Chemistry & Biology *13*, 139–147, February 2006 DOI 10.1016/j.chembiol.2005.10.015

Directed Evolution of ATP Binding Proteins from a Zinc Finger Domain by Using mRNA Display

Glen S. Cho[1,2] and Jack W. Szostak[1,*]
[1] Howard Hughes Medical Institute
Department of Molecular Biology and
Center for Computational and Integrative Biology
Massachusetts General Hospital
Boston, Massachusetts 02114

Summary

Antibodies have traditionally been used for isolating affinity reagents to new molecular targets, but alternative protein scaffolds are increasingly being used for the directed evolution of proteins with novel molecular recognition properties. We have designed a combinatorial library based on the DNA binding domain of the human retinoid-X-receptor (hRXRα). We chose this domain because of its small size, stable fold, and two closely juxtaposed recognition loops. We replaced the two loops with segments of random amino acids, and used mRNA display to isolate variants that specifically recognize adenosine triphosphate (ATP), demonstrating a significant alteration of the function of this protein domain from DNA binding to ATP recognition. Many novel independent sequences were recovered with moderate affinity and high specificity for ATP, validating this scaffold for the generation of functional molecules.

Introduction

In nature, certain protein frameworks, or scaffolds, have been used as starting points for the evolution of many different proteins with novel functions. For example, the immunoglobulin fold is used for molecular recognition of extremely diverse targets, while the $(\beta/\alpha)_8$ barrel fold is shared by hundreds of functionally diverse enzymes [1, 2]. Scaffolded libraries have also proven to be a rich resource for the directed evolution of new functional molecules in the laboratory [3–10]. Novel functional proteins have been recovered from both antibody-based libraries and from libraries based on a number of other protein scaffolds. Several factors are driving the search for alternatives to the antibody scaffold, including the complications associated with producing a multidomain multimeric protein with multiple intra- and interchain disulfide bonds. Single chain antibodies have been used as a scaffold for selections; however, producing large amounts of soluble single chain antibodies remains difficult [11–14]. Simpler scaffolds may lead to recombinant proteins that are easier to produce than antibodies and may be better suited for certain functions, such as catalysis.

*Correspondence: szostak@frodo.mgh.harvard.edu
[2] Present address: Department of Chemistry, Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, Massachusetts 02114.

A number of different display methods have been employed in combination with new scaffolds to evolve interesting and useful new proteins [9, 11, 15–19]. Phage display has been the most popular method for selecting functional molecules from new scaffolds; the following are examples of successful experiments. The lipocalin fold was re-engineered into a scaffolded library by randomizing specific residues that line its ligand binding cavity; proteins that specifically bind to small-molecule targets have been recovered from this library [15, 20]. Libraries generated by randomizing specific surfaces of proteins, such as protein A, tendamistat, and cytochrome b562, have yielded variants with the ability to bind to large target surfaces on proteins, but have been less successful when the target was a small molecule [6, 7, 17]. In addition, phage display has been used to select for variants of Zif268, a natural zinc finger DNA binding protein that can bind to specific trinucleotide sequences in order to build up transcription factors by using a modular approach [21, 22]. The fibronectin type III domain has been used to select affinity reagents to several protein targets [23, 24]. An alternative method, ribosome display, was used to evolve functional ankyrin repeats that could recognize maltose binding protein [25]. Yeast display has been used to optimize the function of an MHC class II molecule. So far, the only scaffold used in conjunction with mRNA display has been the fibronectin type III domain which was used to find affinity reagents to TNF-α [18].

We are interested in extending the discovery of new functional molecules from scaffolds by using mRNA display (Figure 1A) [26, 27]. The large library complexity attainable by this entirely in vitro method may allow the isolation of proteins with properties that are not easily accessible by other currently available methods. We synthesized a new scaffolded library based on the DNA binding domain of the retinoid-X-receptor (hRXRα, residues 132–208) because of the known structure, small size, and stable fold of this domain. hRXRα is a member of a highly conserved family of nuclear receptors that dimerize in response to ligand binding and induce gene expression [28]. The structure of the DNA binding domain has been solved by NMR by using a mutant version (C195A) with increased solubility [29]; we used this version for library construction. The hydrophobic core of the RXR fold is separated from the loops by two zinc ions, each of which is coordinated by four cysteine residues (Figure 1B). In the native protein, the first zinc finger contains a loop that makes nonspecific contacts with the DNA backbone while positioning an α helix to interact with the major groove of DNA. DNA binding induces a rigid conformation in the second zinc finger loop, which acts as a protein-protein dimer interface [30]. The first zinc finger loop was replaced with twelve random amino acid positions, while nine random positions were substituted for the second zinc finger loop. The finished library has a length of 96 amino acids, with 21 randomized positions (Figure 1C). Here, we demonstrate that novel small molecule binding proteins can be recovered from the randomized RXR library,

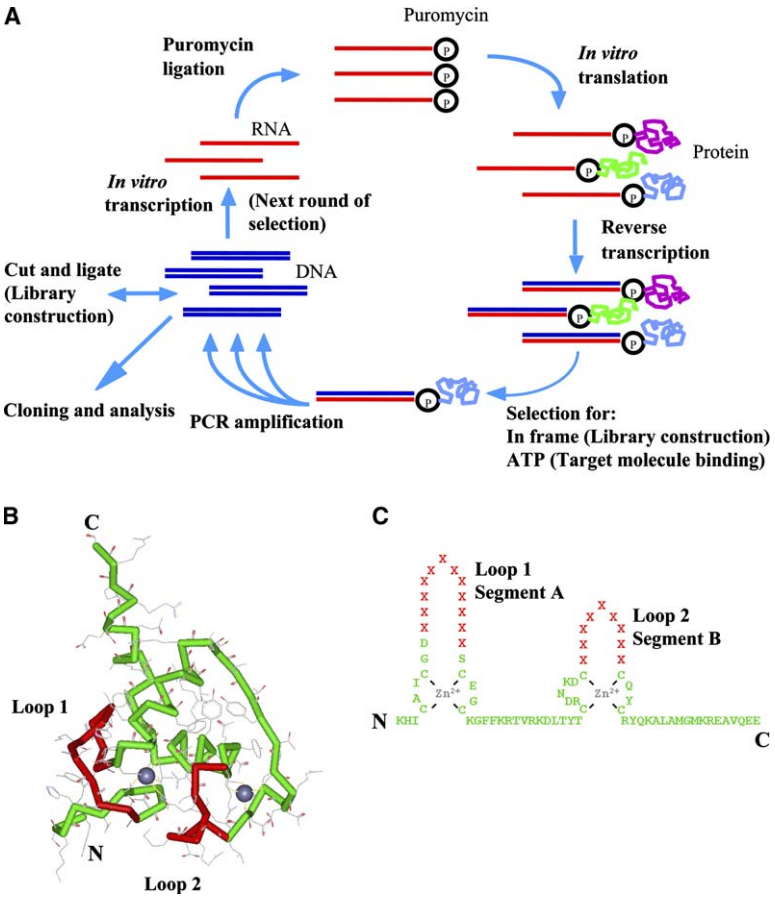


Figure 1. mRNA Display and the RXR-Scaffolded Library

(A) Steps involved in directed protein evolution using mRNA display. DNA is transcribed to RNA, which is then modified with puromycin. In vitro translation generates RNA-protein fusions that are then reverse transcribed. Functional binding proteins are separated based on affinity binding, or for library construction, on the basis of being completely in-frame. DNA sequences encoding the selected proteins are then amplified and used for the next round of selection, or, in the case of library construction, cut with restriction enzymes and ligated together to form the full-length library. The assembled library is then PCR-amplified in order to archive the diversity, and RNA is generated through in vitro transcription. At the end of the selection, individual clones are isolated and analyzed from the resultant DNA pool.

(B) The solution structure of the DNA binding domain of hRXRα (PDB code: 1RXR [27]) as displayed on Webviewer Lite (Molecular Dynamics). Red, loop residues that have been replaced with randomized amino acid positions; green, constant regions.

(C) Representation of the library at the primary sequence level. Gray lines represent the tetrahedral coordination of the two Zn^{2+} ions by the eight fixed cysteines.

showing that it is possible to significantly alter the natural function of a protein domain by a straightforward directed-evolution approach.

Results and Discussion

We created the RXR library by using a previously published protocol for making high-complexity combinatorial protein libraries through preselection of the encoded DNA segments [31]. The library was made in two halves—segments A and B—consisting of residues from the wild-type (WT) hRXRα, but with randomized loop regions along with additional N- and C-terminal tag sequences. Each segment was created from two deoxyribo-oligonucleotides by mutual primer extension. Degenerate positions in the sequence were used to code for random amino acid positions within the loops. N-terminal FLAG and C-terminal His_6 purification tags on the protein products from each segment allowed us to use mRNA display to enrich for open reading frames that are free of deletions, insertions, and stop codons by requiring continuous in-frame translation from N to C terminus (Table 1). Before preselection, only 39% of the sequences for segment A and 11% of segment B sequences were free of such imperfections; after preselection, 95% of segment A and 93% of segment B were error-free. Preselection increased the complexity of the final A-B combinatorial library by more than 20-fold by increasing the fraction of error-free full-length library molecules from ~4% to 88%. Based on the number of fusions generated and recovered, we estimate that the segment A library has a complexity of 3×10^{12} different molecules, while the segment B library contains 5×10^{11} unique molecules. These segments were amplified, cut, and ligated to form the full-length library. The final complexity is estimated as 1.4×10^{14} different molecules, based on the yield of DNA after ligation and assuming that each A-B combination is unique.

Table 1. Preselection Results from the Construction of the RXR Library

	Before Preselection			After Preselection			
	Fraction w/o Frameshift	Fraction w/o Stop Codons	Fraction Perfect	Fraction w/o Frameshift	Fraction w/o Stop Codons	Fraction Perfect	Compexity
Segment A	0.47 (17)	0.82	0.39	0.95 (20)	1.00	0.95	2.9×10^{12}
Segment B	0.15 (20)	0.70	0.11	0.93 (14)	1.00	0.93	5.3×10^{11}
Final library (A + B)			0.04			0.88	1.4×10^{14}

Parentheses denote number of sequences examined.

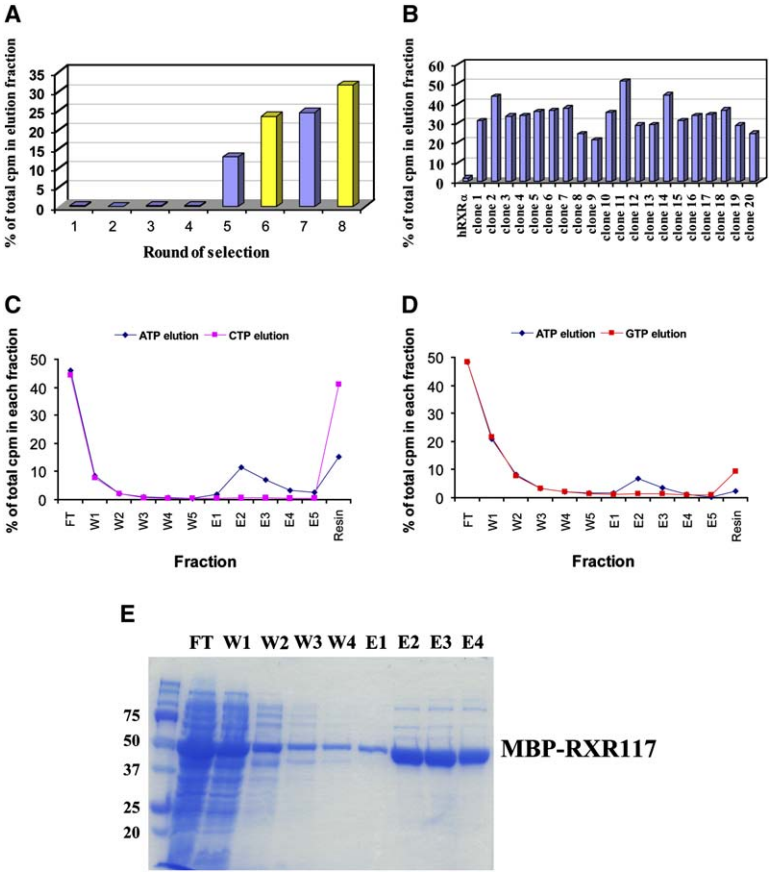


Figure 2. Selection for ATP Binding Proteins from the RXR-Scaffolded Library

(A) The percentage of [^{35}S]methionine-labeled RNA-protein fusions that bound to ATP-agarose and specifically eluted with ATP was measured for each round of in vitro selection. For rounds 6 and 8 (yellow), the output of the previous round was directly applied to the ATP-agarose after removing free ATP by using a FLAG affinity column, without amplification.

(B) Twenty clones from the eluted fraction of round 8 were translated in vitro as free proteins. These proteins were purified by using α-FLAG agarose and then bound to ATP-agarose. The percentage of labeled protein that bound and specifically eluted is shown for each clone and the parental hRXRα.

(C) The pool derived from round 8 was translated to yield free proteins, which were then bound to ATP-agarose. FT = flowthrough; W = wash; E = elution; numbers indicate fraction; Resin = percentage of cpms left on the resin. The bound material was eluted with either 5 mM free ATP with 5 mM $MgCl_2$ or 5 mM free CTP with 5 mM $MgCl_2$.

(D) Same as (C) but 5 mM ATP was compared with 5mM GTP.

(E) MBP-RXR117 was expressed in *E. coli* and the soluble fraction of the crude lysate was incubated with ATP-agarose and then specifically eluted with 5 mM free ATP and 5 mM $MgCl_2$. The sample was resolved by 10% SDS-PAGE and stained with Coomassie blue.

We examined the utility of the RXR-library as a source of proteins capable of specifically recognizing small molecules by using ATP as a selection target. The RXR-library, expressed as mRNA-displayed proteins, was incubated with ATP-derivatized agarose beads, and specifically bound material was eluted with free ATP. For round 1 of the selection, approximately 2.3×10^{13} RNA-protein fusion molecules, purified from a 10 ml translation reaction, were applied to an ATP affinity resin; 1 ml reactions were used in subsequent rounds. We monitored the amount of radioactively labeled protein retained on the column and specifically eluted in each round. In round 5, there was a significant increase in the amount of labeled protein specifically eluting with ATP (Figure 2A). By round 8 the binding and ATP-specific elution had increased to 32% of the total labeled input protein. Random clones from the post-round 8 pool, and WT hRXRα, were translated in vitro as free proteins and tested for their ability to bind to ATP-agarose (Figure 2B). All 20 selected clones, but not the WT protein, show strong ATP-agarose binding activity. We tested the ligand specificity of the selected pool by eluting the bound protein with different nucleotides. Neither cytidine triphosphate nor guanosine triphosphate (GTP) can elute the bound protein, indicating that the tested proteins specifically interact with ATP (Figures 2C and 2D). When these evolved proteins were fused to an unrelated protein, they could be purified from crude *Escherichia coli* cell lysates on the basis of ATP binding (Figure 2E).

The sequences of the clones discussed above are shown in Figure 3, and the multiple alignment reveals clear sequence patterns. The first random loop of every sequence begins with XVXH, where X is any amino acid. The second random loop of every sequence with the exception of clone 19 begins with XXCXX(Ar)H, where Ar is an aromatic residue. We confirmed by sequencing that these positions were truly random in the unselected pool (see Figure S1); thus, the conservation of these positions in the evolved pool is the result of some selective advantage. Apart from these invariant positions, the random loop sequences appear to be quite diverse, and are therefore of independent origin in the starting pool. The mutations in the nonrandomized portion of the scaffold are most likely due to errors that occurred during RT-PCR amplification. The sequences in the loops do not show any significant homology to known nucleotide binding proteins when compared by a BLAST search of the National Center for Biotechnology Information protein sequence database, nor do they contain signature motifs known to interact with nucleotides, such as the glycine-rich elements commonly found in proteins that interact with phosphates in nucleotides [32]. The large fraction of unique sequences observed (20 distinct sequences from 38 clones; 15 sequences occurred only once) suggests that many more active sequences exist in the selected pool.

One possible explanation for the presence of the invariant residues within the selected loop sequences is that these sequences all bind ATP in a similar manner,

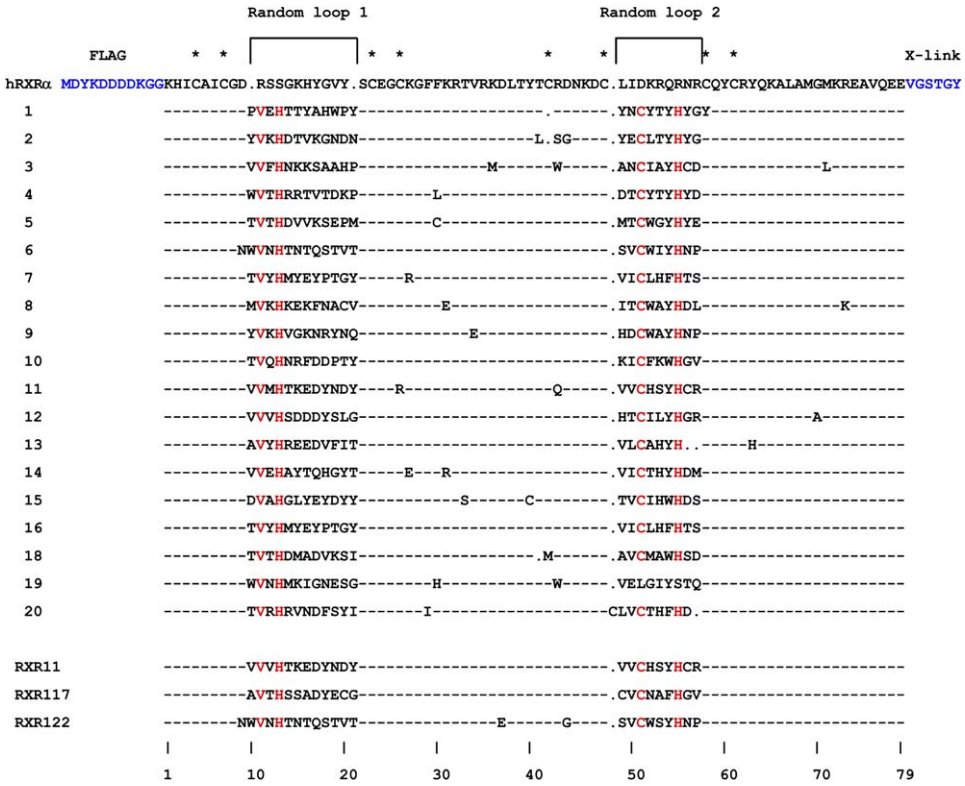


Figure 3. Sequences of Selected ATP Binding Proteins

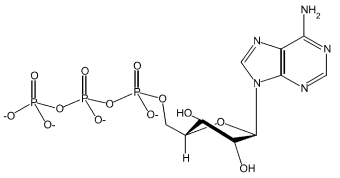
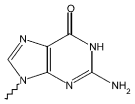
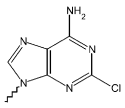
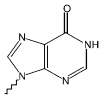
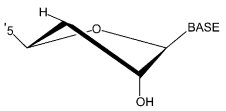
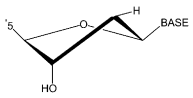
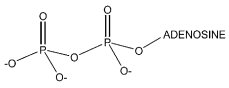
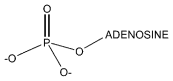
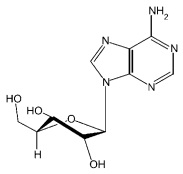
From 38 sequences from the post-round 8 pool, we observed 20 unique sequences. Sequence 17 is not included due to ambiguity of some of the amino acid positions. Blue, constant sequences added for the purpose of performing the selection; red, highly conserved positions among the selected sequences. The asterisks denote the conserved cysteines involved with zinc ion coordination. Numbering begins from the first native residue in the designed library. RXR11, RXR117, and RXR122 are clones that were further characterized.

and the invariant residues are making conserved structural contributions to the binding site. The fixed aromatic position is of interest because a previously selected ATP binding protein was shown to use aromatic residues to make π-π stacking interactions with the purine base of the ATP molecule [33]. This type of interaction is commonly seen in many ATP binding proteins [34]. Alternatively, some of these residues may play a structural role in stabilizing the overall protein fold. It will be interesting to compare selections for different ligands by using this RXR library to see if the same fixed positions are observed again. We examine the role of the cysteine in the second motif later in this article.

We chose three clones, RXR 11, RXR117, and RXR122, for further biochemical characterization based on their binding properties as heterologously expressed proteins. RXR11 is a variant of sequence 11 from the post-round 8 pool that was isolated after several more rounds of selection under more stringent conditions, and RXR122 is a variant of sequence 6 (Figure 3). RXR117 and RXR122 were discovered from a subsequent screen of post-round 8 sequences (see Supplemental Data) for soluble proteins by using fusions to green fluorescent protein (GFP) in *E. coli* [35]. In order to obtain concentrated protein samples for biochemical and biophysical characterization, the evolved sequences were expressed and purified as fusions with maltose binding protein (MBP). Surprisingly, many of the clones that show significant binding when expressed by in vitro translation do not bind when tested as purified MBP-fusions. It is possible that the MBP interferes with the folding of the fused protein, or that these proteins cannot be expressed in native form in bacteria. However, RXR11, RXR117, and RXR122 all demonstrate strong binding to ATP-agarose as MBP-fusions, whereas the WT hRXRα does not show any significant binding.

We characterized the molecular recognition of ATP in solution by MBP-fusion proteins of RXR11, RXR117, and RXR122 (Table 2 and Figure S2), by using spin-filtration competition experiments as described in the Experimental Procedures. All three fusion proteins have micromolar dissociation constants (K_ds) for ATP, consistent with the minimum required affinity for binding to the ATP-agarose column during the selection procedure. RXR11 has the strongest affinity to ATP of the three clones, with a binding constant of 12 μM. RXR122 shows an affinity of 23 μM for ATP and RXR117 has a binding constant of 47 μM. Substitutions in the nucleobase have the most dramatic effect on binding, with no binding observed to GTP, and greatly reduced affinity for inosine triphosphate (ITP) and 2-chloro ATP. RXR117 is the most specific, with its K_d being affected more than 100-fold for both the ITP and 2-chloro ATP analogs. RXR11 shows strong discrimination at the 2-position of at least 200-fold, but only a 20-fold difference at the

Table 2. Dissociation Constants of MBP-RXR Fusion Proteins to ATP and Analogs

		RXR11	RXR117	RXR122
ATP		$1.2 \pm 0.3 \times 10^{-5}$	$4.7 \pm 1.0 \times 10^{-5}$	$2.3 \pm 0.4 \times 10^{-5}$
Base				
GTP		$>5 \times 10^{-3}$	$>5 \times 10^{-3}$	$>5 \times 10^{-3}$
2-chloro ATP		2.3×10^{-3}	$>5 \times 10^{-3}$	5.0×10^{-4}
ITP		3.3×10^{-4}	3.6×10^{-3}	1.7×10^{-4}
Sugar				
3′ deoxy ATP		1.1×10^{-4}	5.8×10^{-4}	3.4×10^{-4}
2′ deoxy ATP		1.6×10^{-4}	1.1×10^{-3}	6.1×10^{-4}
Phosphates				
ADP		2.1×10^{-5}	5.0×10^{-5}	1.9×10^{-5}
AMP		3.8×10^{-5}	9.2×10^{-5}	6.9×10^{-5}
Adenosine		6.4×10^{-5}	1.2×10^{-4}	1.1×10^{-4}

6-position (ITP), while RXR122 is the least specific, with a only a 20-fold discrimination at the 2-position and a 7-fold difference at the 6-position.

The three proteins bind similarly to 2′- or 3′-deoxy-ATP, with all three recognizing 3′-dATP better than 2′-dATP. It is possible that the lower affinity for 2′-dATP is due to the change in the ribose pucker rather than lost interactions with the 2′ hydroxyl, since 2′-dATP tends to adopt a 2′ endo conformation, while 3′-dATP is preferentially 3′ endo [36].

The three proteins display similar discrimination patterns with respect to the phosphates. None show strong discrimination between ATP and adenosine diphosphate (ADP), suggesting that the γ-phosphate is not involved in any significant interactions. There is a gradual weakening of the binding affinity as the phosphates are removed in going from ADP to adenosine monophosphate (AMP) and then to adenosine.

RXR122 differs from the other two clones in that the binding of ATP is not Mg^{2+} dependent, as shown by

ATP-affinity column binding experiments with immobilized ATP (Figure S3A). The fraction of protein that binds to and is recovered from an ATP-agarose resin decreases 2-fold for both RXR11 and RXR117 in the absence of Mg^{2+}, but increases 2-fold for RXR122. These results and the analog affinities reveal that the three independent RXR clones differ in the details of how they recognize ATP. The clearest differences are in the discrimination of the purine ring, whereas more subtle differences are seen in the phosphate recognition patterns.

Given that the RXR-library is based on a two-zinc finger protein, we explored the role of zinc in forming functional ATP binding proteins by examining the sensitivity of binding to the presence of a metal chelator, ethylenediaminetetraacetic acid (EDTA), in amounts that would bind essentially all of the Zn^{2+} ions but not the Mg^{2+} ions. When 1 mM EDTA is added to in vitro translated protein during the binding incubation, no binding of either the pool or the individual cloned ATP binding proteins is detected. However, once a protein-ATP complex has formed, 1 mM EDTA does not disrupt the complex and elute the bound protein after a 1 hr incubation at 4ºC (Figure S3B). EDTA also inhibits the binding of *E. coli*-expressed MBP-RXR fusion proteins when added prior to binding (Figure S3C). Elemental analysis revealed that the MBP-fusion protein of the evolved ATP binding RXR clones and the WT hRXRα all contain similar amounts of zinc, while no zinc was detected in the control MBP (Table S1). These results are consistent with a continued requirement for zinc in ATP binding, as expected if the zinc fingers are still present in the evolved ATP binding proteins.

The presence of a conserved cysteine in the second loop suggests a possible rearrangement of the cysteine ligands for the second zinc ion. The original hRXRα protein uses a noncanonical spacing CX_5C, instead of the more common CXXC, to provide the first two cysteine ligands of the second zinc finger. This unusual structure has been preserved in other glucocorticoid family members, and the extra amino acids between the cysteines (called the D-box) are involved in creating a dimer interface with other DNA binding domains in certain conformations [37]. We tested the possibility that the invariant cysteine in the second randomized loop reconstitutes a canonical CXXC spacing of cysteine ligands for zinc coordination in the second zinc finger, since the selection did not require dimerization (Figure 4A).

In order to examine the roles of both the original cysteine and the new conserved cysteine in the random region, we made a series of site-directed mutations in each of the three characterized MBP-RXR fusions. The original cysteine C42 (corresponding to C171 in hRXRα) provides the first ligand in the coordination of the Zn^{2+} of the second zinc finger. However, mutating C42 to serine has no detrimental effect on ATP binding in any of the three ATP binding proteins (Figure 4B). In contrast, mutating the second cysteine (C51S, from the random region in the second loop) to serine completely abolishes ATP binding in all three evolved proteins (Figure 4B). These results are consistent with the hypothesis that C51 substitutes for C42 as one of the ligands for zinc coordination, such that the loss of binding that we observe is due to the loss of structure. We cannot rule out the possibility that C51 makes direct contact with the ATP

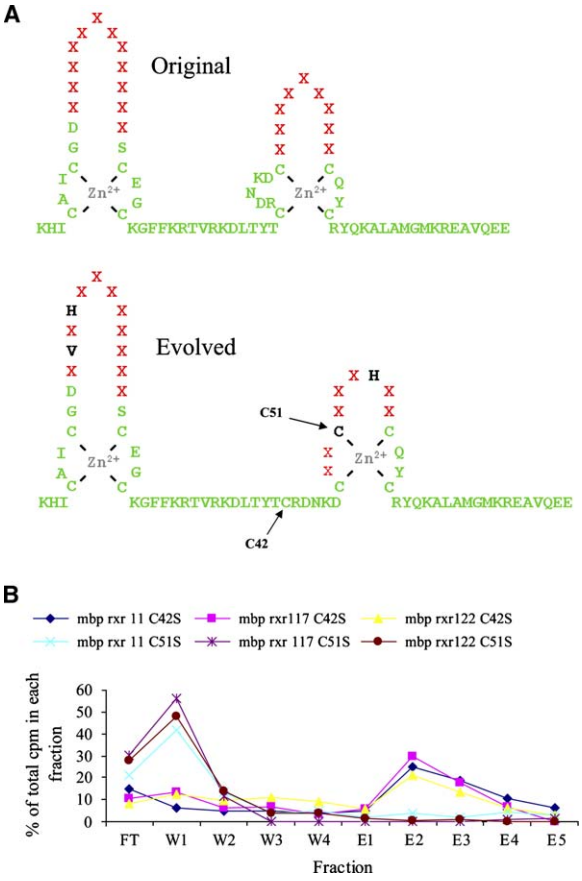


Figure 4. Role of Cysteines in the Selected Proteins

(A) Primary structure of the original library and proposed rearrangement of the selected RXR proteins. Green, regions kept constant in the library; red, random positions; black, positions that have been highly selected for in the random region among the selected sequences. The mutated cysteines are indicated.

(B) ATP binding properties of the C42S and C51S versions of RXR11, RXR117, and RXR122. Each mutant was expressed as an MBP-fusion, purified from *E. coli*, and then bound to ATP-agarose in 1× RSB. Proteins were washed and eluted with 5 mM free ATP and $MgCl_2$.

and that the serine is insufficient as a substitute. However even if this were the case, the second zinc finger would have to be substantially rearranged, given that C42 is not essential for the proper function of the evolved RXR ATP binding proteins.

We compared the biophysical properties of free (i.e., non-MBP-fused) ATP binding proteins expressed in *E. coli* to those of the WT hRXRα DNA binding domain in order to determine whether the evolved proteins retained the structural features of the parent molecule on which they were based. The unfused proteins precipitate over time in selection buffer in the absence of ATP, whereas hRXRα remains soluble. The addition of 5 mM ATP slows this precipitation, presumably by stabilizing the folded state of the proteins, or at least the recognition loops. If these loops are unstructured in the absence of ATP, intermolecular interactions between exposed loop residues may lead to aggregation. Gel filtration analysis of the dissolved protein demonstrates that the majority of the free protein elutes as a 10 kDa monomer for the evolved RXR clones (Figure 5A), but

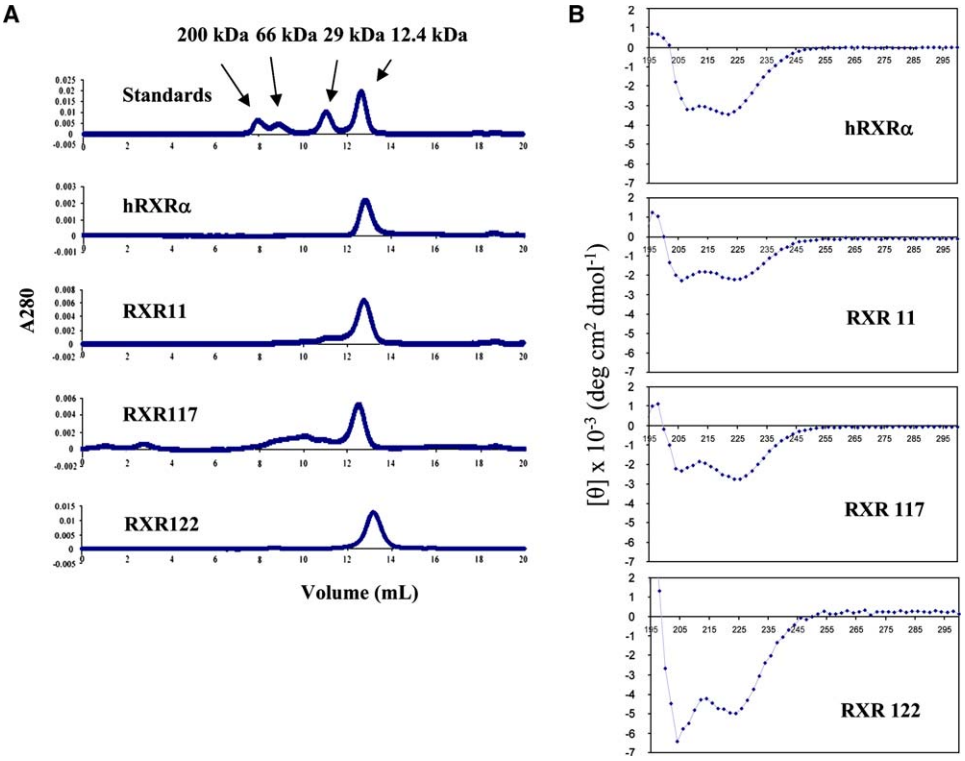


Figure 5. Biophysical Analysis of RXR Proteins

(A) Size-exclusion chromatography on Sephadex-75 of purified free RXR proteins. Monomer molecular weights are ~11.5 kDa.
(B) CD spectra reported in molar ellipticity, normalized to blank buffer trace.

some higher molecular weight peaks of aggregated material are also observed, particularly for RXR117. If the second zinc finger is indeed rearranged, this may also contribute to the tendency of the evolved proteins to aggregate. The new proposed configuration extends the linker region between the two zinc fingers by an additional six amino acids in the evolved RXR proteins (Figure 4A). This region also contains a free unbound cysteine residue that could be involved in intermolecular disulfide bond formation.

We used circular dichroism (CD) spectroscopy to compare the secondary structure of the evolved proteins and the WT hRXRα protein. Despite rearrangements in the zinc coordination, it appears that the α-helical properties of the proteins are preserved (Figure 5B). Unfortunately, we were unable to measure the thermal stabilities of the evolved RXR proteins, since they denature irreversibly at elevated temperatures.

When we compare the results of our current selection for ATP binding proteins with a previous selection experiment in which ATP binding proteins were selected from a completely random library [38], there are some notable differences. Selection from the scaffolded library resulted in approximately 32% of input protein binding to the ATP-agarose after 8 rounds of selection, while selection from the initially random library reached a plateau at only 6% after 9 rounds of selection. Only after mutagenesis and an additional 9 rounds of selection did the column binding reach 30%–40%. In addition, selected sequences from the RXR library are soluble enough to perform some biophysical characterization, whereas no soluble protein could be obtained from unoptimized proteins derived from the random library. These results suggest that proteins selected from the RXR library tend to fold more consistently than the initial proteins derived from the random library, as expected from the stable fold of the RXR domain. We have also found that there are many more independent binding clones in the output of the RXR selection (>> 20) than were obtained from the random sequence library (~4), suggesting that the RXR-scaffolded library has a higher frequency of active sequences than the random library. An experiment in which both libraries are mixed in equal proportions and subjected to selection under the same conditions, without replication bias, would be required to test the theory that scaffolds present a clear advantage over completely random libraries. In addition, considerable structural characterization will be required to assess the structural diversity of the solutions obtained from these distinct libraries.

Significance

The rapid evolution of new functional molecules in the laboratory is a promising approach to the creation of new binding surfaces and enzymes. The use of alternative scaffolds may enhance the range of activities that can be discovered by using directed evolution. In this report, we describe the design and use of a novel scaffold based on the zinc finger DNA binding domain, hRXRα. The library was generated by randomizing two loop sequences, which were then assembled

in a combinatorial manner. The use of mRNA display allowed us to sample a much greater number of sequence combinations than can be sampled by using other methods of directed evolution. We have used directed evolution to alter the natural function of a protein domain. We were able to select functional molecules from this library with moderate affinity and high specificity for ATP. As with previous selections from both RNA and protein libraries, the greatest discrimination between ATP and analogs occurs with substitutions in the nucleobase, while changes in the sugar and triphosphate have smaller effects. There may be more protein-ligand contacts with the nucleobase, possibly reflecting the conformational rigidity of the nucleobase as opposed to the more flexible sugar and triphosphate moieties. Our study has uncovered unexpected structural changes in the RXR scaffold that were strongly selected for during the selection for ATP binding; mutational analysis suggests that there has been a rearrangement of the cysteine ligands for the second zinc finger. This change may improve the stability of the selected clones. If so, it is possible that an improved RXR-scaffold library, built with this modification, would have an increased effective diversity of functional molecules. Selections for RXR variants that can recognize proteins, peptides, and DNA, as well as enzymatic selections, are underway in order to further explore the potential of this scaffold.

Experimental Procedures

ATP Binding Selection

Preparation of RNA for mRNA display and RNA-protein fusion formation were as previously described [26, 27, 31, 39]. In round 1, a splinted ligation method was used for the generation of puromycin-modified RNAs, while for round 2 and every other round, a psoralen crosslinking strategy was used [40]. For round 1, fusions from a 10 ml translation reaction were purified on oligo(dT) cellulose (New England Biolabs) and then on M2 α-FLAG agarose (Sigma) [31]. The purified fusions were transferred to reverse transcription buffer by using a NAP-10 gel filtration column (Amersham Pharmacia) and subsequently reverse-transcribed with Superscript II (Invitrogen). Fusions were then exchanged into selection buffer (50 mM Tris-HCl, 250 mM KCl, 5 mM $MgCl_2$, 75 uM $ZnCl_2$, 1 mM DTT, 0.1% Triton X-100, pH 7.4). Fusions were incubated with C-8-linked ATP agarose (Sigma) for 1 hr at 4ºC. The column was drained and washed with selection buffer five times with two column volumes each after five minute incubations. The column was then eluted five times with two column volumes of elution buffer (selection buffer + 5 mM ATP + 5 mM $MgCl_2$). The elution fractions were pooled and exchanged into H_2O by using a NAP column before amplification.

Column Binding Experiments

In vitro translated protein labeled with [^{35}S]methionine was purified on M2 α-FLAG agarose. A total of 50 μl of the labeled protein was incubated with 450 ul of selection buffer and 100 μl of ATP-agarose for 1 hr. The resin was washed five times with 2 column volumes of selection buffer and then eluted five times with 2 column volumes with selection buffer supplemented with 5 mM ATP and 5 mM $MgCl_2$. The amount of protein in each fraction was determined by radioactive detection.

The same protocol was followed for MBP-RXR proteins expressed in *E. coli*, except that a different buffer was used (1× RSB + 5 mM $MgCl_2$: 50 mM Tris-HCl, 250 mM KCl, 100 μM $ZnCl_2$, 5 mM 2-mercaptoethanol [pH 8.3]). The resin was washed with 10 column volumes and eluted with 1× RSB supplemented with 5 mM ATP and 5 mM $MgCl_2$. Protein was quantified by mixing an aliquot of each fraction with protein assay reagent (BioRad) and reading absorbance at 595 nm.

Competition Assays

Approximately 20 μM of refolded MBP-fusion protein was incubated with 0.4 nM ^{32}P α-ATP and varying amounts of unlabeled ligand in 1× RSB. For RXR11 and RXR117, an additional 5 mM $MgCl_2$ was added, while $MgCl_2$ was completely omitted for the measurements with RXR122. After incubating for 2 hr at 4ºC, the mixture was spun in a Microcon-30 spin concentrator unit (Amicon) [41]. Aliquots from the top and the bottom chambers were counted in a Beckman LS 6500 scintillation counter or Packard Top Count NXT in order to determine the percentage of ATP bound in the presence of different concentrations of competitor. The K_d for each inhibitor was determined by the following equation [42–44]:

$$\%\ \text{bound} = \frac{1}{(1 + K_{dp}/([P]_t\, f - [P\ I]))}$$

$$[P\ I] = \left(([P]_t\, f + [I]_t + K_{di}) - \left(([P]_t\, f + [I]_t + K_{di})^2 - 4[P]_t\, f[L]_t\right)^{0.5}\right)/2$$

where $[P]_t$ = total protein concentration, $[I]_t$ = total inhibitor concentration in each incubation, f = fraction of active binding sites, K_{di} = dissociation constant of the inhibitor protein complex, and K_{dp} = dissociation constant of the protein to the labeled probe. We fitted the data from the binding titrations to this equation by nonlinear regression with Deltagraph 4.0 and found values to K_{dp}, K_{di}, and f for each of the MBP-RXR clones for ATP and its analogs.

Gel Filtration

A Sephadex-75 column (Amersham Pharmacia) was equilibrated with 1× RSB. Free (non-MBP-fused) proteins were exchanged into 1× RSB by using a NAP column and concentrated with a Biomax concentrator (Millipore). A total of 100 μl of protein was injected, and the column was run for 40 min at 0.5 ml/min using a Biocad FPLC (Perceptive Biosystems).

Circular Dichroism

Free RXR protein was exchanged into CD buffer (10 mM Tris-HCl, 250 mM KCl, 100 μM $ZnCl_2$, 5 mM $MgCl_2$, 5 mM 2-mercaptoethanol [pH 8.3]). CD readings were measured on an Aviv Circular Dichroism Spectrometer Model 202. Wavelength scans were performed with a 0.1 cm cell length and 1 nm bandwidth in 1–2 nm increments with an averaging time of 4 s.

Supplemental Data

Supplemental Data, including library construction, deoxy-oligonucleotide sequences, cloning, protein expression, purification, refolding, column binding experiments, GFP solubility screen, elemental analysis, and three supplemental figures, are available at http://www.chembiol.com/cgi/content/full/13/2/139/DC1/.

Acknowledgments

We thank Glenn Short, Burckhard Seelig, James Carothers and Justin Ichida for critical reading of this manuscript, and Ivan Lessard and Christopher T. Walsh for kindly providing the pIADL14 version of the maltose-binding protein plasmid. We are grateful to Geoffrey Waldo and Thomas Terwilliger for providing the reporter plasmid for performing the GFP solubility screen. G.S.C. was funded in part by the Howard Hughes Medical Research Fund for predoctoral students. J.W.S. is an investigator of the Howard Hughes Medical Institute. This work was funded in part by a grant from the NASA Astrobiology Institute.

Received: March 1, 2005
Revised: August 15, 2005
Accepted: October 27, 2005
Published: February 24, 2006

References

1. Davies, D.R., and Metzger, H. (1983). Structural basis of antibody function. Annu. Rev. Immunol. *1*, 87–117.
2. Farber, G.K., and Petsko, G.A. (1990). The evolution of alpha/beta barrel enzymes. Trends Biochem. Sci. *15*, 228–234.
3. Ladner, R.C., and Ley, A.C. (2001). Novel frameworks as a source of high-affinity ligands. Curr. Opin. Biotechnol. *12*, 406–410.
4. Nygren, P.A., and Uhlen, M. (1997). Scaffolds for engineering novel binding sites in proteins. Curr. Opin. Struct. Biol. *7*, 463–469.
5. Doi, N., and Yanagawa, H. (2001). Genotype-phenotype linkage for directed evolution and screening of combinatorial protein libraries. Comb. Chem. High Throughput Screen. *4*, 497–509.
6. Smith, G. (1998). Patch engineering: a general approach for creating proteins that have new binding activities. Trends Biochem. Sci. *23*, 457–460.
7. Eklund, M., Axelsson, L., Uhlen, M., and Nygren, P.A. (2002). Anti-idiotypic protein domains selected from protein A-based affibody libraries. Proteins *48*, 454–462.
8. Skerra, A. (2000). Engineered protein scaffolds for molecular recognition. J. Mol. Recognit. *13*, 167–187.
9. Malabarba, M.G., Milia, E., Faretta, M., Zamponi, R., Pelicci, P.G., and Di Fiore, P.P. (2001). A repertoire library that allows the selection of synthetic SH2s with altered binding specificities. Oncogene *20*, 5186–5194.
10. Weiss, G.A., and Lowman, H.B. (2000). Anticalins versus antibodies: made-to-order binding proteins for small molecules. Chem. Biol. *7*, R177–R184.
11. Boder, E.T., and Wittrup, K.D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. Nat. Biotechnol. *15*, 553–557.
12. Griffiths, A.D., Malmqvist, M., Marks, J.D., Bye, J.M., Embleton, M.J., McCafferty, J., Baier, M., Holliger, K.P., Gorick, B.D., Hughes-Jones, N.C., et al. (1993). Human anti-self antibodies with high specificity from phage display libraries. EMBO J. *12*, 725–734.
13. Hanes, J., and Pluckthun, A. (1997). In vitro selection and evolution of functional proteins by using ribosome display. Proc. Natl. Acad. Sci. USA *94*, 4937–4942.
14. Pluckthun, A., and Pack, P. (1997). New protein engineering approaches to multivalent and bispecific antibody fragments. Immunotechnology *3*, 83–105.
15. Beste, G., Schmidt, F.S., Stibora, T., and Skerra, A. (1999). Small antibody-like proteins with prescribed ligand specificities derived from the lipocalin fold. Proc. Natl. Acad. Sci. USA *96*, 1898–1903.
16. Hanes, J., Schaffitzel, C., Knappik, A., and Pluckthun, A. (2000). Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. Nat. Biotechnol. *18*, 1287–1292.
17. Ku, J., and Schultz, P.G. (1995). Alternate protein frameworks for molecular recognition. Proc. Natl. Acad. Sci. USA *92*, 6552–6556.
18. Xu, L., Aha, P., Gu, K., Kuimelis, R.G., Kurz, M., Lam, T., Lim, A.C., Liu, H., Lohse, P.A., Sun, L., et al. (2002). Directed evolution of high-affinity antibody mimics using mRNA display. Chem. Biol. *9*, 933–942.
19. Benhar, I. (2001). Biotechnological applications of phage and cell display. Biotechnol. Adv. *19*, 1–33.
20. Schlehuber, S., Beste, G., and Skerra, A. (2000). A novel type of receptor protein, based on the lipocalin scaffold, with specificity for digoxigenin. J. Mol. Biol. *297*, 1105–1120.
21. Greisman, H.A., and Pabo, C.O. (1997). A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. Science *275*, 657–661.
22. Liu, Q., Segal, D.J., Ghiara, J.B., and Barbas, C.F., 3rd. (1997). Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. Proc. Natl. Acad. Sci. USA *94*, 5525–5530.
23. Karatan, E., Merguerian, M., Han, Z., Scholle, M.D., Koide, S., and Kay, B.K. (2004). Molecular recognition properties of FN3 monobodies that bind the Src SH3 domain. Chem. Biol. *11*, 835–844.
24. Batori, V., Koide, A., and Koide, S. (2002). Exploring the potential of the monobody scaffold: effects of loop elongation on the stability of a fibronectin type III domain. Protein Eng. *15*, 1015–1020.
25. Binz, H.K., Amstutz, P., Kohl, A., Stumpp, M.T., Briand, C., Forrer, P., Grutter, M.G., and Pluckthun, A. (2004). High-affinity binders selected from designed ankyrin repeat protein libraries. Nat. Biotechnol. *22*, 575–582.
26. Roberts, R.W., and Szostak, J.W. (1997). RNA-peptide fusions for the in vitro selection of peptides and proteins. Proc. Natl. Acad. Sci. USA *94*, 12297–12302.
27. 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.
28. Mangelsdorf, D.J., and Evans, R.M. (1995). The RXR heterodimers and orphan receptors. Cell *83*, 841–850.
29. Holmbeck, S.M., Foster, M.P., Casimiro, D.R., Sem, D.S., Dyson, H.J., and Wright, P.E. (1998). High-resolution solution structure of the retinoid X receptor DNA-binding domain. J. Mol. Biol. *281*, 271–284.
30. Zhao, Q., Chasse, S.A., Devarakonda, S., Sierk, M.L., Ahvazi, B., and Rastinejad, F. (2000). Structural basis of RXR-DNA interactions. J. Mol. Biol. *296*, 509–520.
31. Cho, G., Keefe, A.D., Liu, R., Wilson, D.S., and Szostak, J.W. (2000). Constructing high complexity synthetic libraries of long ORFs using in vitro selection. J. Mol. Biol. *297*, 309–319.
32. Bellamacina, C.R. (1996). The nicotinamide dinucleotide binding motif: a comparison of nucleotide binding proteins. FASEB J. *10*, 1257–1269.
33. Surdo, P.L., 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.
34. Mao, L., Wang, Y., Liu, Y., and Hu, X. (2004). Molecular determinants for ATP-binding in proteins: a data mining and quantum chemical analysis. J. Mol. Biol. *336*, 787–807.
35. Waldo, G.S., Standish, B.M., Berendzen, J., and Terwilliger, T.C. (1999). Rapid protein-folding assay using green fluorescent protein. Nat. Biotechnol. *17*, 691–695.
36. Saenger, W. (1984). Principles of Nucleic Acid Structure (New York: Springer-Verlag).
37. Holmbeck, S.M., Dyson, H.J., and Wright, P.E. (1998). DNA-induced conformational changes are the basis for cooperative dimerization by the DNA binding domain of the retinoid X receptor. J. Mol. Biol. *284*, 533–539.
38. Keefe, A.D., and Szostak, J.W. (2001). Functional proteins from a random-sequence library. Nature *410*, 715–718.
39. 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.
40. Kurz, M., Gu, K., and Lohse, P.A. (2000). Psoralen photocrosslinked mRNA-puromycin conjugates: a novel template for the rapid and facile preparation of mRNA-protein fusions. Nucleic Acids Res. *28*, E83.
41. Jenison, R.D., Gill, S.C., Pardi, A., and Polisky, B. (1994). High-resolution molecular discrimination by RNA. Science *263*, 1425–1429.
42. Kyte, J. (1995). Mechanism in Protein Chemistry (New York: Garland Publishing, Inc.).
43. Nemeria, N., Yan, Y., Zhang, Z., Brown, A.M., Arjunan, P., Furey, W., Guest, J.R., and Jordan, F. (2001). Inhibition of the *Escherichia coli* pyruvate dehydrogenase complex E1 subunit and its tyrosine 177 variants by thiamin 2-thiazolone and thiamin 2-thiothiazolone diphosphates: evidence for reversible tight-binding inhibition. J. Biol. Chem. *276*, 45969–45978.
44. Hodel, M.R., Corbett, A.H., and Hodel, A.E. (2001). Dissection of a nuclear localization signal. J. Biol. Chem. *276*, 1317–1325.