ARTICLE

ACS chemical biology

A Synthetic Protein Selected for Ligand Binding Affinity Mediates ATP Hydrolysis

Chad R. Simmons[†,‡], Joshua M. Stomel[†,§], Michael D. McConnell[†,‡], Daniel A. Smith[†,‡], Jennifer L. Watkins[†,‡], James P. Allen[‡], and John C. Chaput[†,‡,*]

[†]Center for BioOptical Nanotechnology, The Biodesign Institute, [‡]Department of Chemistry and Biochemistry, and [§]School of Life Sciences, Arizona State University, Tempe, Arizona 85287-5201

ABSTRACT How primitive enzymes emerged from a primordial pool remains a fundamental unanswered question with important practical implications in synthetic biology. Here we show that a *de novo* evolved ATP binding protein, selected solely on the basis of its ability to bind ATP, mediates the regiospecific hydrolysis of ATP to ADP when crystallized with 1 equiv of ATP. Structural insights into this reaction were obtained by growing protein crystals under saturating ATP conditions. The resulting crystal structure refined to 1.8 Å resolution reveals that this man-made protein binds ATP in an unusual bent conformation that is metal-independent and held in place by a key bridging water molecule. Removal of this interaction using a null mutant results in a variant that binds ATP in a normal linear geometry and is incapable of ATP hydrolysis. Biochemical analysis, including high-resolution mass spectrometry performed on dissolved protein crystals, confirms that the reaction is accelerated in the crystalline environment. This observation suggests that proteins with weak chemical reactivity can emerge from high affinity ligand binding sites and that constrained ligand-binding geometries could have helped to facilitate the emergence of early protein enzymes.

Primordial enzymes presumably evolved from pools of random sequences, but it is not known how these molecules achieved catalytic function (*1*). Recent insight into enzyme evolution has been obtained by resurrecting extant proteins from ancient organisms (*2, 3*) and by engineering novel functions into known protein scaffolds (*4–6*). Both approaches provide evidence for how enzymes evolve new types of catalytic function but fail to describe how the first progenitor enzymes came into being. Because the paleobiological record has long since been erased (*7, 8*), any understanding of the chemical constraints that led to protein biocatalysis must now be inferred using synthetic methods that mimic events that happened over 3 billion years ago when primordial proteins first appeared on the primitive Earth (*9*). Here we report the discovery that a synthetic ATP binding protein, evolved *de novo* from a random pool of protein sequences, mediates the regiospecific hydrolysis of ATP to ADP when crystallized with stoichiometric amounts of ATP.

The original *in vitro* selection experiment randomly sampled different regions of protein sequence space for polypeptides that folded themselves into three-dimensional structures that were capable of binding ATP (*10*). Starting from a pool of 6×10^{12} unique protein sequences, each containing 80 contiguous random amino acid residues, four families of proteins were identified that bound ATP. None of the four proteins identified in this selection shared any significant sequence similarity with one another or any known protein in the NCBI protein sequence database. Using directed evolution, we optimized the highest affinity variant (clone 18-19) of the Family B group of related protein sequences for improved folding stability and solved the X-ray crystal structure of this protein to a resolution limit of 1.65 Å

*Corresponding author, john.chaput@asu.edu.

Received for review December 8, 2008 and accepted June 11, 2009.

Published online June 12, 2009

10.1021/cb900109w CCC: $40.75

© 2009 American Chemical Society

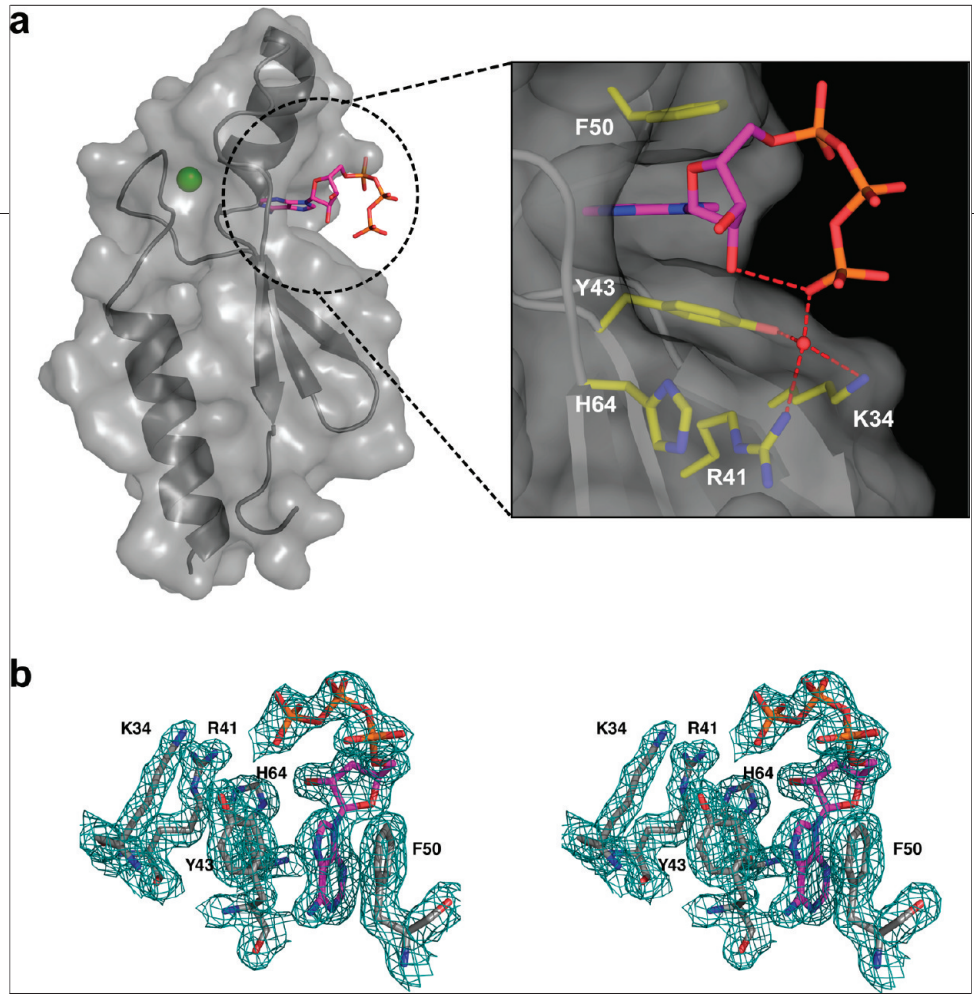

Figure 1. Adenosine binding site of protein DX crystallized in the presence of a high molar excess of ATP. a) Ribbon representation of the X-ray crystal structure of protein DX with translucent surface rendering. A bound zinc ion (green sphere) is shown at ~8.4 Å from the adenine nucleobase of the bound nucleotide (atom colors: magenta, carbon; red, oxygen; blue, nitrogen; and orange, phosphate). Inset is the structure of the ligand binding pocket of protein DX bound to ATP. Depicted beneath the surface are important side chains (atom colors: yellow, carbon; red, oxygen; and blue, nitrogen) that directly interact with the adenosine nucleotide. Polar contacts are drawn with red dashes, and an important structural water is shown as a red sphere. b) Stereoview of the ligand-binding site of DX with $2F_o - F_c$ electron density contoured at 1.6 σ. Stick representations of select ligand binding site residues are shown with atom coloring of protein residues (gray, carbon; red, oxygen; and blue, nitrogen). Also shown are Lys34, Arg41, Tyr43, Phe50, and His64.

(*11*). The evolutionary optimized version of the Family B protein differed from its synthetic progenitor by two amino acid substitutions and was given the name double mutant protein (DX). Our lab and another lab independently solved the X-ray crystal structure of protein 18-19, the evolutionary progenitor to protein DX, to resolution limits of 2.8 and 1.9 Å, respectively (*11, 12*). Together, the three structures reveal a novel zinc-nucleated α/β fold with a topology that has not yet been observed in nature.

Perhaps even more striking than the fold was the observation that each of these protein crystal structures contained electron density that was consistent with the presence of ADP in the ligand-binding pocket even though 1 molar equiv of ATP was present in the crystallization liquor (*11, 12*). Since the half-life of ATP for spontaneous hydrolysis is no shorter than ~12 months (*13*) and protein crystals were obtained after 3 days of growth, it was unclear how ADP became bound to the protein. We reasoned that the presence of ADP in the protein crystal structure could be explained by one of three possible scenarios: (i) the γ-phosphate of the ATP molecule is highly disordered and therefore the third phosphate of ATP is simply not visible in any of the electron density maps; (ii) the Family B protein binds preferentially to ADP, which is present as a minor contaminant from the disproportionation of ATP to ADP and adenosine tetraphosphate (*14*); or (iii) the Family B protein binds the ATP ligand in such a way that it is able to lower the activation barrier (~8.5 kcal mol^{-1}) and allow ATP hydrolysis to proceed at a rate that is faster than the uncatalyzed reaction. Intrigued by the possibility

TABLE 1. Summary of X-ray crystallization data[a]

	Protein		
	DX	**Y43F**	**DX**
Ligand (crystal drop)	ATP	ATP	ADP
Protein:ligand	1:100	1:10	1:100
Ligand (observed)	ATP	ATP	ADP
	Data collection		
Experimental station	ASU-XRD	ASU-XRD	ASU-XRD
Wavelength (Å)	1.54	1.54	1.54
Exposure time (s)	180	180	180
Oscillation range (deg)	0.5	0.5	0.5
Cell dimensions (Å)	$a = b = 73.70$, $c = 54.76$	$a = b = 71.74$, $c = 55.49$	$a = b = 72.38$, $c = 54.77$
Space group	$P3_221$	$P3_221$	$P3_221$
Resolution (Å)	25−1.8	25−2.5	25−2.85
Total observations	208646	33143	13692
Unique reflections	16214	5940	4018
Multiplicity	12.9 (12.5)	5.6 (5.2)	3.4 (3.3)
I/σ	15.5	18.9	6.2
R_{sym} (%)[b]	6.9 (44.5)	8.4 (62.5)	22.5 (59.1)
Completeness (%)	100	99.7 (100)	98.6 (100)
	Refinement		
R_{cryst} (%)[c]/R_{free} (%)[d]	17.18/20.00	18.39/23.05	18.14/26.66
rmsd bonds (Å)[e]	0.011	0.024	0.021
rmsd angles (deg)[e]	1.423	2.298	1.986
Residues modeled	5−73	5−73	5−73
Average *B*-factor (Å^2)	24.36	42.52	24.2

[a]The numerical values in parentheses refer to the highest-resolution shells. [b]$R_{sym} = \Sigma_h\Sigma_i(|I_i(h) - <I(h)>|)/\Sigma_h\Sigma_i I_i(h)$, where $I_i(h)$ is the ith intensity measurement and $<I(h)>$ is the weighted mean of all measurements of $I(h)$. [c]$R_{cryst} = \Sigma_h||F_{obs}| - |F_{calc}||/\Sigma_h|F_{obs}|$, where R_{cryst} is evaluated by the summation of all reflections used in refinement, and $|F_{obs}|$ and $|F_{calc}|$ are the observed and calculated structure factor amplitudes. [d]$R_{free} = \Sigma_h||F_{obs}| - |F_{calc}||/\Sigma_h|F_{obs}|$, where R_{free} is evaluated by randomly choosing 5% of the diffraction data not included in refinement. [e]rmsd = root-mean-square deviation from ideal values.

that a protein selected entirely on the basis of its ability to bind ATP might also function as a primitive catalyst, we initiated a series of structural and biochemical experiments designed to explain the presence of ADP in our protein crystal structures.

RESULTS AND DISCUSSION

Structural Analysis of Protein DX. We began by crystallizing protein DX in the presence of saturating concentrations of free ATP ligand to determine whether the γ-phosphate of ATP might become visible in electron density maps obtained under conditions that favored ATP binding. We determined the structure of protein DX by X-ray crystallography (Figure 1, panel a and Table 1) and refined the atomic models against diffraction data extending to 1.8 Å resolution. As expected, the protein adopts the same zinc-nucleated α/β fold previously observed for this class of synthetic ATP binding proteins (*11, 12*). However, unlike the earlier structures, this new structure now contains definitive electron density in the ligand-binding pocket that maps to an entire ATP molecule (Figure 1, panel b). Close analysis of the protein reveals that ATP is bound in an unusual bent conformation (Figure 1, panel a, inset) that allows the phosphate

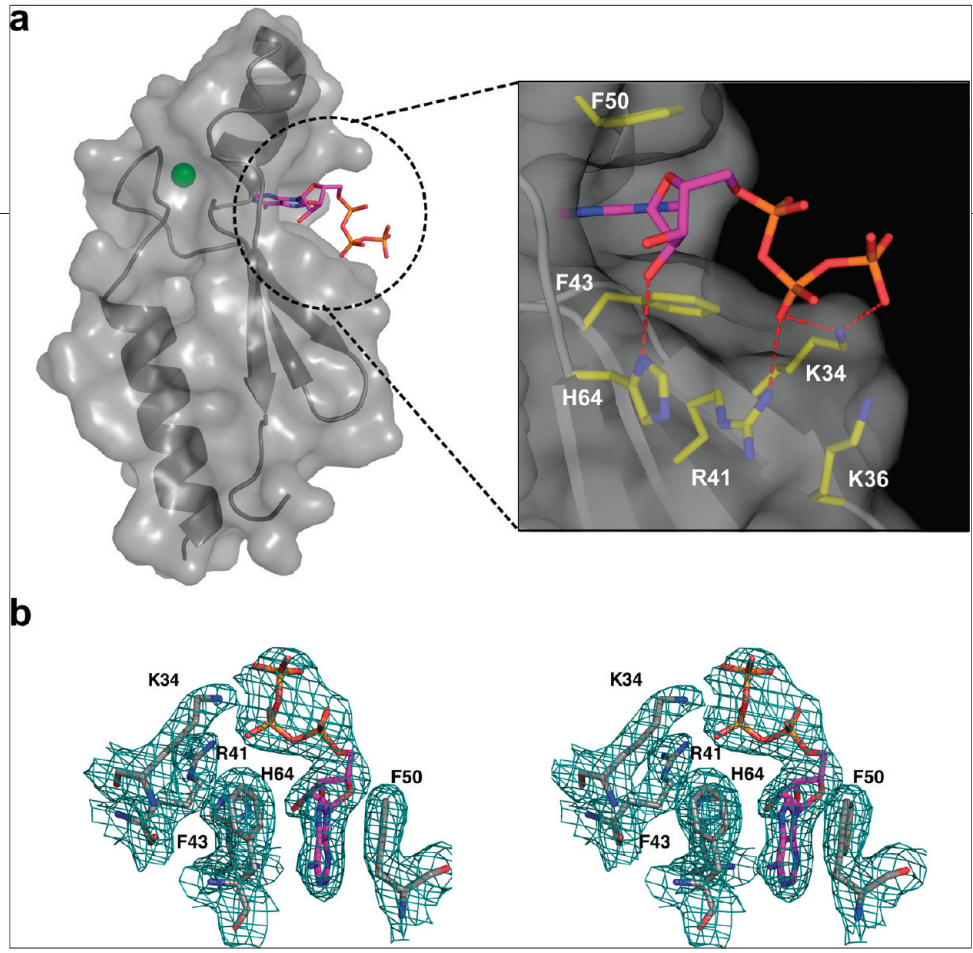

Figure 2. Adenosine binding site of the Y43F variant. a) Ribbon representation of the X-ray crystal structure of Y43F with translucent surface rendering. The bound zinc ion (green sphere) is shown at ~8.4 Å from the adenine nucleobase of the bound nucleotide (atom colors: magenta, carbon; red, oxygen; blue, nitrogen; and orange, phosphate). Inset is the structure of the ligand binding pocket of the protein. Depicted beneath the surface are important side chains (atom colors: yellow, carbon; red, oxygen; and blue, nitrogen) that directly interact with the adenosine nucleotide. Polar contacts are drawn with red dashes. b) Stereoview of the ligand-binding site of DX with $2F_o - F_c$ electron density contoured at 1.0 σ. Stick representations of select ligand binding site residues are shown with atom coloring of protein residues (gray, carbon; red, oxygen; and blue, nitrogen). Also shown are Lys34, Arg41, Phe43, Phe50, and His64.

backbone to form a strong intramolecular hydrogen bond (2.4 Å) between the γ-phosphate and 2′-OH on the sugar ring. Because of the unusual geometry of this ligand, we compared the bent ATP conformation to all ATP-bound proteins whose structures are available in the protein databank (PDB). This analysis revealed that the bound conformation differs significantly from the typical linear conformation observed in most natural proteins that bind ATP (*15*). Only a few structures, such as aspartyl-tRNA synthetase, contain an ATP molecule in a bent conformation that bears a resemblance to our bent ATP ligand (*16*). However, these structures require divalent metal ions and numerous hydrophobic and electrostatic contacts to constrain ATP in a bent geometry.

In contrast to natural ATP binding proteins, protein DX binds ATP in a conformation that exposes the sugar-phosphate backbone to bulk solvent. A key structural feature of this unusual ligand geometry (Figure 1) is the presence of a bound water molecule that forms a tight polar contact to the γ-phosphate of ATP and is stabilized with hydrogen bonds to amino acid side chains Tyr43, Lys34, and Arg41. This feature suggests a possible involvement of the coordinated water molecule in organizing and maintaining the bent ATP geometry. To test the hypothesis that removal of this interaction would allow us to obtain an ATP bound structure under conditions that previously yielded ADP in the ligand binding pocket, we mutated Tyr43 to phenylalanine (Y43F mutant) and crystallized the mutant protein in the presence of low concentrations of ATP. Protein crystals were obtained that diffracted to a resolution limit of 2.5 Å (Table 1). The structural model (Figure 2) contained electron density in the ligand-binding pocket consistent with a traditional linear conformation of ATP with all three phosphates clearly visible in the electron density map. This

result demonstrates that the Y43F mutation in protein DX produces a null mutant that is unable to bind ATP in the bent geometry and is incapable of ATP hydrolysis.

Mass Spectrometry Analysis of Protein DX Crystals. We used high-resolution mass spectrometry to demonstrate that ADP is produced when protein DX is crystallized with 1 equiv of ATP. Crystals of protein DX were grown for 3 days in the presence of 1 equiv of ATP, washed with mother liquor, dissolved in water, and analyzed by MALDI-TOF mass spectrometry. The resulting mass spectrum (Figure 3) shows a strong molecular ion peak for ADP, which demonstrates that ADP is produced when protein DX is crystallized with 1 equiv of ATP. In addition to the ADP product, a small amount of ATP is also observed in the spectrum, which might represent the fraction of unreacted ATP or the portion of ATP that remains in the solvent channels of the crystals. While it is difficult to distinguish these two possibilities, the high occupancy of ADP in the crystal structures of proteins 18-19 and DX indicate that ADP is the dominant ligand bound to the protein (*11, 12*). We also analyzed the high mass portion of the spectrum, which reveals that protein DX remains unmodified after crystallization. This demonstrates that the γ-phosphate is not transferred to a side chain in the ATP binding pocket of protein DX. To eliminate the possibility that ATP hydrolysis was an artifact of ionization or the crystallization process, control experiments were performed in which the Y43F null mutant was crystallized in the presence of 1 equiv of pure ATP and protein DX was crystallized with 1 equiv of pure ADP. The resulting spectra show that both proteins remain bound to ATP and ADP, respectively (Figure 3). The absence of any significant ADP in the mass spectrum of the Y43F mutant supports the hypothesis that the null mutant is incapable of ATP hydrolysis.

We crystallized protein DX bound to pure ADP and solved its three-dimensional structure by Fourier synthesis (Table 1). As expected, the electron density of the bound ADP ligand is consistent with the presence of ADP in the ligand binding pocket (Figure 4, left panel) and the entire ligand is clearly visible in the electron density map. We then compared the ligand binding mode that results when pure ADP is bound to protein DX with the ADP conformation that is obtained when protein DX is crystallized with 1 equiv of ATP. The ADP molecule observed in the pure ADP bound structure adopts the same conformation and forms the same hydrophobic and electrostatic contacts to the protein as the ADP bound structure obtained when protein DX is crystallized with low concentrations of ATP (Figure 4, right panel). This comparison demonstrates that the two

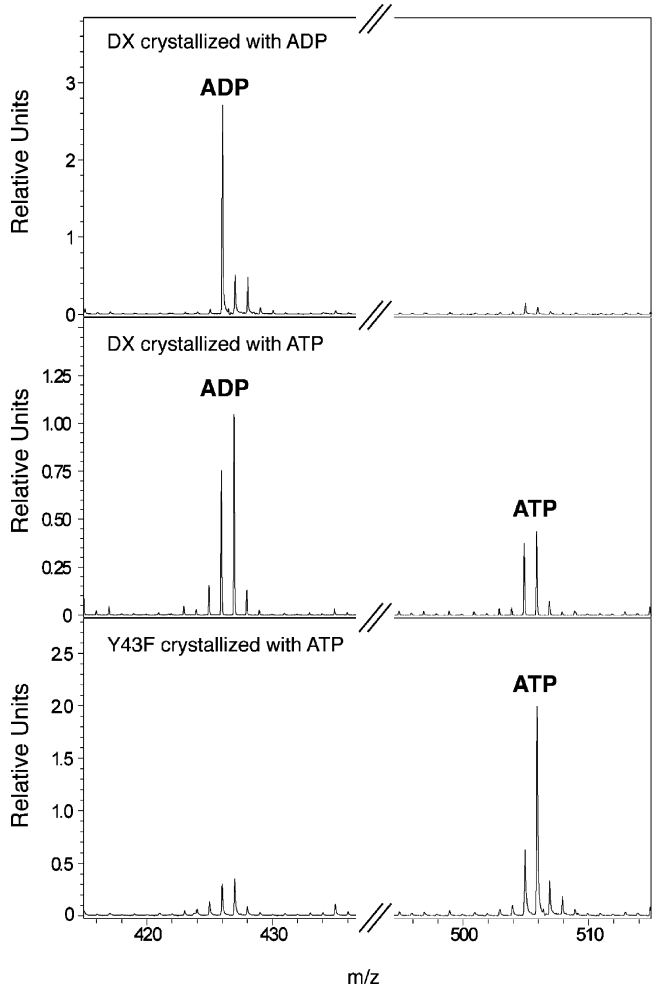

Figure 3. High-resolution MALDI-TOF mass spectrometry analysis of dissolved protein crystals. Protein DX co-crystals obtained in the presence of 1 molar equiv of pure ADP (top) and pure ATP (middle) and Y43F crystallized with pure ATP (bottom) were dissolved and analyzed by MALDI-TOF mass spectrometry. The middle panel shows that a substantial amount of ADP (calculated MH^-, 426.022) is present in the protein crystal when protein DX is crystallized with 1 equiv of ATP (calculated MH^-, 505.988). The ratio of ATP to ADP in this experiment is not 1:1 as a result of differences in the ionization potential of ATP and ADP. The presence of a small amount of ADP in the Y43F control is consistent with a small amount of ADP contaminant in pure ATP solutions, and a no protein control reveals a comparable amount of ADP in the mass spectrum.

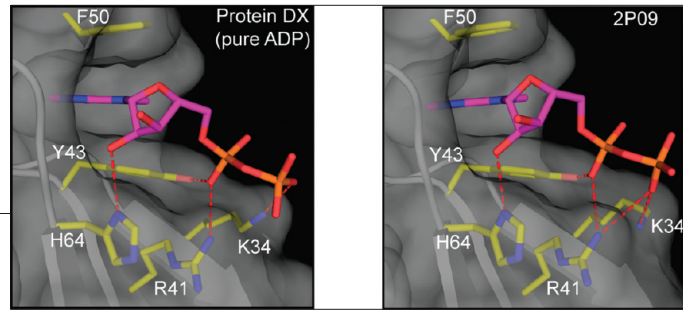

Figure 4. Protein DX crystallized in the presence of pure ADP yields a crystal structure identical to the same protein crystallized in the presence of low ATP. The bound ADP (atom colors: magenta, carbon; red, oxygen; blue, nitrogen; orange, phosphate) is shown. Depicted beneath the translucent gray surface of the protein are important side chains (atom colors: yellow, carbon; red, oxygen; blue, nitrogen) that are critical for ADP binding. Relevant polar interactions with the ligand are shown by red dashes. The right panel shows the original DX structure (PDB accession code 2P09) for purposes of comparison.

ADP molecules are indistinguishable by protein X-ray crystallography.

Solution Binding Affinity and Specificity. To examine the possibility that the ADP-bound crystal structures were a result of preferential binding of protein DX to ADP, we determined the ligand binding preferences of our synthetic proteins for ATP and ADP. Using equilibrium filtration (*17*), we measured the solution binding affinity (K_d) of proteins 18-19, DX, and DX (Y43F) as maltose binding protein fusions (Table 2). We have used this technique previously to evaluate the binding affinity of other *de novo* evolved ATP binding proteins (*11, 18*). Proteins 18-19 and DX bind ATP with K_d values of 1200 and 250 nM, respectively, and both proteins favor ATP over ADP by up to 2.5-fold in specificity. This result shows that unless a large amount of ADP is produced during the crystallization period, it is unlikely that the resulting ADP bound structure is due to preferential binding of ADP by either protein.

Comparative ligand binding experiments performed on protein DX (Y43F) in which the critically important tyrosine residue is replaced with a phenylalanine side chain indicate that the null mutant has high affinity and good selectivity for ATP. Protein DX (Y43F) binds ATP with a K_d of 180 nM and discriminates against ADP by 23-fold (Table 2). The higher affinity and selectivity of the Y43F mutant protein is due to the linear conformation of the ATP molecule in the ligand binding site (Figure 2), which in addition to several new contacts to the γ-phosphate, maintains all of the electrostatic contacts previously observed for ADP. Thus, the specificity observed for the Y43F mutant protein is due entirely to the presence of new electrostatic contacts formed between the protein and the γ-phosphate of ATP.

TABLE 2. Selectivity of binding and dissociation constants (K_d) of Family B variants

	K_d (nM)[a]		
Variant	ATP	ADP	Selectivity[b]
18-19	1200 ± 200	1400 ± 180	1.2
DX	250 ± 9	620 ± 81	2.5
DX (Y43F)	180 ± 30	4200 ± 760	23

[a]Standard deviation was calculated using three or more measurements. [b]Calculated using the ratio of K_d for ADP *versus* ATP.

Solution Activity. To examine whether protein DX could engage ATP in a hydrolysis reaction in solution, we incubated trace amounts of α-P^{32}-labeled ATP in crystallization buffer that contained protein DX. A no protein control was performed in parallel to examine the composition of the ATP molecule under the conditions and time required for protein crystallization. After 3 days, aliquots were removed from each reaction mixture, and the contents were analyzed by thin-layer chromatography (TLC). The resulting image (Figure 5) shows that ATP obtained from the no protein control migrates as a single spot with the expected mobility of free ATP

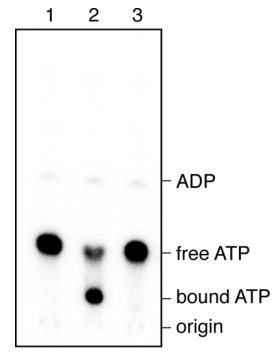

Figure 5. Analysis of ATP by thin-layer chromatography. Thin-layer chromatography analysis of α-P^{32}-labeled ATP incubated for 3 days in protein crystallization buffer (100 mM Na_2HPO_4 (pH 8.5), 250 mM sodium citrate, 300 mM NaCl, and 19–24% PEG-400). P^{32}-labeled ATP was incubated in the absence (lane 1) and presence of protein DX (lanes 2 and 3). In lane 2, the reaction mixture was spotted directly onto the TLC plate. In lane 3, the reaction mixture was incubated with 1 mM cold ATP prior to spotting on the plate. The absence of ADP in the TLC indicates that ADP is not an artifact of a nonenzymatic or disproportionation reaction.

(lane 1). In contrast, ATP that was incubated with the protein produces two spots on the TLC plate. The more polar spot exhibits a mobility that is similar to the free ATP spot, while the less polar spot migrates more slowly (lane 2). When an aliquot from this mixture is combined with excess nonradioactive ATP, the slower moving spot disappears and the free ATP spot increases in abundance (lane 3). We determined that the slower moving spot was ATP bound to the protein by analyzing the spot by mass spectrometry. This experiment validates the mass spectroscopy data obtained on dissolved protein crystals by demonstrating that the ADP ligand observed in the crystal structure of protein DX was not the result of nonenzymatic hydrolysis or disproportionation of ATP since ADP was not observed under the conditions used to crystallize the protein.

The contrasting behavior of the protein when it is in solution *versus* when it is in the crystalline form suggests that the protein's environment plays an important role in the hydrolysis reaction. Close inspection of the crystal lattice reveals that the ATP ligand is located in the solvent channels, which suggests that the packing interactions between neighboring proteins are not responsible for ATP hydrolysis. One interpretation of this result is that, in solution, ATP is able to adopt multiple conformations when bound to protein DX, one of which is the bent conformation whose lifetime is short relative to the time required for hydrolysis. During protein crystallization, the equilibrium between the different bound states shifts to favor the bent conformation, which is observed in the crystal structure of protein DX obtained in the presence of saturating amounts of ATP. Since the bent conformation is believed to be essential for hydrolysis, it is possible that this reaction is limited to the crystalline environment where the bent conformation is stabilized by a network of well-ordered water molecules. Because the reaction is very slow relative to natural enzymes, it is difficult to detect ADP when the protein is crystallized in the presence of higher ATP concentrations, as any ADP product formed would be displaced by excess free ATP present in the mother liquor.

Significance of ATP Hydrolysis. *In vitro* selection experiments provide a powerful method for reconstructing scenarios that describe the evolutionary path that enabled primordial proteins to achieve catalytic function. Up until now, most studies have focused on the role of conformational diversity and functional promiscuity in the evolution of divergent functions from existing protein scaffolds (*19, 20*). By applying the principles of Darwinian evolution to large pools of random sequences it is possible to survey the functional landscape of the protein universe in a way that is unbiased by biology and therefore better able to answer questions about what exists outside of nature (*21*). While the immensity of protein sequence space (20^L, where L is the chain length) suggests that the emergence of a functional protein from a stochastic pool of sequences is a highly improbable event, the *de novo* evolution of this ATP binding protein demonstrates that folded proteins with desired functions occur more frequently than previously thought (*10*). Through a combination of directed evolution and structure determination, we have used this class of synthetic proteins as a model for studying the diversity of folds that are available in the protein universe, and the role that subtle mutations play in improving protein folding stability and solubility (*11, 18, 22*). These studies led to the observation that ADP is observed in our crystal structures whenever equivalent amounts of ATP were added to the mother liquor (*11*).

In the present study, we provide structural and biochemical evidence that suggests that the bound ADP molecule is a result of the protein's ability to bind ATP in a bent conformation in which the terminal phosphate is within van der Waals contact of the adenosine ring. Surprisingly, this conformation is maintained by a key water molecule that bridges the γ-phosphate of ATP with the amino acid residues Tyr43, Lys34, and Arg41 located in the ligand binding pocket of the protein. We demonstrate that the bent conformation is important for hydrolysis and show that a null mutant, which removes the bridging water interaction from the structure is incapable of ATP hydrolysis. We speculate that the bent conformation causes the γ-phosphate to be more electrophilic than normal, which makes ATP susceptible to attack by a nearby water molecule. While many additional experiments are needed to determine the precise reaction mechanism, a superposition of the linear and bent ATP conformations suggests that protein DX binds ATP in a constrained geometry.

The crystal structure of protein DX showing ATP bound to the protein in an unusual bent conformation (Figure 6, panel a, left) suggests that constrained ligand binding geometries might represent one of the simplest ways that primordial enzymes could have achieved primitive catalytic function. Indeed, many natural proteins bind their substrates in high-energy

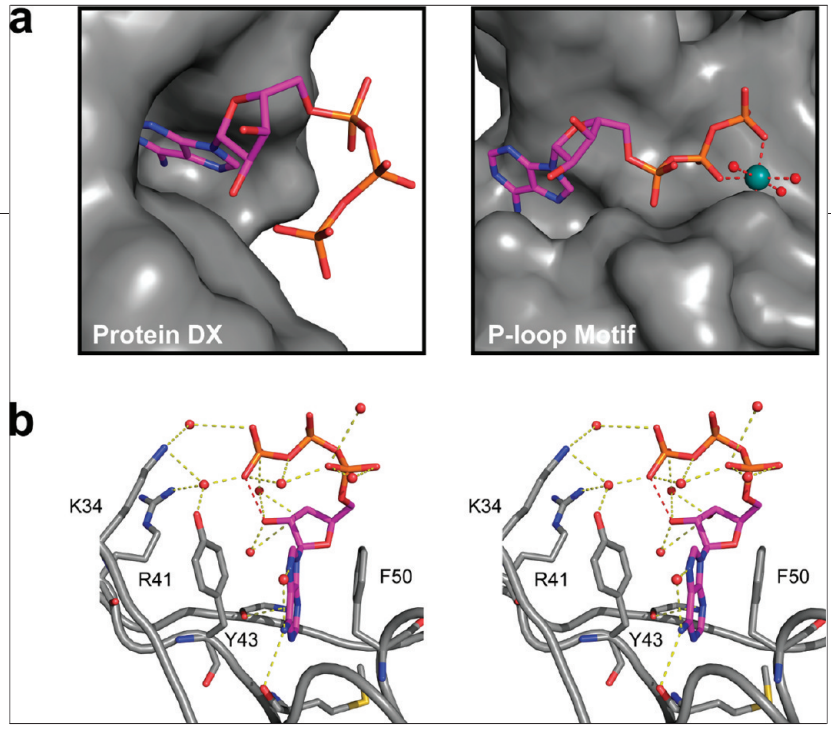

Figure 6. Comparison of the ATP binding site of protein DX with a modern P-loop motif. a) Surface rendering of the solvent-exposed ligand binding pocket of protein DX (left) compared with the canonical P-loop ATP binding domain of human thymidylate kinase (right; PDB accession 1E2Q). In contrast to modern ATP processing enzymes, protein DX maintains a solvent-exposed binding site devoid of divalent metal ions. Also shown is the hexacoordinated magnesium ion (teal), chelated by ligand, protein, and water (red spheres) interactions (red dashes). b) Stereoview of solvent shell surrounding the bent ATP conformer (atom colors: magenta, carbon; red, oxygen; blue, nitrogen; orange, phosphate) with main and side chains represented with atom coloring (gray, carbon), and relevant polar interactions indicated with yellow dashes. Structural solvent molecules are shown (red spheres). Loop regions containing the ligand binding residues are shown as a cartoon representation.

conformations that differ significantly from their ground state conformations (*23*). Critical to this hypothesis is the observation that the hydrolysis step proceeds without assistance from a divalent metal ion cofactor, which to our knowledge are always required of modern enzymes that hydrolyze ATP (*24, 25*). This suggests that early enzymes may have utilized water molecules to attain primitive catalytic functions that were later superseded by more complex metal-dependent mechanisms. Modern ATP binding motifs, such as the common P-loop motif (Figure 6, panel a, right), bind ATP in a metal-dependent linear conformation that is often buried inside a hydrophobic cleft. In contrast, our synthetic protein binds ATP in a bent geometry using an intricate network of water molecules (Figure 6, panel b) to constrain the ligand outside the confines of a hydrophobic cleft and without the aid of divalent metal ions. These features highlight some of the possible structural changes that early proteins might have undergone in their evolution to contemporary biological catalysts (*26*).

It is interesting to note that the ligand binding motif observed in the structure of protein DX is qualitatively similar to the ATP binding motif found in the Sassanfar aptamer (*27, 28*). This is an RNA aptamer that was evolved from a similar size pool of random sequences to bind ATP. Both structures bind ATP in a conformation that buries the nucleobase inside a hydrophobic pocket and exposes the sugar and phosphate moieties to bulk solvent. The parallel between the evolutionary history and ligand binding motifs of these two structures indicate that early macromolecular structures likely adopted very simple folds with very simple functions. Then, over the course of time these small scaffolds were able to recombine in different ways to create larger, more elegant structures with very sophisticated functional properties. Evidence to support such a progression can be found in the large number of protein structures available in the PDB that derive from a small set of unique single-domain protein folds (*29*).

One implication of the current study is that within the context of sequence space and the fitness landscape certain catalytic functions might not be that much harder for an amino acid sequence to achieve than binding. This hypothesis is consistent with the general history of nucleic acid selections, which have shown that ribozymes with relatively undemanding functions occur in pools of sequences with frequencies that are similar to many aptamers (*30, 31*). Because we were able to identify a protein with weak chemical reactivity from a selection in which no intentional selective pressure was applied for catalysis, it is possible that proteins that fold into structures with catalytic activity might be as common or nearly as common as proteins that fold into structures with ligand binding activity. However, the challenges of developing efficient selection strategies

needed to isolate these catalysts emphasizes the magnitude of the problem for those attempting to evolve protein enzymes *de novo*.

Conclusion. The combination of *de novo* protein evolution with structure determination provides a powerful approach to establishing an unbiased view of the structural and chemical diversity available in the protein universe. Our characterization of a synthetic protein that derives entirely from random-sequence origin demonstrates that design-free methods can be used to generate proteins with novel functions. We show that a protein selected solely on the basis of its ability to bind ATP emerged with the ability to hydrolyze ATP. This achievement suggests that relatively undemanding catalytic reactions may not have been that much harder for primordial proteins to attain than ligand binding. This indicates that it should be possible to evolve proteins with simple catalytic functions from large pools of random sequences in a manner similar to the way in which RNA enzymes are selected today (*32*). Initial progress toward this goal has already been achieved with the successful evolution of an RNA ligase from a nonfunctional, zinc-finger protein scaffold (*4*). Further experiments to examine the generality of this observation provide exciting opportunities to search protein sequence space for new protein folds and catalysts.

METHODS

Expression, Purification, and Crystallization. Protein DX was purified from *E. coli* as described previously (*11*). The Y43F point mutation was introduced using a QuikChange site-directed mutagenesis kit (Stratagene). Pure DX protein was concentrated using a centricon filter device (Millipore) to a final volume of ~250–300 μL, which gave a final concentration of ≥20 mg mL^{-1}. Synthetic ATP binding proteins were crystallized using the sitting drop vapor diffusion method by combining 1.5 μL of protein solution with an equivalent volume of reservoir solution containing 0.1 M sodium phosphate pH 8.5, 0.25 M sodium citrate, 0.3 M sodium chloride, and 19–24% polyethylene glycol 400. Co-crystals of protein DX and ATP were obtained by crystallizing the protein in drops containing 100 mM ATP. Co-crystals of protein DX and ADP were obtained by crystallizing the protein in drops containing 100 mM ADP. Co-crystals of protein Y43F with ATP were obtained by crystallizing Y43F with 10 mM ATP. Small hexagonal crystals appeared within hours and grew to full size (0.075 mm × 0.075 mm) within 72 h.

Crystallographic Data Collection and Refinement. Data collection was performed at the Arizona State University X-ray Diffraction Facility home source on an RAXIS IV detector at the Cu Kα wavelength of 1.54 Å. The crystals were determined to be isomorphic with the DX protein crystals described previously and belong to the space group $P3_22_1$ (*11*). All data were processed using the HKL package (*33*). Data quality statistics are summarized in Table 1. All protein structures were determined by Fourier synthesis with model building performed in Coot (*34*), with alternating rounds of refinement using the CCP4 suite of programs (*35*). Coordinates and structure factors have been deposited at the RCSB for immediate release upon publication. Protein DX crystallized with ATP and pure ADP were assigned the PDB codes 3DGL and 3DGN. Protein Y43F crystallized with ATP was assigned the PDB code and 3DGO.

Mass Spectrometry. Proteins DX and Y43F were crystallized with 1 mM ATP. Protein DX was also crystallized with 1 mM ADP. Drops that contained a uniform shower of crystals were selected. All mother liquor was removed from the drop, and the crystals were washed three times with 3 μL of mother liquor devoid of ATP and dissolved in water and spotted using 2,5-dihydroxybenzoic acid as the matrix. Linear mass spectra were acquired in negative reflectron mode in the low molecular weight regime and positive reflectron mode in the high mass (protein) regime using an Ultraflex III TOF/TOF MS instrument (Bruker Daltonics, Billerica, MA). The resulting mass spectra were analyzed using Flex Analysis software (Bruker Daltonics).

Determination of Equilibrium Dissociation Constants. Equilibrium dissociation constants (K_d) for purified MBP fusion proteins were measured by equilibrium ultrafiltration as described previously (*11*). Apparent K_d's were measured using trace amounts of γ-[^{32}P]-ATP (Amersham Biosciences) and a series of concentrations of MBP fusion protein spanning the $K_{d\ app}$. Apparent K_d's for ADP were measured using trace amounts of α-[^{32}P]-ADP produced by incubating γ-[^{32}P]-ATP with T4 polynucleotide kinase and DNA for 1 h at 37 °C. The α-[^{32}P]-ADP product was separated from the reaction mixture by passing the solution through a microcon spin filter with a 2 kDa membrane. Quantitative conversion of ATP to ADP was verified by TLC.

Thin-Layer Chromatography Analysis. Solution-based activity was measured under standard crystallographic conditions in 5 μL of buffer that contained 0.2 μL of α-[^{32}P]-ATP, 100 mM sodium phosphate (pH 8.5), 250 mM sodium citrate, 250 mM NaCl, 20% PEG-400, and protein DX (20 mg mL^{-1}). A no protein control was performed in the absence of protein DX to examine possible nonenzymatic hydrolysis or disproportionation. After 3 days, an aliquot (1 μL) was removed from both reaction mixtures and diluted 10-fold in water, and 1 μL of this solution was spotted onto silica gel TLC plates (Sorbent Technologies). An additional aliquot (1 μL) was removed from the protein DX reaction and diluted 10-fold in water containing 1 mM cold ATP, and 1 μL of this solution was spotted onto the TLC plate. ATP was analyzed by running the TLC in 55% *n*-propanol, 10% ammonium hydroxide and 35% water, drying the plate at room temperature, and visualizing the image by phosphorimaging (GE Healthcare).

Acknowledgment: We would like to thank Jack Szostak, Steven Benner, David Eisenberg, and Sid Hecht for helpful comments and discussions. This work was supported by grants from the National Science Foundation (MCB 0821032) and the New Frontiers Program at the Biodesign Institute at ASU. We would like to acknowledge the protein structure facility and Craig Magee at ASU for help in data collection and Matthew Greving for assistance with mass spectrometry.

REFERENCES

1. Petsko, G. A., Kenyon, G. L., Gerlt, J. A., Ringe, D., and Kozarich, J. W. (1993) On the origin of enzymatic species, *Trends Biochem. Sci. 18*, 372–376.

2. Gaucher, E. A., Thomson, J. M., Burgan, M. F., and Benner, S. A. (2003) Inferring the paleoenvironment of ancient bacteria on the basis of resurrected proteins, *Nature 425*, 285–288.
3. Ortlund, E. A., Bridgham, J. T., Redinbo, M. R., and Thornton, J. W. (2007) Crystal structure of an ancient protein: evolution by conformational epistasis, *Science 317*, 1544–1548.
4. Seelig, B., and Szostak, J. W. (2007) Selection and evolution of enzymes from a partially randomized non-catalytic scaffold, *Nature 448*, 828–831.
5. Rothlisberger, D., Khersonsky, O., Wollacott, A. M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J. L., Althoff, E. A., Zanghellini, A., Dym, O., Albeck, S., Houk, K. N., Tawfik, D. S., and Baker, D. (2008) Kemp elimination catalysts by computational enzyme design, *Nature 453*, 190–195.
6. Park, H.-S., Nam, S.-H., Lee, J. K., Yoon, C. N., Mannervik, B., Benkovic, S. J., and Kim, H.-S. (2006) Design and evolution of new catalytic activity with an existing protein scaffold, *Science 311*, 535–538.
7. Benner, S. A., Caraco, M. D., Thomson, J. M., and Gaucher, E. A. (2002) Planetary biology-paleontological, geological, and molecular histories of life, *Science 296*, 864–868.
8. Joyce, G. F. (2002) The antiquity of RNA-based evolution, *Nature 418*, 214–221.
9. Szostak, J. W., Bartel, D. P., and Luisi, P. L. (2001) Synthesizing life, *Nature 409*, 387–390.
10. Keefe, A. D., and Szostak, J. W. (2001) Functional proteins from a random-sequence library, *Nature 410*, 715–718.
11. Smith, M. D., Rosenow, M. A., Wang, M., Allen, J. P., Szostak, J. W., and Chaput, J. C. (2007) Structural insights into the evolution of a non-biological protein: importance of surface residues in protein fold optimization, *PLoS ONE 2*, e467.
12. 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.
13. Hulett, H. R. (1970) Non-enzymatic hydrolysis of adenosine phosphates, *Nature 225*, 1248–1249.
14. Holler, E. (1992) The Chemistry of Dinucleotide Polyphosphates, in *Ap4A and Other Dinucleotide Polyphosphates* (McLennan, A. G., Ed.), p 14, CRC Press, Ann Arbor.
15. Stockwell, G. R., and Thornton, J. M. (2006) Conformational diversity of ligands bound to proteins, *J. Mol. Biol. 356*, 928–944.
16. Schmitt, E., Delarue, M., Poch, O., Gangloff, J., and Moras, D. (1998) Crystal structure of aspartyl-tRNA synthetase from *Pyrococcus kodakaraensis* KOD: archaeon specificity and catalytic mechanism of adenylate formation, *EMBO J. 17*, 5227–5237.
17. Carothers, J. M., Oestreich, S. C., and Szostak, J. W. (2006) Aptamers selected for higher-affinity binding are not more specific for the target ligands, *J. Am. Chem. Soc. 128*, 7929–7937.
18. Mansy, S. S., Zhang, J., Kummerle, R., Nilsson, M., Chou, J. J., Szostak, J. W., and Chaput, J. C. (2007) Structure and evolutionary analysis of a non-biological ATP-binding protein, *J. Mol. Biol. 371*, 501–513.
19. Tokuriki, N., and Tawfik, D. S. (2009) Protein dynamism and evolvability, *Science 324*, 203–207.
20. Chaput, J. C., Woodbury, N. W., Stearns, L. A., and Williams, B. A. R. (2008) Creating protein biocatalysts as tools for future industrial applications, *Expert Opin. Biol. Ther. 8*, 1087–1098.
21. Watkins, J. L., and Chaput, J. C. (2008) Searching combinatorial libraries for native proteins with novel folds, *ChemBioChem 9*, 1361–1363.
22. Chaput, J. C., and Szostak, J. W. (2004) Evolutionary optimization of a nonbiological ATP binding protein for improved folding stability, *Chem. Biol. 11*, 865–874.
23. Perola, E., and Charifson, P. S. (2004) Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding, *J. Med. Chem. 47*, 2499–2510.
24. Matte, A., Tari, L. W., and Delbaere, L. T. J. (1998) How do kinases transfer phosphoryl groups? *Structure 6*, 413–419.
25. Vetter, I. R., and Wittinghofer, A. (1999) Nucleoside triphosphate-binding proteins: different scaffolds to achieve phosphoryl transfer, *Q. Rev. Biophys. 32*, 1–56.
26. Narlikar, G., and Herschlag, D. (1997) Mechanistic aspects of enzymatic catalysis: lessons from comparison of RNA and protein enzymes, *Annu. Rev. Biochem. 66*, 19–59.
27. Dieckmann, T., Suzuki, E., Nakamura, G. K., and Feigon, J. (1996) Solution structure of an ATP-binding RNA aptamer reveals a novel fold, *RNA 2*, 628–640.
28. Sassanfar, M., and Szostak, J. W. (1993) An RNA motif that binds ATP, *Nature 364*, 550–553.
29. Chothia, C., Gough, J., Vogel, C., and Teichmann, S. A. (2003) Evolution of the protein repertoire, *Science 300*, 1701–1703.
30. Beaudry, A. A., and Joyce, G. F. (1992) Directed evolution of an RNA enzyme, *Science 257*, 635–641.
31. Ellington, A. D., and Szostak, J. W. (1990) *In vitro* selection of RNA molecules that bind specific ligands, *Nature 346*, 818–822.
32. Wilson, D. S., and Szostak, J. W. (1999) *In vitro* selection of functional nucleic acids, *Annu. Rev. Biochem. 68*, 611–647.
33. Otwinowski, Z., and Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307–326.
34. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr., Sect. D 60*, 2126–2132.
35. Project, C. C. (1994) Number 4. The CCP4 suite-programs for protein crystallography, *Acta Crystallogr., Sect. D 50*, 760–763.