Supplemental Data

Directed Evolution of ATP Binding Proteins

from a Zinc Finger Domain Using mRNA-Display

Glen S. Cho and Jack W. Szostak

Supplemental Experimental Procedures

Library Construction

For segment A, 10^{14} molecules of gc128 (5' ATG GAC TAC AAG GAC GAC GAC GAC AAG GGC GGA AAG CAC ATC TGC GCC ATC TGT GGA GA 3') and gc132 (5' ACC TAG TCT CCC CTT TCT CAC GGT GCG CTT GAA GAA GCC TTT ACA GCC CTC ACA GGA 765 765 765 765 765 765 765 765 765 765 765 765 ATC TCC ACA GAT GGC GCA G 3'; 5 is a mixture of bases in these ratios, A: 24.6, T:27.1, G:11.3, C:37, 6 is a mixture of bases in these ratios, A:26.9, T:34, G:18, C:21.1; 7 is a mixture of bases in these ratios, A:32.1, T:0, G:21.7, C:46.2) were extended in a mutually primed reaction using two cycles of PCR with Taq (Roche). The resultant molecules were amplified with gc127 (5'GCT AAT ACG ACT CAC TAT AGG GAC AAT TAC TAT TTA CAA TTA CAA TGG ACT ACA AGG ACG ACG A 3') and gc129 (5' AGA ACC GCC ATG ATG GTG GTG GTG ACT ACC TAG TCT CCC CTT TCT CAC 3'). For segment B, 10^{14} molecules of gc130 (5' ATG GAC TAC AAG GAC GAC GAC GAC AAG GGC GTC TCC AAG GAC CTG ACC TAC ACC TGT CGG 3') and gc133 (5' TG CAC GGC CTC CCT TTT CAT GCC CAT GGC GAG GGC CTT CTG GTA CCT ACA GTA CTG GCA 765 765 765 765 765 765 765 765 765 ACA ATC CTT GTT GTC CCG ACA GGT GTA GGT CAG 3') were mixed and extended. The product was amplified with gc127 and gc131 (5'AGA ACC GCC ATG ATG ATG GTG GTG GTG ACT ACC TAC CTC TTC CTG CAC GGC CTC CCT TTT C 3'). Oligos were synthesized at the W.M Keck Facility at Yale University. RNA was produced from these constructs and pre-selection was performed as described [S1]. Splint gc134 (5' TTT TTT TTT TTT TTT NAG AAC CGC CAT GAT G 3', N refers to a equal mixture of A,T,G, and C) was used in the splinted ligation step and the fusions were also reverse transcribed with primer gc134 using Superscript II (Invitrogen). Segment A yielded approximately 3 x 10^{12} molecules and segment B yielded ~5 x 10^{11}. After PCR amplification of the selected material, the PCR products were cut with BsmaI (New England Biolabs) to produce asymmetric sticky ends in the sequence context of the original RXR sequence. After ligation, around 1.4 x 10^{14} unique molecules were formed from the combinatorial union of segment A and segment B.

Selection

For round 1, the fusions were reverse transcribed using primer gc134 and Superscript II. Fusions in round 1 elution were amplified with gc185 (5' GCT AAT ACG ACT ACT ATA GGG ACA ATT ACT ATT TAC AAT TAC A 3') and gc135 (5' AGA ACC GCC ATG ATG ATG 3'). For round 2 and all subsequent rounds, several changes were made. A crosslink site was introduced at the 3' end in order to use the psoralen-crosslinking protocol with primer gc186 (5' CTA ATA GCC GGT GCT ACC TAC CTC TTC CTG 3') [S2]. Reverse transcription was performed directly after oligo(dT) purification with primer gc192 (5' TTT TTT TTT TTT TTT TTT CTA

CCT ACT CTT TCC TG 3'). Denaturing Ni-NTA was performed on the ATP-column elution for round 2, and then FLAG purification after round 2. Samples were exchanged into H_2O before PCR amplification using a NAP column. The pool was regenerated for the next round of selection using primers gc185 and gc186.

After eight rounds of selections, six more rounds were done under more stringent conditions. The pool was not incubated with the resin, but instead passed directly over the immobilized ATP. The selection was performed at room temperature instead of at 4ºC. Only sequences that eluted after a pre-elution of 2 hours in the presence of free ATP were retrieved and amplified. The pool was dominated by variants of sequence 11 observed in the post-round 8 pool. RXR11 was selected from the round 14 pool.

Cloning

The DNA-binding domain of *hrxrα* (Genbank, BC063827) was amplified from a placental-cDNA library (Ambion) with primers gc112 (5'AAG CAC ATC TGCG CC ATC TGC 3') and gc213 (5' CTC CTC CTG CAC GGC TTC CCG 3'). A C195A mutation was made by subsequent amplification with gc215 (5' GCA TGG CTC GAG GAT CCG GGC GGA AAG CAC ATC TGC GGG GAC CGC 3') and gc216 (5'TAC TCG GAT CCC TAC TCC TCC TGC ACG GCT TCC CGC TTC ATG CCC ATG CCC AGG GCC TT 3'). A 5' FLAG tag was added to the sequence through amplification with gc128 and gc224 (3' CTA ATA GCC GGT GCT ACCT AC CTC CTC CTG CAC GGC TTC 3'). T7 promoter was added by amplifying with gc127 and gc224.

MBP-fusion proteins were made by ligating each sequence into pIADL14 (courtesy of Ivan Lessard and Chris Walsh, Harvard Medical School) [S3]. *hrxrα, rxr11, rxr117*, and *rxr122* were amplified with gc201 (5' CAT GGC TCG AGG ATC CGC GGA AAG CAC ATC TGC 3') and gc232 (5' TAG ACC CTC GAG TGA GCC CCT CTC TT CCT GCA CGG CCT C 3') and cut with BamHI and XhoI. The inserts were ligated with T4 DNA ligase (New England Biolabs) into a similarly cut vector. The resultant protein has a N-terminal MBP fusion and a C-terminal His_6 tag. A version without the His_6 tag was cloned into pIADL14 by amplifying the clones with gc236 (5' TGT GAT CTC GAG TCA CCT CTC TTC CTG CAC GGC CTC 3'), which places a stop codon before the His_6 tag in the vector. A control sequence was created by ligating the annealed oligos gc241 (5' GAT CCG GCG GCT CTT GAT GAT AGC 3') and gc242 (5'TCG AGC TAT TAT CAA GAG CGC CG 3') into the BamHI/XhoI cloning site.

We prepared free-protein (non-MBP fused) versions of the RXR proteins by cloning into pET24a (Novagen). Each of the sequences were amplified with gc235 (5' TGT GAT ATA GCT AGC GGC GAA AGC ACA TCT GC 3') and gc240 (5' GTC TAC TCG AGA GAA CCG CGT GGC ACC AGA CCA GAA GAC CTC TCT TCC TGC ACG GCC TC 3') and then cut with NheI and XhoI and ligated into pET24a. This construct has the RXR variant proceeded by a thrombin cleavage site and a His_6 tag.

Site directed mutations for C42S and C51S were made using the Quick Change Mutagenesis Method (Stratagene) from the pIADL-rxr vectors. The following primers were used for site directed mutagenesis: rxr11 C42S (gc255 - 5' GAC CTG ACC TAC ACC AGT CGG CAC AAC AAG GAT 3', gc256 - 5' ATC CTT GTT GTG CCG ACT GGT GTA GGT CAG GTC 3'); rxr11 C51S (gc261 - 5' AAG GAT TGT GTG GTG AGT CAC TCT TAT CAC TGC 3', gc262 - 5' GCA GTG ATA AGA GTG ACT CAC CAC ACA ATC CTT 3'); rxr117 C42S (gc267 - 5' GAC CTG ACC TAC ACC AGT CGG GAC AAC AAG GAT 3', gc268 - 5' ATC CTT GTT GTC CCG ACT GGT GTA GGT CAG GTC 3'); rxr117 C51S (gc269 - 5' GAC

AAC AAG GAT TGT TGC GTG AGT AAT GCT TTC CAC GGC 3', gc270 - 5' GCC GTG GAA AGC ATT ACT CAC GCA ACA ATC CTT GTT GTC 3'); rxr122 C42S (gc259 - 5' GAG CTG ACC TAC ACC AGT CGG GGC AAC AAG GAT 3', gc260 - 5' ATC CTT GTT GCC CCG ACT GGT GTA GGT CAG CTC 3'); rxr122 C51S (gc265 - 5' AAG GAT TGT TCC GTG AGC TGG AGT TAT CAT AAT 3', gc266 - 5' ATT ATG ATA ACT CCA GCT CAC GGA ACA ATC CTT 3').

Round 9 pool of the RXR selection was cloned into GFP reporter vector after amplification of the selected pool with gc228 (5'ATA GCG ATA CAT ATG GAC TAC AAG GAC GAC GAC 3') and gc229 (5'TAG ACC GGA TCC CCT CTC TTC CTG CAC GGC CTC 3'). The PCR products were cut with NdeI and BamHI and cloned into the reporter vector to create a C-terminal GFP fusion protein. The GFP screen was performed as described [S4, S5]. RXR117 and RXR122 were retrieved from a visual selection of colonies with the most intense green color.

Protein Purification

MBP-RXR fusion proteins were expressed in BL21star cells (Invitrogen) at 20°C and purified under native or denaturing conditions depending on the downstream application. For protein purification under native conditions, cells were lysed in 50 mM Tris-HCl, 300 mM NaCl, 100 μM $ZnCl_2$, 5 mM 2-mercaptoethanol, pH 8.0, with 1mg/mL lysozme and 10 μL of RQ1 DNAse (Promega) after freeze/thaw treatment and sonication. The cleared lysate was incubated with Ni-NTA (Qiagen) for 1 hr at 4°C. The matrix was washed with ten column volumes HS wash buffer (50 mM Tris-HCl, 1 M NaCl, 100 μM $ZnCl_2$, 5 mM 2-mercaptoethanol, pH 8.0) and then with three column volumes of 10 mM imidazole wash (50 mM Tris HCl, 300 mM NaCl, 100 μM $ZnCl_2$, 5 mM 2-mercaptoethanol, 10 mM imidazole, pH 8.0). The proteins were eluted with the addition of 30 mM imidazole to the buffer. Protein was transferred to 1X RSB (50 mM Tris-HCl, 250 mM KCl, 100 μM $ZnCl_2$, 5 mM 2-mercaptoethanol, pH 8.3) through dialysis or a NAP desalting column. For protein purification under denaturing conditions, cells were lysed with 50 mM Tris HCl, 300 mM NaCl, 6 M guanidinium chloride, 100 μM $ZnCl_2$, 5 mM 2-mercaptoethanol, pH 8.0. The cleared lysate was bound to the Ni-NTA beads for 1 hour, washed with HS urea buffer (50 mM Tris HCl, 8 M urea, 1 M NaCl, 100 μM $ZnCl_2$, 5 mM 2-mercaptoethanol, pH 8.0), and then washed with 10mM imidazole urea wash (50 mM Tris-HCl, 8 M Urea, 300 mM NaCl, 100 μM $ZnCl_2$, 5 mM 2-mercaptoethanol, 10 mM imidzaole, pH 8.0). The proteins were eluted with the same buffer but with 100 mM imidazole added. The protein was refolded by dialyzing three times against 100 volumes 1X RSB.

Free RXR was expressed from pET24 clones in BL21star cells overnight at 20°C, then purified under denaturing conditions by Ni-NTA affinity chromatography. The protein was renatured on the column through stepwise reduction of the urea concentration from 8 M to 0 M in half molar steps. The resin was washed with HS wash buffer and then with 10 mM imidazole wash buffer and then eluted with 200 mM imidazole buffer. The protein was exchanged into 10 mM Tris-HCl, 250 mM KCl, 100 μM $ZnCl_2$, 5 mM 2-mercaptoethanol, 5 mM $MgCl_2$, pH 8.0 using a NAP-5 gel filtration column.

Elemental Analysis

MBP-RXR clones with C-terminal His_6 tags were purified using native Ni-NTA purification protocol. Purified proteins were dialyzed three times each with 100-fold dilutions against 1X RSB buffer lacking $ZnCl_2$. 10 μM samples of protein were analyzed by Proton Induced X-ray

Emission (PIXE; Elemental Analysis Corporation). The non-His_6 tag versions of the protein were purified on amylose resin and washed with 30 column volumes with 1X RSB lacking $ZnCl_2$ and then eluted with the same buffer containing 10mM maltose.

Supplemental References

S1. 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.

S2. Kurz, M., Gu, K., and Lohse, P.A. (2000). Psoralen photo-crosslinked mRNA-puromycin conjugates: a novel template for the rapid and facile preparation of mRNA-protein fusions. Nucleic Acids Res 28, E83.

S3. McCafferty, D.G., Lessard, I.A., and Walsh, C.T. (1997). Mutational analysis of potential zinc-binding residues in the active site of the enterococcal D-Ala-D-Ala dipeptidase VanX. Biochemistry 36, 10498-10505.

S4. 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.

S5. Pedelacq, J.D., Piltch, E., Liong, E.C., Berendzen, J., Kim, C.Y., Rho, B.S., Park, M.S., Terwilliger, T.C., and Waldo, G.S. (2002). Engineering soluble proteins for structural genomics. Nat Biotechnol 20, 927-932.

Table S1. Elemental Analysis of MBP-RXR Fusion Protein

	Zn	[Zn]/[protein]
hRXRa	2.18 ± 0.39 ppm	3.72 ± 0.66
RXR11	3.05 ± 0.41 ppm	5.21 ± 0.69
RXR117	2.50 ± 0.38 ppm	4.26 ± 0.64
RxR122	1.85 ± 0.38 ppm	2.98 ± 0.65
MBP	none detected	N/A

Figure S1. Confirmation of Random Positions

The unselected pool was cloned and sequenced. A multiple alignment was made with the consensus sequence from the evolved loops and the loop regions from the unselected randomized pool.

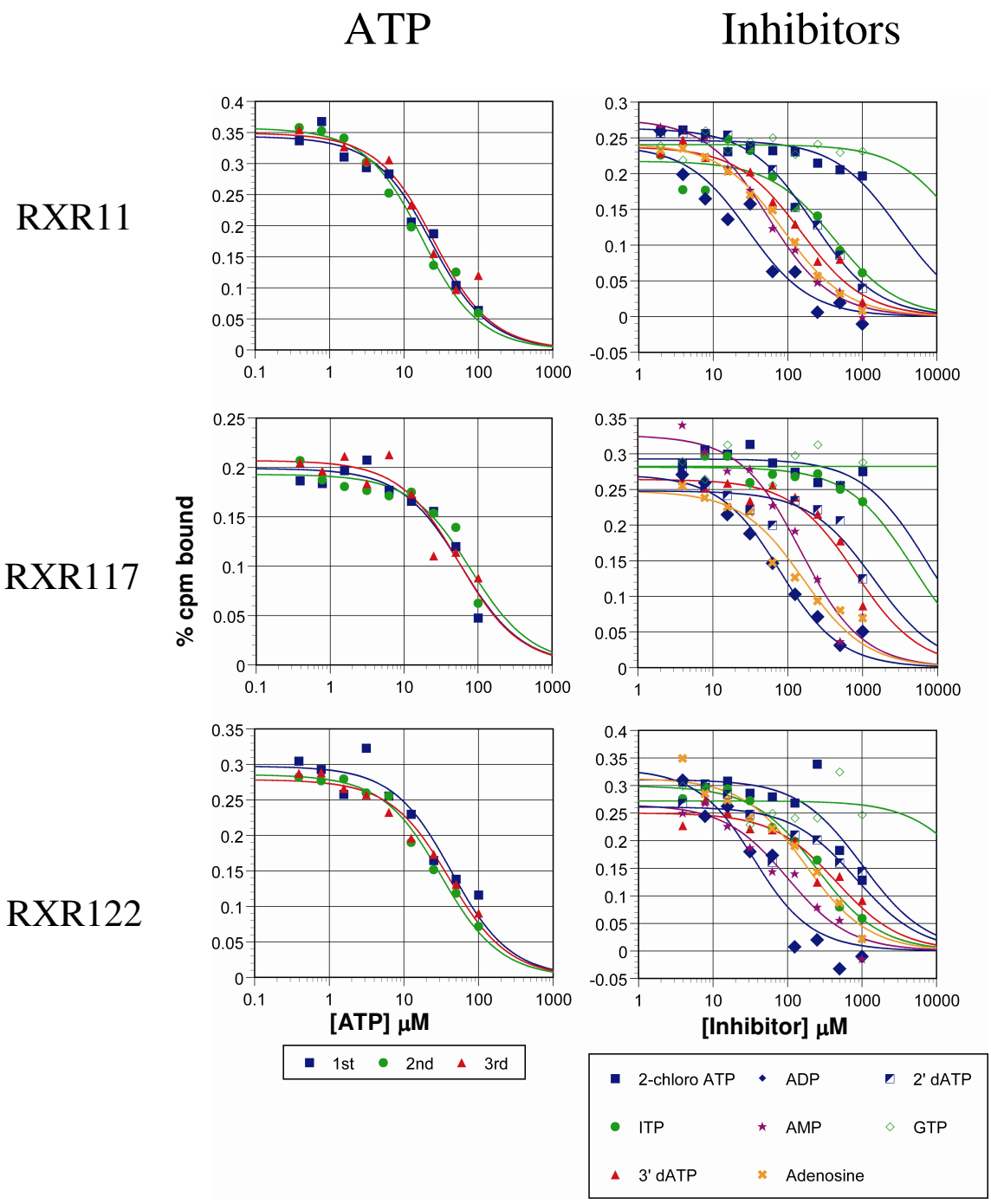

Figure S2. Raw Data from Competition Experiments in Solution

The fraction of bound ^{32}P-αATP (trace concentration) was measured in the presence of varying competitor ligands and a binding curve was fitted to the data using non-linear regression (see experimental procedures).

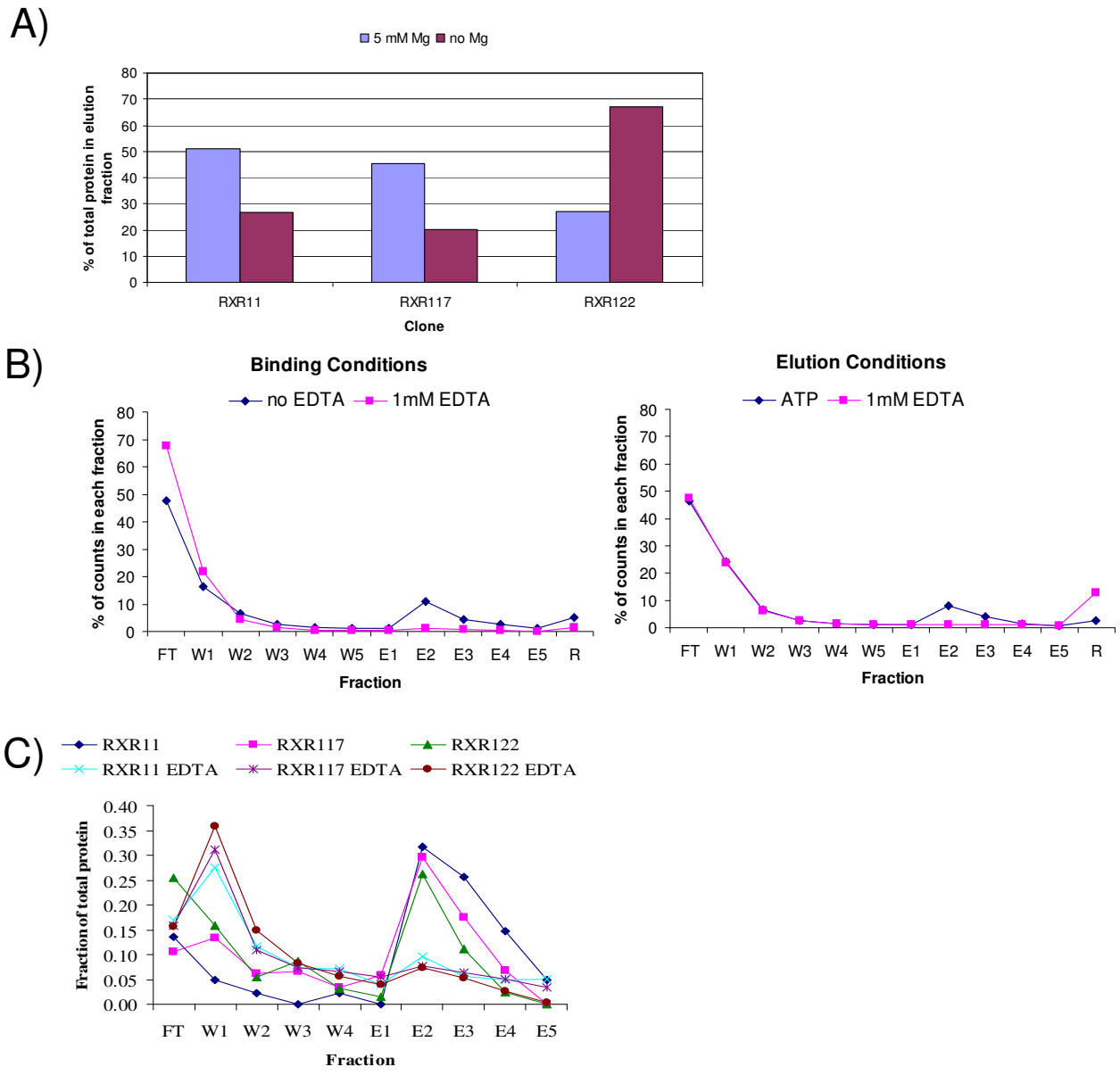

Figure S3. Effect of Divalent Metals on ATP Binding

Purified MBP-fusion proteins were incubated with ATP-agarose in the presence or absence of 5 mM $MgCl_2$. The column was washed and eluted with 5mM ATP with or without 5 mM $MgCl_2$; the percentage of protein in the elution fraction is shown. A) 1 mM EDTA was added (a 10-fold excess over Zn^{2+}), to *in vitro* translated protein before binding to the ATP column. Both columns were washed, and then eluted with 5 mM ATP. In the second experiment, protein bound to the resin under normal conditions was then eluted with either 5 mM ATP and 5 mM $MgCl_2$ or with 1 mM EDTA. B) MBP-fused protein was incubated with or without 1mM EDTA prior to binding to ATP-resin. The column was washed and eluted with 5 mM ATP and 5 mM $MgCl_2$.