receptor identification from multiple tissues using cellular-derived mRNA display libraries. Chem. Biol. *9*, 691–698.
29. Wilson, D.S., Keefe, A.D., and Szostak, J.W. (2001). The use of mRNA display to select high-affinity protein-binding peptides. Proc. Natl. Acad. Sci. USA *98*, 3750–3755.
30. Mammond, P.W., Alpin, J., Rise, C.E., Wright, M., and Kreider, B.L. (2001). In vitro selection and characterization of Bcl-XL-binding proteins from a mix of tissue-specific mRNA display libraries. J. Biol. Chem. *276*, 20898–20906.
31. Cujec, T.P., Medeiros, P.F., Hammond, P.W., Rise, C.E., and Kreider, B.L. (2002). Selection of v-Abl tyrosine kinase substrate sequences from randomized peptide and cellular proteomic libraries using mRNA display. Chem. Biol. *9*, 253–264.
32. Kurz, M., Gu, K., and Lohse, P.A. (2000). Psoralen photo-cross-linked mRNA-puromycin conjugates: a novel template for the rapid and facile preparation of MRA-protein fusions. Nucleic Acids Res. *28*, e83.
33. Jenison, R.D., Gill, S.C., Pardi, A., and Polisky, B. (1994). High resolution molecular discrimination by RNA. Science *263*, 1425–1429.
34. Eaton, B.E., Gold, L., and Zichi, D.A. (1995). Let's get specific: the relationship between specificity and affinity. Chem. Biol. *2*, 633–638.
35. Pace, C.N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. Methods Enzymol. *131*, 266–280.
36. Greene, R.F., and Pace, N.C. (1974). Urea and guanidine hydrochloride determination of ribonuclease, lysozyme, α-chymotrypsin, and β-lactoglobulin. J. Biol. Chem. *249*, 5388–5393.
37. Ahmad, F., and Bigelow, C.C. (1982). Estimation of the free energy of stabilization of ribonuclease A, lysozyme, α-lactalbumin, and myoglobin. J. Biol. Chem. *257*, 12935–12938.
38. Woody, R.W. (1995). Circular dichroism. Methods Enzymol. *246*, 34–70.
39. Liu, R., Barrick, J.E., Szostak, J.W., and Roberts, R.W. (2000). Optimized synthesis of RNA-protein fusions for in vitro protein selection. Methods Enzymol. *318*, 268–293.
40. Cho, G., Keefe, A.D., Wilson, D.S., Liu, R., and Szostak, J.W. (2000). Construction of high complexity synthetic libraries of long ORFs using in vitro selection. J. Mol. Biol. *297*, 309–319.
41. Keefe, A.D. (2000). In Current Protocols in Molecular Biology, Unit 24.5, R.M. Ausubel et. al., eds. (New York: Wiley).
42. McCafferty, D.G., Lessard, I.A.D., and Walsh, C.T. (1997). Mutational analysis of potential zinc-binding residues in the active site of the enterococcal D-Ala-D-Ala dipeptidase VanX. Biochemistry *36*, 10498–10505.
43. Vaish, N.K., Larralde, R., Fraley, A.W., Szostak, J.W., and McLaughlin, L.W. (2003). Biochemistry *42*, 8842–8851.
44. Oestereich, S., and Szostak, J.W. (2004).
45. Nozaki, Y. (1972). The preparation of guanidine hydrochloride. Methods Enzymol. *26*, 43–50.

Note Added in Proof

While this paper was in revision, the structure of the 18-19 variant of the ATP binding protein was solved by X-ray crystallography and shown to represent a novel protein fold (Lo Surdo, P., Walsh, M.A., and Sollazzo, M. (2004). A novel ADP- and zinc-binding fold from function-directed in vitro evolution. Nat. Struct. Mol. Biol. *11*, 382–383).