doi:10.1016/j.jmb.2007.05.062 J. Mol. Biol. (2007) 371, 501–513

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Structure and Evolutionary Analysis of a Non-biological ATP-binding Protein

Sheref S. Mansy[1], Jinglei Zhang[2], Rainer Kümmerle[3], Mikael Nilsson[2] James J. Chou[4], Jack W. Szostak[1] and John C. Chaput[2]*

[1]*Howard Hughes Medical Institute, and Department of Molecular Biology Massachusetts General Hospital, Boston, MA 02114 USA*

[2]*Center for BioOptical Nanotechnology, The Biodesign Institute, and Department of Chemistry and Biochemistry Arizona State University Tempe, AZ 85287-1604, USA*

[3]*Bruker BioSpin AG Fällanden, Switzerland*

[4]*Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston MA 02115, USA*

**Corresponding author*

We present a structural and functional analysis of the evolutionary optimization of a non-biological protein derived from a library of random amino acid sequences. A series of previously described *in vitro* selection experiments transformed a low-affinity ancestral sequence into a stably folded, high affinity ATP binding protein structure. While the evolutionarily optimized protein differs from its ancestral sequence through the accumulation of 12 amino acid mutations, the means by which those mutations enhance the stability and functionality of the protein were not well understood. We used a combination of mutagenesis, biochemistry, and NMR spectroscopy to investigate the structural and functional significance of each mutation. We solved the three-dimensional structure of the folding optimized protein by solution NMR, which revealed a fourth strand of the β-sheet of the α/β-fold that was not observed in an earlier crystallographic analysis of a less stable version of the protein. The structural rigidity of the newly identified β-strand was confirmed by T_1, T_2, and heteronuclear nuclear Overhauser enhancement (NOE) measurements. Biochemical experiments were used to examine point mutations that revert the optimized protein back to the ancestral residue at each of the 12 sites. A combination of structural and functional data was then used to interpret the significance of each amino acid mutation. The enhanced ATP affinity was largely due to the emergence of a patch of positive charge density on the protein surface, while the increased solubility resulted from several mutations that increased the hydrophilicity of the protein surface, thereby decreasing protein aggregation. One mutation may stabilize the hydrophobic face of the β-sheet.

Keywords: mRNA display; solution NMR; in vitro evolution

Introduction

We previously used mRNA display[1] to isolate a non-biological ATP-binding protein from an unbiased random-sequence library.[2] The library from which our ancestral ATP-binding protein was isolated consisted of $>10^{12}$ unique proteins each containing a central random region of 80 residues. After eight rounds of mRNA display based *in vitro* selection and amplification, four families of ATP binding proteins were identified from the starting library. None of these four families exhibited any significant sequence homology to any other known amino acid sequence found in nature. One of these protein families, family B, was further optimized for improved binding activity through mutagenesis and several additional rounds of selection and amplification.[2] Analysis of the protein sequences from this selection revealed two conserved CXXC motifs that bound a single zinc-metal ion. Complete biophysical characterization of the binding optimized proteins from the output of this selection was complicated by the presence of additional amino acid sequences outside the core-binding domain of the selected ATP-binding protein.

Abbreviations used: NOE, nuclear Overhauser enhancement; MBP, maltose binding protein; ABP, ATP binding protein; RCD, residual dipolar coupling; FOB, folding optimized family B.

E-mail address of the corresponding author: john.chaput@asu.edu

Identification of the core-binding domain by the characterization of truncation constructs allowed for the *in vivo* expression and isolation of monomeric protein in pure form. The X-ray crystal structure of one highly divergent variant (protein 18–19) from this selection was reported[3] and reveals a novel α/β fold not found in biology with a metal binding site remarkably similar to the treble clef zinc binding motif.[4]

All of the variants examined required high concentrations of free ATP in order to remain stably folded and soluble. In an attempt to overcome this limitation, mRNA display was used to select for new protein variants that remained stably folded in the absence of excess ATP ligand.[5] The selection strategy involved using guanidine hydrochloride to enrich for protein variants that remained folded, and therefore functional, in the presence of increasing concentrations of chemical denaturant. For each round of selection, the guanidine hydrochloride concentration was gradually increased to ensure that less than 10% of the input into each round of selection was recovered from the previous round. After six rounds of selection, many of these folding optimized ATP binding proteins remained water soluble and monomeric when expressed in *Escherichia coli*, and bound ATP with high affinity and specificity. We were then able to identify a 62 amino acid residue core domain of this protein that retained the same ATP binding function as the full-length construct. Comparison of the consensus sequences for each stage in the evolution of this protein (Figure 1(a)), from ancestral sequence to the folding optimized variant, revealed the location and identity of 12 amino acid substitutions that arose during the *in vitro* evolution of this protein from a low affinity ancestral state to a well folded high affinity structure.

We used structural and functional studies to investigate this *in vitro* evolutionary processes in greater detail. We solved the tertiary structure of the core ATP binding domain of the folding optimized protein by nuclear magnetic resonance spectroscopy. Analysis of the folded structure revealed a fourth strand in the β-sheet that was not present in the crystal structure of a less stable and highly divergent variant. The rigidity of the fourth β-strand was confirmed by T_1, T_2, and heteronuclear nuclear Overhauser enhancement (NOE) measurements. To investigate the functional properties of the individual mutations, we constructed single-point revertants for each mutation present in the core-binding domain of the protein and evaluated each revertant for its affect on ATP binding affinity, protein solubility, and thermal stability. We then used the structure of the folding optimized protein to interpret the effect of each mutation on ATP affinity and protein solubility. Taken together, the data suggest that a poorly functional, artificial protein can be significantly optimized through the accumulation of single-point mutations that do not dramatically affect the fold of the protein or the mode of ligand binding. This suggests that it might be possible to evolve many biologically relevant proteins that are structurally unstable or poorly soluble to a state of increased folding stability.

(a)

		Apparent Solution ATP Kd (nM)
Ancestral	GSYRVKPCVICKVAPRDWWVENRHLRIYTMCKTCFSNCINYGDDTYYGHDDWLMYTDCKEFSNT	>2.5 x 10^5
Binding Opt.	-----R----------R---------N--------S--------------------------	~1.0 x 10^5
Clone 18-19	-----R---K------N-K-K-K-----N------N-S-DI-----H--------A-S--I---	1051 ± 120
Folding Opt.	--F------V--------R-K-------N------N-S-KS-----H--V------------S-	494 ± 96
	sssss sssssss hhhhhhhhhh ssssssss	
	1 10 20 30 40 50 60	

(b)

		Apparent Solution ATP Kd (nM)
Folding Opt.	GSFRVKPCVVCKVAPRDWRVKNRHLRIYNMCKTCFNNSIKSGDDTYHGHVDWLMYTDAKEFSST	494 ± 96
F3Y	--Y----------	458 ± 10
R4Q (B6-64)	---Q----------	471 ± 28
K6R	-----R----------	753 ± 180
V10I	---------I----------	707 ± 256
R19W	------------------W----------	5626 ± 513
K21E	--------------------E----------	2623 ± 609
N29T	----------------------------T----------	759 ± 64
N36S	------------------------------------S----------	427 ± 41
S38C	--------------------------------------C----------	1057 ± 164
K40N	--N----------	460 ± 13
S41Y	---Y----------	855 ± 197
H47Y	--Y----------	555 ± 54
V50D	--D----------	666 ± 53
S63N	--N-	487 ± 134
	1 10 20 30 40 50 60	

Figure 1. Sequence alignment of ATP binding protein variants of family B. (a) Alignment of primordial, binding-optimized, and folding-optimized consensus sequences from rounds 8 and 18 of our previous selection[2] and round 6 of our denaturing selection,[5] respectively, and protein 18–19. (b) Amino acid sequences of individual single-point mutant proteins used to identify genetic mutations with structural and/or functional significance. In each case, the selected mutation was reverted back to its ancestral residue. Invariant residues are indicated by dashes and selected mutations are shown in bold. Apparent solution binding affinity for ATP (K_d) was determined by displacement equilibrium filtration.[14]