Published as: *Nature*. 2007 August 16; 448(7155): 828–831.

Selection and Evolution of Enzymes from a Partially Randomized Non-Catalytic Scaffold

Burckhard Seelig and **Jack W. Szostak**
Howard Hughes Medical Institute, Department of Molecular Biology, and Center for Computational and Integrative Biology (CCIB), 7215 Simches Research Center, Massachusetts General Hospital, 185 Cambridge Street, Boston, MA, 02114, USA

Abstract

Enzymes are exceptional catalysts that facilitate a wide variety of reactions under mild conditions, achieving high rate enhancements with excellent chemo-, regio- and stereoselectivities. There is considerable interest in developing new enzymes for the synthesis of chemicals and pharmaceuticals[1–3] and as tools for molecular biology. Methods have been developed for modifying and improving existing enzymes through screening, selection and directed evolution[4,5]. However, the design and evolution of truly novel enzymes has relied upon extensive knowledge of the mechanism of the reaction[6–10]. Here we show that genuinely new enzymatic activities can be created *de novo* without the need for prior mechanistic information by selection from a naive protein library of very high diversity with product formation as the sole selection criterion. We used mRNA-display, in which proteins are covalently linked to their encoding mRNA[11], to select for functional proteins from an *in vitro* translated protein library of high complexity (>10^{12}), without the constraints imposed by any *in vivo* step. This technique has been used to evolve new peptides and proteins that can bind a specific ligand[12–18], from both random-sequence libraries[12,14–16] and libraries based on a known protein fold[17,18]. We now describe the isolation of novel RNA ligases from a library based on a zinc finger scaffold[18,19], followed by *in vitro* directed evolution to further optimize these enzymes. The resulting ligases exhibit multiple turnover with rate enhancements of more than two million-fold.

We have devised a general scheme for the direct selection of enzymes catalyzing bond-forming reactions from mRNA-displayed protein libraries (Fig. 1a). To demonstrate that new protein catalysts can be created using this scheme, we chose, as a model reaction, the ligation of two RNA molecules aligned on a template, with one RNA activated with a 5′-triphosphate (Fig. 1b). For our selection we used a library in which two loops of the small, stable protein retinoid-X-receptor (hRXRα) domain were randomized; this library has previously been used for the isolation of novel ATP-binding proteins[18] (Fig. 1c). We

Correspondence and requests for materials should be addressed to J.W.S. (szostak@molbio.mgh.harvard.edu).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions All experiments were performed by B.S. Both authors designed the experiments, discussed the results and wrote the paper.

The authors declare no competing financial interests.

transcribed and translated this synthetic DNA library[18] to generate mRNA-displayed proteins (Fig. 1a), which we then reverse transcribed with a primer joined at its 5′-end to the triphosphorylated RNA (PPP-substrate, Fig. 1b). We incubated the library of 4×10^{12} unique mRNA-displayed proteins with the biotinylated oligonucleotide (HO-substrate) and the complementary splint oligonucleotide that aligns the two substrate oligonucleotides. Proteins that catalyzed the ligation of the two substrates covalently attached the biotin moiety to their own cDNA, which we captured on streptavidin-coated agarose beads. After washing, we eluted the cDNA by cleaving the photocleavable linker between the HO-substrate and the biotin. We amplified the cDNA by PCR and used it as input for the next round of selection and amplification. Over 9 rounds, the fraction of the input library immobilized on the streptavidin beads and photoreleased increased from 0.01% to 0.3%, and after 12 rounds it increased to 2.3% (Fig. 2). To increase the activity of the selected ligases, we returned to the DNA library after round 8 and performed recombination and random mutagenesis[20,21] by restriction enzyme digestion and ligation of the DNA, and by subjecting the input DNA for rounds 9* through 11* to error-prone PCR amplification. We then continued the cycles of selection and amplification without further recombination or mutagenesis until round 17, while increasing the selection pressure by gradually decreasing the reaction time from overnight to 5 minutes (Fig. 2).

The evolved pool of enzymes contained several families of closely related sequences as well as multiple unrelated single isolates (Supplementary Information). Of the nine amino acids in loop 2, four positions were absolutely conserved in all sequences, four other sites were conserved in 86–90% of the clones and one position was conserved in 50% of the sequences. In contrast, we observed the motif “DYKXXD” at varying positions in the 12 originally randomized positions of loop 1 in 57% of the clones. This motif was probably enriched because it resembles the recognition site for the anti-Flag antibody M2[22] that we used for purification of the mRNA-displayed proteins. These results suggest that the highly conserved loop 2 may play an essential role in ligase activity, while loop 1 may not be as important.

Analysis of the non-loop regions revealed a low conservation of specific cysteines of the original scaffold structure (Fig. 3). After 17 rounds of selection, just 16% of the clones (8 of 49) retained the cysteine pattern as originally designed and were free of major deletions. The first and the fourth CX_nC sequences were highly conserved (47 and 48 clones out of 49, respectively), but the second and third CX_nC motifs were retained in only 24% and 20% of the clones, respectively. In addition, two deletions (of 17 and 13 amino acids) were frequently observed (Fig. 3). Because of the mutation of up to half of the eight original zinc-coordinating cysteines and the frequent deletion of significant segments of the protein during the selection and evolution process, we believe that the majority of the proteins have undergone a substantial structural rearrangement in comparison to the original scaffold. We chose 18 clones from the final evolved library and screened them as mRNA-displayed proteins for ligation activity. All of the clones, including those with mutated cysteines or deleted regions, showed activity. We then expressed the 7 most active ligases (#1–7, Fig. 3) in *Escherichia coli* as C-terminal fusions with maltose binding protein (MBP) or without any fusion partner. All 7 enzymes were soluble when fused to MBP (≥ 3 mg/ml for several

weeks at 4°C). When expressed on their own, two of the ligases were soluble (#6 and #7) whereas the other five precipitated or aggregated.

We chose the MBP-fusion of the most active enzyme (ligase #4) for more detailed characterization (Fig. 4). Incubation of the purified MBP-fusion enzyme with the PPP-substrate, the HO-substrate, and the splint oligonucleotide yielded the desired ligation product (Fig. 4a, b) as well as the expected inorganic pyrophosphate by-product (Fig. 4c). We did not detect any product when we substituted the PPP-substrate with an oligonucleotide of identical sequence but with either a 5′-monophosphate or a 5′-hydroxyl instead of the 5′-triphosphate group (Fig. 4b). Preliminary experiments show that the enzyme catalyzes the ligation equally well for all four nucleobases at the 3′-terminal base of the HO-substrate as long as they are correctly base-paired to the splint oligonucleotide. A mismatch at this position reduces the ligation efficiency several fold. Enzymatic digestion of the ligated product confirmed the 3′-5′ regiospecificity of the ligase reaction (Fig. 4d).

Because the RXR-library was based on a zinc finger protein, we examined the role of zinc and other cations in catalysis. The reaction required Zn^{2+} and monovalent cations (K^+ or Na^+) with optima of 100 μM and 80 mM respectively. The rate of the catalyzed reaction showed a strong pH dependence with an optimum at pH 7.6. The optimal ligation conditions with regard to Zn^{2+}, monovalent cation and pH coincide with the conditions used during the selection. In contrast to the enzymatic reaction described here, the non-enzymatic template-directed ligation is inhibited by Zn^{2+} and shows a linear increase in reaction rate with increasing pH[23]. Incubating the ligase with chelating resin (Chelex 100) resulted in an almost complete loss of activity; activity could be restored by the addition of Zn^{2+}, but not by the addition of Cu^{2+}, Ni^{2+}, Co^{2+}, Mn^{2+}, Cd^{2+} or Mg^{2+} ions. Elemental analysis by inductively coupled plasma optical emission spectroscopy revealed 2.6 ± 0.4 equivalents (± s.d.) of bound zinc per ligase molecule, whereas the original wild-type hRXRα protein contained 2.1 ± 0.1 molecules of zinc (± s.d.). The strong zinc dependence of the ligase enzymes could be due to either the continued existence of structural zinc sites or the presence of a catalytic zinc in the molecule.

To quantify the rate acceleration achieved by the selected ligase, we determined the rates of the catalyzed as well as the uncatalyzed RNA-RNA ligation reactions. We could not detect any uncatalyzed formation of product in the absence of Mg^{2+}, consistent with previous work on the Mg^{2+}-dependence of the nonenzymatic ligation reaction[23]. During the selection process, Mg^{2+} was present at a concentration of 5 mM, yet we found that the catalyzed reaction did not require magnesium ions, and indeed was faster in its absence. By quantifying the detection limit of our assay we determined that the upper limit of the rate of the uncatalyzed reaction of the pseudo-intramolecular complex of two substrate oligonucleotides pre-aligned on the template oligonucleotide in the absence of Mg^{2+} was $k_{obs\ (uncatalyzed)} < 3 \times 10^{-7}\ h^{-1}$. We measured the rate of the catalyzed ligation in the absence of Mg^{2+}, at a subsaturating substrate concentration of 10 μM as $k_{obs\ (catalyzed)} = 0.65 \pm 0.11\ h^{-1}$ (± s.d.), which is at least 2×10^6-fold faster than the uncatalyzed reaction. For the wild-type hRXRα protein domain, we could not detect any ligated product (Fig. 4b).

We found that the evolved enzyme catalyzed the ligation reaction with multiple turnover (Fig. 4e), although the selection scheme we employed utilized a single-turnover strategy that did not exert any selective pressure for product release. The intramolecular single-turnover design of the mRNA-display selection scheme used here facilitates the isolation of enzymes, even if the rate acceleration is low or the substrate affinity is weak[24].

Preliminary biophysical studies suggest that the ligase possesses a folded structure. We chose ligase #6 for the following experiments because of its superior solubility in the absence of a fusion protein partner. Circular dichroism (CD) spectroscopy revealed an α–helical component of the secondary structure (Supplementary Fig. 1), and thermal denaturation indicated cooperative thermal unfolding (Fig. 4f). The two-dimensional $^{1}H^{15}N$ heteronuclear single-quantum coherence (HSQC) NMR spectrum showed about 67 well-resolved peaks with a good chemical shift dispersion in the amide region of the spectrum, which indicates that a significant portion of the ligase protein is well folded (Supplementary Fig. 2a). A similar HSQC experiment with selectively ^{15}N-cysteine labelled protein suggests that all six cysteines of ligase #6 are well structured (Supplementary Fig. 2b).

No natural enzyme is known to catalyze the ligation of a 5′-triphosphorylated RNA oligonucleotide to the terminal 3′-hydroxyl group of a second RNA, the reaction catalyzed by the enzymes described here. An enzyme catalyzing a similar reaction, the T4 RNA ligase, joins a 3′-hydroxyl group to a 5′-monophosphorylated RNA with the concomitant conversion of ATP to AMP and inorganic pyrophosphate via a covalent AMP-ligase intermediate[25]. The reaction catalyzed by the ligase described in this paper is more closely related to chain elongation by one nucleotide during RNA polymerization: in both cases, the growing strand and the triphosphate-containing substrate base pair to a template, the 3′-hydroxyl of the growing strand attacks the α–phosphate of a 5′-triphosphate, and a pyrophosphate is released in concert with the formation of a 3′-5′ phosphodiester bond (Fig. 4a). RNA polymerases can be very fast, for example T7 RNA polymerase catalyzes chain elongation at 240 nts/s[26]. Preliminary results with our selected ligase #4 did not show any polymerase activity with nucleoside triphosphates.

Ribozymes[27] and deoxyribozymes[28] previously selected from random oligonucleotide libraries catalyze the same reaction as our protein ligases. These ribozymes and deoxyribozymes have rate enhancements over the uncatalyzed background reaction of the same order of magnitude as our protein ligase and in the case of the ribozymes, these rates were significantly increased by further *in vitro* evolution (up to 10^9-fold rate acceleration)[29]. While our protein ligase was dependent on Zn^{2+} and inhibited by Mg^{2+}, the ribozyme-catalyzed ligation is strongly dependent on Mg^{2+} with an optimum at ~60 mM[27]. The deoxyribozymes have been selected as Mg^{2+}-dependent variants and also as Zn^{2+}-dependent variants. The pH dependence of our ligase enzyme suggests that the catalytic mechanism involves acid-base catalysis by amino acid residues of the enzyme; in contrast, the pH-dependence of the ribozyme and deoxyribozyme ligases is more consistent with a catalytic role for one or more bound metal ions.

Our results represent the first use of mRNA-display to select for a novel enzyme activity. The general scheme could be readily adapted to a selection for other bond-making enzymes,

and bond-breaking enzymes could be selected by immobilizing the mRNA-displayed proteins via the substrate to be cleaved and then enriching those molecules that liberate their encoding RNA. Alternatively, in order to isolate enzymes that catalyze other covalent modification reactions, the substrate could be attached to the cDNA and the active molecules separated with product-specific agents such as antibodies or aptamers. Thus, we believe that mRNA-display represents a broadly applicable route to the isolation of novel enzymatic activities that are otherwise difficult to generate without explicit knowledge of structure or mechanism.

Online Methods

Sequences of oligonucleotides are described in Supplementary Information.

Preparation of primer for reverse transcription (RT)

The RT-primer was a chimeric oligonucleotide made from a 5′-triphosphate RNA oligonucleotide and a DNA oligonucleotide at the 3′-end. The PPP-substrate (5′-pppGGAGACUCUUU) was synthesized by T7 RNA polymerase from a double stranded template of BS47 and BS48 and purified by denaturing polyacrylamide gel electrophoresis (PAGE). The PPP-substrate was then ligated to BS50 in the presence of BS56 as template by T4 DNA ligase[31] and the product was purified by denaturing PAGE to yield the RT-primer. 5′-pppGGAGACUCUUUTTTTTTTTTTTTTTTTTTCCCAGATCCAGACATTC

In vitro selection and evolution

The DNA library, designed and synthesized by Cho[18], was PCR amplified with primers BS3long and BS24RXR2 to introduce a cross-link site at the 3′-end in order to use the psoralen-crosslinking protocol[32] (for DNA library sequence see Supplementary Information). RNA was produced from the DNA library with T7 RNA polymerase. After purification by denaturing polyacrylamide gel electrophoresis the RNA was photo-crosslinked[32] with the XL-PSO oligonucleotide and ethanol precipitated. The mRNA-displayed proteins were generated as previously described[11,12,18,33,34] with the following modifications. In the first round of selection a 10 ml translation was incubated at 30°C for one hour (200 nM psoralen cross-linked RNA template, Red Nova Rabbit Reticulocyte Lysate (Novagen, Madison, WI), used according to the manufacturer's instructions with an additional 100 mM KCl / 0.9 mM $Mg(OAc)_2$ and 69 nM ^{35}S-methionine). After addition of 600 mM KCl and 25 mM $MgCl_2$ the translation reaction was incubated at room temperature for five minutes and then diluted ten-fold into oligo(dT)cellulose binding buffer (10 mM EDTA, 1 M NaCl, 10 mM 2-mercaptoethanol, 20 mM Tris(hydroxymethyl) amino methane, pH 8.0, 0.2% w/v Triton X-100) and this mixture was incubated with 10 mg/ml oligo(dT)cellulose (New England Biolabs, Beverly, MA) for fifteen minutes at 4°C with rotation. The oligo(dT)cellulose was washed on a chromatography column (Bio-Rad, Hercules, CA) with the same oligo(dT)cellulose binding buffer, then with oligo(dT)cellulose wash buffer (300 mM KCl, 5 mM 2-mercaptoethanol, 20 mM Tris(hydroxymethyl) amino methane, pH 8.0) and then eluted with oligo(dT)cellulose elution buffer (5 mM 2-mercaptoethanol, 2 mM Tris(hydroxymethyl) amino methane, pH 8.0) to yield 4×10^{13} mRNA-displayed proteins. The eluate was mixed with 10x Flag binding buffer (1x is 150

mM KCl, 5 mM 2-mercaptoethanol, 50 mM HEPES, pH 7.4, 0.01% w/v Triton X-100) and then incubated with 50 µl Anti-Flag M2-agarose affinity gel (Sigma, St. Louis, MO; prewashed with Flag clean buffer (100 mM glycine, pH 3.5, 0.25% w/v Triton X-100) and Flag binding buffer) for two hours at 4°C with rotation. The Anti-Flag M2-agarose affinity gel was then washed with Flag binding buffer and eluted with Flag binding buffer containing two equivalents of Flag peptide (Sigma, St. Louis, MO; one equivalent of Flag peptide saturates both antigen sites of the antibody resin) for twenty minutes at 4°C with rotation.

The eluate was diluted two-fold with an additional 50 mM Tris(hydroxymethyl) amino methane, pH 8.3, 3 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 0.5 mM each of dCTP, dGTP, dTTP, 5 µM dATP, 50 nM α-^{32}P dATP and used for the reverse transcription of the mRNA-displayed proteins with the RT-primer and Superscript II (Gibco BRL, Rockville, MD) at 42°C for 30 minutes. This sample was then dialyzed twice against Flag binding buffer at a ratio of 1/1000 and then incubated with 100 µl Anti-Flag M2-agarose affinity gel and processed as described for the first Flag affinity purification above. Zinc chloride and 5x selection buffer (1x is 400 mM KCl, 5 mM $MgCl_2$, 20 mM HEPES, pH 7.4, 0.01% w/v Triton X-100) was added to the Flag elution in order to make a final concentration of 100 µM and 1x, respectively. The mixture was incubated with 2 µM HO-substrate (PC-biotin) and 3 µM splint for the indicated times (Fig. 2) at room temperature. After quenching the reaction with 10 mM EDTA, the solution was incubated with 700 µl ImmunoPure immobilized streptavidin agarose [Pierce, Rockford, IL; prewashed with PBS buffer (138 mM NaCl, 2.7 mM KCl, 10 mM potassium phosphate, pH 7.4) including 2 mg/ml t-RNA (from baker's yeast, Sigma, St. Louis, MO)] and then washed with PBS alone at room temperature for twenty minutes with rotation. The streptavidin agarose was washed on a chromatography column (Bio-Rad, Hercules, CA) with SA binding buffer (1 M NaCl, 10 mM HEPES, pH 7.2, 5 mM EDTA), with SA urea wash buffer (8 M urea, 0.1 M Tris(hydroxymethyl) amino methane, pH 7.4), with SA basic wash buffer (20 mM NaOH, 1 mM EDTA) and with water. For the first round of selection the streptavidin agarose was used directly in the PCR amplification reaction (50 µl streptavidin agarose beads per 1 ml PCR). Every round was assayed by scintillation counting of the ^{35}S-methionine-labelled proteins (from translation to reverse transcription) or of the ^{32}P-labelled cDNA (after reverse transcription) to measure the efficiencies of the various steps. These data were then used to determine that the number of purified individual protein sequences introduced into the round 1 ligation reaction step (incubation with biotin-PC-RNA and splint RNA) was 4×10^{12}, based on the proportion of total methionine (translation) and total dATP (reverse transcription) incorporated into the mRNA-displayed proteins, and the efficiency of each of the subsequent purification steps.

This procedure was repeated for 17 rounds except for the following changes: in round 2 and in all subsequent rounds the translation reaction was 2 ml, only 400 µl of streptavidin agarose were used and directly before the PCR amplification, the streptavidin agarose beads were aliquoted in a 50% PBS slurry to 100 µl open wells. The slurry was irradiated with a UV lamp (4 Watts) at 360 nm from a 1 cm distance for fifteen minutes while shaking in order to release the cDNA. The beads were filtered off and the solution was used for PCR amplification. Before round 9* and 11* the DNA was digested with restriction endonuclease

AvaII, which recognizes a unique restriction site between the two zinc fingers, and then ligated back together with T4 DNA ligase to achieve a recombination of the two halves of the proteins. The input DNA in rounds 9*, 10* and 11* were further mutagenized by error prone PCR[20,21] at an average mutagenic rate of 3.8% at the amino acid level.

Cloning

Cloning was done as in[18] with some changes. In order to analyze the results of the selection, the cDNA of the respective round was cloned into the pCR®-TOPO vector (TOPO TA Cloning, Invitrogen, Carlsbad, CA) and the individual clones were sequenced. To express the proteins in *E. coli*, the ligase genes were amplified with primers BS63 and BS65, and the wild-type hRXRα motif (courtesy of Glen S. Cho) was amplified with primers BS68 and BS70. The PCR products were digested with NdeI and XhoI and cloned into the pIADL14 vector[35] (courtesy of Ivan Lessard and Chris Walsh, Harvard Medical School) to yield the MBP-fusion proteins or into the pET24a vector (Novagen, Madison, WI) to yield the protein without any fusion partner.

Sequence analysis

For sequence alignments the following software was used: Seqlab of the GCG Wisconsin Package™, BioEdit[36], MultAlin[37].

Ligation activity assay of mRNA-displayed ligases by gel shift

The sequences of 18 individual ligases were amplified from their respective pCR®-TOPO vector with primers BS3long / BS24RXR2 and separately subjected to one round of selection as described above. After the incubation with HO-substrate (PC-biotin) and splint for 5 h the ligation reaction mixture was quenched with 10 mM EDTA / 8 M urea and was then mixed with an excess of streptavidin (Pierce, Rockford, IL) and separated by denaturing PAGE. The gel was analyzed using a GE Healthcare (Amersham Bioscience, Piscataway, NJ) phosphorimager and ImageQuant software.

Expression of ligases and wild-type hRXRα in *E. coli*

All proteins were expressed in Rosetta BL21 (DE3) cells (Novagen, Madison, WI) containing the recombinant plasmids at 37°C in LB broth containing 50 μg/ml kanamycin. Cells were harvested, resuspended in lysis buffer (400 mM NaCl, 5 mM 2-mercaptoethanol, 20 mM HEPES, pH 7.5, 100 μM $ZnCl_2$, 10% glycerol) and sonicated. After centrifugation the supernatant was applied to an amylose resin column (New England Biolabs, Beverly, MA) in the case of the MBP-fusion proteins. The immobilized protein was washed and then eluted with amylose elution buffer (150 mM NaCl, 5 mM 2-mercaptoethanol, 20 mM HEPES, pH 7.5, 100 μM $ZnCl_2$, 10 mM maltose) and stored at 4°C for further use.

In order to purify the proteins lacking the MBP-fusion the supernatant after centrifugation was applied to a Ni-NTA resin column (Quiagen, Hilden, Germany) instead. The immobilized protein was washed and then eluted with acidic Ni-NTA elution buffer (20 mM NaOAc, pH 4.5, 400 mM NaCl, 5 mM 2-mercaptoethanol, 100 μM $ZnCl_2$) directly into a 1 M HEPES, pH 7.5 solution to yield a final concentration of 100 mM HEPES. For use in CD and NMR spectroscopy experiments the protein was further purified by FPLC (BioCAD

Sprint Perfusion System) using a Sephadex-200 gel filtration column (Pharmacia Biotech, Uppsala, Sweden) with isocratic elution in 150 mM NaCl, 5 mM 2-mercaptoethanol, 20 mM HEPES, pH 7.4, 100 μM $ZnCl_2$ at 4°C. The proteins were stored at 4°C for further use. Protein concentration was determined by the Bradford method.

Ligation activity assay of free ligases

20 μM PPP-substrate (11mer), 15 μM splint and 10 μM 5′-^{32}P-labelled HO-substrate (11mer) were incubated with 5 μM ligase in reaction buffer (100 mM NaCl, 20 mM HEPES, pH 7.5, 100 μM $ZnCl_2$) for the indicated time and separated and analyzed as above.

The k_{obs} values were determined by fitting the ratio of product concentration divided by enzyme concentration against time to a linear equation, and are the average of three independent experiments measured at less than 10% product formation. The standard deviation is provided.

Detection of pyrophosphate

The MBP fusion of ligase #4 (purified on amylose column) was immobilized on Ni-NTA resin (Quiagen, Hilden, Germany), washed with buffer (150 mM KCl, 5 mM 2-mercaptoethanol, 50 mM HEPES, pH 7.4, 0.01% w/v Triton X-100, 100 μM $ZnCl_2$) and eluted in acidic elution buffer (50 mM NaOAc, pH 4.5, 150 mM NaCl, 5 mM 2-mercaptoethanol, 100 μM $ZnCl_2$). The ligase was then dialyzed against 150 mM NaCl, 5 mM 2-mercaptoethanol, 20 mM HEPES, pH 7.5, 100 μM $ZnCl_2$. The ligase (3 μM) was incubated with 6 μM γ-^{32}P-labelled PPP-substrate (11mer), 9 μM splint and 12 μM HO-substrate (11mer). The reactions were separated by thin-layer chromatography on PEI cellulose plates, developed in 0.5 M KH_2PO_4 at pH 3.4.

Analysis of metal content

The MBP-fusion proteins of ligase #4 and wild-type hRXRα (purified on amylose column) were dialyzed three times against buffer (100 mM NaCl, 5 mM 2-mercaptoethanol, 20 mM HEPES at pH 7.5; pre-treated with Chelex 100 beads (BioRad) for 3 h and filtered) at a ratio of 1/1000. The metal content of 4 μM samples was measured with an Inductively Coupled Plasma Emission Spectrometer (Jarrell-Ash 965 ICP, University of Georgia).

Circular dichroism spectroscopy

CD spectra were recorded on an Aviv CD Spectrometer Model 202. Wavelength scans were performed in 15 mM NaCl, 0.5 mM 2-mercaptoethanol, 2 mM HEPES at pH 7.5, 10 μM $ZnCl_2$ and 100 μM ligase #6 at 25°C in a 0.1 mm cuvette at 1 nm bandwidth in 1 nm increments with an averaging time of 4 s. Thermal denaturation of 324 μM ligase #6 in 150 mM NaCl, 5 mM 2-mercaptoethanol, 50 mM HEPES, pH 7.4, 100 μM $ZnCl_2$ was monitored at 222 nm from 5°C to 90°C in 4°C increments and an equilibration time of 2 min at each temperature step in a 1 mm cuvette at 1.5 nm bandwidth with an averaging time of 10 s.

NMR spectroscopy

$^{1}H^{15}N$-NMR spectra were recorded on Bruker 500 MHz and 600 MHz NMR instruments with either uniformly ^{15}N-labelled or selectively ^{15}N-cysteine labelled protein (0.3 mM) in 10% D_2O, 150 mM NaCl, 5 mM 2-mercaptoethanol, 50 mM HEPES, pH 7.4, 100 μM $ZnCl_2$. Protein samples were prepared from minimal media cultures using ^{15}N-labelled NH_4Cl as the sole source of nitrogen or ^{15}N-labelled cysteine as sole source of cysteine, respectively.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank G.S. Cho for providing the RXR library prior to its publication along with valuable advice; A.D. Keefe for critical input during the initiation of this project; C. Mueller-Vahl for help with protein expression and purification; G.-P. Zhou, J.J. Chou for NMR measurements; A.D. K., G.S.C., G.F. Short, R. Larralde, J.M. Carothers, J. K. Ichida, F.P. Seebeck, S.S. Mansy, C. Del Bianco, D.A. Treco, D.S. Wilson, A.J. Bell, A. Luptak, R. Bruckner, Z. Sachs for helpful discussions. This work was supported by a grant from the NASA Astrobiology Institute. B.S. was supported in part by the Emmy Noether-Programm of the Deutsche Forschungsgemeinschaft. J.W.S. is an Investigator of the Howard Hughes Medical Institute.

References

1. Schmid A, et al. Industrial biocatalysis today and tomorrow. Nature. 2001; 409:258–268. [PubMed: 11196655]
2. Cherry JR, Fidantsef AL. Directed evolution of industrial enzymes: an update. Curr Opin Biotech. 2003; 14:438–443. [PubMed: 12943855]
3. Panke S, Held M, Wubbolts M. Trends and innovations in industrial biocatalysis for the production of fine chemicals. Curr Opin Biotech. 2004; 15:272–279. [PubMed: 15357999]
4. Bloom JD, et al. Evolving strategies for enzyme engineering. Curr Opin Struct Biol. 2005; 15:447–452. [PubMed: 16006119]
5. Aharoni A, Griffiths AD, Tawfik DS. High-throughput screens and selections of enzyme-encoding genes. Curr Opin Chem Biol. 2005; 9:210–216. [PubMed: 15811807]
6. Bolon DN, Mayo SL. Enzyme-like proteins by computational design. Proc Natl Acad Sci USA. 2001; 98:14274–14279. [PubMed: 11724958]
7. Dwyer MA, Looger LL, Hellinga HW. Computational design of a biologically active enzyme. Science. 2004; 304:1967–1971. [PubMed: 15218149]
8. Park HS, et al. Design and evolution of new catalytic activity with an existing protein scaffold. Science. 2006; 311:535–538. [PubMed: 16439663]
9. Cesaro-Tadic S, et al. Turnover-based in vitro selection and evolution of biocatalysts from a fully synthetic antibody library. Nature Biotech. 2003; 21:679–685.
10. Xu Y, Yamamoto N, Janda KD. Catalytic antibodies: hapten design strategies and screening methods. Bioorg Med Chem. 2004; 12:5247–5268. [PubMed: 15388154]
11. Roberts RW, Szostak JW. RNA-peptide fusions for the in vitro selection of peptides and proteins. Proc Natl Acad Sci USA. 1997; 94:12297–12302. [PubMed: 9356443]
12. Keefe AD, Szostak JW. Functional proteins from a random-sequence library. Nature. 2001; 410:715–718. [PubMed: 11287961]
13. Wilson DS, Keefe AD, Szostak JW. The use of mRNA display to select high-affinity protein-binding peptides. Proc Natl Acad Sci USA. 2001; 98:3750–3755. [PubMed: 11274392]
14. Baggio R, et al. Identification of epitope-like consensus motifs using mRNA display. J Mol Recognit. 2002; 15:126–134. [PubMed: 12203838]

15. Cujec TP, Medeiros PF, Hammond P, Rise C, Kreider BL. Selection of v-abl tyrosine kinase substrate sequences from randomized peptide and cellular proteomic libraries using mRNA display. Chem Biol. 2002; 9:253–264. [PubMed: 11880040]
16. Raffler NA, Schneider-Mergener J, Famulok M. A novel class of small functional peptides that bind and inhibit human alpha-thrombin isolated by mRNA display. Chem Biol. 2003; 10:369–369.
17. Xu LH, et al. Directed evolution of high-affinity antibody mimics using mRNA display. Chem Biol. 2002; 9:933–942. [PubMed: 12204693]
18. Cho GS, Szostak JW. Directed evolution of ATP binding proteins from a zinc finger domain by using mRNA display. Chem Biol. 2006; 13:139–147. [PubMed: 16492562]
19. Holmbeck SMA, et al. High-resolution solution structure of the retinoid X receptor DNA-binding domain. J Mol Biol. 1998; 281:271–284. [PubMed: 9698548]
20. Cadwell RC, Joyce GF. Randomization of genes by PCR mutagenesis. PCR Methods Appl. 1992; 2:28–33. [PubMed: 1490172]
21. Wilson, DS.; Keefe, AD. Curr Prot Mol Biol. Ausubel, FM.; Brent, R.; Kingston, RE.; Moore, DD.; Seidman, JG.; Smith, JA.; Struhl, K., editors. Wiley; New York: 2000.
22. Slootstra JW, Kuperus D, Pluckthun A, Meloen RH. Identification of new tag sequences with differential and selective recognition properties for the anti-FLAG monoclonal antibodies M1, M2 and M5. Mol Divers. 1997; 2:156–164. [PubMed: 9238646]
23. Rohatgi R, Bartel DP, Szostak JW. Kinetic and mechanistic analysis of nonenzymatic, template-directed oligoribonucleotide ligation. J Am Chem Soc. 1996; 118:3332–3339. [PubMed: 11539267]
24. Griffiths AD, Tawfik DS. Man-made enzymes - from design to in vitro compartmentalisation. Curr Opin Biotech. 2000; 11:338–353. [PubMed: 10975453]
25. Bryant FR, Benkovic SJ. On the mechanism of T4 RNA ligase. J Am Chem Soc. 1981; 103:696–697.
26. Bonner G, Lafer EM, Sousa R. Characterization of a set of T7 RNA-polymerase active site mutants. J Biol Chem. 1994; 269:25120–25128. [PubMed: 7929200]
27. Bartel DP, Szostak JW. Isolation of new ribozymes from a large pool of random sequences. Science. 1993; 261:1411–1418. [PubMed: 7690155]
28. Purtha WE, Coppins RL, Smalley MK, Silverman SK. General deoxyribozyme-catalyzed synthesis of native 3′-5′ RNA linkages. J Am Chem Soc. 2005; 127:13124–13125. [PubMed: 16173722]
29. Ekland EH, Szostak JW, Bartel DP. Structurally complex and highly-active RNA ligases derived from random RNA sequences. Science. 1995; 269:364–370. [PubMed: 7618102]
30. Volckaert G, Fiers W. Micro thin-layer techniques for rapid sequence-analysis of P-32-Labeled RNA - double digestion and pancreatic ribonuclease analyses. Anal Biochem. 1977; 83:228–239. [PubMed: 920942]
31. Moore MJ, Sharp PA. Site-specific modification of pre-messenger-RNA - the 2′-hydroxyl groups at the splice sites. Science. 1992; 256:992–997. [PubMed: 1589782]
32. Kurz M, Gu K, Lohse PA. Psoralen photo-crosslinked mRNA-puromycin conjugates: a novel template for the rapid and facile preparation of mRNA-protein fusions. Nucl Acids Res. 2000; 28:e83. [PubMed: 10982894]
33. Cho G, Keefe AD, Liu RH, Wilson DS, Szostak JW. Constructing high complexity synthetic libraries of long ORFs using in vitro selection. J Mol Biol. 2000; 297:309–319. [PubMed: 10715203]
34. Liu RH, Barrick JE, Szostak JW, Roberts RW. Optimized synthesis of RNA-protein fusions for in vitro protein selection. Methods Enzymol. 2000; 318:268–293. [PubMed: 10889994]
35. McCafferty DG, Lessard IAD, Walsh CT. Mutational analysis of potential zinc-binding residues in the active site of the enterococcal D-Ala-D-Ala dipeptidase VanX. Biochemistry. 1997; 36:10498–10505. [PubMed: 9265630]
36. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser. 1999; 41:95–98.
37. Corpet F. Multiple sequence alignment with hierarchical-clustering. Nucl Acids Res. 1988; 16:10881–10890. [PubMed: 2849754]

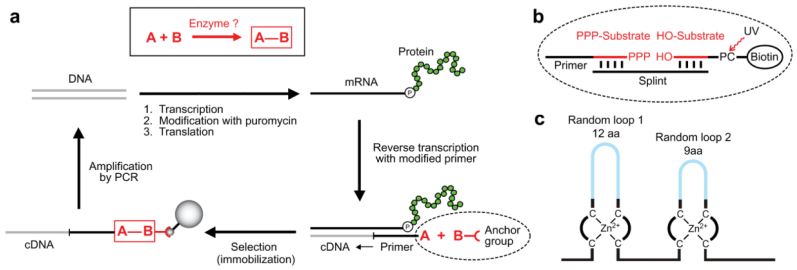

Figure 1. ***In vitro*** **selection of enzymes by mRNA-display**

a, General selection scheme for enzymes for bond-forming reactions. A DNA library is transcribed into RNA, cross-linked to a 3′-puromycin oligonucleotide, and *in vitro* translated. The library of mRNA-displayed proteins is reverse transcribed with a primer bearing substrate A. Substrate B, which carries an anchor group, is added. Proteins that join A and B attach the anchor group to their encoding cDNA. Selected cDNA sequences are then amplified by PCR, and used as input for the next round. **b,** Selection of enzymes that perform template-dependent ligation of a 5′-triphosphate-activated RNA (PPP-substrate) to a second RNA (HO-substrate). The PPP-substrate is ligated to the primer and then used in the reverse transcription reaction. The cDNA of the catalytically active molecules is immobilized on streptavidin-coated beads via biotin, washed, and released by UV-irradiation of the photocleavable linker (PC). **c,** The scaffolded library[18] is based on a two zinc finger domain with two loop regions (light blue) that are replaced by segments of 12 or 9 random amino acids.

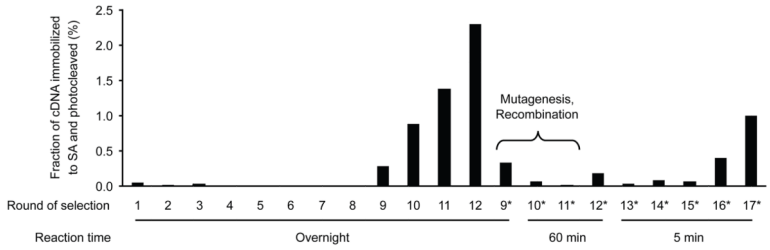

Figure 2. Progress of the selection

The fraction of ^{32}P-labelled cDNA that bound to streptavidin agarose (SA) and eluted after photocleavage at each round of selection is shown. The input DNA into rounds 9*, 10* and 11* was subjected to mutagenic PCR amplification and, in addition, a recombination procedure was performed before rounds 9* and 11*. The selection pressure was increased by decreasing the time of the reaction as indicated. Asterisks indicate selection rounds after mutagenesis and recombination.

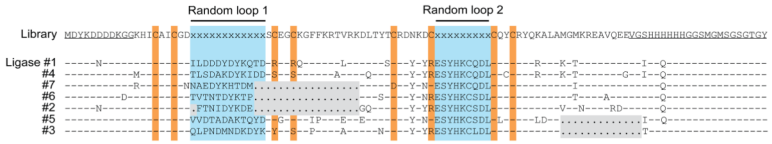

Figure 3. Sequences of the starting library and selected ligases

Loop regions are highlighted in light blue. The cysteines highlighted in orange constitute the two pairs of CX_nC (n = 2 or 5) motifs that coordinate zinc ions in the original hRXR domain. Randomized amino acids in the library are shown as x. Dashes indicate amino acids that are the same as in the starting library, whereas periods highlighted in grey symbolize deletions. The underlined flanking regions were not part of the hRXRα domain but were added to contain a Flag epitope tag, a hexahistidine tag and a linker region.

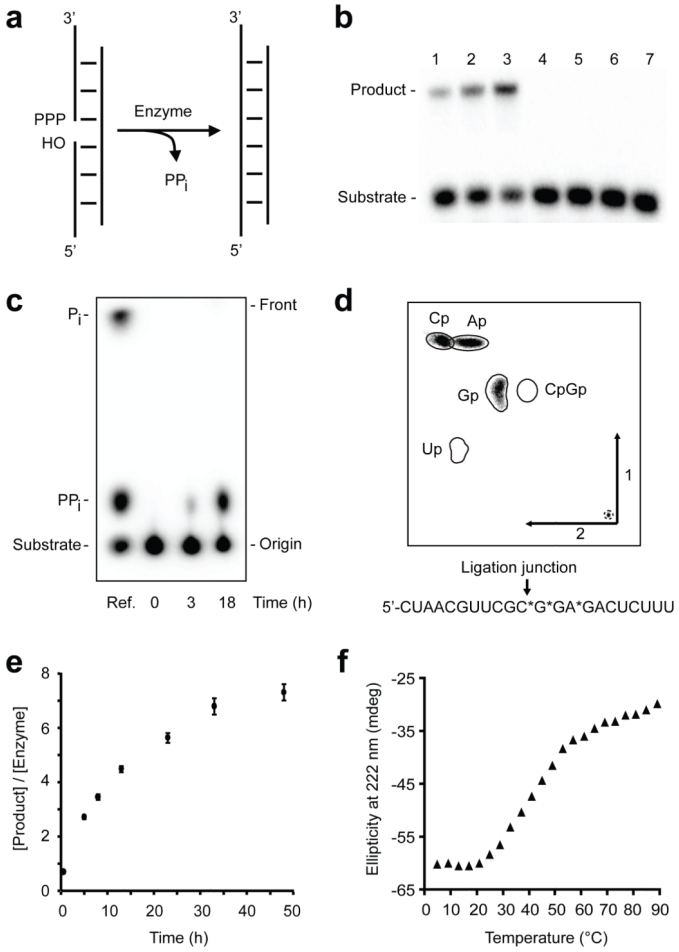

Figure 4. Characterization of ligase enzyme

a, The RNA ligation reaction. **b,** Reaction catalyzed by ligase #4 after 1, 3 and 10 hrs (lanes 1,2,3). Lanes 4–7: 10 hrs with no splint, 5′-monophosphate instead of PPP-substrate; 5′-hydroxyl instead of PPP-substrate, and wild-type hRXRα protein domain instead of ligase #4. **c,** Release of inorganic pyrophosphate during ligation. Ligation reactions with γ-^{32}P GTP-labelled PPP-substrate were separated by thin-layer chromatography. A mixture of inorganic ^{32}P-phosphate (Pi), ^{32}P-pyrophosphate (PPi) and 5′-γ-^{32}P-labelled PPP-substrate was run for reference (Ref.). **d,** 3′-5′ regiospecificity of ligation. Ligation of α-^{32}P GTP body-labelled PPP-substrate yielded product with ^{32}P at the indicated (*) positions. The product was digested to nucleoside monophosphates with ribonuclease T2 (which does not efficiently digest 2′,5′ RNA linkages) in the presence of a 22-nt chemically synthesized RNA identical in sequence to the predicted ligation product but which contains a 2′-5′ linkage at the ligation junction (5′-CUAACGUUCGC$^{2'}$p$^{5'}$GGAGACUCUUU). Digestion products were separated by two-dimensional thin-layer chromatography[30]. Ultraviolet shadowing revealed the carrier RNA digestion products (Ap, Cp, Gp, Up), including the 2′-linked GpCp dinucleotide (encircled spots). Black spots represent the overlaid autoradiograph. The small dashed circle indicates the origin. **e,** Multiple turnover ligation. Substrate oligonucleotides and splint (each 20 μM) were incubated with ligase #4 (1 μM) for

the indicated times and the ligation product was quantified. Error bars indicate s.d. **f**, Thermal unfolding of ligase #6 monitored by circular dichroism spectroscopy.