Vol 448|16 August 2007|doi:10.1038/nature06032

LETTERS

Selection and evolution of enzymes from a partially randomized non-catalytic scaffold

Burckhard Seelig[1] & Jack W. Szostak[1]

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 on 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 messenger RNA 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 $>10^{12}$ independent sequences 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 that is 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 catalysing 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 human protein domain retinoid-X-receptor (RXRα) were randomized; this library has previously been used for the isolation of new ATP-binding proteins[18] (Fig. 1c). We transcribed and translated this synthetic DNA library[18] to generate mRNA-displayed proteins (Fig. 1a), which we then reverse-transcribed with a primer joined to the 3′-end of the 5′-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, which aligns the two substrate oligonucleotides. Proteins that catalysed the ligation of the two substrates

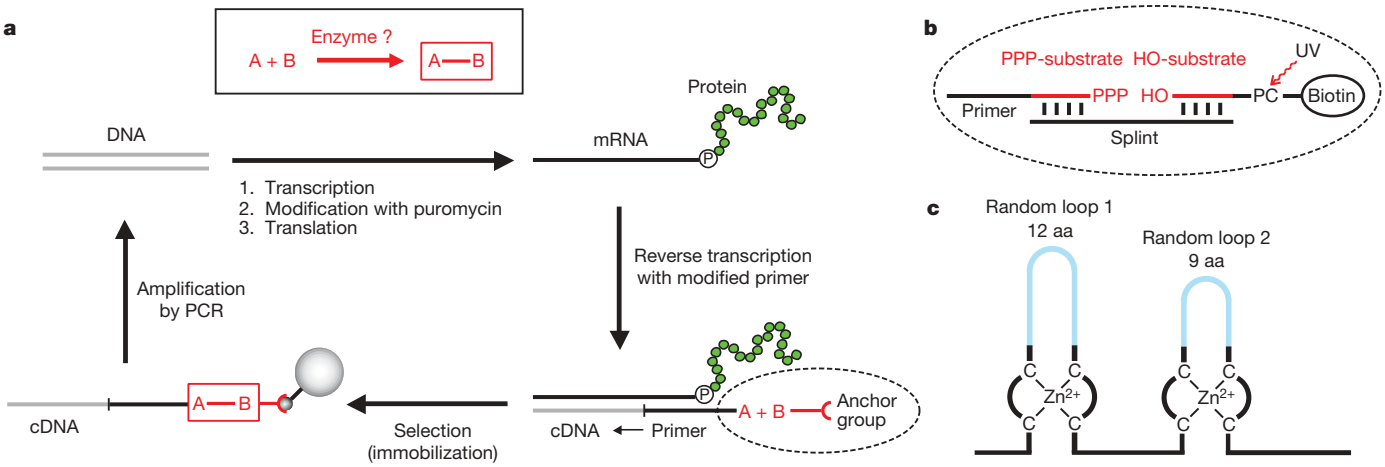

Figure 1 | *In vitro* selection of enzymes by mRNA display. a, General selection scheme for enzymes that catalyse bond-forming reactions. A DNA library is transcribed into RNA, cross-linked to a 3′-puromycin oligonucleotide, and *in vitro* translated. The library of mRNA-displayed proteins is reverse 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 with a 3′-hydroxyl (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 ultraviolet (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 (aa).

[1]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, Massachusetts 02114, USA.

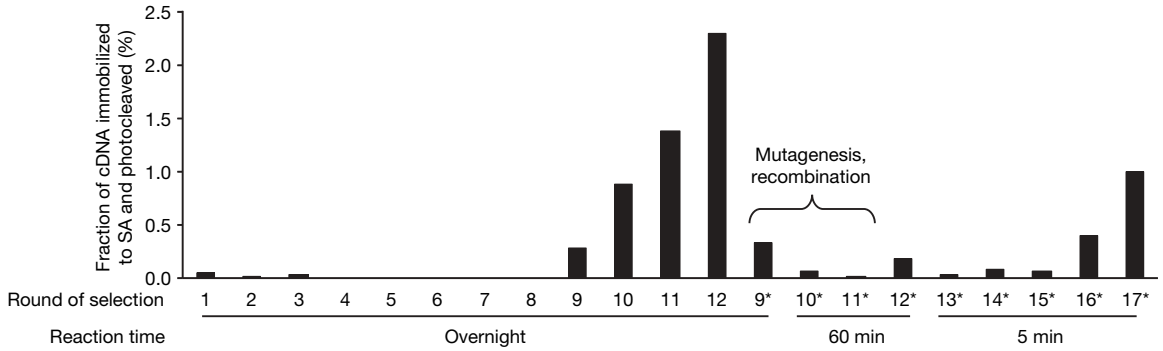

Figure 2 | Progress of the selection. The fraction of ^{32}P-labelled cDNA that bound to streptavidin agarose (SA) and eluted after photocleavage at each round of selection is shown. The input DNA into rounds 9*, 10* and 11* was subjected to mutagenic PCR amplification and, in addition, a recombination 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.

covalently attached the biotin moiety to their own complementary DNA, 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 then 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* (asterisks indicate selection rounds after mutagenesis and recombination) 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 min (Fig. 2).

The evolved pool of enzymes contained several families of closely related sequences as well as multiple unrelated single isolates (see 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 (ref. 22), which we used for purification of the mRNA-displayed proteins. These results indicate that the highly conserved loop 2 may have an essential role in ligase activity, whereas 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 out 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 (Fig. 3) in *Escherichia coli* as carboxy-terminal fusions with maltose binding protein (MBP) or without any fusion partner. All 7 enzymes were soluble when fused to MBP ($\geq$3 mg ml^{-1} 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 catalyses 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

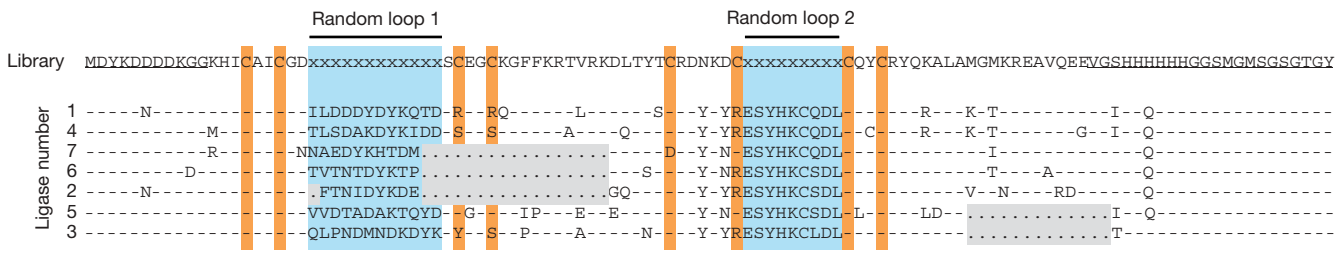

Figure 3 | Sequences of the starting library and selected ligases. Loop regions are highlighted in light blue. The cysteines highlighted in orange constitute the two pairs of CX_nC (n = 2 or 5) motifs that coordinate zinc ions in the original RXRα domain. Randomized amino acids in the library are shown as x. Dashes indicate amino acids that are the same as in the starting library, whereas full-stops highlighted in grey symbolize deletions. The underlined flanking regions were not part of the RXRα domain but were added to form a Flag epitope tag, a hexahistidine tag and a linker region.

required Zn^{2+} and monovalent cations (K^+ or Na^+) with optima of 100 μM and 80 mM, respectively. The rate of the catalysed 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 (ref. 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 ($\pm$s.d.) of bound zinc per ligase molecule, whereas the original wild-type RXRα protein contained 2.1 ± 0.1 molecules of zinc ($\pm$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.

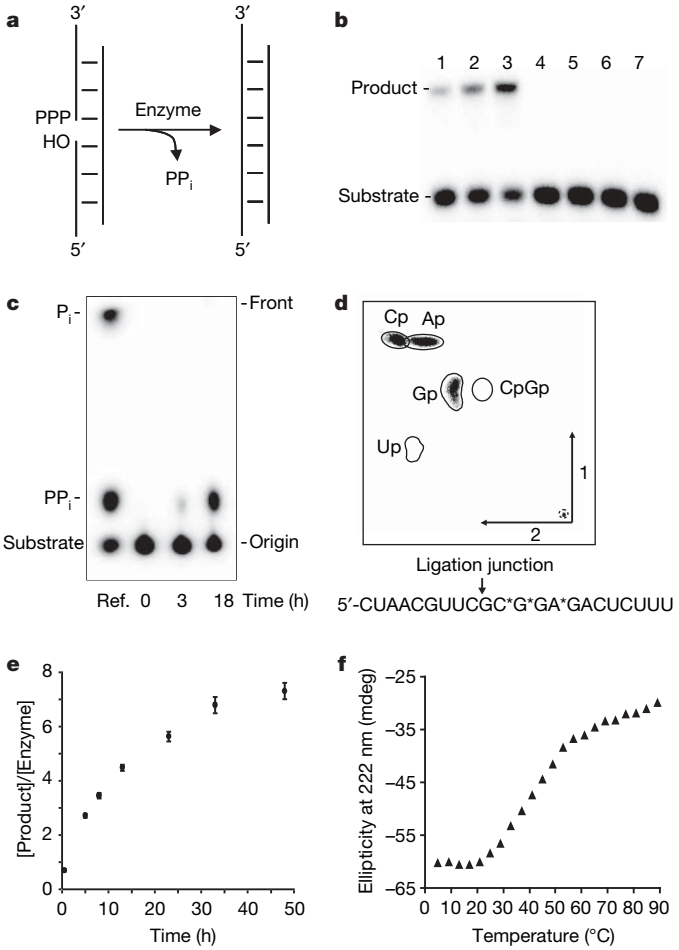

Figure 4 | Characterization of ligase enzyme. a, The RNA ligation reaction. **b**, Reaction catalysed by ligase 4 after 1, 3 and 10 h (lanes 1, 2, 3, respectively). Lanes 4–7, 10 h with no splint; 5′-monophosphate instead of PPP-substrate; 5′-hydroxyl instead of PPP-substrate; and wild-type RXRα protein 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 (P_i), ^{32}P-pyrophosphate (PP_i) 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 RNase T2 (which does not efficiently digest 2′–5′ RNA linkages) in the presence of a chemically synthesized RNA oligonucleotide that is 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 CpGp 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.

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

We found that the evolved enzyme catalysed the ligation reaction with multiple turnover (Fig. 4e), although the selection scheme we employed used 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 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 $^1H^{15}N$ 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 catalyse the ligation of a 5′-triphosphorylated RNA oligonucleotide to the terminal 3′-hydroxyl group of a second RNA, the reaction catalysed by the enzymes described here. An enzyme catalysing 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 catalysed 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 catalyses chain elongation at 240 nucleotides per second[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 catalyse the same reaction as our

protein ligases. These ribozymes and deoxyribozymes have rate enhancements over the uncatalysed 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]. Although our protein ligase was dependent on Zn^{2+} and inhibited by Mg^{2+}, the ribozyme-catalysed 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 new 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, to isolate enzymes that catalyse 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.

METHODS SUMMARY

Ligation activity assay of mRNA-displayed ligases by gel shift. Eighteen individual ligases were expressed separately as mRNA-displayed proteins, and incubated with HO-substrate and splint. After 5 h, the ligation reaction mixture was quenched, mixed with excess streptavidin, and separated by denaturing polyacrylamide gel electrophoresis (PAGE). The substrate and gel-shifted product bands were quantified.

Expression of ligases in *E. coli*. All proteins were expressed in Rosetta BL21 (DE3) cells and purified on either an amylose resin column (MBP-fused proteins) or a Ni-NTA resin column.

Ligation activity assay of free ligases. PPP-substrate (20 μM), 15 μM of splint and 10 μM of radiolabelled HO-substrate were incubated with 5 μM ligase, separated by PAGE and analysed. 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.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 February; accepted 19 June 2007.

1. Schmid, A. *et al.* Industrial biocatalysis today and tomorrow. *Nature* **409,** 258–268 (2001).
2. Cherry, J. R. & Fidantsef, A. L. Directed evolution of industrial enzymes: an update. *Curr. Opin. Biotechnol.* **14,** 438–443 (2003).
3. Panke, S., Held, M. & Wubbolts, M. Trends and innovations in industrial biocatalysis for the production of fine chemicals. *Curr. Opin. Biotechnol.* **15,** 272–279 (2004).
4. Bloom, J. D. *et al.* Evolving strategies for enzyme engineering. *Curr. Opin. Struct. Biol.* **15,** 447–452 (2005).
5. Aharoni, A., Griffiths, A. D. & Tawfik, D. S. High-throughput screens and selections of enzyme-encoding genes. *Curr. Opin. Chem. Biol.* **9,** 210–216 (2005).
6. Bolon, D. N. & Mayo, S. L. Enzyme-like proteins by computational design. *Proc. Natl Acad. Sci. USA* **98,** 14274–14279 (2001).
7. Dwyer, M. A., Looger, L. L. & Hellinga, H. W. Computational design of a biologically active enzyme. *Science* **304,** 1967–1971 (2004).
8. Park, H. S. *et al.* Design and evolution of new catalytic activity with an existing protein scaffold. *Science* **311,** 535–538 (2006).
9. Cesaro-Tadic, S. *et al.* Turnover-based *in vitro* selection and evolution of biocatalysts from a fully synthetic antibody library. *Nature Biotechnol.* **21,** 679–685 (2003).
10. Xu, Y., Yamamoto, N. & Janda, K. D. Catalytic antibodies: hapten design strategies and screening methods. *Bioorg. Med. Chem.* **12,** 5247–5268 (2004).
11. Roberts, R. W. & Szostak, J. W. RNA–peptide fusions for the *in vitro* selection of peptides and proteins. *Proc. Natl Acad. Sci. USA* **94,** 12297–12302 (1997).
12. Keefe, A. D. & Szostak, J. W. Functional proteins from a random-sequence library. *Nature* **410,** 715–718 (2001).
13. Wilson, D. S., Keefe, A. D. & Szostak, J. W. The use of mRNA display to select high-affinity protein-binding peptides. *Proc. Natl Acad. Sci. USA* **98,** 3750–3755 (2001).
14. Baggio, R. *et al.* Identification of epitope-like consensus motifs using mRNA display. *J. Mol. Recognit.* **15,** 126–134 (2002).
15. Cujec, T. P., Medeiros, P. F., Hammond, P., Rise, C. & Kreider, B. L. Selection of v-abl tyrosine kinase substrate sequences from randomized peptide and cellular proteomic libraries using mRNA display. *Chem. Biol.* **9,** 253–264 (2002).
16. Raffler, N. A., 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.* **10,** 69–79 (2003).
17. Xu, L. H. *et al.* Directed evolution of high-affinity antibody mimics using mRNA display. *Chem. Biol.* **9,** 933–942 (2002).
18. Cho, G. S. & Szostak, J. W. Directed evolution of ATP binding proteins from a zinc finger domain by using mRNA display. *Chem. Biol.* **13,** 139–147 (2006).
19. Holmbeck, S. M. A. *et al.* High-resolution solution structure of the retinoid X receptor DNA-binding domain. *J. Mol. Biol.* **281,** 271–284 (1998).
20. Cadwell, R. C. & Joyce, G. F. Randomization of genes by PCR mutagenesis. *PCR Methods Appl.* **2,** 28–33 (1992).
21. Wilson, D. S. & Keefe, A. D. in *Current Protocols in Molecular Biology* (eds Ausubel, F. *et al.*) Suppl. 51, unit 8.3 (Wiley, New York, 2000).
22. Slootstra, J. W., Kuperus, D., Pluckthun, A. & Meloen, R. H. Identification of new tag sequences with differential and selective recognition properties for the anti-FLAG monoclonal antibodies M1, M2 and M5. *Mol. Divers.* **2,** 156–164 (1997).
23. Rohatgi, R., Bartel, D. P. & Szostak, J. W. Kinetic and mechanistic analysis of nonenzymatic, template-directed oligoribonucleotide ligation. *J. Am. Chem. Soc.* **118,** 3332–3339 (1996).
24. Griffiths, A. D. & Tawfik, D. S. Man-made enzymes — from design to *in vitro* compartmentalisation. *Curr. Opin. Biotechnol.* **11,** 338–353 (2000).
25. Bryant, F. R. & Benkovic, S. J. On the mechanism of T4 RNA ligase. *J. Am. Chem. Soc.* **103,** 696–697 (1981).
26. Bonner, G., Lafer, E. M. & Sousa, R. Characterization of a set of T7 RNA-polymerase active site mutants. *J. Biol. Chem.* **269,** 25120–25128 (1994).
27. Bartel, D. P. & Szostak, J. W. Isolation of new ribozymes from a large pool of random sequences. *Science* **261,** 1411–1418 (1993).
28. Purtha, W. E., Coppins, R. L., Smalley, M. K. & Silverman, S. K. General deoxyribozyme-catalyzed synthesis of native 3′–5′ RNA linkages. *J. Am. Chem. Soc.* **127,** 13124–13125 (2005).
29. Ekland, E. H., Szostak, J. W. & Bartel, D. P. Structurally complex and highly-active RNA ligases derived from random RNA sequences. *Science* **269,** 364–370 (1995).
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.* **83,** 228–239 (1977).

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

Acknowledgements We thank G. S. Cho for providing the RXR library before 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 and J. J. Chou for NMR measurements; A. D. Keefe, G. S. Cho, 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 and 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.

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

Author Information The DNA sequences encoding the ligase enzymes 1-7 have been deposited in GenBank under the accession numbers EU019543 to EU019549, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.W.S. (szostak@molbio.mgh.harvard.edu).

METHODS

Sequences of oligonucleotides are described in Supplementary Information.

Preparation of primer for reverse transcription. The reverse transcription primer was a chimaeric 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 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 reverse transcription 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 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 PAGE 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), used according to the manufacturer's instructions with an additional 100 mM KCl per 0.9 mM $Mg(OAc)_2$ and 69 nM ^{35}S-methionine). After addition of 450 mM KCl and 50 mM $MgCl_2$, the translation reaction was incubated at room temperature for 5 min and then diluted 10-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^{-1} oligo(dT)cellulose (New England Biolabs) for 15 min at 4 °C with rotation. The oligo(dT)cellulose was washed on a chromatography column (Bio-Rad) 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 10× Flag binding buffer (1× 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; prewashed with Flag clean buffer (100 mM glycine, pH 3.5, 0.25% w/v Triton X-100) and Flag binding buffer) for 2 h 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; one equivalent of Flag peptide saturates both antigen sites of the antibody resin) for 20 min at 4 °C with rotation. The eluate was diluted to a concentration of 5 nM mRNA-displayed proteins 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 50 nM reverse transcription primer and Superscript II (Gibco BRL) at 42 °C for 30 min. This sample was then dialysed twice against Flag binding buffer at a ratio of 1/1,000 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 5× selection buffer (1× 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 to make a final concentration of 100 µM and 1×, 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; prewashed with PBS buffer (138 mM NaCl, 2.7 mM KCl, 10 mM potassium phosphate, pH 7.4), including 2 mg ml^{-1} tRNA (from baker's yeast, Sigma), and then washed with PBS alone) at room temperature for 20 min with rotation. The streptavidin (SA) agarose was washed on a chromatography column (Bio-Rad) 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 W) at 360 nm from a 1 cm distance for 15 min, while shaking 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 *Ava*II, 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 ref. 18 with some changes. To analyse the results of the selection, the cDNA of the respective round was cloned into the pCR-TOPO vector (TOPO TA Cloning) 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 *RXRA* motif (courtesy of G. S. Cho) was amplified with primers BS68 and BS70. The PCR products were digested with *Nde*I and *Xho*I and cloned into the pIADL14 vector[35] (courtesy of I. Lessard and C. Walsh) to yield the MBP-fusion proteins or into the pET24a vector (Novagen) 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] and 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) and separated by denaturing PAGE. The gel was analysed using a GE Healthcare (Amersham Bioscience) phosphorimager and ImageQuant software.

Expression of ligases and wild-type RXRα in *E. coli*. All proteins were expressed in Rosetta BL21 (DE3) cells (Novagen) containing the recombinant plasmids at 37 °C in LB broth containing 50 µg ml^{-1} 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) 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.

To purify the proteins lacking the MBP-fusion the supernatant after centrifugation was applied to a Ni-NTA resin column (Qiagen) 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 circular dichroism and NMR spectroscopy experiments, the protein was further purified by FPLC (BioCAD Sprint Perfusion System) using a Sephadex-200 gel filtration column (Pharmacia Biotech) 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. PPP-substrate (20 µM; 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 analysed as above.

Detection of pyrophosphate. The MBP fusion of ligase 4 (purified on an amylose column) was immobilized on Ni-NTA resin (Qiagen), 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 dialysed 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, which were 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 RXRα (purified on amylose column) were dialysed 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 1/1,000. 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. Circular dichroism spectra were recorded on an Aviv circular dichroism 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 the sole source of cysteine, respectively.

31. Moore, M. J. & Sharp, P. A. Site-specific modification of pre-messenger-RNA — the 2′-hydroxyl groups at the splice sites. *Science* **256,** 992–997 (1992).
32. Kurz, M., Gu, K. & Lohse, P. A. Psoralen photo-crosslinked mRNA-puromycin conjugates: a novel template for the rapid and facile preparation of mRNA–protein fusions. *Nucleic Acids Res.* **28,** e83 (2000).
33. Cho, G., Keefe, A. D., Liu, R. H., Wilson, D. S. & Szostak, J. W. Constructing high complexity synthetic libraries of long ORFs using *in vitro* selection. *J. Mol. Biol.* **297,** 309–319 (2000).
34. Liu, R. H., Barrick, J. E., Szostak, J. W. & Roberts, R. W. Optimized synthesis of RNA–protein fusions for *in vitro* protein selection. *Methods Enzymol.* **318,** 268–293 (2000).
35. McCafferty, D. G., Lessard, I. A. D. & Walsh, C. T. Mutational analysis of potential zinc-binding residues in the active site of the enterococcal D-Ala-D-Ala dipeptidase VanX. *Biochemistry* **36,** 10498–10505 (1997).
36. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.* **41,** 95–98 (1999).
37. Corpet, F. Multiple sequence alignment with hierarchical-clustering. *Nucleic Acids Res.* **16,** 10881–10890 (1988).