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

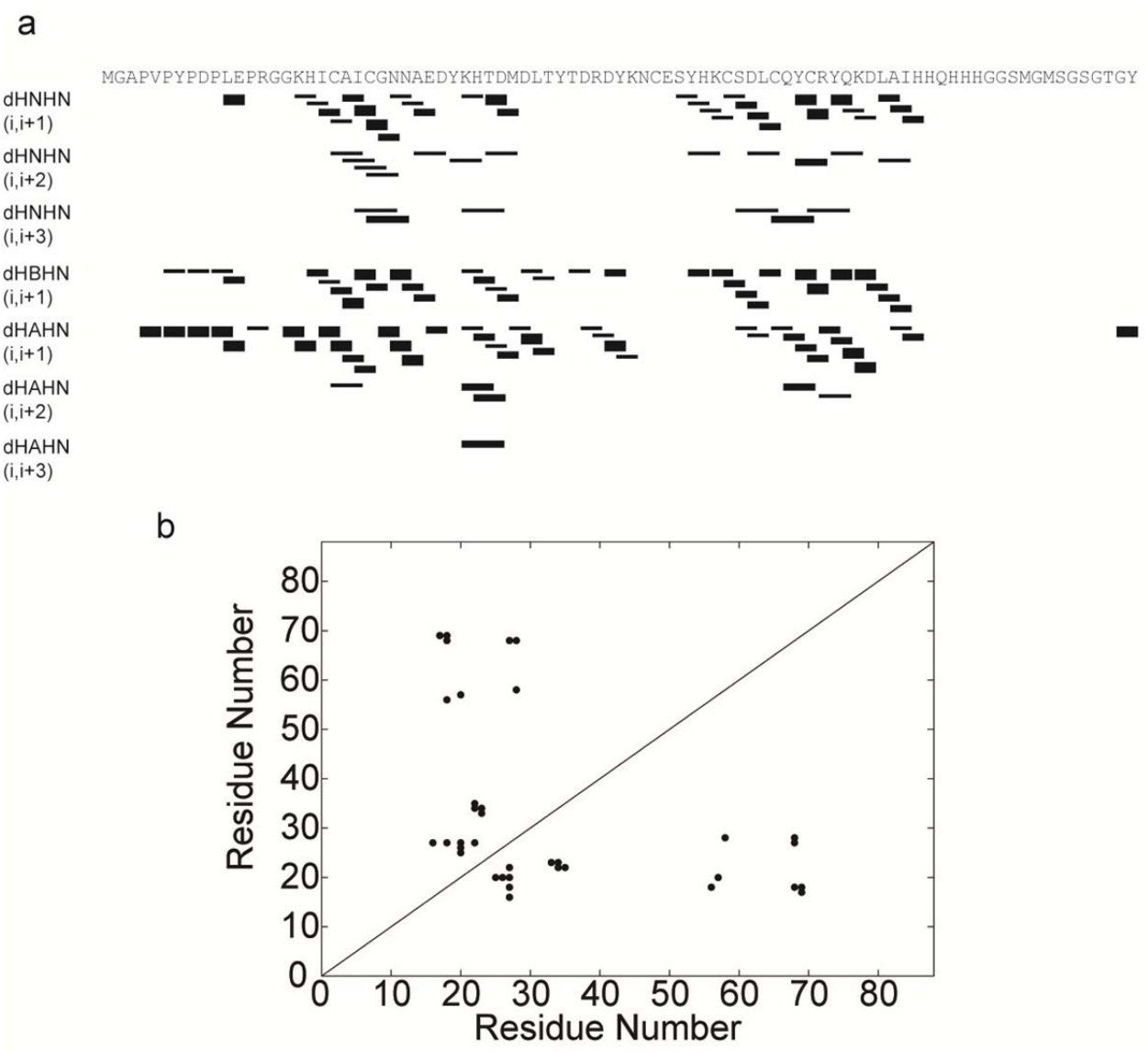
Supplementary Table 2. Thermodynamic parameters for Zn^{2+} binding determined by Isothermal Titration Calorimetry. After completely removing the Zn^{2+} from the protein with Chelex 100 chelating ion exchange resin (BioRad), 5 µM ligase enzyme was slowly titrated with 400 µM $ZnCl_2$ solution, and the 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

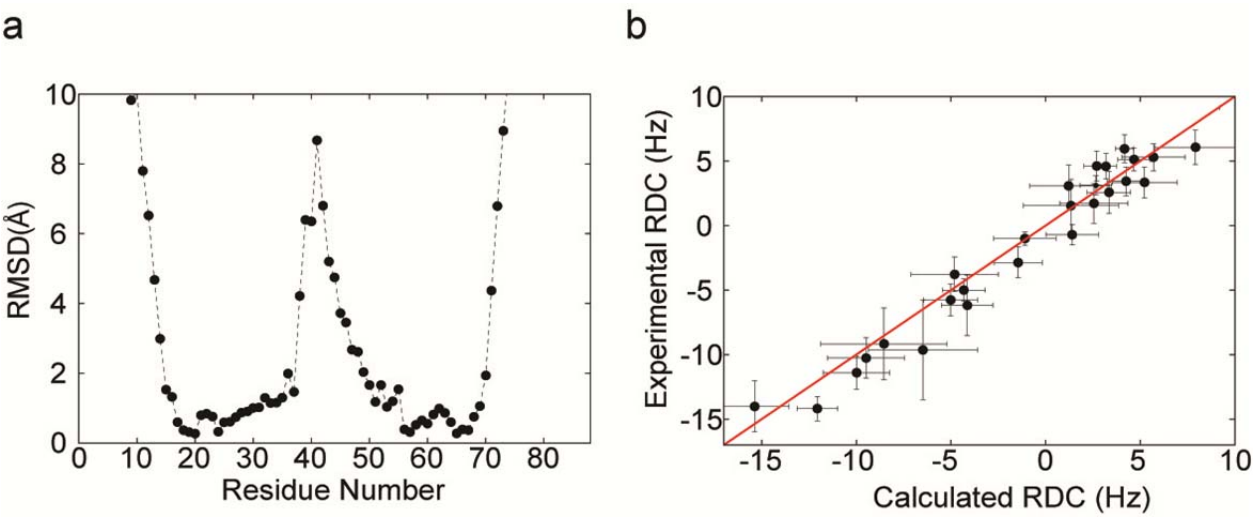
Supplementary Table 3. EXAFS least squares fitting results for ligase 10C.

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]		

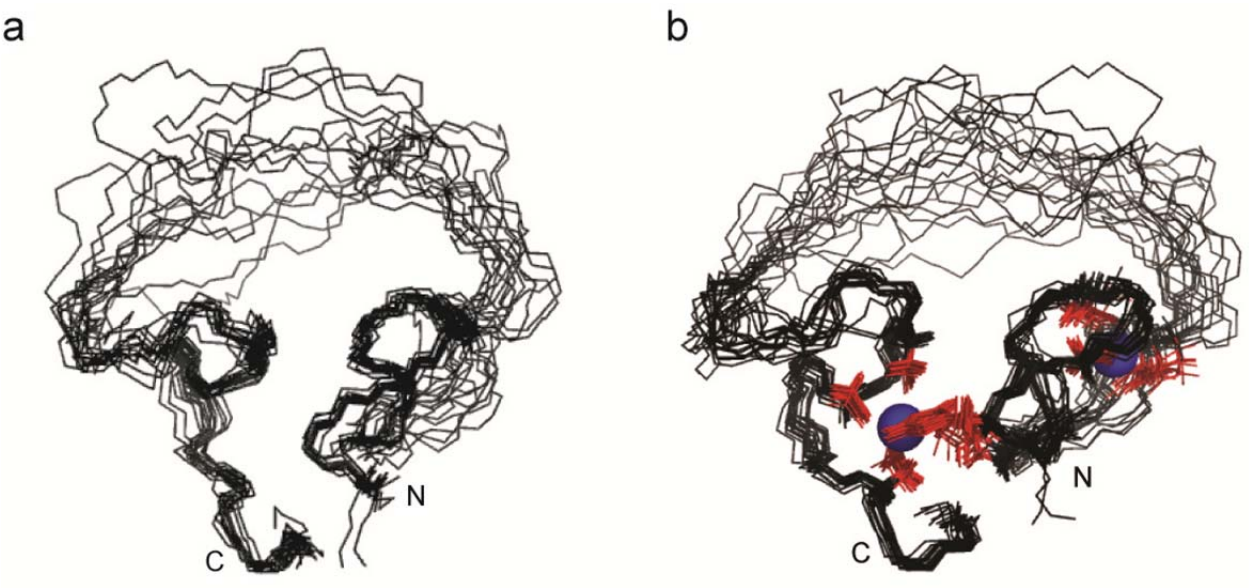
[a]The estimated standard deviations for the distances were calculated by EXAFSPAK and are given in parentheses (see also **Supplementary Figure 8** caption). [b]The σ^2 values are multiplied by 10^5. [c]Error is given by $\Sigma[(\chi_{obsd} - \chi_{calcd})^2 k^6]/\Sigma[(\chi_{obsd})^2 k^6]$. [d]The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



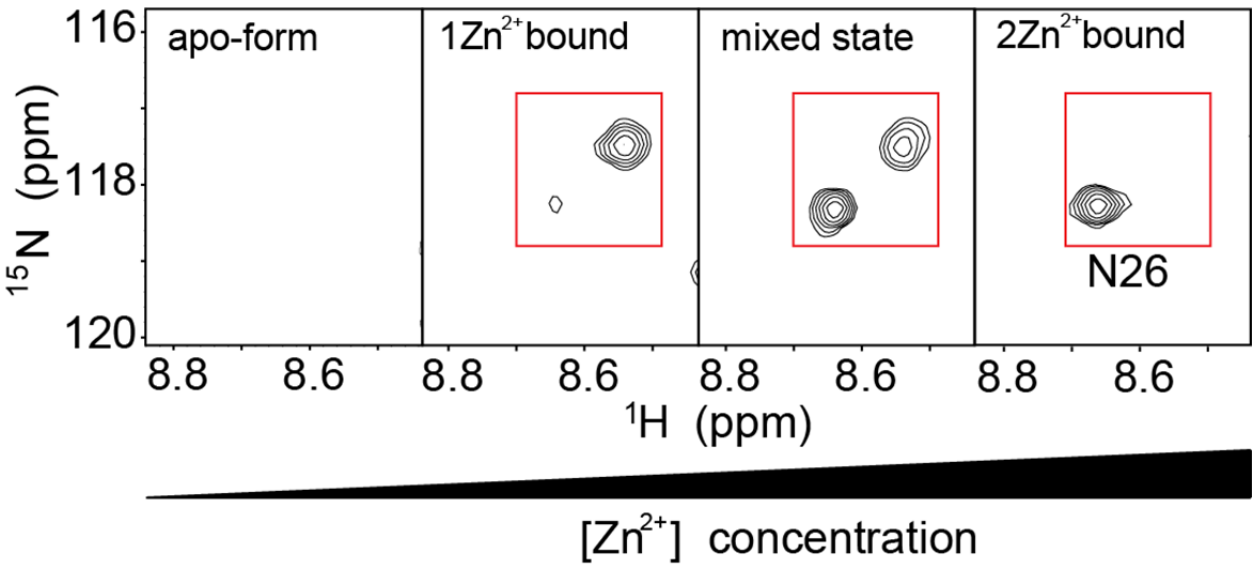
Supplementary Figure 1. Summary of the NOEs observed from NOESY spectra. (**a**) Horizontal bars show the presence of NOE signals between residues. Bar thickness corresponds to the NOE intensity. (**b**) Long-range NOEs observed in ligase 10C.



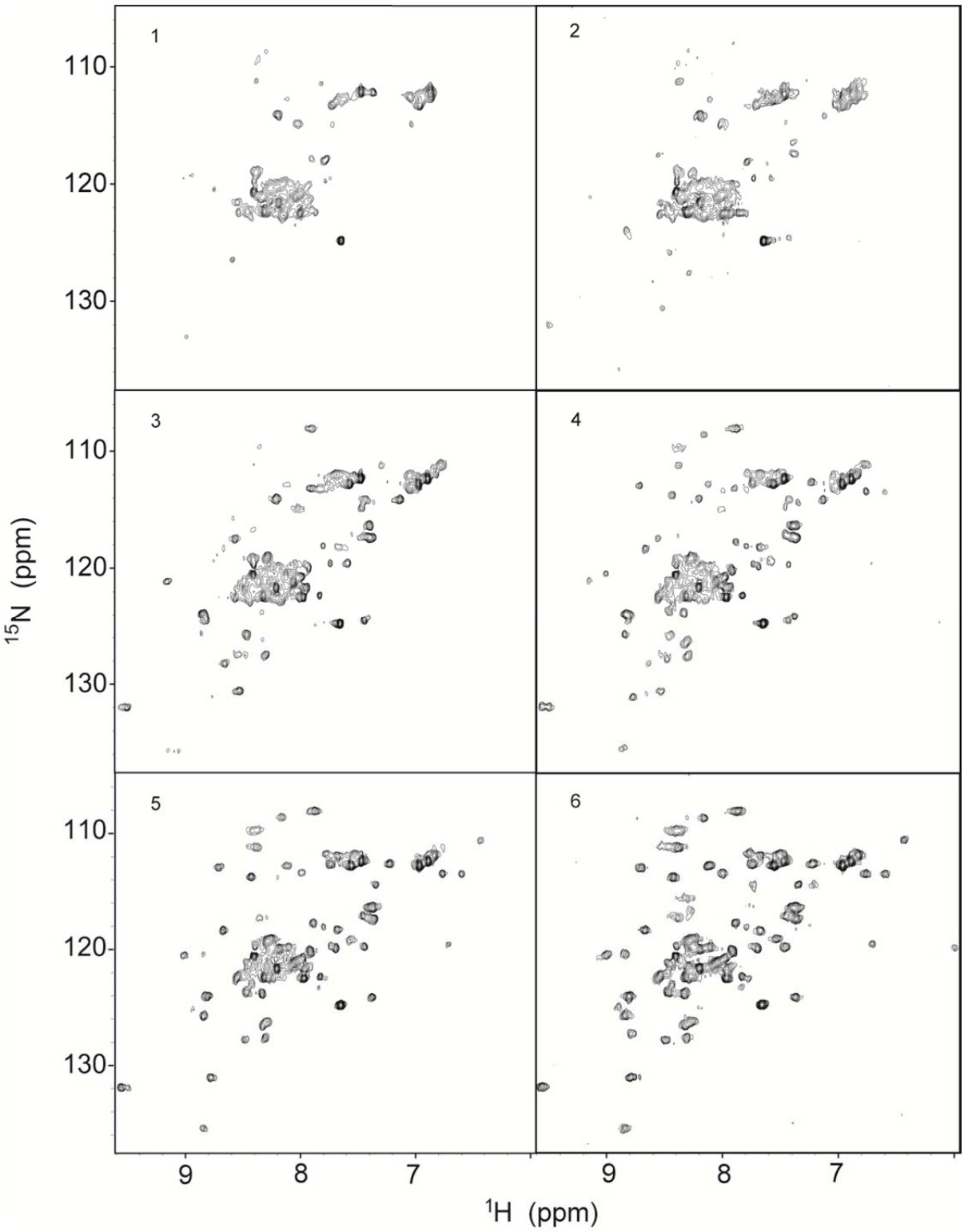
Supplementary Figure 2. Convergence of the structural ensemble of 20 conformers. (**a**) Average backbone RMSD of the conformational ensemble. (**b**) Correlation between experimental RDC values and average back-calculated RDC values from the ensemble.



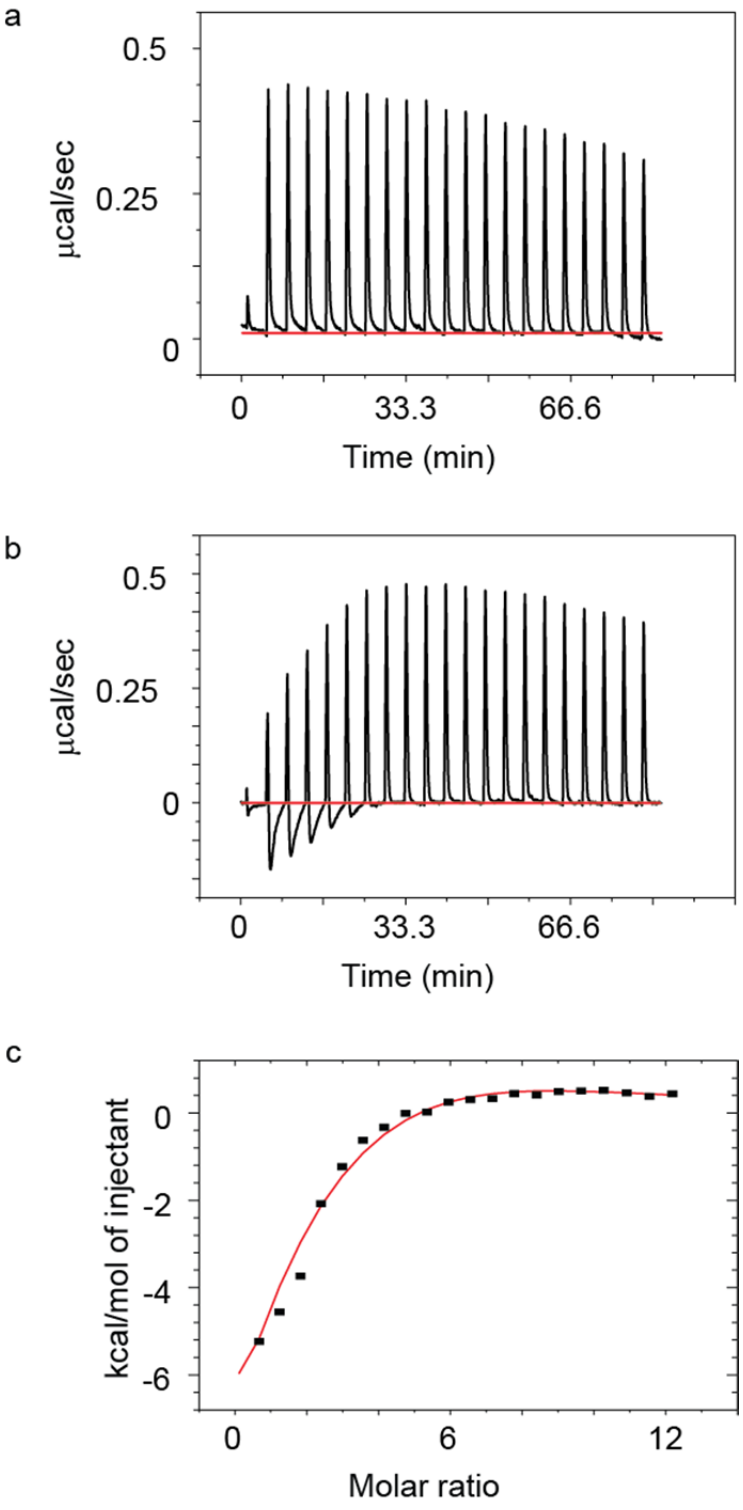
Supplementary Figure 3. The structural ensembles are calculated before and after incorporating Zn^{2+} ions into the coordinates. (**a**) Ensemble of 20 lowest energy conformers (residues 17-69) selected from 100 structures. The NOEs, torsion angles, and RDC values were included in these calculations. 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.



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+}.

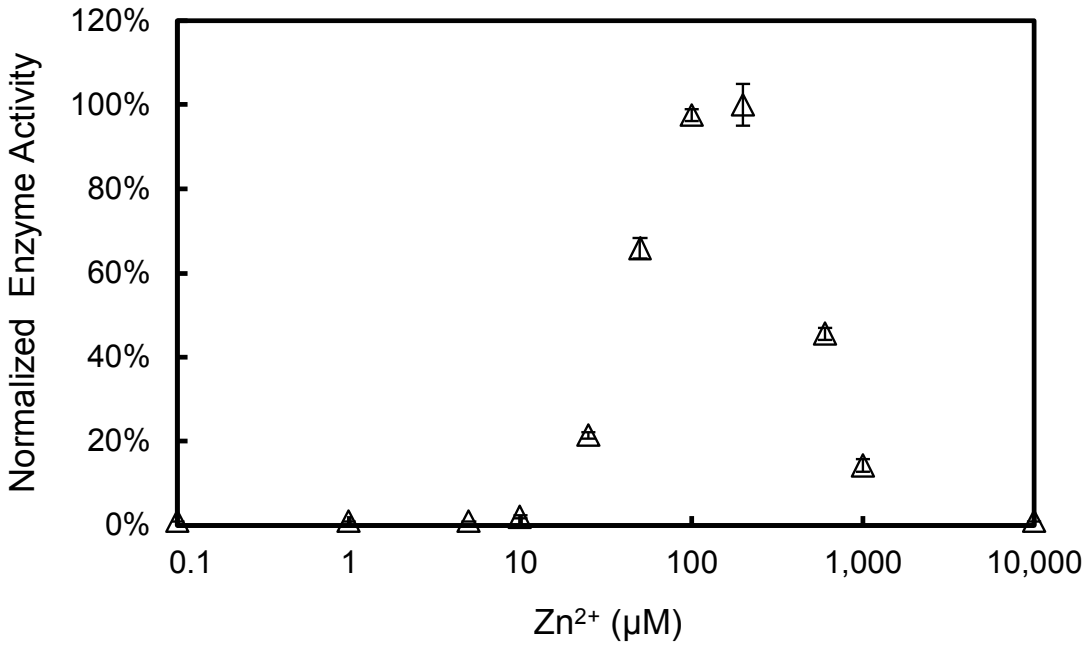


Supplementary Figure 5. HSQC spectra recorded upon Zn^{2+} titration. Molar ratios of ligase 10C to zinc were: 1) 10C:Zn=1:0, 2) 10C:Zn=1:1, 3) 10C:Zn=1:2, 4) 10C:Zn=1:3, 5) 10C:Zn=1:4, 6) 10C:Zn=1:6.

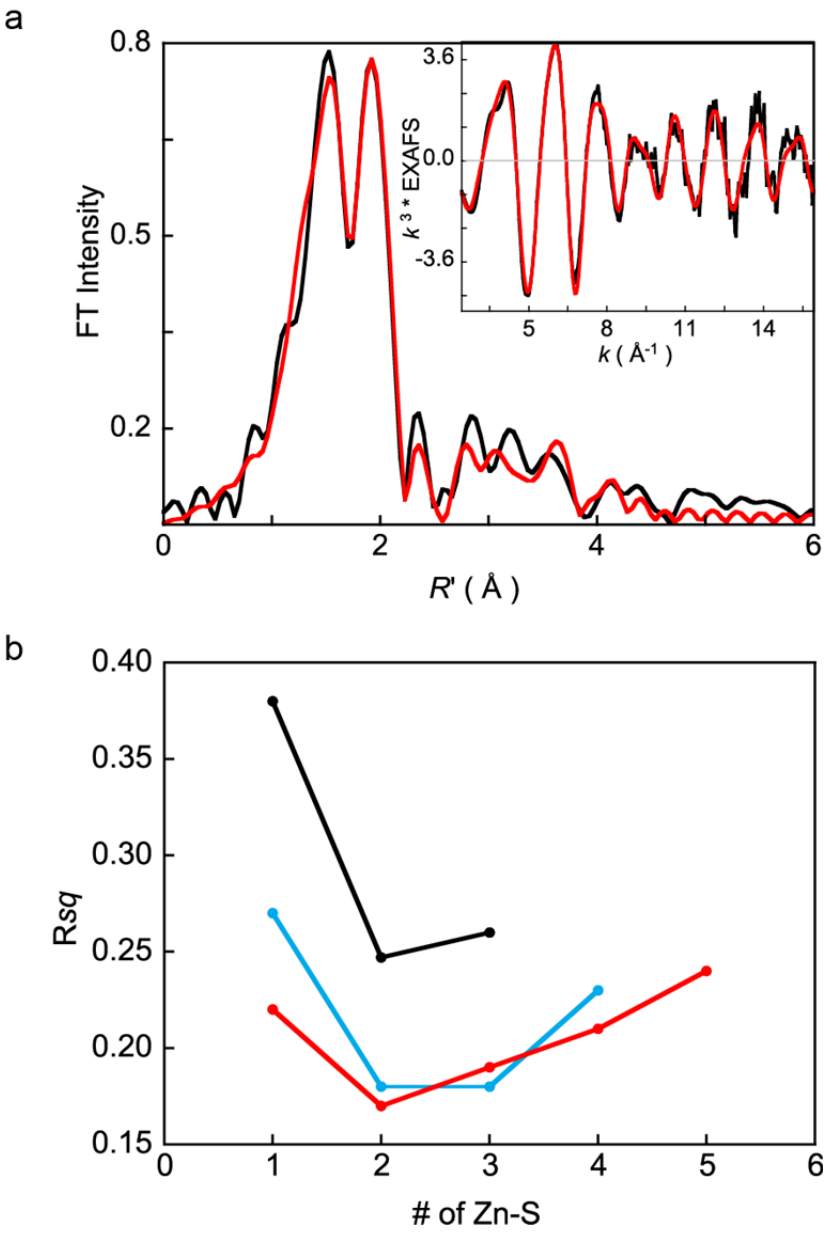


Supplementary Figure 6. Zn^{2+} titration into the ligase enzyme monitored by ITC. Samples contained 5 µM ligase 10C, 150 mM NaCl, 20 mM HEPES, 10 mM β-mercaptoethanol, pH 7.5 and were measured by Isothermal Titration Calorimetry using a MicroCal VP-ITC instrument (GE

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**.



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%.



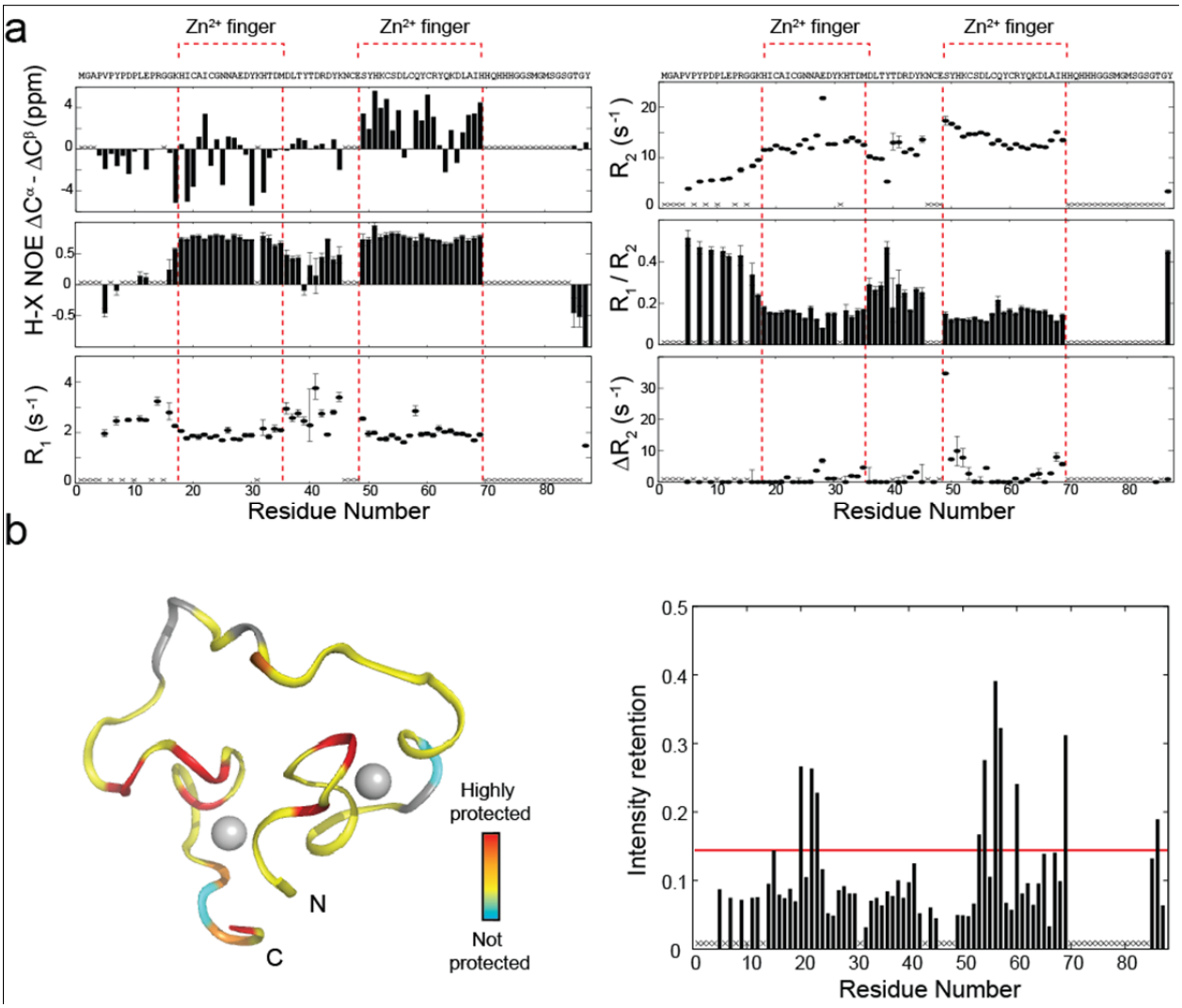
Supplementary Figure 8. Analysis of zinc coordination by EXAFS spectroscopy (Extended X-ray Absorption Fine Structure). (**a**) The k^3 weighted Zn K-edge EXAFS (inset) and their corresponding non-phase shift corrected Fourier transforms for ligase 10C are presented. The experimental data are shown as 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.

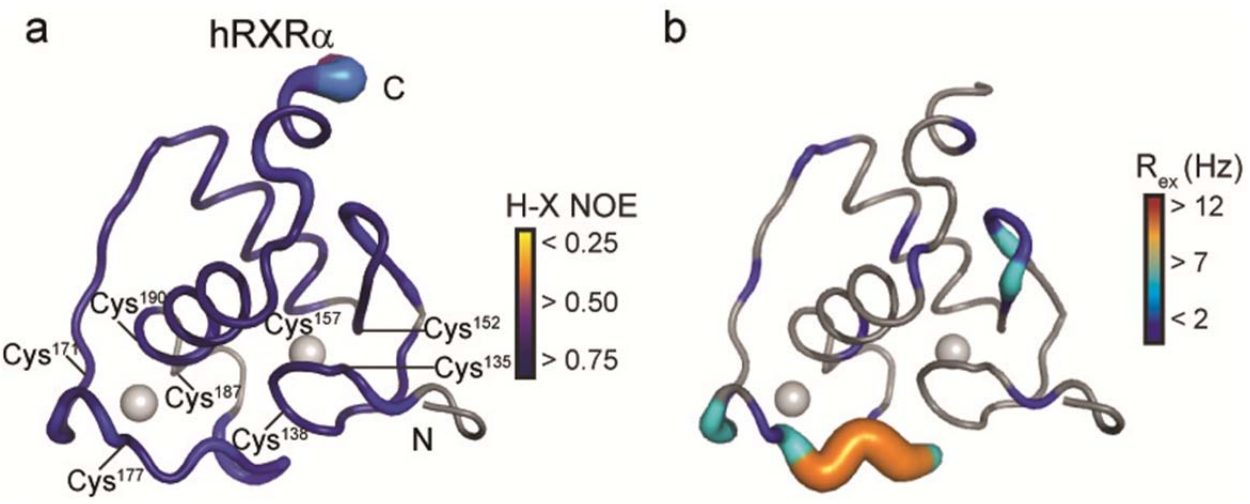
The EXAFS data are dominated by first shell Zn-N/O and Zn-S, while second and third shells are significantly weaker. The second and third shells were fit with single (Zn-C) and multiple-scattering (Zn-C-N) theoretical paths generated using a representative Zn-N(His) model. These weak features are due 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.

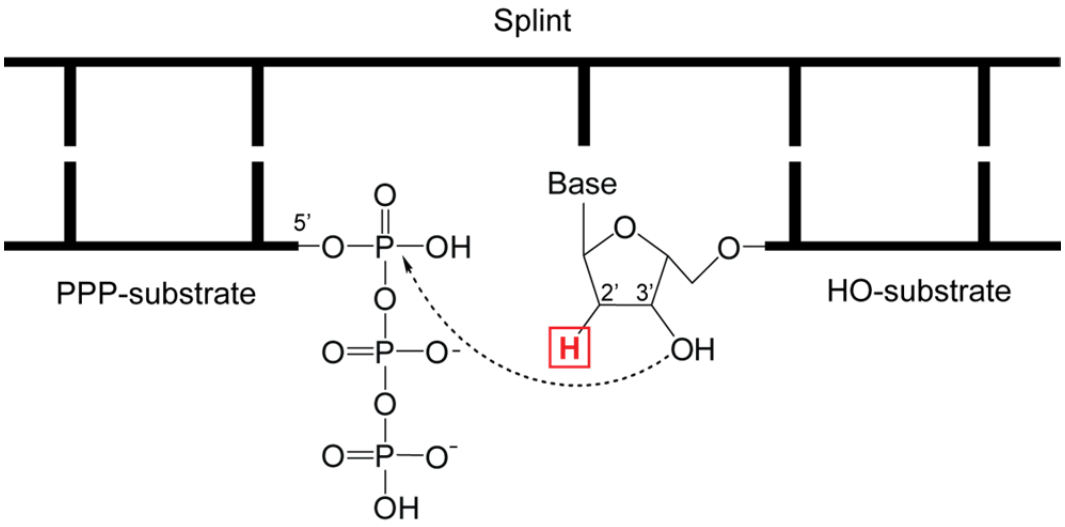
(**b**) The R_{sq} values ($\Sigma[(\chi_{obsd} - \chi_{calcd})^2 k^6]/\Sigma[(\chi_{obsd})^2 k^6]$) for four- to six- coordinate first shell fits are presented as a function of increasing number of Zn-S ligands with concomitant decrease in the number of Zn-N/O ligands. (—) four-coordinate, (—) five-coordinate. (—) 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 a meaningful comparison of EXAFS data between two systems with different coordination numbers difficult[5].



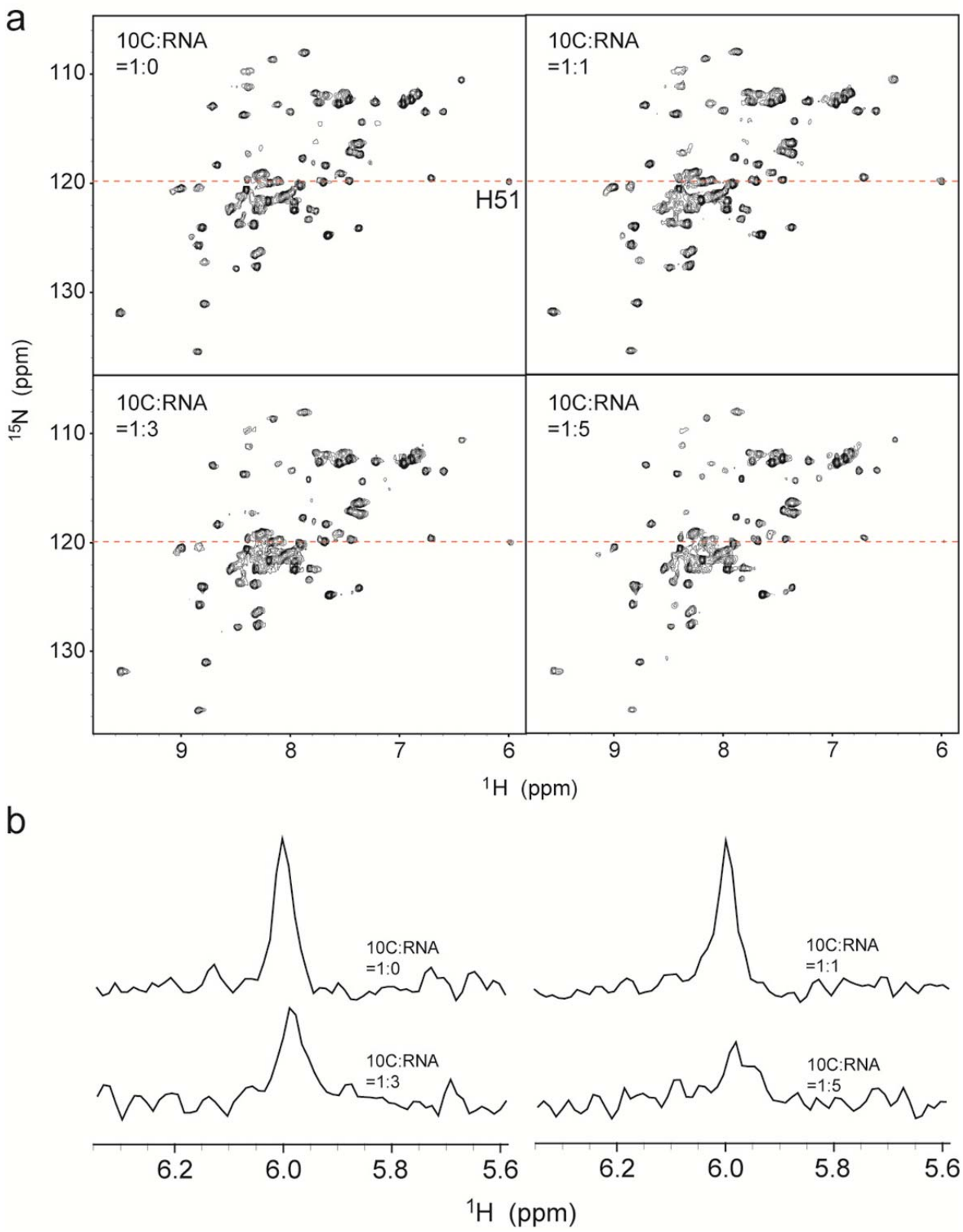
Supplementary Figure 9. Structure and conformational dynamics probed by NMR experiments. (**a**) Chemical shift indexes (ΔCα-ΔCβ), steady-state NOE, longitudinal relaxation rates (R_1), transverse relaxation rates (R_2), and R_1/R_2 ratios as determined by NMR spectroscopy (unassigned residues are marked 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.



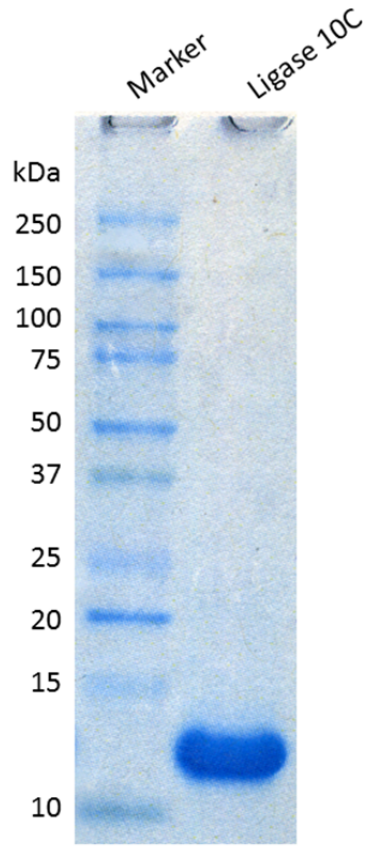
Supplementary Figure 10. Mapping of the conformational dynamics of the DNA binding domain (hRXRα)[7]. (**a**) Heteronuclear NOEs (proxy for fast dynamics on a picosecond-nanosecond time scale) on the structure of hRXRα[7]. (**b**) Exchange rates (R_{ex}) obtained from relaxation dispersion measurements 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.


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.



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).