nature chemical biology

BRIEF COMMUNICATION

PUBLISHED ONLINE: 9 DECEMBER 2012 | DOI: 10.1038/NCHEMBIO.1138

Structure and dynamics of a primordial catalytic fold generated by *in vitro* evolution

Fa-An Chao[1], Aleardo Morelli[1,2], John C Haugner III[1,2], Lewis Churchfield[1,2], Leonardo N Hagmann[1,2], Lei Shi[3], Larry R Masterson[1], Ritimukta Sarangi[4], Gianluigi Veglia[1,3] & Burckhard Seelig[1,2]*

Engineering functional protein scaffolds capable of carrying out chemical catalysis is a major challenge in enzyme design. Starting from a noncatalytic protein scaffold, we recently generated a new RNA ligase by *in vitro* directed evolution. This artificial enzyme lost its original fold and adopted an entirely new structure with substantially enhanced conformational dynamics, demonstrating that a primordial fold with suitable flexibility is sufficient to carry out enzymatic function.

The known structures of naturally occurring proteins can be assigned to an apparently finite number of different fold families[1,2]. Starting from an existing fold, divergent evolution through a combination of gene duplication and mutations is a common path for proteins to acquire new functions while retaining their original fold[3,4]. However, the origin of those biological folds remains subject to debate[5,6]. Only a few examples have been described in which new function acquisition is accompanied by a simultaneous change in the protein fold. Those examples have largely been generated by rational design or involve protein binders[7–13].

Recently, we created artificial RNA ligase enzymes by *in vitro* evolution[14,15]. These enzymes catalyze the joining of a 5′-triphosphorylated RNA to the 3′-hydroxyl group of a second RNA, a reaction for which no natural enzyme catalysts have been found. We began with a small noncatalytic protein domain consisting of two zinc finger motifs from the DNA-binding domain of human retinoid X receptor (hRXRα)[16] (**Fig. 1**). Two adjacent loops of this protein were randomized to generate a combinatorial library of mutants as input for the selection and evolution process[17]. Although zinc fingers are common structural motifs, they are not known to take part in catalysis in natural proteins. In contrast, we isolated from the zinc finger library active enzymes that have rate accelerations of more than 2-million-fold[14]. Sequence analysis of the artificial enzyme showed that several amino acids essential to maintaining zinc finger structure integrity were mutated or deleted, suggesting that the original scaffold may have been abandoned during the process of mutagenesis and evolution. The original hRXRα scaffold consisted of two loop-helix domains, each containing a zinc ion tetrahedrally coordinated by four cysteines[16]. However, during evolution of the ligase enzyme, only half of the zinc-coordinating cysteines had been conserved. In the starting scaffold, two helices were packed perpendicularly to form the globular fold and build the hydrophobic core, and an additional helix was located at the C terminus (**Fig. 1b**). In the ligase enzyme, seven residues of the former DNA recognition helix and ten residues of the C-terminal helix were deleted from the original hRXRα scaffold.

NMR structural analyses of the ligase 10C (Online Methods), chosen for its superior solubility and thermostability, revealed that

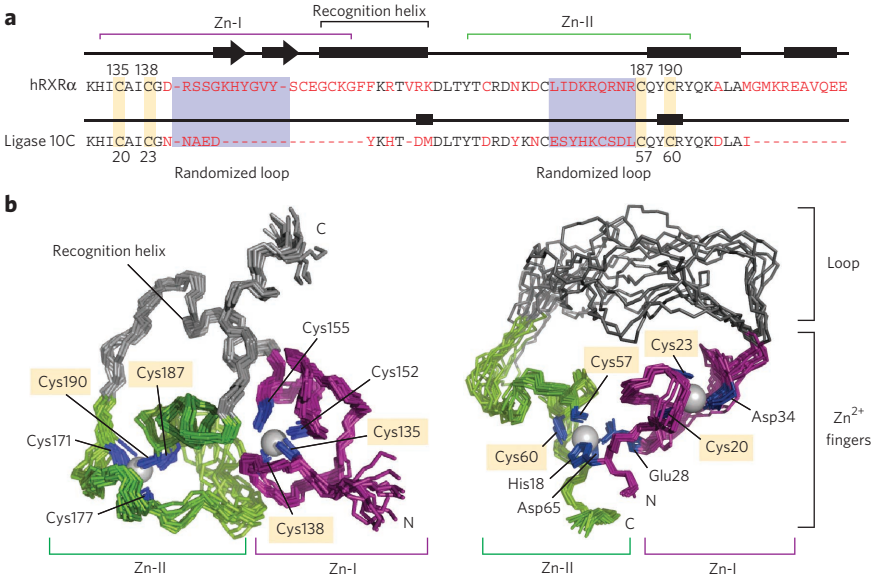


Figure 1 | Changes in primary sequence and 3D structure upon directed evolution of the hRXRα scaffold to the ligase enzyme 10C. (**a**) Comparison of the primary sequences of hRXRα[16] (residues 132–208) and the artificially evolved ligase (residues 17–68). The two zinc finger regions are highlighted with purple and green brackets. Red letters denote residues not conserved between the two sequences. (**b**) 3D structure of hRXRα[16] (left) and NMR ensemble of ligase 10C (right; flexible termini omitted for clarity). Although both proteins contain two zinc fingers, the overall structures are substantially different. Only two zinc-coordinating cysteines of each zinc finger in hRXRα are still coordinating zinc in ligase 10C (highlighted in yellow; also shown in **a**), whereas all other ligands differ in the two structures. Zinc-coordinating residues are labeled and shown in blue. In contrast to that in hRXRα, zinc finger Zn-II in ligase 10C comprises residues of both N- and C-terminal sequences, imposing a cyclic structure to the enzyme. Notably, the new ligase lost both helical domains of hRXRα (gray), replacing the recognition helix with a long unstructured loop (gray).

[1]Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, Minnesota, USA. [2]BioTechnology Institute, University of Minnesota, St. Paul, Minnesota, USA. [3]Department of Chemistry, University of Minnesota, Minneapolis, Minnesota, USA. [4]SLAC National Accelerator Laboratory, Stanford Synchrotron Radiation Lightsource, Menlo Park, California, USA. *e-mail: seelig@umn.edu

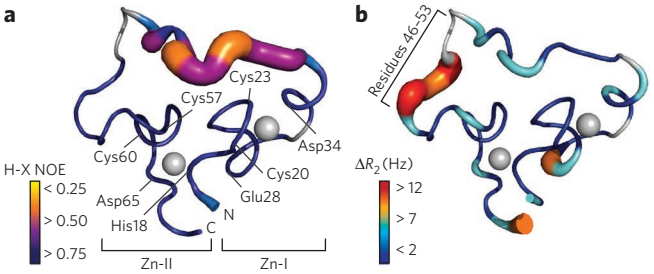


Figure 2 | Conformational dynamics of ligase enzyme 10C. (**a**) Mapping of heteronuclear (H-X) NOEs (proxy for fast dynamics on a ps-ns timescale) on ligase 10C. Zinc-coordinating residues are labeled. (**b**) Mapping of exchange rates (R_{ex} and ΔR_2 (difference between R_2 = 5 ms and R_2 = 0.2 ms)) obtained from relaxation dispersion measurements as proxy for slow dynamics (μs-ms timescale). Color gradient and thickness of backbone indicate that the ligase fast dynamics is located mostly in the unstructured loop, whereas the slow dynamics is located mostly in the region N-terminal to the Zn-II site (residues 46–53) and is potentially correlated to catalytic activity.

the evolved ligase lost the original zinc finger scaffold, adopting an entirely new structure (**Fig. 1b**). This new 3D structure still contained two zinc sites that constitute the folding core of the protein; however, the two Zn^{2+} ions were coordinated by several new ligands with a different register. The deletion of two N-terminal cysteines during directed evolution resulted in the concomitant rearrangement of the local geometry of the zinc-binding loop. Additionally, the short stretch of antiparallel β-sheet within the first zinc finger (Zn-I) was also deleted. The C-terminal loop-helix domains and the recognition helix of hRXRα responsible for binding to the DNA groove[18] were lost completely; the latter was replaced by an unstructured loop of twenty amino acids connecting the two new zinc fingers. The zinc fingers made up the most structured region, as demonstrated by the presence of short- and long-range NOE contacts. Moreover, several long-range NOEs indicated that the two metal-binding loops are in close proximity, whereas most of the protein presented only short-range NOE contacts (**Supplementary Results**, **Supplementary Fig. 1** and **Supplementary Table 1**). The conformational ensemble resulting from simulated annealing calculations showed two well-defined regions (residues 17–35 and 49–69) with r.m.s. deviation from the average of less than 1 Å, whereas the large loop encompassing residues 36–48 was completely unstructured (r.m.s. deviation greater than 6 Å). The 3D structure of the enzyme was compounded by residual dipolar coupling measurements, which also helped to better define the local geometry around the zinc-binding sites (**Supplementary Figs. 2** and **3**).

The two metal centers were responsible for the overall fold of the ligase. In the absence of Zn^{2+}, the NMR fingerprint spectrum of the enzyme showed broad and mostly unresolved resonances, typical of a molten globule. Titration of Zn^{2+} to the metal-free protein first saturated the C-terminal Zn^{2+}-binding site (Zn-II) and induced a substantial structural rearrangement with sharper and more dispersed resonances (**Supplementary Figs. 4** and **5**). The transition between the unfolded and folded states of the ligase involved multiple intermediate species. For selected resonances, we could discern the presence of two distinct states in slow exchange in the NMR timescale. Complete saturation with Zn^{2+} funneled the enzyme into a more defined structure, with the complete resolution of fingerprint resonances showing only one population of peaks. Elemental analysis by inductively coupled plasma MS revealed 2.74 ± 0.01 equivalents (± s.d.) of bound zinc per ligase molecule. We were able to fit the thermocalorimetry data using models with two or more Zn^{2+}-binding sites; the fit, however, does not improve considerably with $n > 2$ (**Supplementary Fig. 6** and **Supplementary Table 2**). Assigning two sites in accordance with the NMR titration data leads to one binding site, Zn-II, with higher affinity (K_d ~3 μM) and a second binding site, Zn-I, with lower affinity for Zn^{2+} (K_d ~93 μM). These values were further supported by the zinc concentration dependence of the enzyme activity, showing a steep drop in activity at concentrations below 100 μM Zn^{2+} (**Supplementary Fig. 7**). Notably, the ligase affinity for Zn^{2+} was substantially lower than those reported for natural zinc-containing proteins, which commonly have dissociation constants of 10^{-8} to10^{-13} M[19]. Structure calculations were carried out in the absence of explicit Zn^{2+} ions to avoid conformational search bias and converged toward a structural ensemble with two distinct Zn^{2+}-binding sites: the tetracoordinated N-terminal site (Zn-I) with weaker binding affinity and the hexacoordinated C-terminal loop (Zn-II) with higher binding affinity (**Supplementary Figs. 2** and **3**). Extended X-ray absorption fine structure (EXAFS) data corroborated these results, showing that both Zn sites coordinated with two sulfur ligands (cysteines) with a Zn-S distance of 2.3 Å, and at least one site had four Zn-N or Zn-O ligands, whereas the other site had two to four. These atom ligands can be either protein-based molecules or water molecules (**Supplementary Fig. 8**).

The directed evolution process that yielded the artificial enzyme was based only on product formation without structural constraints[14]. As a result, the ligase enzyme evolved into a new structure with substantially increased conformational dynamics compared to the original DNA-binding scaffold[20]. In fact, ligase 10C showed an overall increase in structural plasticity and malleability. Although the two zinc fingers had heteronuclear NOEs similar to those of the original scaffold, the loop region that replaced the recognition helix had much higher flexibility, with heteronuclear NOEs below 0.5. These data indicate augmentation of conformational dynamics in the ps-ns timescale supported by longitudinal (T_1) and transverse (T_2) nuclear spin-relaxation measurements as well as hydrogen-deuterium exchange data (**Fig. 2a** and **Supplementary Figs. 9** and **10a**). A distinct signature of the hRXRα structure is slow (μs-ms) conformational dynamics[20] (**Supplementary Fig. 10b**), which may be correlated with the protein's ability to optimize protein-DNA interactions.

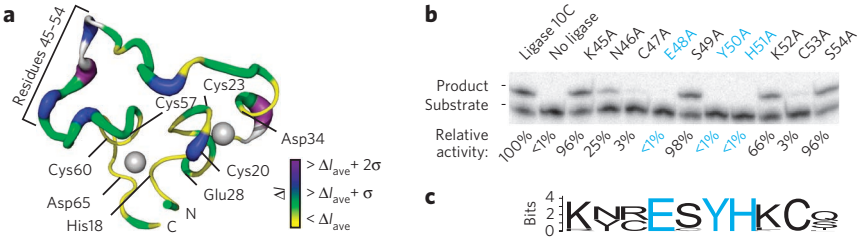


Figure 3 | Substrate-binding surface of ligase 10C probed by NMR and alanine scanning. (**a**) Mapping of NMR chemical shift perturbations (intensity changes, Δ*I*) for ligase 10C shows regions affected by formation of the complex with RNA substrate. The most perturbed regions are indicated by thicker lines and darker colors. Zinc-coordinating residues are labeled. (**b**) Activity assay of ligase 10C and alanine mutants by gel shift. Ligation activity is normalized to the activity of ligase 10C and represents the mean value from two independent experiments. Residues with activity below detection limit are shown in blue. (**c**) Sequence conservation analysis by WebLogo (V2.8.2) of ligase residues 45–54 from 49 enzyme sequences generated by directed evolution[14]. Residues Glu48, Tyr50 and His51 (blue), which completely lost activity during alanine scanning (described above), were conserved among all sequences. The two residues with 97% reduced activity of their alanine mutant were either conserved (Cys53) or had one alternative amino acid (Cys47). None of the other residues in this region were conserved.

The *in vitro* evolution of hRXRα into the RNA ligase redistributed those conformational dynamics, particularly in the region N-terminal to the Zn^{2+}-binding site Zn-II (residues 46–53; **Fig. 2b**).

To probe the substrate-binding surface, we carried out an NMR titration with a pseudosubstrate that lacked the 2′-hydroxyl group, preventing enzyme turnover (**Supplementary Fig. 11**). Chemical shift perturbation mapping of the ligase structure (**Fig. 3a** and **Supplementary Fig. 12**) indicated that one of the highly perturbed regions in the substrate-bound form (residues 46–53) corresponds to high values of chemical exchange (slow conformational dynamics) in the substrate-free form (**Fig. 2b**). Notably, most alanine mutations in this region decreased or completely obliterated the enzyme's activity (**Fig. 3b**). Specifically, mutations E48A, Y50A and H51A abolished enzymatic activity, whereas the C47A and C53A mutations caused a 97% reduction in ligase function. These residues' high conservation among evolved ligase variants further demonstrated their importance (**Fig. 3c**). The combined results suggest that this protein region (residues 46–53) is important for substrate recognition and binding and may contain the active site of the enzyme. Four of those five mutation-sensitive residues (Cys47, Glu48, His51 and Cys53) are good potential metal ligands. Many natural enzymes, such as polymerases, that catalyze chemical reactions similar to the specific RNA ligation described here use a mechanism involving catalytic divalent metal ion cofactors, which are coordinated jointly by the nucleic acid substrates and active site residues of the enzyme[21]. One may speculate that, upon forming the enzyme-substrate complex, some of the mutation-sensitive residues in ligase 10C are involved in binding additional Zn^{2+} ions that facilitate catalysis but are not bound by the protein alone. However, additional experiments studying the enzyme in complex with substrate are needed to elucidate the catalytic mechanism of our artificial enzyme.

The increased flexibility of the new ligase structure relative to the parent hRXRα could potentially originate from their different functional roles. hRXRα is a DNA binder and has been proposed to work through an induced-fit mechanism[16]. In contrast, the RNA ligase has evolved to function as a catalyst. This role requires additional flexibility to optimize interactions with a target molecule and carry out chemical catalysis using transient interactions that occur in excited conformational states rather than through a stable, low-energy complex[22–24]. This argument is supported by an independent directed evolution experiment in which the same hRXRα library yielded proteins that bind ATP and maintain the original, noncatalytic DNA-binding scaffold but have no catalytic function[17]. In contrast, the evolution of the ligase enzyme resulted in a different structure and increased dynamics.

Compared to natural enzymes, which evolved over billions of years, the laboratory-evolved ligase enzyme contains substantially fewer secondary structure elements, such as α-helices and β-strands, and instead has increased flexibility. The complete reorganization of the starting scaffold during *in vitro* evolution may have led to the loss of these structural elements. This new structure has not been subjected to the extensive selection pressure that shaped contemporary enzymes during their natural evolution and can therefore be considered an early or primordial catalytic fold. Further evolution of this enzyme *in vitro* or inside a cell will allow us to explore whether incremental mutations lead to structural and dynamic properties more similar to natural enzymes. Although flexibility has been suggested to increase the probability of developing new functions[5], it also reduces overall protein stability—a trade-off that enzymes must balance during evolution.

This report describes what is to our knowledge the first new protein structure emerging simultaneously with a new enzymatic function. This ligase evolved in the absence of selection pressure to maintain the protein's original function (DNA binding). Would proteins evolving in nature also more readily adopt new folds and functions if they were freed from having to maintain their original function? Although the search for such examples in nature is still ongoing, the simplified environment of *in vitro* evolution enables us to generate precedents and study basic principles of complex natural evolution. Finally, *in vitro* directed evolution has the potential to produce new biocatalysts for a wide range of applications. The unique structure of the artificial ligase enzyme demonstrates that this approach can successfully generate new enzymes without being limited to known biological folds[25].

Received 21 February 2012; accepted 1 November 2012; published online 9 December 2012

Methods

Methods and any associated references are available in the online version of the paper.

Accession codes. Protein Data Bank: the accession code for ligase 10C is 2LZE. Biological Magnetic Resonance Data Bank: the accession number for ligase 10C is 18749.

References

1. Chothia, C. *Nature* **357**, 543–544 (1992).
2. Murzin, A.G., Brenner, S.E., Hubbard, T. & Chothia, C. *J. Mol. Biol.* **247**, 536–540 (1995).
3. Ohno, S. *Evolution by Gene Duplication* (Springer-Verlag, New York, 1971).
4. Chothia, C., Gough, J., Vogel, C. & Teichmann, S.A. *Science* **300**, 1701–1703 (2003).
5. James, L.C. & Tawfik, D.S. *Trends Biochem. Sci.* **28**, 361–368 (2003).
6. Tokuriki, N. & Tawfik, D.S. *Science* **324**, 203–207 (2009).
7. Cordes, M.H.J., Walsh, N.P., McKnight, C.J. & Sauer, R.T. *Science* **284**, 325–328 (1999).
8. Kaplan, J. & DeGrado, W.F. *Proc. Natl. Acad. Sci. USA* **101**, 11566–11570 (2004).
9. Tuinstra, R.L. *et al. Proc. Natl. Acad. Sci. USA* **105**, 5057–5062 (2008).
10. Bryan, P.N. & Orban, J. *Curr. Opin. Struct. Biol.* **20**, 482–488 (2010).
11. Smith, B.A. & Hecht, M.H. *Curr. Opin. Chem. Biol.* **15**, 421–426 (2011).
12. Keefe, A.D. & Szostak, J.W. *Nature* **410**, 715–718 (2001).
13. Mansy, S.S. *et al. J. Mol. Biol.* **371**, 501–513 (2007).
14. Seelig, B. & Szostak, J.W. *Nature* **448**, 828–831 (2007).
15. Seelig, B. *Nat. Protoc.* **6**, 540–552 (2011).
16. Holmbeck, S.M.A. *et al. J. Mol. Biol.* **281**, 271–284 (1998).
17. Cho, G.S. & Szostak, J.W. *Chem. Biol.* **13**, 139–147 (2006).
18. Zhao, Q. *et al. J. Mol. Biol.* **296**, 509–520 (2000).
19. Maret, W. & Li, Y. *Chem. Rev.* **109**, 4682–4707 (2009).
20. van Tilborg, P.J. *et al. Biochemistry* **39**, 8747–8757 (2000).
21. Yang, W., Lee, J.Y. & Nowotny, M. *Mol. Cell* **22**, 5–13 (2006).
22. Bhabha, G. *et al. Science* **332**, 234–238 (2011).
23. Baldwin, A.J. & Kay, L.E. *Nat. Chem. Biol.* **5**, 808–814 (2009).
24. Henzler-Wildman, K. & Kern, D. *Nature* **450**, 964–972 (2007).
25. Golynskiy, M.V. & Seelig, B. *Trends Biotechnol.* **28**, 340–345 (2010).

Acknowledgments

We thank M. Golynskiy and A. Pohorille for helpful discussions; Z. Sachs, F.P. Seebeck, J.W. Szostak and F. Hollfelder for comments on the manuscript; and R. Majerle for isothermal titration calorimetry instrument use. This work was supported by the US National Aeronautics and Space Administration (NASA) Agreement no. NNX09AH70A through the NASA Astrobiology Institute–Ames Research Center (to F.-A.C., A.M., L.C. and B.S.); the Minnesota Medical Foundation (to B.S.) and the US National Institutes of Health (NIH) (T32 GM08347 to J.C.H., T32 DE007288 to L.R.M., GM100310 to G.V. and P41 RR001209). Stanford Synchrotron Radiation Lightsource (SSRL) operations are funded by the US Department of Energy (DOE)–Basic Energy Sciences. The SSRL Structural Molecular Biology program is supported by NIH–National Center for Research Resources and DOE–Biological Environmental Resarch.

Author contributions

G.V. and B.S. designed the project; A.M., J.C.H., L.C. and L.N.H. expressed and assayed proteins; F.-A.C. carried out all NMR and isothermal titration calorimetry experiments; F.-A.C. and L.S. calculated the structure; R.S. performed the EXAFS measurements, all authors analyzed the data; and F.-A.C., L.R.M., G.V. and B.S. wrote the paper.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to B.S.

ONLINE METHODS

All chemical compounds used in this study were purchased from Sigma-Aldrich unless noted otherwise, of molecular biology grade and certified for the absence of RNases when used for ligation reactions.

Sequence of RNA ligase 10C. MGAPVPYPDPLEPRGGKHICAICGNNA EDYKHTDMDLTYTDRDYKNCESYHKCSDLCQYCRYQKDLAIHHQ HHHGGSMGMSGSGTGY

All ligase protein preparations consisted of the sequence above except for point mutations in the case of ligase mutants. Note that the sequence HHQHHH functions similarly to a His_6 tag.

Expression and purification of ^{15}N-labeled ligase protein for NMR studies. Ligase samples were expressed in *Escherichia coli* BL21-DE3 Rosetta strain cells (Novagen). Cells were grown in LB medium with 36 μg/mL kanamycin overnight at 37 °C. This culture was then used to inoculate 1 l of LB medium containing 36 μg/mL kanamycin. The cultures were grown to a $D_{600\ nm}$ of 0.6–0.8 at 37 °C, spun down and resuspended in M9 minimal medium (50 mM Na_2PHO_4, 22 mM KH_2PHO_4, 8.5 mM NaCl, 2 mM $MgSO_4$, 1 mg/L thiamine, 1 mg/L biotin, 60 μM $ZnSO_4$, 10 g/L dextrose, 1 g/L $^{15}NH_4Cl$ and 36 μg/mL kanamycin, pH = 7.3). Cultures were shaken for 1 h at 37 °C, induced with 1 mM IPTG and shaken overnight at room temperature before being spun down and stored at −20 °C.

Frozen cell pellets were resuspended in lysis buffer (20 mM HEPES, 400 mM NaCl, 100 μM $ZnCl_2$, 100 mg/L Triton X-100, 5 mM β-mercaptoethanol, pH = 7.4) and lysed using a S-450D Digital Sonifier (Branson). Cell debris was removed by centrifugation, and the His_6-tagged ligase protein was purified by affinity chromatography using Ni-NTA Superflow resin (Qiagen). The protein was eluted with acidic elution buffer (20 mM NaOAc, 400 mM NaCl, 0.1 mM $ZnCl_2$, 100 mg/L Triton X-100, 5 mM β-mercaptoethanol, pH = 4.5) into 1 M HEPES at pH 7.5 and immediately mixed to adjust the pH. Protein purification was evaluated by SDS-PAGE on Ready Gel precast gels (Bio-Rad). Elution fractions containing ligase protein were concentrated under high pressure in a stirred-cell concentrator unit with a 5,000-MWCO Ultracel Ultrafiltration cellulose membrane (Millipore) and were dialyzed into FPLC buffer (20 mM HEPES, 150 mM NaCl, 0.1 mM $ZnCl_2$ and 0.5 mM β-mercaptoethanol, pH = 7.5).

Monomer ligase protein was isolated by size-exclusion chromatography using the AKTA FPLC system (GE Healthcare) equipped with a 10 mm × 300 mm column (Tricorn) and Superdex 75 resin (GE Healthcare). The separation was carried out in FPLC buffer. Fractions containing monomer protein were pooled and concentrated using 10,000-MWCO Ultra-4 Centrifugal Filter units (Millipore). Purity was assessed by SDS-PAGE gel (**Supplementary Fig. 13**).

Expression and purification of ^{15}N/^{13}C-labeled ligase samples for NMR studies. Ligase samples were expressed in *E. coli* BL21-DE3 Rosetta strain cells (Novagen). Cells were grown in LB medium with 36 μg/mL kanamycin overnight at 37 °C, spun down and resuspended in M9 minimal medium (contents as described above, except with 2 g/L [^{13}C]dextrose). The resuspended cells were used to inoculate 100 mL of M9 minimal medium and were grown to a $D_{600\ nm}$ of 0.6 at 37 °C, at which time the culture was used to inoculate 900 mL of M9 minimal medium. The 1-l culture was grown to a $D_{600\ nm}$ of 1.0 at 37 °C, induced with 1 mM IPTG and shaken overnight at 37 °C before being spun down and stored at −20 °C. The ^{15}N,^{13}C-labeled protein was purified in the same manner as the ^{15}N-labeled protein samples.

Expression of selectively labeled ligase protein for NMR studies. Ligase samples were expressed in *E. coli* BL21-DE3 Rosetta strain cells (Novagen). Cells were grown in LB medium with 36 μg/mL kanamycin overnight at 37 °C, spun down and used to inoculate 1 l of selectively labeled M9 medium (40 mM Na_2PHO_4, 22 mM KH_2PHO_4, 8.5 mM NaCl, 1 mM $MgSO_4$, 50 μM $CaCl_2$, essential vitamins and minerals and 36 μg/mL kanamycin, pH 7.0). To the medium was also added 250 mg of a single ^{15}N-labeled amino acid (cysteine, leucine, lysine or tyrosine), 600 mg of the remaining 19 unlabeled amino acids and, except when labeling [^{15}N]cysteine, one of the following additional amino acid supplements: 900 mg glutamine, asparagine and arginine when labeling [^{15}N]lysine; 900 mg valine and isoleucine when labeling [^{15}N]leucine; and 900 mg phenylalanine, tryptophan, alanine, serine, glycine and cysteine when labeling [^{15}N]tyrosine. Cultures were grown to a $D_{600\ nm}$ of 1.0 at 37 °C, induced with 1 mM IPTG and shaken for 6 h at 37 °C before being spun down and stored at −20 °C. Selectively labeled protein was purified in the same manner as the ^{15}N-labeled protein samples.

Generation of ligase mutants. Ligase mutants were obtained by site-directed mutagenesis (QuikChange Lightning, Agilent). Plasmid DNA was purified using the QIAprep Spin Miniprep kit (Qiagen). The ligase mutants were verified by DNA sequencing. The primer sequences used to generate the indicated mutations in the ligase were designed in accordance with the QuikChange Primer Design tool (Agilent) and were as follows:

K45A F: 5′-CTACACCGATCGAGACTACGCGAATTGTGAGAGCTA CC-3′

K45A R: 5′-GGTAGCTCTCACAATTCGCGTAGTCTCGATCGGTGT AG-3′

N46A F: 5′-CCGATCGAGACTACAAGGCTTGTGAGAGCTACCATAA GTG-3′

N46A R: 5′-CACTTATGGTAGCTCTCACAAGCCTTGTAGTCTCGAT CGG-3′

C47A F: 5′-CCGATCGAGACTACAAGAATGCTGAGAGCTACCATAA-3′

C47A R: 5′-TTATGGTAGCTCTCAGCATTCTTGTAGTCTCGATCGG-3′

E48A F: 5′-GACTACAAGAATTGTGCGAGCTACCATAAGTGCTCGG-3′

E48A R: 5′-CCGAGCACTTATGGTAGCTCGCACAATTCTTGTAGTC-3′

S49A F: 5′-AGACTACAAGAATTGTGAGGCCTACCATAAGTGCTCG GAC-3′

S49A R: 5′-GTCCGAGCACTTATGGTAGGCCTCACAATTCTTGTAGT CT-3′

Y50A F: 5′-CTACAAGAATTGTGAGAGCGCCCATAAGTGCTCGGAC TTGTG-3′

Y50A R: 5′-CACAAGTCCGAGCACTTATGGGCGCTCTCACAATTCT TGTAG-3′

H51A F: 5′-CTACAAGAATTGTGAGAGCTACGCTAAGTGCTCGGAC TTGTG-3′

H51A R: 5′-CACAAGTCCGAGCACTTAGCGTAGCTCTCACAATTCT TGTAG-3′

K52A F: 5′-ACAAGAATTGTGAGAGCTACCATGCGTGCTCGGACTT GTGC-3′

K52A R: 5′-GCACAAGTCCGAGCACGCATGGTAGCTCTCACAATTC TTGT-3′

C53A F: 5′-GTGAGAGCTACCATAAGGCCTCGGACTTGTGCCAGT-3′

C53A R: 5′-ACTGGCACAAGTCCGAGGCCTTATGGTAGCTCTCAC-3′

S54A F: 5′-GTGAGAGCTACCATAAGTGCGCGGACTTGTG-3′

S54A R: 5′-CACAAGTCCGCGCACTTATGGTAGCTCTCAC-3′

Expression and purification of ligase mutants. Ligase mutants were expressed in *E. coli* BL21-DE3 Rosetta cells (Novagen). Cells were cultured in 1 l of LB medium with 36 μg/mL kanamycin to a D_{600} of 0.8–1.0 at 37 °C. Cultures were induced with 1 mM IPTG and shaken for 6 h at 37 °C before being spun down and stored at −20 °C. Ligase mutant proteins were purified by Ni-NTA affinity chromatography in the same manner as described for the ^{15}N-labeled protein samples.

Analysis of metal content by ICP-MS. Ligase 10C was purified as described for the ^{15}N-labeled protein samples and then dialyzed three times against buffer (100 mM NaCl, 10 mM β-mercaptoethanol and 20 mM TrisHCl at pH 7.5; pretreated with Chelex 100 beads (Bio-Rad) for 2 h and filtered) at a ratio of 1/1,000. The metal content of 14 μM protein was measured by ICP MS (Thermo Scientific XSERIES 2 ICP-MS with ESI PC3 Peltier-cooled spray chamber at the Department of Earth Sciences, University of Minnesota).

Ligase activity assay for zinc dependence. Ligase 10C was purified as reported previously[14]. Zinc was removed from ligase 10C by treatment with ion-exchange resin (Chelex 100, Bio-Rad). Ligase 10C (5 μM) was incubated with 20 μM HO-substrate, 10 μM ^{32}P-labeled PPP-substrate base-paired to splint, 20 mM HEPES (pH 7.5), 150 mM NaCl, 500 μM β-mercaptoethanol and the concentrations of $ZnCl_2$ as indicated in **Supplementary Figure 7** for 6 h at room temperature. The ligation reactions were quenched with 20 mM EDTA in 8 M urea, heated to 95 °C for 4 min and separated by denaturing PAGE gel. The gel was analyzed using a GE Healthcare (Amersham Bioscience) Phosphorimager and ImageQuant software (Amersham Bioscience).

doi:10.1038/nchembio.1138

Ligase activity assay of 10C and alanine mutants. Ligase 10C (5 μM) (or alanine mutant) was incubated for 6 h at room temperature in the presence of 20 μM HO-substrate, 10 μM ^{32}P-labeled PPP-substrate base-paired to splint, 24 mM HEPES (pH 7.5), 130 mM NaCl, 100 μM β-mercaptoethanol and 120 μM $ZnCl_2$. The ligation reactions were quenched with 20 mM EDTA and 8 M urea, heated to 95 °C for 4 min and separated by denaturing PAGE gel. The gel was analyzed using a GE Healthcare (Amersham Bioscience) Phosphorimager and ImageQuant software (Amersham Bioscience).

Resonance assignments. All NMR data were collected at the University of Minnesota NMR Center. NMR spectra were acquired at 298 K on a Bruker spectrometer equipped with a cryoprobe at 700 MHz and a Varian spectrometer at 600 MHz. The samples were in a buffer of 150 mM NaCl, 20 mM HEPES and 10 mM β-mercaptoethanol, pH 7.5. Moreover, all protein samples were saturated with $ZnCl_2$ by observing changes in HSQC spectra before other NMR experiments. Triple resonance spectra such as CBCA(CO)NH and HNCACB[26–28] were used to assign peaks on ^{15}N-HSQC. All resonances in these two three-dimensional spectra and ^{15}N-HSQC were picked and fed into the PISTACHIO program (National Magnetic Resonance Facility, Madison, Wisconsin, USA)[29] to obtain preliminary assignments. Final complete assignments were done by manual checks and searches. Carbonyl groups and others side chain carbons were assigned by HNCO and C(CO)NH-TOCSY[30]; side chain protons were assigned by ^{15}N-NOESY-HSQC, ^{15}N-TOCSY-HSQC and HC(CO)NH-TOCSY experiments[30], with mixing times of 150 ms, 60 ms and 12 ms, respectively.

Distance restraints. All proton distance restraints were determined from the cross-peak intensities in the NOESY spectra by calibration with HN(i)Hα(i–1) distances located at the C-terminal region[31], whose helix propensity was shown by chemical shift index and $^3J_{HNH\alpha}$ coupling values[31,32]. The cross-peaks from HN(i)Hα(i–1) distances in that region were categorized as medium NOEs, so the intensities of other cross-peaks smaller than this intensity range were defined as weak NOEs, and those larger than this range belonged to strong NOEs. The upper bounds of distance restraints of strong, medium and weak NOEs were given as 2.9 Å, 3.5 Å and 5 Å, respectively, and lower bounds were set to 1.8 Å in all cases. Starting from unambiguously assigned NOEs at the beginning of calculation, miscalibrated NOEs were adjusted, and then ambiguously assigned NOEs were gradually added into the restraint table during iterative calculation.

Torsion angle restraints. Backbone φ angle restraints were acquired from the HNHA experiment, and the quantitative $^3J_{HNH\alpha}$ coupling values were calculated from the intensity ratios of cross-peaks to diagonal peaks and corrected by 3.7% to account for relaxation[33]. The correction is proportional to the rotational correlation time of the protein (3 ns), which was measured from one-dimensional TRACT experiments[34]. The φ angle of residue i with J-coupling larger than 8.5 Hz was restrained from −160° to −80°, and that with J-coupling smaller than 6 Hz was restrained from −90° to −40°. Moreover, the ψ angle restraints were derived from the ^{15}N-NOESY data. If the intensity ratio of the HN(i)Hα(i) cross-peak to the HN(i)Hα(i–1) peak is smaller than one, the ψ(i-1) is restrained from 20° to 220°; otherwise, the ψ(i–1) is restrained from 80° to −140° (refs. 35,36).

RDC measurement. The stability of several alignment media for ligase 10C was tested. Using 5% neutral and negatively charged acrylamide gel was first attempted, but only weak residual dipolar couplings (absolute values <5 Hz) were obtained. Additionally, the samples precipitated in both DMPC-D7PC and DMPC-D6PC bicelle preparations. We also tested the liquid crystalline medium formed by cetylpyridinium chloride (CPCl) and 1-hexanol but had poor results in terms of sample stability. The sample was finally aligned in the other liquid crystalline medium made by the mixture of C12E5 (5% alkylpoly(ethylene glycol)) and 1-hexanol (r = 0.85)[37]. The residual dipolar couplings of amide groups were obtained by measuring the splitting difference between a decoupling HSQC peak and a TROSY peak in isotropic solution and anisotropic medium.

Structure calculations. Simulating annealing protocols were performed in the XPLOR package[38]. An extended structure was first generated, and the initial temperature was set at 3,500 K, then the temperature was cooled down to 0 K with 15,000 steps. The structure with the lowest energy was used for refinement with the initial temperature of 5,000 K and 30,000 steps. The resulting structure was further refined with RDC data after optimization of the parameters Da and Rh. The angle restraints of the zinc coordination geometry were based on ideal geometries derived from X-ray data[39,40], which are in quantitative agreement with the EXAFS experiments. Distances derived from EXAFS have previously been used as restraints in NMR refinement[41,42]. Here, we report a structural ensemble of 20 conformers. The PROCHECK statistics show that 76.4% of residues are in most favored regions and that 21.1% of residues are in allowed regions.

Zn K-edge EXAFS. Ligase 10C protein was fully saturated with excess Zn^{2+} and then dialyzed to remove excess Zn^{2+}. The final protein sample was 1.39 mM in 15 mM Tris, pH 7.5 and 112.5 mM NaCl. Glycerol (20% v/v) was added to the protein samples to form a glass required for the EXAFS experiments. The Zn K-edge X-ray absorption spectra of ligase 10C were measured at the Stanford Synchrotron Radiation Lightsource (SSRL) on the 16-pole, 2-T Wiggler beamline 9-3 under standard ring conditions of 3 GeV and ~200-mA ring current. A Si(220) double-crystal monochromator was used for energy selection. Another optical component used for the experiment was a cylindrical Rh-coated bent focusing mirror. Spectra were collected in the fully tuned configuration of the monochromator. The solution samples were immediately frozen after preparation and stored under liquid N_2 until measurement. During data collection, the samples were maintained at a constant temperature of ~6 K using an Oxford Instruments CF 1208 liquid helium cryostat. Data were measured to k = 16 Å^{-1} by using a Canberra Ge 100-element monolith detector. Internal energy calibration was accomplished by simultaneous measurement of the absorption of a Zn foil placed between two ionization chambers situated after the sample. The first inflection point of the foil spectrum was fixed at 9,660.7 eV. Data presented here are a 15-scan average. The data were processed by fitting a second-order polynomial to the pre-edge region and subtracting this from the entire spectrum as background. A five-region spline of orders 2, 3, 3, 3 and 3 was used to model the smoothly decaying post-edge region. The data were normalized by subtracting the cubic spline and assigning the edge jump to 1.0 at 9,680 eV using the Pyspline program[43]. Theoretical EXAFS signals $\chi(k)$ were calculated by using FEFF (Macintosh version 8.4)[44–46]. The initial model was based on the $Zn(Cys)_2(His)_2$ active site in a zinc finger protein (Protein Data Bank code 1MEY)[47]. On the basis of the preliminary fits, the models were modified to accommodate a six-coordinate active site (four Zn-N/O and two Zn-S(Cys)).

The theoretical models were fit to the data using EXAFSPAK[48]. The structural parameters varied during the fitting process were the bond distance (R) and the bond variance σ^2, related to the Debye-Waller factor resulting from thermal motion and static disorder of the absorbing and scattering atoms. The nonstructural parameter E_0 (the energy at which k = 0) was also allowed to vary but was restricted to a common value for every component in a given fit. Coordination numbers were systematically varied in the course of the fit but were fixed within a given fit.

26. Grzesiek, S. & Bax, A. *J. Magn. Reson.* **96**, 432–440 (1992).
27. Muhandiram, D.R. & Kay, L.E. *J. Magn. Reson. B.* **103**, 203–216 (1994).
28. Wittekind, M. & Mueller, L. *J. Magn. Reson. B.* **101**, 201–205 (1993).
29. Eghbalnia, H.R., Bahrami, A., Tonelli, M., Hallenga, K. & Markley, J.L. *J. Am. Chem. Soc.* **127**, 12528–12536 (2005).
30. Grzesiek, S., Anglister, J. & Bax, A. *J. Magn. Reson. B.* **101**, 114–119 (1993).
31. Wuthrich, K. *NMR of Proteins and Nucleic Acids* (John Wiley and Sons, New York, 1986).
32. Wishart, D.S., Sykes, B.D. & Richards, F.M. *J. Mol. Biol.* **222**, 311–333 (1991).
33. Vuister, G.W. & Bax, A. *J. Am. Chem. Soc.* **115**, 7772–7777 (1993).
34. Lee, D., Hilty, C., Wider, G. & Wuthrich, K. *J. Magn. Reson.* **178**, 72–76 (2006).
35. Gagné, S.M. *et al. Protein Sci.* **3**, 1961–1974 (1994).
36. Wang, Y., Zhao, S., Somerville, R.L. & Jardetzky, O. *Protein Sci.* **10**, 592–598 (2001).
37. Rückert, M. & Otting, G. *J. Am. Chem. Soc.* **122**, 7793–7797 (2000).
38. Schwieters, C.D., Kuszewski, J.J., Tjandra, N. & Clore, G.M. *J. Magn. Reson.* **160**, 65–73 (2003).
39. Alberts, I.L., Nadassy, K. & Wodak, S.J. *Protein Sci.* **7**, 1700–1716 (1998).
40. Viles, J.H. *et al. J. Mol. Biol.* **279**, 973–986 (1998).
41. Ohlenschläger, O. *et al. Oncogene* **25**, 5953–5959 (2006).
42. Banci, L., Bertini, I., Del Conte, R., Mangani, S. & Meyer-Klaucke, W. *Biochemistry* **42**, 2467–2474 (2003).
43. Tenderholt, A. *Pyspline* (Stanford University, Stanford, 2007).
44. Mustre de Leon, J., Rehr, J.J., Zabinsky, S.I. & Albers, R.C. *Phys. Rev. B. Condens. Matter* **44**, 4146–4156 (1991).

45. Rehr, J.J. & Albers, R.C. *Rev. Mod. Phys.* **72**, 621–654 (2000).
46. Rehr, J.J., Deleon, J.M., Zabinsky, S.I. & Albers, R.C. *J. Am. Chem. Soc.* **113**, 5135–5140 (1991).
47. Kim, C.A. & Berg, J.M. *Nat. Struct. Biol.* **3**, 940–945 (1996).
48. George, G.N. *EXAFSPAK and EDG-FIT* (Stanford Synchrotron Radiation Lightsource, Menlo Park, 2000).

doi:10.1038/nchembio.1138

SUPPLEMENTARY INFORMATION

Structure and dynamics of a primordial catalytic fold generated by *in vitro* evolution

Fa-An Chao[1], Aleardo Morelli[1,3], John C. Haugner III[1,3], Lewis Churchfield[1,3], Leonardo N. Hagmann[1,3], Lei Shi[2], Larry R. Masterson[1], Ritimukta Sarangi[4], Gianluigi Veglia[1,2] & Burckhard Seelig[1,3*]

[1]Department of Biochemistry, Molecular Biology, and Biophysics, [2]Department of Chemistry, University of Minnesota, Minneapolis, Minnesota, USA, [3]BioTechnology Institute, University of Minnesota, St. Paul, Minnesota, USA, [4]SLAC National Accelerator Laboratory, Stanford Synchrotron Radiation Lightsource, Menlo Park, California 94025, USA.

**email: seelig@umn.edu*

Content

Supplementary Table 1. Summary of NMR structural statistics of 20 conformers. The RMSD of the structural ensemble is calculated within well-structured regions (residues 17-35 and 49-69).

	Protein
NMR distance and dihedral constraints	
Distance constraints	
Total NOE	354
Intra-residue	106
Inter-residue	248
Sequential ($\|i-j\| = 1$)	162
Medium-range ($\|i-j\| < 4$)	34
Long-range ($\|i-j\| > 5$)	52
Intermolecular	0
Hydrogen bonds	0
Total dihedral angle restraints	34
ϕ	15
ψ	19
Total RDCs	26
Q (%)	14.6
Structure statistics	
Violations (mean and s.d.)	
Distance constraints (Å)	0.1 (0.01)
Dihedral angle constraints (º)	1.3 (0.4)
Max. distance constraint violation (Å)	0.8 (0.4)
Max. dihedral angle violation (º)	5.7 (1.6)
Deviations from idealized geometry	
Bond lengths (Å)	0.008
Bond angles (º)	1.0
Impropers (º)	0.5
Average pairwise r.m.s. deviation** (Å)	
Heavy	1.4
Backbone	0.8

Supplementary Table 2. Thermodynamic parameters for Zn^{2+} binding determined by Isothermal Titration Calorimetry. After completely removing the Zn^{2+} from the protein with Chelex 100 chelating ion exchange resin (BioRad), 5 µM ligase enzyme was slowly titrated with 400 µM $ZnCl_2$ solution, and the heat release was monitored using a MicroCal VP-ITC instrument (GE Healthcare). All samples contained at 150 mM NaCl, 20 mM HEPES, 10 mM β-mercaptoethanol, and pH 7.5, and the data were fitted with a sequential two-site binding model. The values represent the average of three measurements.

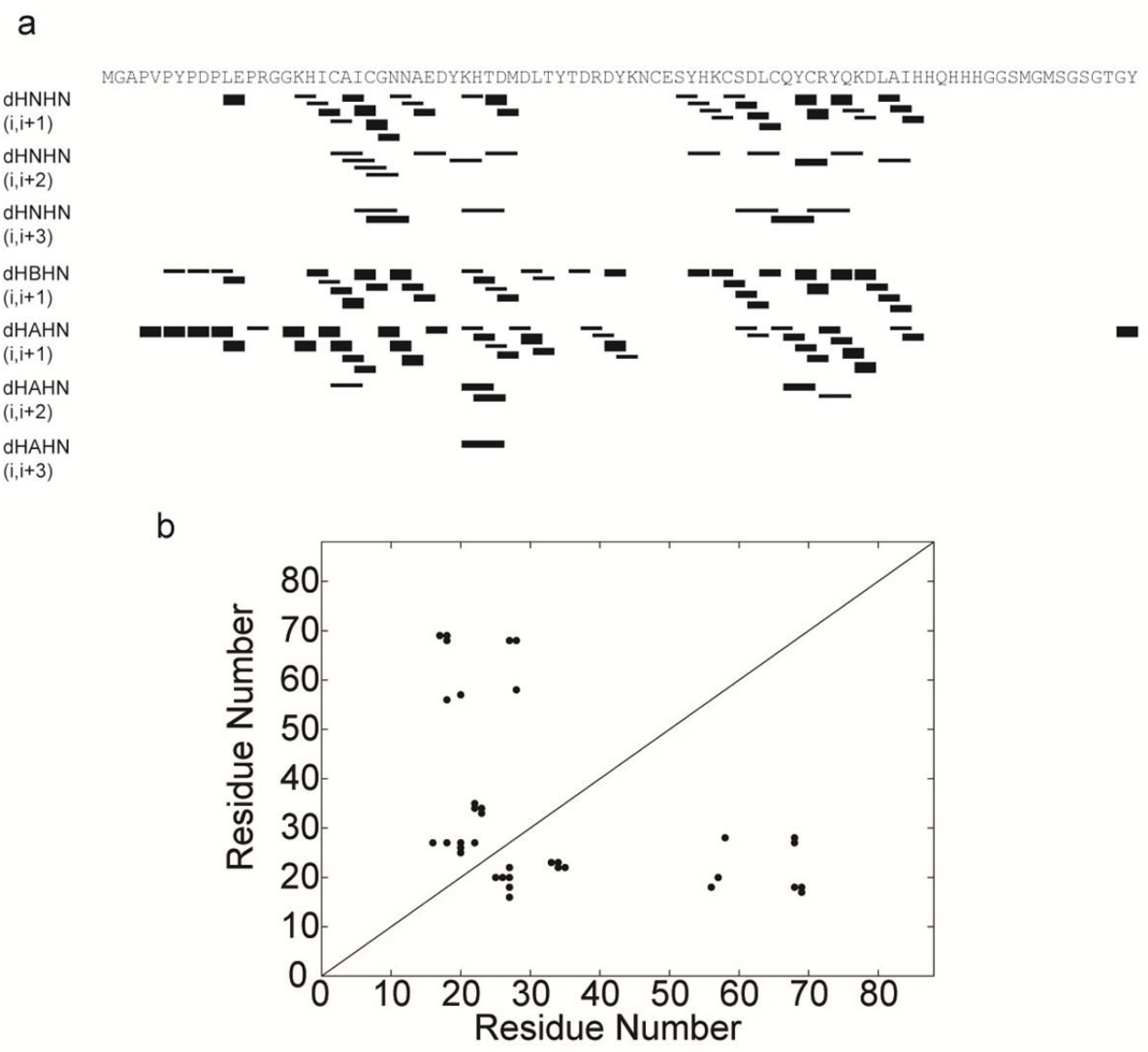
	Average value	Standard deviation
K_d1 (µM)	3.0	0.6
ΔH1 (kcal/mole)	122.9	14.8
ΔS1 (cal/mole/°)	437.7	49.7
K_d2 (µM)	92.8	8.9
ΔH2 (kcal/mole)	-123.7	13.3
ΔS2 (cal/mole/°)	-396.3	45.0

Nature Chemical Biology: doi:10.1038/nchembio.1138

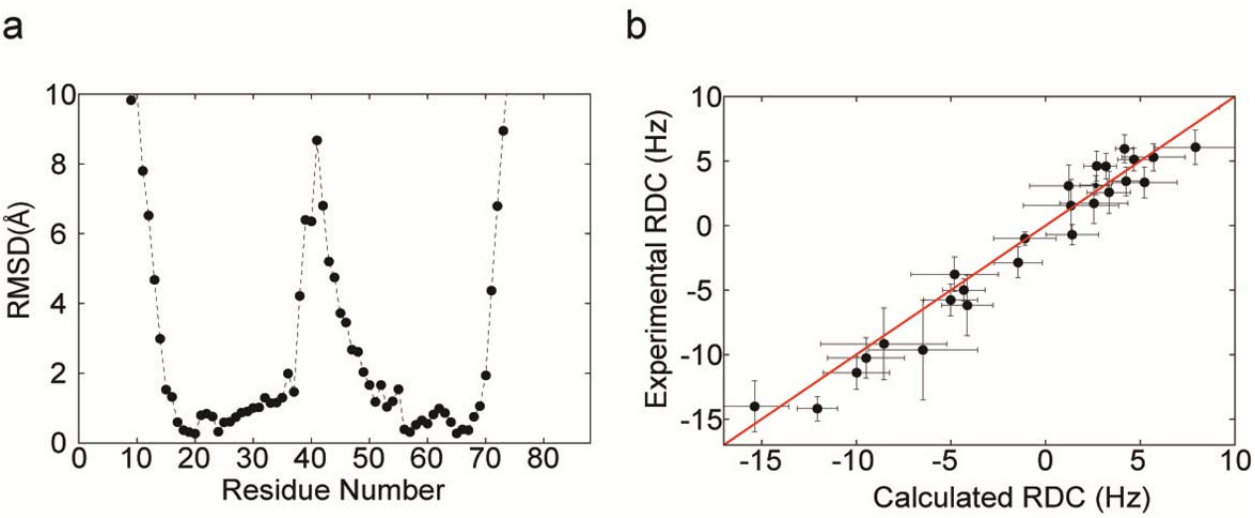
Supplementary Table 3. EXAFS least squares fitting results for ligase 10C.

Coordination/Path	R(Å)[a]	σ^2(Å^2)[b]	E_0 (eV)	*F*[c]
4 Zn-N	2.00 (0.005)	731		
2 Zn-S	2.30 (0.003)	453		
6 Zn-C	3.00 (0.015)	1,077		
12 Zn-C-N	3.09 (0.034)	1,077[d]	-12.3	0.18
6 Zn-C	4.16 (0.012)	169		
6 Zn-C-N	4.19 (0.013)	169[d]		
6 Zn-C-N	4.30 (0.014)	169[d]		

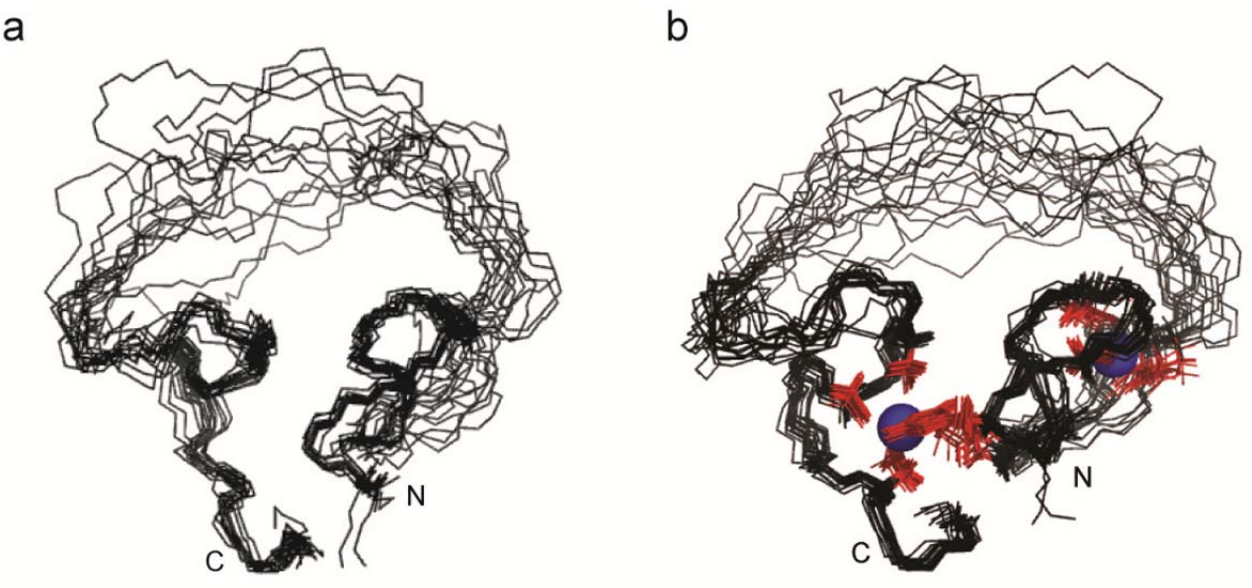
[a]The estimated standard deviations for the distances were calculated by EXAFSPAK and are given in parentheses (see also **Supplementary Figure 8** caption). [b]The σ^2 values are multiplied by 10^5. [c]Error is given by $\Sigma[(\chi_{obsd} - \chi_{calcd})^2 k^6]/\Sigma[(\chi_{obsd})^2 k^6]$. [d]The σ^2 value for the Zn-C (single scattering) and Zn-C-N (multiple scattering) paths were linked to be the same value.



Supplementary Figure 1. Summary of the NOEs observed from NOESY spectra. (**a**) Horizontal bars show the presence of NOE signals between residues. Bar thickness corresponds to the NOE intensity. (**b**) Long-range NOEs observed in ligase 10C.

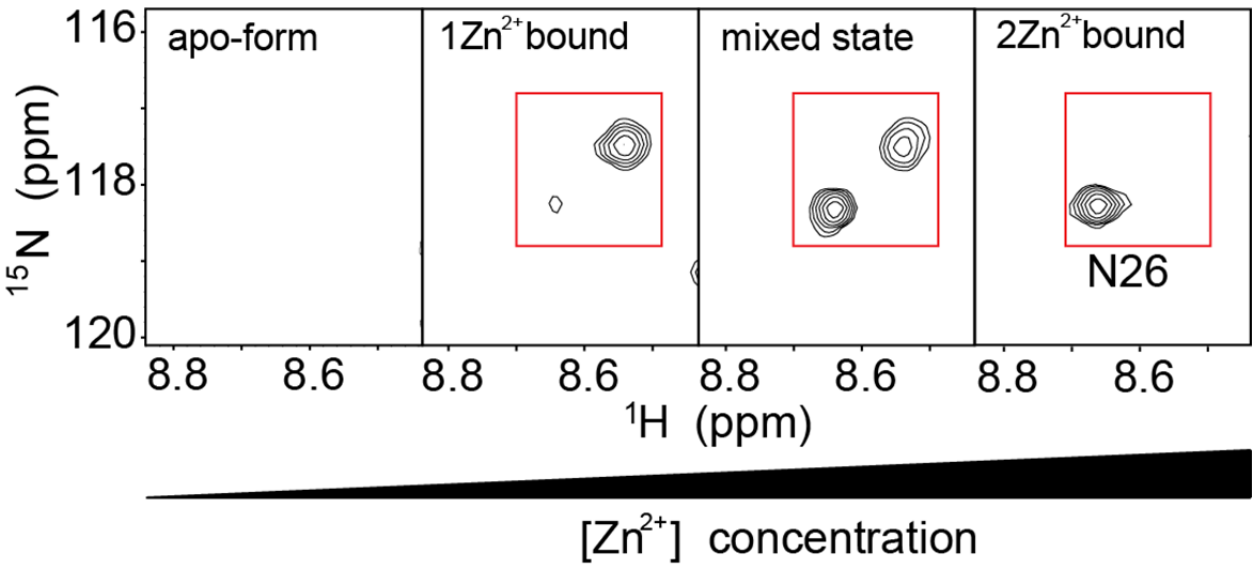


Supplementary Figure 2. Convergence of the structural ensemble of 20 conformers. (**a**) Average backbone RMSD of the conformational ensemble. (**b**) Correlation between experimental RDC values and average back-calculated RDC values from the ensemble.

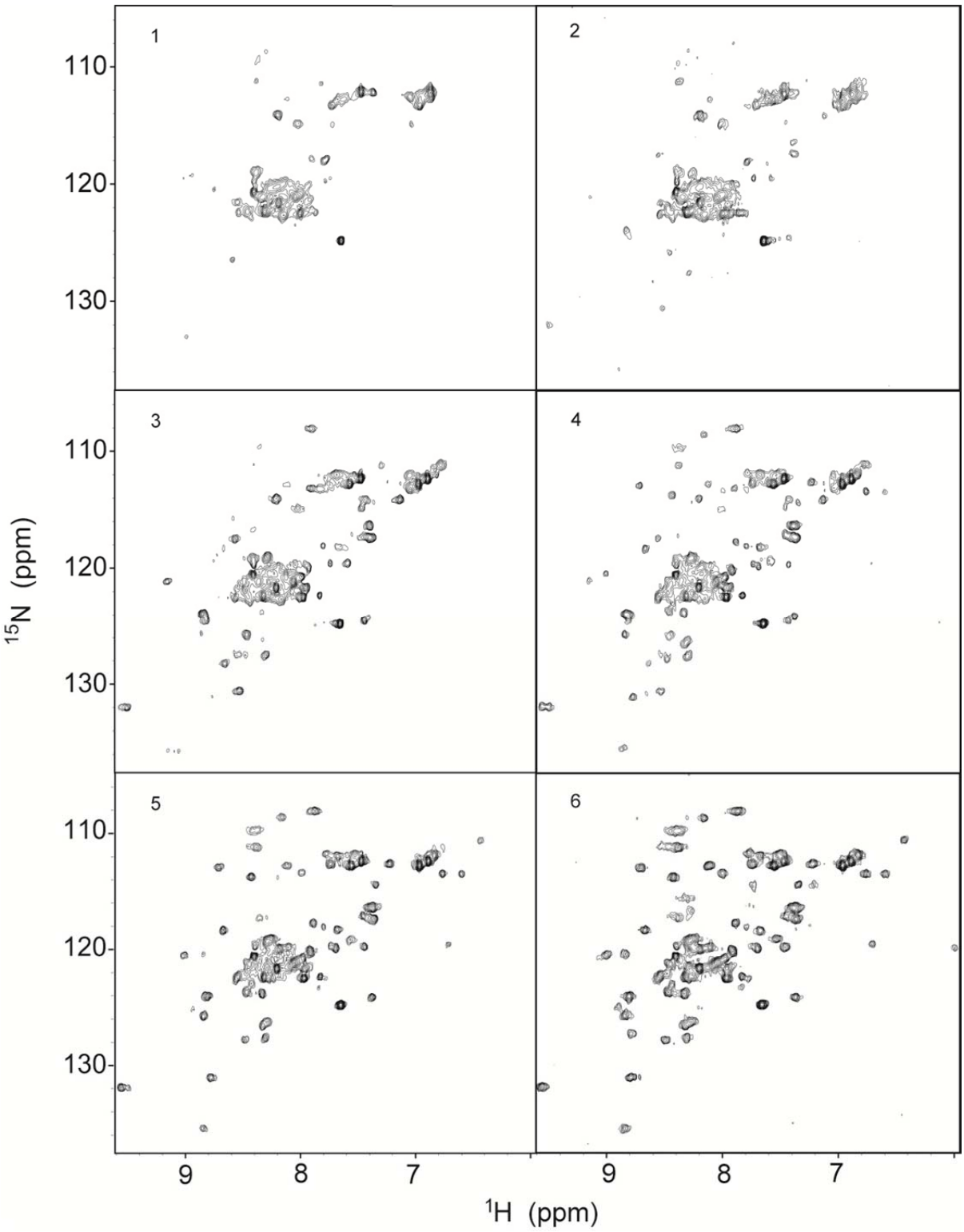


Supplementary Figure 3. The structural ensembles are calculated before and after incorporating Zn^{2+} ions into the coordinates. (**a**) Ensemble of 20 lowest energy conformers (residues 17-69) selected from 100 structures. The NOEs, torsion angles, and RDC values were included in these calculations. Note that the two Zn^{2+} ions and coordination were not included. (**b**) Ensemble of 20 lowest energy conformers (residues 17-69) obtained including Zn^{2+} ions and coordination. The two Zn^{2+} ions are shown as blue spheres. The side chains involved in the coordination are displayed in red (H18, E28(OD2), C57, C60, D65(OD1), and C20, C23, D34(OD1)) and, additionally, both zinc binding sites each contain a single water ligand (not shown) resulting in a hexacoordinated and tetracoordinated sites, respectively. The backbone RMSD between the well-structured regions (residues 17-35 and 49-69) of the ensembles with Zn^{2+} and without Zn^{2+} is 0.52 Å.

Nature Chemical Biology: doi:10.1038/nchembio.1138

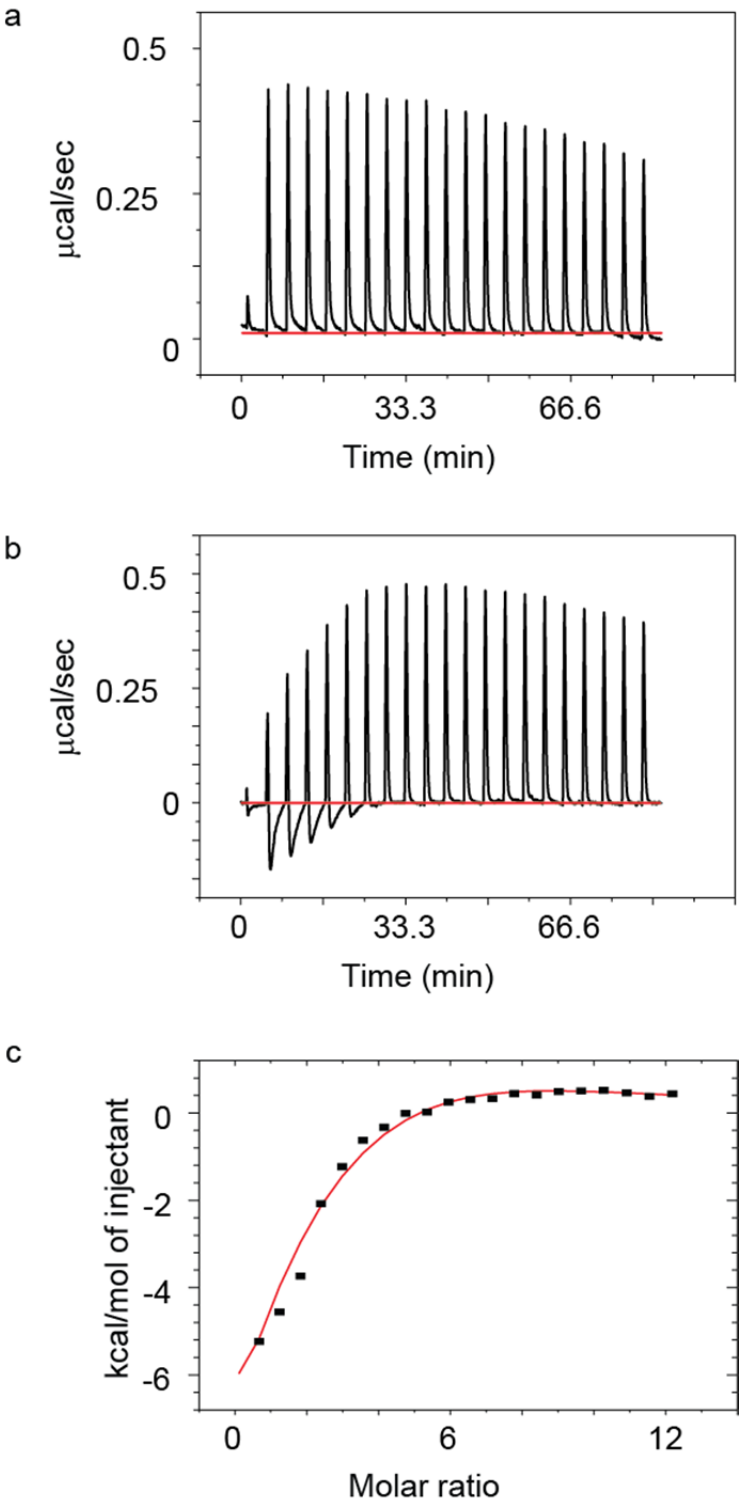


Supplementary Figure 4. Zn^{2+} titration into ^{15}N-labeled ligase 10C monitored by NMR. A selected region of HSQC spectra recorded during Zn^{2+} titration is shown. Residues of partially Zn^{2+}-saturated sample displayed slow exchange on the NMR time scale between forms bound to one or two Zn^{2+}.



Supplementary Figure 5. HSQC spectra recorded upon Zn^{2+} titration. Molar ratios of ligase 10C to zinc were: 1) 10C:Zn=1:0, 2) 10C:Zn=1:1, 3) 10C:Zn=1:2, 4) 10C:Zn=1:3, 5) 10C:Zn=1:4, 6) 10C:Zn=1:6.

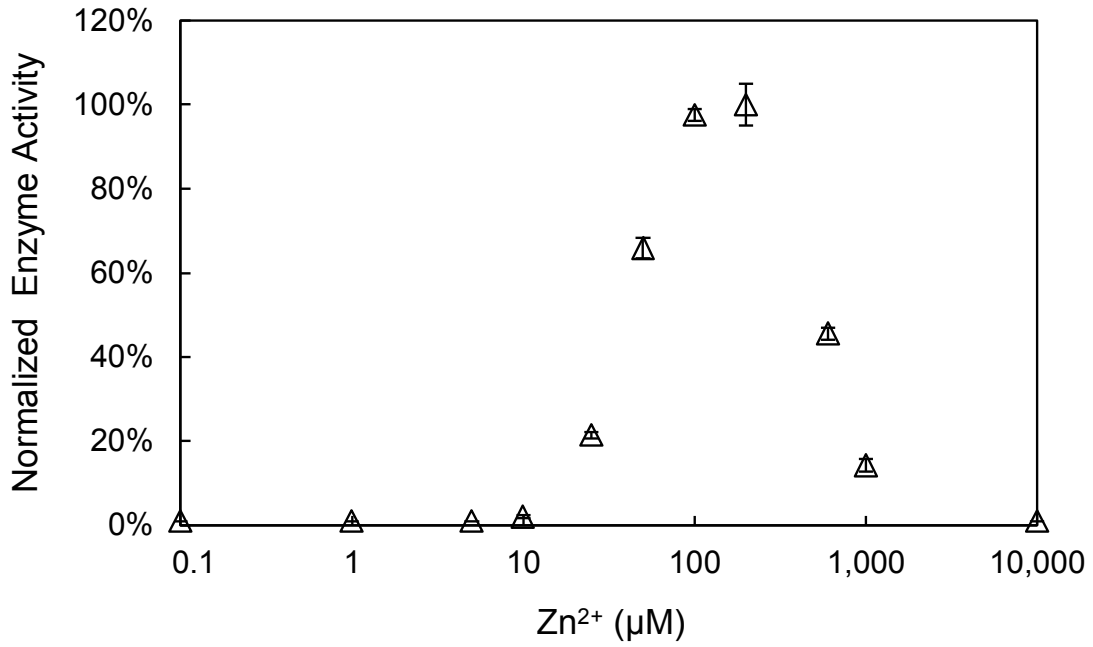
Nature Chemical Biology: doi:10.1038/nchembio.1138



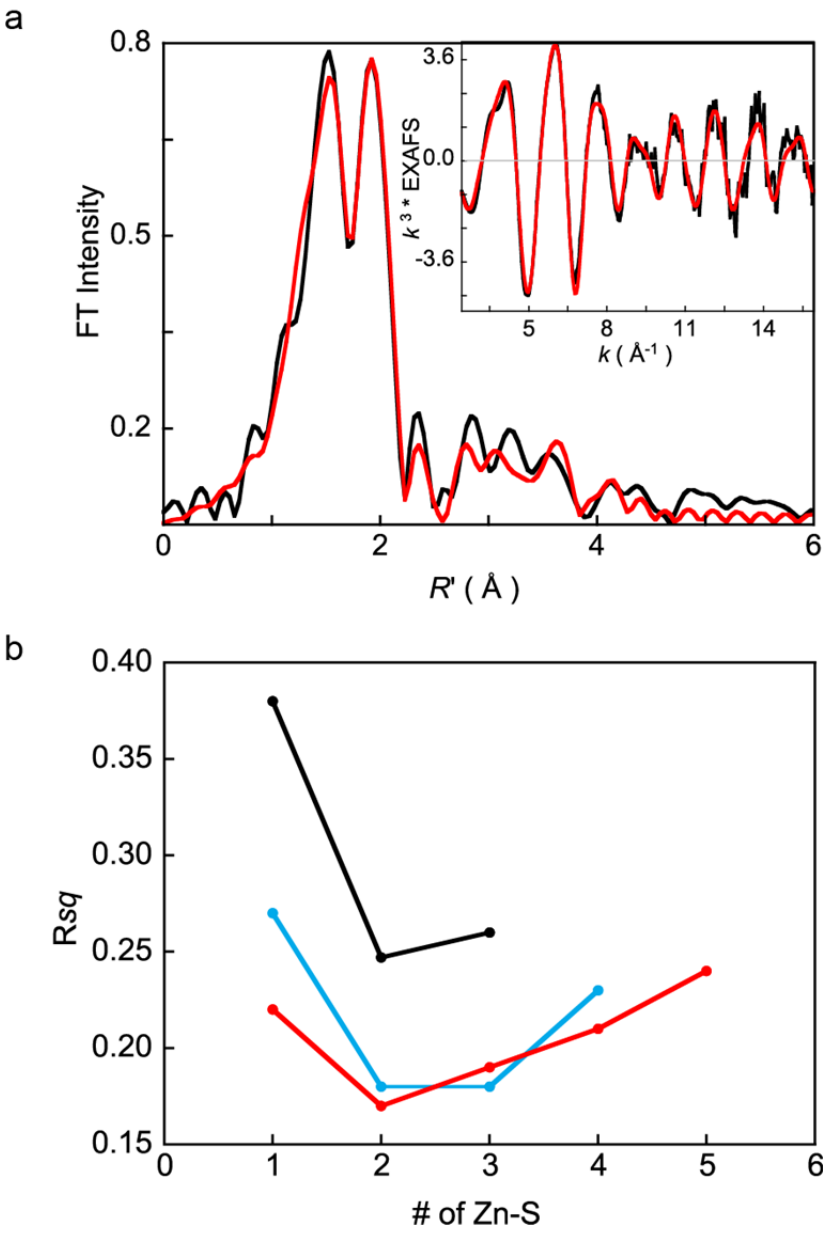
Supplementary Figure 6. Zn^{2+} titration into the ligase enzyme monitored by ITC. Samples contained 5 µM ligase 10C, 150 mM NaCl, 20 mM HEPES, 10 mM β-mercaptoethanol, pH 7.5 and were measured by Isothermal Titration Calorimetry using a MicroCal VP-ITC instrument (GE

Healthcare). (**a**) The graph represents the raw data for the blank titration (buffer without ligase). (**b**) The graph represents raw data for the Zn^{2+} titration of ligase 10C. (**c**) The figure shows the heat release of the Zn^{2+} titration of ligase 10C after subtracting the blank titration. The data is fit to a model of two binding sites. The data can be fitted to models with two or more Zn^{2+} binding sites, however, the fit does not improve significantly with n>2. The Zn^{2+} titration was carried out in triplicate and the errors are summarized in the **Supplementary Table 2**.

Nature Chemical Biology: doi:10.1038/nchembio.1138



Supplementary Figure 7. Zn^{2+} dependence of ligase activity. The maximum activity was observed at 145 μM Zn^{2+}. Towards lower Zn^{2+} concentrations the activity sharply drops, matching the expected behavior predicted from the dissociation constants measured by Thermocalorimetry. Towards higher Zn^{2+} concentrations, the activity also decreases but more slowly. One possible explanation is that Zn^{2+} at high concentrations might also bind to additional sites with lower affinity thereby reducing the activity. Error bars represent one standard deviation. Ligation activity for samples at 0.1, 1, 5 and 10,000 μM $ZnCl_2$ was below the detection limit of 1%.



Supplementary Figure 8. Analysis of zinc coordination by EXAFS spectroscopy (Extended X-ray Absorption Fine Structure). (**a**) The k^3 weighted Zn K-edge EXAFS (inset) and their corresponding non-phase shift corrected Fourier transforms for ligase 10C are presented. The experimental data are shown as black lines and the fit as red lines. The best-fit parameters are given in **Supplementary Table 3**. While a coordination with four ligands is most commonly observed for zinc ions, a coordination geometry including six ligands has been observed in natural proteins numerous times[1,2]. The first shell coordination number was varied from four-coordinate to six-coordinate. In each case the number of Zn-S and Zn-N/O components was systematically varied to obtain the best *F* value. These fits show that the data are most consistent with a six-coordinate site with 2 Zn-S and 4 Zn-N/O components. A 1:1 occupation of the two sites modeled from NMR analysis would have resulted in best-fit with 3 Zn-N/O and 2 Zn-S coordination. However, the process of dialysis (removal of excess Zn is necessary for EXAFS

experiments) may lead to stripping of some Zn from the weakly bound N-terminal site. This leads to an increase in the number of six-coordinate sites over four-coordinate sites in the protein and results in a best-fit first shell with more than 3 Zn-N/O paths.

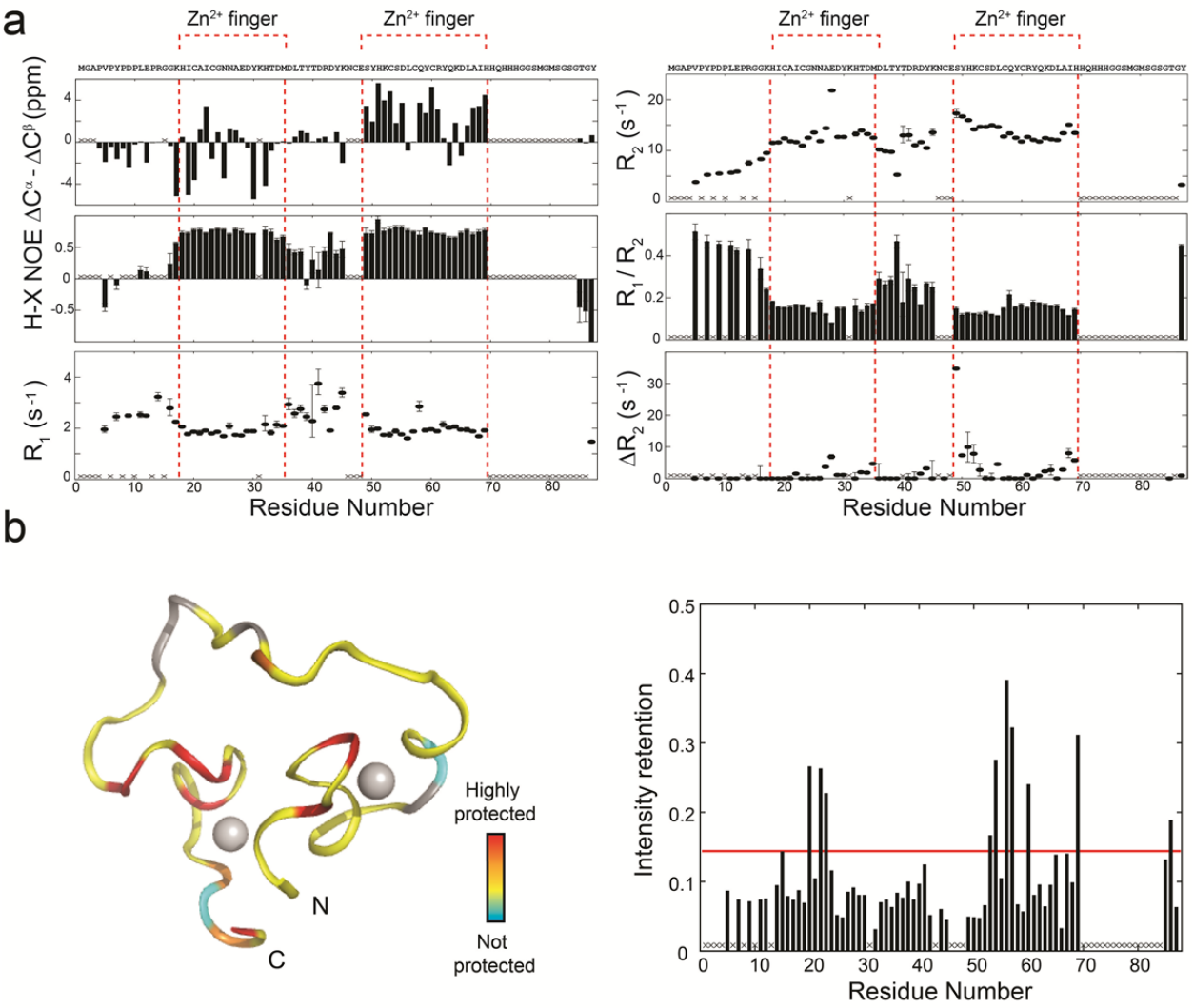
The EXAFS data are dominated by first shell Zn-N/O and Zn-S, while second and third shells are significantly weaker. The second and third shells were fit with single (Zn-C) and multiple-scattering (Zn-C-N) theoretical paths generated using a representative Zn-N(His) model. These weak features are due to a combination of single and multiple scattering from the amino acid ligands. The multiple scattering features are different from characteristic Zn-N(His)n[3] or ZnS(Cys)n[4] systems due to interference between second shell components of Cys and His ligands. Note that standard deviations in bond distances obtained from EXAFSPAK assume the use of raw, low-noise data. Although the data quality presented here are quite high, it is important to note that in the presence of several single and multiple scattering paths, the choice of a specific path to represent an average of multiple paths will also affect the standard deviations. Typically second shell paths have errors of the order 0.05 to 0.1 Å. Furthermore these standard deviations do not reflect the fact EXAFS analysis typically underestimates bond distances (relative to crystallography). The protein samples used for EXAFS analysis were extensively dialyzed and had no extraneous source of sulfur, precluding non-protein based Zn-S ligation.

A visual inspection of the FEFF fit presented here shows that the first peak (corresponding to Zn-N/O paths) in the Fourier Transform is a slightly poorer fit relative to the second peak (corresponding to Zn-S paths). In an attempt to improve the fits and to differentiate between 3 Zn-N/O and 4 Zn-N/O fits, split first shell fits were performed. Significant statistical improvement was not observed.

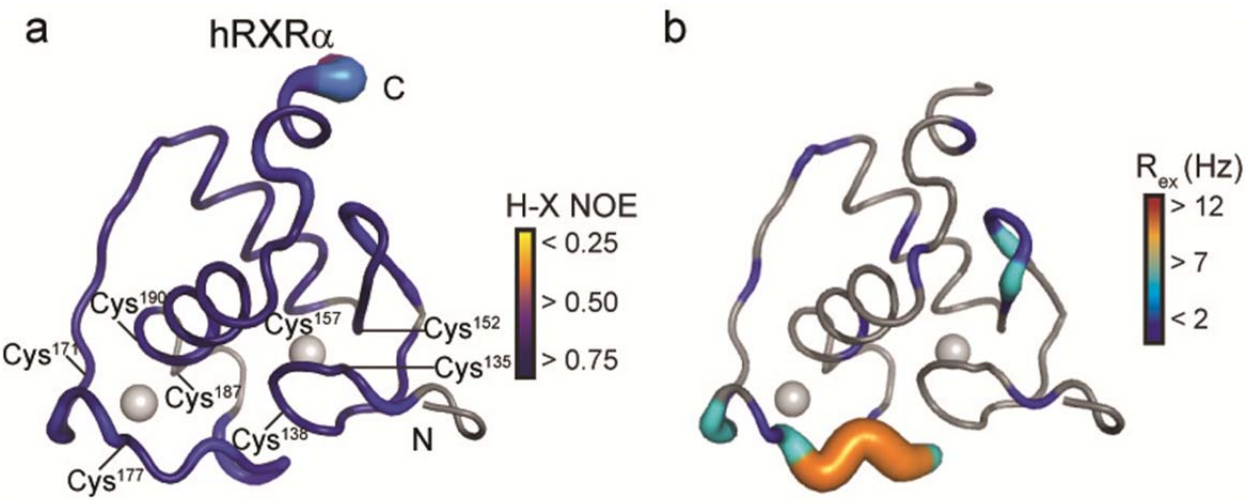
(**b**) The R_{sq} values ($\Sigma[(\chi_{obsd} - \chi_{calcd})^2 k^6]/\Sigma[(\chi_{obsd})^2 k^6]$) for four- to six- coordinate first shell fits are presented as a function of increasing number of Zn-S ligands with concomitant decrease in the number of Zn-N/O ligands. (—) four-coordinate, (—) five-coordinate. (—) six-coordinate. The R_{sq} of the four-coordinate fit is significantly worse than that of the five- or six- coordinate fits. Note that although the best R_{sq} value is obtained with 4 Zn –N/O and 2 Zn-S ligands, the five coordinate fits with either 3 Zn-N/O and 2 Zn-S or 2 Zn-N/O and 3 Zn-S ligands also have reasonably low R_{sq} values. For the 2 Zn-N/O and 3 Zn-S fit to be correct, the two Zn sites need to have 2 Zn-S and 4 Zn-S ligands, respectively. Such a structure is ruled out by NMR data, which show that the sites do not have more than two S-based ligands. Since the first shell coordination number error can be up to 20%, it is difficult to differentiate between the 4 Zn-O/N and 3 Zn/O fits with a high level of statistical confidence. Note that both the 4 Zn-N/O and 3 Zn-N/O fits indicate that the high Zn-affinity site is six-coordinate. Six-coordinate Zn sites account for at least 11% of all Zn sites in biology based on NMR and crystallography studies[2]. EXAFS studies with cysteine ligands are typically limited to four- and five-coordinate sites[5]. Studies have been performed on six-coordinate sites, but typically with all light atom ligands[6]. In general, a comparison of total EXAFS intensity can give an insight into the coordination number but the presence of two different first shell ligands (N/O and S) modulates the EXAFS data strongly, making an accurate comparison of EXAFS data between two systems with different coordination numbers difficult[5]. In such a situation, the error in first shell coordination number determination can be greater than 20%. Since the EXAFS data are best fit with between

3 and 4 light atom ligands, the higher error indicates that the second site can be between tetra- and septa-coordinate. Since, a seven coordinate site has no biological precedence, the second site is between tetra- and hexa-coordinate.

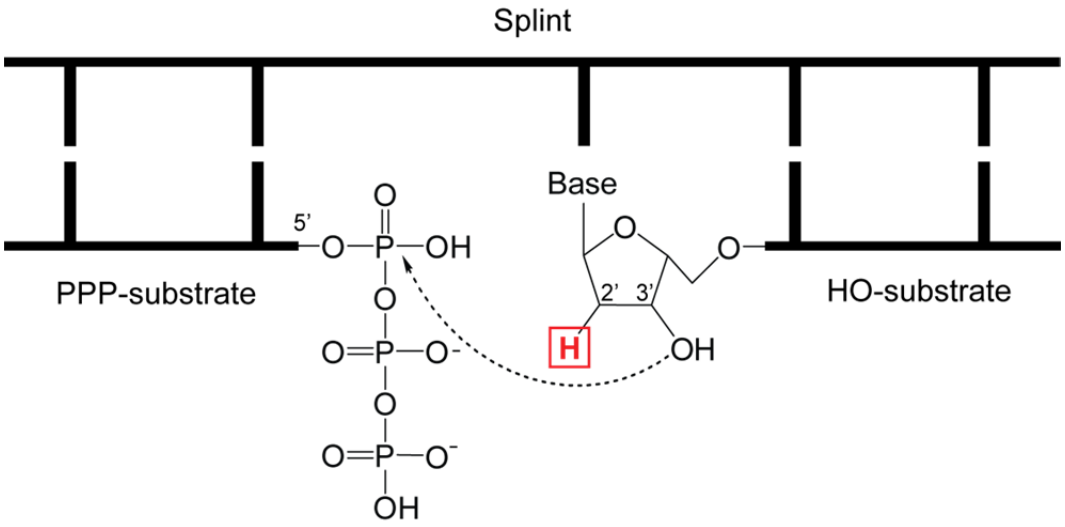
Nature Chemical Biology: doi:10.1038/nchembio.1138



Supplementary Figure 9. Structure and conformational dynamics probed by NMR experiments. (**a**) Chemical shift indexes (ΔCα-ΔCβ), steady-state NOE, longitudinal relaxation rates (R_1), transverse relaxation rates (R_2), and R_1/R_2 ratios as determined by NMR spectroscopy (unassigned residues are marked with an "X"). The errors are estimated by the signal-to-noise (H-X NOE), standard deviations of the fitting (R_1, R_2, and R_1/R_2), or duplicate experiments (ΔR_2). The two zinc fingers are highlighted with dashed red lines. (**b**) The decrease of peak intensities due to H/D exchange was mapped onto one NMR conformer (residues 17-69 are displayed). The intensity of the peaks was normalized to a reference HSQC spectrum of the ligase 10C in 10% D_2O. The sample was lyophilized and dissolved in the same volume of 80% D_2O. After 6 minutes, the HSQC spectrum was acquired and compared with the initial spectrum to monitor solvent exposed amide groups. The solid red line in the diagram represents the average intensity retention plus 2σ.

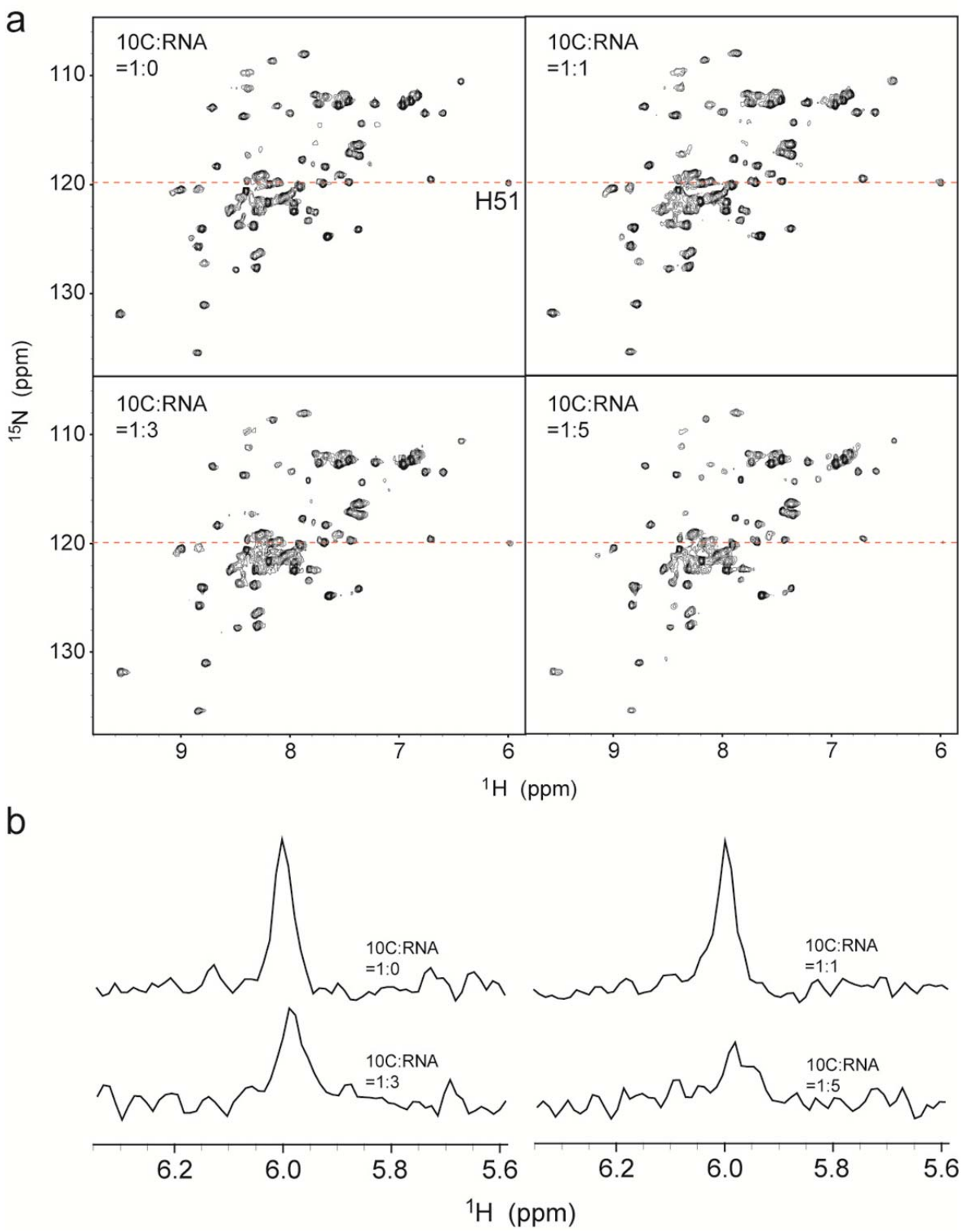


Supplementary Figure 10. Mapping of the conformational dynamics of the DNA binding domain (hRXRα)[7]. (**a**) Heteronuclear NOEs (proxy for fast dynamics on a picosecond-nanosecond time scale) on the structure of hRXRα[7]. (**b**) Exchange rates (R_{ex}) obtained from relaxation dispersion measurements as a proxy for slow dynamics (microsecond-millisecond time scale). The color gradient and the thickness of the backbone indicate the intensity of the motions.



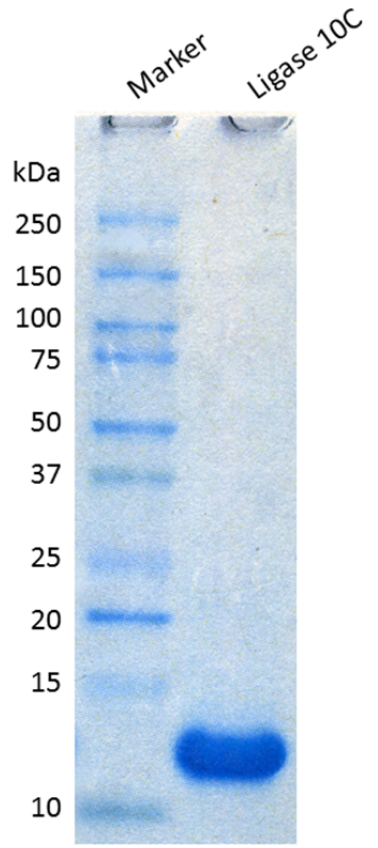
Supplementary Figure 11. Chemical structure of inactive ligation substrate. Substitution of the 2'-hydroxyl group of the terminal nucleotide in the HO-substrate with a 2'-deoxy modification (red box) results in inactivation of the ligation reaction. Ligation of active substrates occurs between a 5'-triphosphorylated RNA (PPP-substrate) and the 3'-hydroxyl group of the second RNA (HO-substrate) while both RNAs are base-paired to a complementary oligonucleotide (splint). The dashed arrow symbolizes the proposed bond formation.

Nature Chemical Biology: doi:10.1038/nchembio.1138


a
10C:RNA
=1:0
10C:RNA
=1:1
10C:RNA
=1:3
10C:RNA
=1:5
H51
110
120
130
^{15}N (ppm)
9
8
7
6
^{1}H (ppm)
b
10C:RNA
=1:0
10C:RNA
=1:1
10C:RNA
=1:3
10C:RNA
=1:5
6.2
6.0
5.8
5.6
^{1}H (ppm)

Supplementary Figure 12. Titration of RNA substrate into ligase 10C monitored by NMR spectroscopy. (**a**) The ligase enzyme (300 μM) was titrated with the inactive RNA ligand in 150 mM NaCl, 20 mM HEPES, 10 mM β-mercaptoethanol, and pH 7.5. The HSQC spectra during the titration showed no significant changes in chemical shifts. (**b**) Slices of a selected peak (H51) in HSQC spectra during ligand titration showed significant line-broadening.

Nature Chemical Biology: doi:10.1038/nchembio.1138



Supplementary Figure 13. Purity and identity of purified ligase 10C. SDS-PAGE gel (NuPAGE 4-12% Bis-Tris gel, Invitrogen) Coomassie stained of ligase 10C purified by nickel affinity chromatography and size exclusion chromatography and 10-250 kDa ladder P7703S (New England Biolabs) used as a marker. The identity of ligase 10C was confirmed by MALDI mass spectrometry yielding a characteristic $[M+H]^+$ signal at m/z = 9,648 $\pm$ 1.4 ($\pm$ s.d. from five independent measurements), which is consistent with the expected mass of the ligase 10C without the N-terminal methionine (MW = 9,648.7 Da). The purity of labeled constructs and all mutants matched that of the purified ligase 10C shown here.

Supplementary References

1. Alberts, I.L., Nadassy, K. & Wodak, S.J. *Protein Sci.* **7**, 1700-1716 (1998).
2. Patel, K., Kumar, A. & Durani, S. *BBA-Proteins Proteom.* **1774**, 1247-1253 (2007).
3. Kupper, H., Mijovilovich, A., Meyer-Klaucke, W. & Kroneck, P.M.H. *Plant Physiol.* **134**, 748-757 (2004).
4. Clark-Baldwin, K. et al. *J. Am. Chem. Soc.* **120**, 8401-8409 (1998).
5. Penner-Hahn, J.E. *Coord. Chem. Rev.* **249**, 161-177 (2005).
6. Bobyr, E. et al. *J. Mol. Biol.* **415**, 102-117 (2012).
7. Holmbeck, S.M.A. et al. *J. Mol. Biol.* **281**, 271-284 (1998).

Nature Chemical Biology: doi:10.1038/nchembio.1138