Supporting Online Material for

A Cytidine Deaminase Edits C to U in Transfer RNAs in Archaea

Lennart Randau,* Bradford J. Stanley, Andrew Kohlway, Sarah Mechta, Yong Xiong,* Dieter Söll*

*To whom correspondence should be addressed. E-mail: lennart.randau@yale.edu (L.R.); dieter.soll@yale.edu (D.S.); yong.xiong@yale.edu (Y.X.)

Published 1 May 2009, *Science* **324**, 657 (2009)
DOI: 10.1126/science.1170123

This PDF file includes:

Materials and Methods

General. Oligonucleotide synthesis and DNA sequencing was performed by the Keck Foundation Biotechnology Resource Laboratory at Yale University. *M. kandleri* cells were obtained from K. O. Stetter and M. Thomm (University of Regensburg, Regensburg, Germany).

Sequence determination of *M.kandleri* RNA. 4 g of a *M. kandleri* cell pellet was disrupted by using a mortar and pestle and 2% SDS, followed by the isolation and purification of total RNA as described (*1*). The tRNA was further purified by anion-exchange chromatography on a MonoQ HR 5/5 with a linear 40-ml gradient of 0–1 M NaCl in 20 mM MOPS pH 6.2 to eliminate residual genomic DNA contamination. Circularization of total tRNA from*M. kandleri* was performed with Thermophage ssDNA ligase (Prokaria) at 65 °C according to the manufacturer's directions. The following primers were used for the amplification of $tRNA^{Asp}$:

forward 5'-TGTCGAGCCCGGGACCCGGGTTCAAATCCC-3',
reverse 5'-CGACAGGCCCGGATGATAGACCGGGCTACAC-3'; $tRNA^{Cys}$:
forward 5'-TGCAGATCCCGTTACCCGGGTTCAAATCCC-3',
reverse 5'-CTGCAGTCCCGCGCCCAGCCAGGCTAGGC-3'; and $tRNA^{His}$:
forward 5'-TGTGGAGCCGGTGGCCCCGGGTTCAAATCCC-3',
reverse 5'- CCACAGGCCGGCGCTCTAACCAGGCTGAGC-3'.

Preparation and purification of RNA transcripts. The *M. kandleri* $tRNA^{His}$ and $tRNA^{Asp}$ fragments indicated in Figure 2C were cloned into a pUC19 vector that allowed for *in vitro* T7 RNA polymerase run-off transcription after plasmid cleavage with *BstNI*. The tRNAs were internally labelled during T7 RNA polymerase *in vitro* transcription in the presence of [α-^{32}P]CTP or [α-^{32}P]ATP and purified by 12% polyacrylamide gel electrophoresis in the presence of 8M Urea as described (*2*).

Preparation and purification of MK0935. The MK0935 gene was amplified from *M. kandleri* genomic DNA and cloned into the NdeI/HindIII site of pET-20b(+) vector to facilitate expression of the proteins in the E. coli BL21-codon plus (DE3)-RIL strain (Stratagene). Cultures were grown at 37 °C in Luria–Bertani medium supplemented with 100 µg/ml ampicillin and 34 µg/ml chloramphenicol and the recombinant protein were produced by autoinduction as described (*3*). Cells were resuspended in buffer containing 50 mM Tris-HCl pH 7.5 and 300 mM NaCl, and broken by sonication. The fractions were extensively flocculated at 80 °C for 30 min, and then centrifuged for 30 min at 20,000 g. RNA bound to the enzyme was removed by anion-exchange chromatography and a final gel filtration step was used to remove soluble aggregates (Fig. S4).

C-to-U editing assays. A typical C-to-U editing reaction contained purified and radioactively labeled RNA substrate (1.1 µM of tRNA and 1.6 µM of tRNA fragments) that was reacted with 10 µM CDAT8 or BSA as negative control for 30min at 70°C in a reaction buffer containing 50 mM Tris-HCl pH 7.9, 10 mM $MgCl_2$, 100 mM NaCl and 1 mM DTT. The reaction was stopped by phenol/chloroform extraction and the products were digested for 2 hours with nuclease P1 (0.66 mg/mL nuclease P1 (Sigma) in 100 mM sodium citrate (pH 4.74) , and separated by TLC on PEI-Cellulose coated glass

plates (20 × 20 cm) with either solvent C: isopropanol:concentrated HCl:water (68:18:14

[v:v:v]) or in two-dimensional TLCs with solvent A: isobutyric acid-concentrated ammonia-water (66:1:33 [v:v:v]) or solvent B: phosphate buffer:NH4 sulfate:*n*-propanol (100:60:2 [v:w:v]) as described (*4*). The nuclease P1 digest was supplemented with a 10mM nucleotide-5'-monophosphates standard. Marker nucleotides were detected by UV shadowing and the radioactively labelled nucleotides derived from RNA cleavage were visualized by autoradiography. Minisequencing assays were conducted with 5'[^{32}P]-labelled oligonucleotides (tRNA^{Asp}: 5'-GACAGGCCCGGATGATAGACCGGGCTACACC-3', tRNA^{His}: 5'-CCACAGGCCGGCGCTCTAACCAGGCTGAGCT-3') annealed to the sequence one nucleotide downstream of base 8. The primers were extended by 10 U of Thermo Script reverse transcriptase (Invitrogen) in the presence of either 20 μM each of dGTP, ddATP, ddCTP, ddTTP (detecting C8) or 20 μM each of dATP, ddCTP, ddGTP, ddTTP (detecting U8).

Crystallization and Structure Determination. Crystals of CDAT8 were obtained by sitting drop vapor diffusion in two different crystal forms: a $P2_12_12$ crystal form in 0.1M MES (pH 6.5) and 18% PEG 20K; and a C2 form in 0.1 M sodium cacodylate (pH 7.2), 32-40% PEG 400, and 0.1 M calcium acetate. Each crystal form had four copies of CDAT8 (two dimers) in the asymmetric unit. Initial phases for the $P2_12_12$ crystal form were obtained by multi-wavelength anomalous diffraction (MAD) and single wavelength anomalous diffraction (SAD) using signals from the intrinsic Zn ions at the enzyme active sites. The Zn ions were located using the program Shelx (*5*) and the phases were calculated using Solve (*6*). Initial phases were not good enough to yield an interpretable map, so additional heavy atom sites from platinum and mercury soaks were used for multiple isomorphous replacement (MIR) phasing (*6*) and non-

crystallographic symmetry (NCS) detection. After two-fold NCS averaging, we were able to build a low resolution backbone model into the experimental density in the $P2_12_12$ crystal form. Initial phases for the higher resolution C2 crystal form were obtained by MIR with the program Solve (*6*). Phased molecular replacement using the program Molrep (*7*) with the low resolution model built in the $P2_12_12$ crystal form allowed us to determine four-fold domain-wise NCS. The domain orientations between the catalytic deaminase domain and the THUMP domain were different between copies in the asymmetric unit, so masks were made around each domain prior to solvent flattening and four-fold multi-domain NCS averaging in the program DM (*8*). After four-fold averaging, we obtained density maps of good quality for automated building of greater than 80% of the protein in ARP/WARP (*9*). The remaining sequence was built manually in the program Coot (*10*). Refinement was carried out with iterative rounds of TLS and restrained refinement in Refmac5 (*11*) followed by rebuilding the model to the experimental and the 2Fo-Fc maps in Coot. The final model has an R_{work}/R_{free} of 21.6%/26.2%. Data collection and refinement statistics are shown in Table S1.

SOM Text

Analysis of the *Methanopyrus kandleri* genome revealed the presence of a complete set of 34 non-redundant tRNA genes (http://gtrnadb.ucsc.edu/Meth_kand/)(*12, 13*). Only four tRNA species are not edited at position 8 as they contain already a T8 in their gene. These four tRNA species are $tRNA^{Gln}_{UUG}$, $tRNA^{Pro}_{GGG}$, $tRNA^{Pro}_{UGG}$ and $tRNA^{Thr}_{GGU}$. The second $tRNA^{Thr}$ gene contains C8 and the sequence of both $tRNA^{Thr}$ isoacceptors differs only in the anticodon stem/loop and base 13.

U8 is conserved in cytoplasmic tRNAs from all three domains of life and is therefore an ancient feature. It is plausible that in the course of evolution U8 was gradually substituted with C8 in *M. kandleri* tRNA genes until nearly all tRNA genes contained a C8. Such mutations could not be selectively allowed without the presence of a cytidine deaminase that would generate the important U8 base. The modular architecture of CDAT8 suggests that this enzyme was created by the fusion of a cytidine deaminase to a tRNA-binding THUMP domain. We searched for the origin of the cytidine deaminase domain by both sequence similarity analyses and structural phylogenies. CDAT8 is in a unique family within the SCOP ‘cytidine deaminase like’ superfamily (sunid: 53927) (*14*) perhaps resulting from duplication of an early ancestor of extant cytidine deaminases.

This observation is in accordance with its currently unique editing position in the tRNA substrate.

The structure of CDAT8 was solved in two different crystal forms. In each form, the asymmetric unit was composed of two identical CDAT8 dimers. The highest resolution crystal diffracted to 2.4 Å and the final model was refined to R_{work}/R_{free} = 21.6%/26.2%. The interface at the CDD is unlike APOBEC2 as the CDAT8 CDDs dimerize using the opposite side of the central β-sheets. The dimer interface is extensive and over 1800 $Å^2$ per monomer is buried.

Figure S1

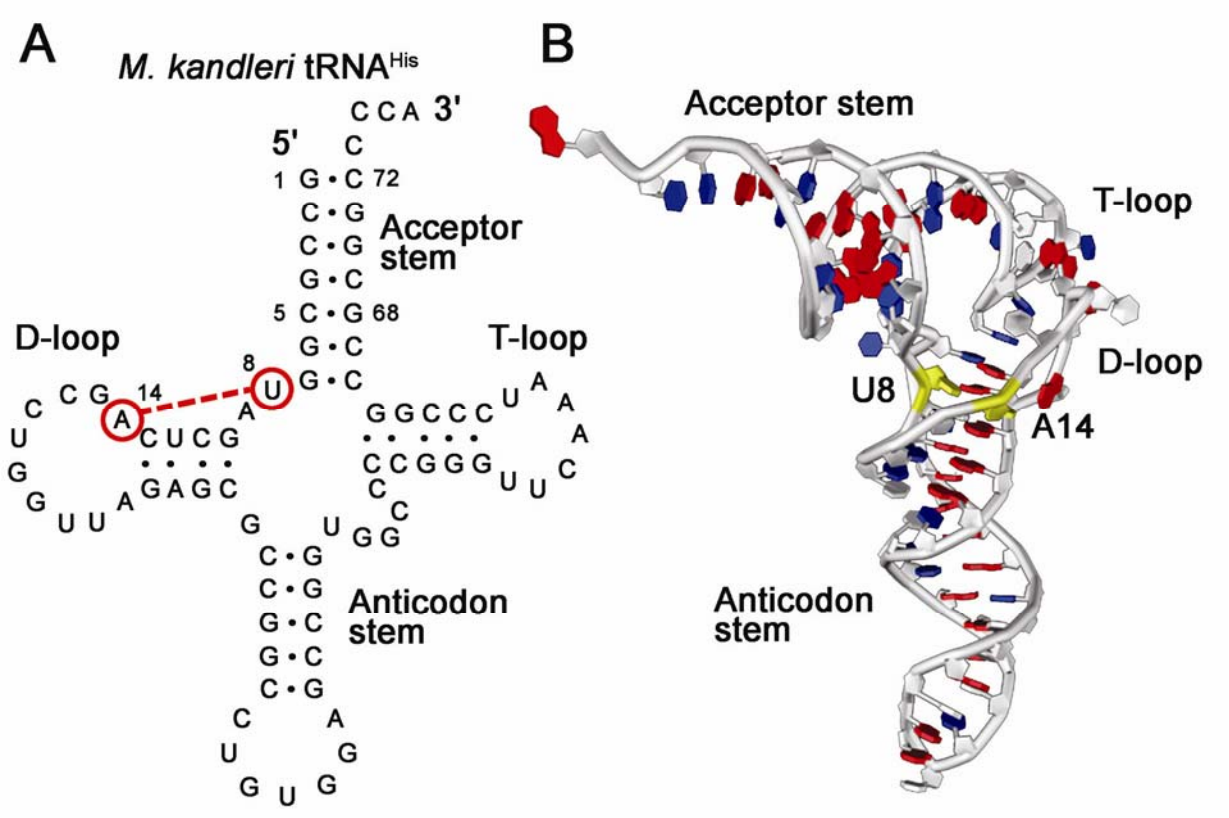

Figure S1. Structural importance of the A14-U8 interaction. The tertiary A14-U8 interaction is indicated **(A)** in the 2D representation of the edited *M. kandleri* $tRNA^{His}$ (red) and **(B)** in the crystal structure of *E. coli* $tRNA^{Phe}$ (yellow, PDB ID 1EVV).

Figure S2

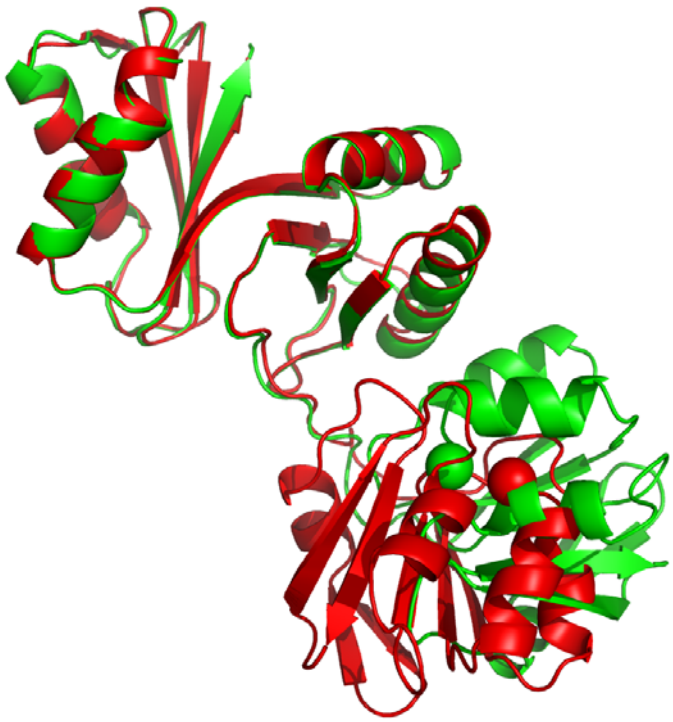

Figure S2. Overlay of the two monomers (red and green) in the asymmetric CDAT8 dimer by superpositioning the THUMP domains only. The spheres represent the catalytic zinc.

Figure S3

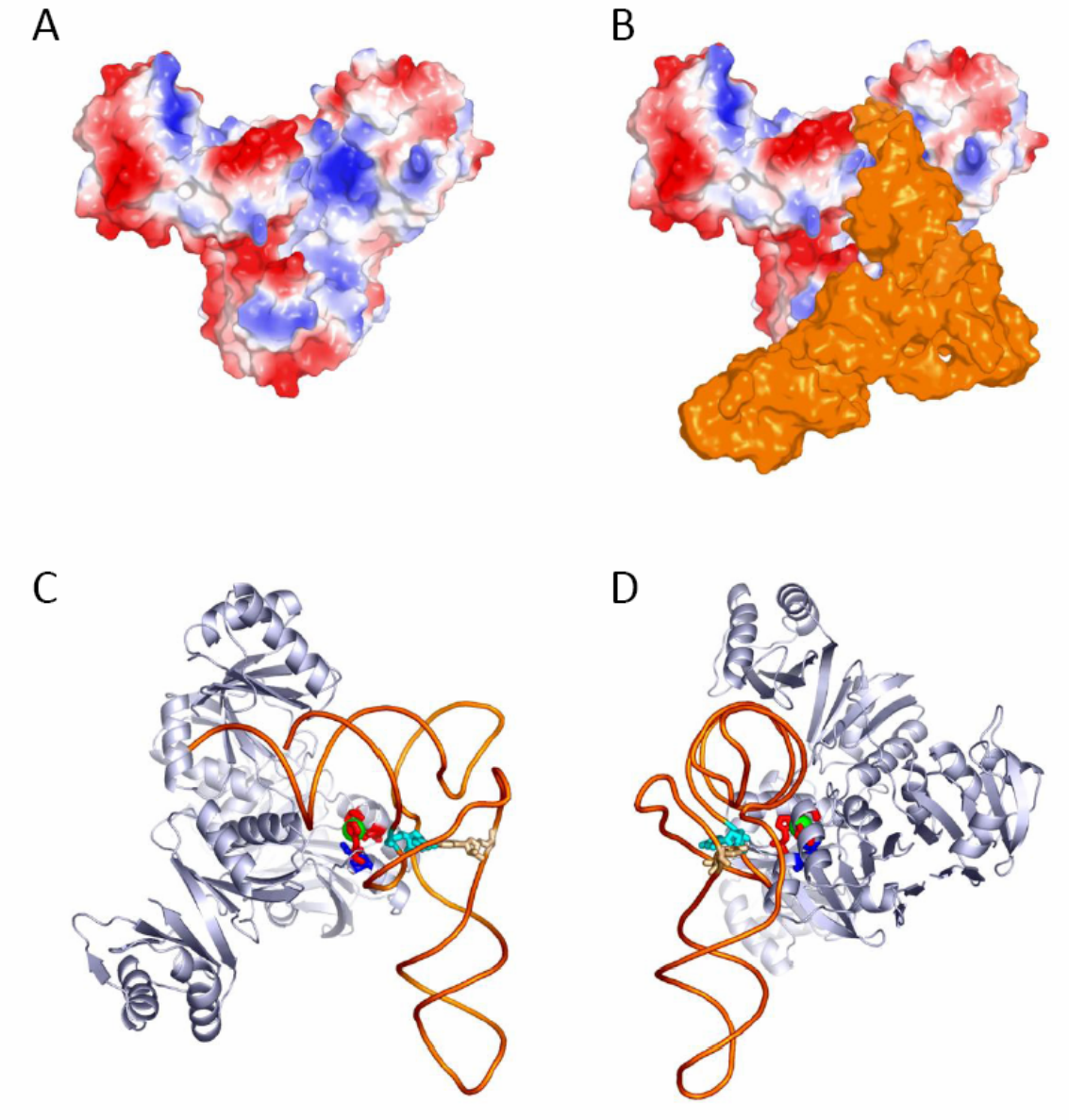

Figure S3. Model for tRNA binding to CDAT8. **(A)** Surface electrostatics for the CDAT8 dimer show a positively charged groove at the interface of the two monomers. **(B)** Surface representation of $tRNA^{Phe}$ (PDB ID 1EVV) modelled with the CDAT8 dimer. **(C)** tRNA position 8 (cyan) is adjacent to the active site. Zinc coordinating residues are shown in red, the catalytic glutamic acid is shown in blue, and the zinc is in green. **(D)** Same as C, but rotated 90 degrees to show the groove complementary to the acceptor stem.

Figure S4

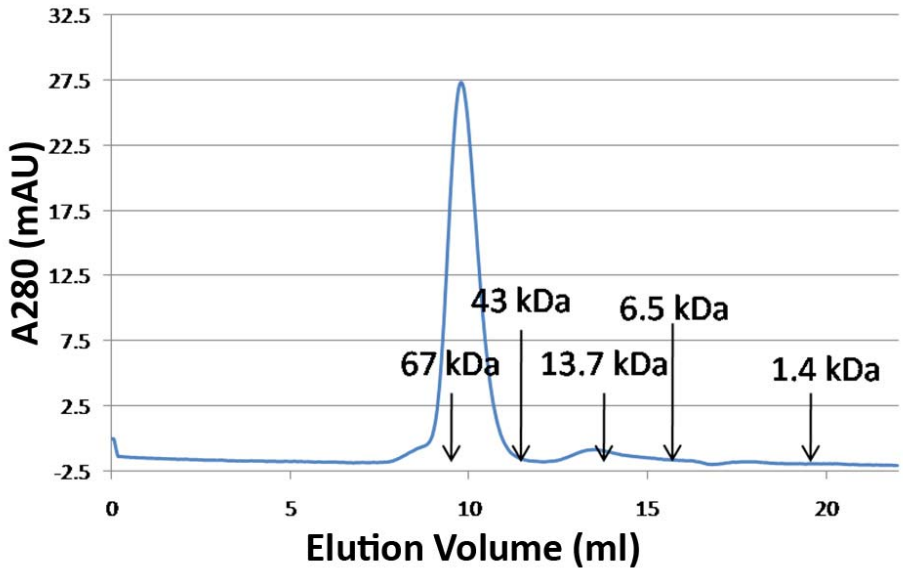

Figure S4. CDAT8 gel filtration chromatogram indicates dimer formation. CDAT8 was applied to a Superdex 75 GL size exclusion column and elutes as a single species of approximately 60 kDa relative to gel filtration standards. The molecular weight of the CDAT8 monomer is 30.6 kDa.

Table S1

Data collection, phasing, and refinement statistics

	C2 Crystal Form				
Cell Dimensions (Å)	Native	KAu(CN)4	PIP	Hg2Cl2	K2PtCl2
A	184.465	185.633	186.227	186.149	184.540
B	77.060	77.327	77.427	77.299	77.356
C	109.082	109.070	109.300	109.434	108.975
Wavelength (nm)		1.038	1.072	1.007	1.072
Resolution	50-2.4	50-3.5	50-3.2	50-3.7	50-3.4
Rmerge	7.1 (64.7)	17.4 (49.4)	10.6 (65.1)	18.9 (73.7)	8.4 (19.0)
I/σ I	14.7 (1.9)	6.1 (1.9)	11.5 (1.6)	6.7 (1.2)	8.8 (4.0)
Completeness	99.4 (99.8)	97.3 (90.3)	98.9 (93.2)	97.2 (85.5)	99.0 (96.5)
Redundancy	3.2 (3.1)	3.2 (2.8)	3.5 (3.1)	3.6 (3.2)	1.9 (1.8)

	P21212 Crystal Form			
	MAD			SAD
Cell Dimensions (Å)	Peak	Inflection	Remote	Peak
A	117.857	117.877	117.883	118.152
B	150.145	150.100	150.202	150.125
C	101.495	101.480	101.460	102.405
Wavelength (nm)	1.282	1.283	1.219	1.282
Resolution	50-3.4	50-3.4	50-3.4	50-3.2
Rmerge	12.0 (>100)	11.8 (>100)	11.0 (>100)	9.6 (>100)
I/σ I	17.3 (1.2)	15.5 (1.1)	19.3 (1.6)	28.9 (1.5)
Completeness	99.9 (99.1)	99.9 (97.4)	100 (100)	100 (100)
Redundancy	9.4 (8.5)	7.1 (6.3)	9.6 (9.6)	15.2 (11.1)

Refinement	
Resolution (Å)	50-2.40
Rwork/Rfree	21.6/26.2
rms Deviations	
Bond Lengths	0.0152
Bond Angles	1.598

Values in parentheses are for the highest resolution shell

Supplementary references

1. A. W. Curnow, D. L. Tumbula, J. T. Pelaschier, B. Min, D. Söll, *Proc. Natl. Acad. Sci. USA* **95**, 12838 (1998).
2. L. Randau *et al.*, *Proc. Natl Acad. Sci. USA* **120**, 17934 (2005).
3. F. W. Studier, *Protein Expr. Purif.* **41**, 207 (2005).
4. H. Grosjean, G. Keith, L. Droogmans, in *Methods in Molecular Biology, RNA Interference, Editing, and Modification, Methods and Protocols*. (Humana Press, 2004), vol. 265, pp. 357-391.
5. G. M. Sheldrick, *Acta Crystallogr. A* **64**, 112 (2008).
6. T. C. Terwilliger, J. Berendzen, *Acta Crystallogr. D Biol. Crystallogr.* **55**, 849 (1999).
7. A. Vagin, A. Teplyakov, *Acta Crystallogr. D Biol. Crystallogr.* **56**, 1622 (2000).
8. K. D. Cowtan, P. Main, *Acta Crystallogr. D Biol. Crystallogr.* **52**, 43 (1996).
9. A. Perrakis, M. Harkiolaki, K. S. Wilson, V. S. Lamzin, *Acta Crystallogr. D Biol. Crystallogr.* **57**, 1445 (2001).
10. G. N. Murshudov, A. A. Vagin, E. J. Dodson, *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240 (1997).
11. P. Emsley, K. Cowtan, *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126 (2004).
12. P. P. Chan, T. M. Lowe, *Nucleic Acids Res.* **37**, D93 (2009).
13. F. Jühling *et al.*, *Nucleic Acids Res.* **37**, D159 (2009).
14. A. Andreeva *et al.*, *Nucleic Acids Res.* **32**, D226 (2004).