X-ray structure of tRNA pseudouridine synthase TruD reveals an inserted domain with a novel fold

Ulrika B. Ericsson, Pär Nordlund, B. Martin Hallberg*

Department of Biochemistry and Biophysics, Stockholm University, Roslagstullsbacken 15, SE-114 21 Stockholm, Sweden

Received 26 February 2004; accepted 22 March 2004

First published online 13 April 2004

Edited by Lev Kisselev

Abstract Pseudouridine synthases catalyse the isomerisation of uridine to pseudouridine in structural RNA. The pseudouridine synthase TruD, that modifies U13 in tRNA, belongs to a recently identified and large family of pseudouridine synthases present in all kingdoms of life. We report here the crystal structure of *Escherichia coli* TruD at 2.0 Å resolution. The structure reveals an overall V-shaped molecule with an RNA-binding cleft formed between two domains: a catalytic domain and an insertion domain. The catalytic domain has a fold similar to that of the catalytic domains of previously characterised pseudouridine synthases, whereas the insertion domain displays a novel fold.

© 2004 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

Keywords: Pseudouridine synthase; RNA modification; tRNA; TruD; Pus7

1. Introduction

Nucleotides of RNA undergo chemical modification in all organisms and the most abundant modification is the conversion of specific uridine (U) residues to pseudouridine (5-ribosyluracil; ψ). Pseudouridine is unique in possessing a C–C bond rather than the N–C glycosidic bond usually present in nucleotides. The post-transcriptional isomerisation of U to ψ is catalysed by pseudouridine synthases in a variety of structural RNAs, including tRNA, rRNA and sn(o)RNA. Site specificity is obtained through the presence of a set of pseudouridine synthases, each specific for one or, at the most, three modification sites.

Four distinct families of pseudouridine synthases have been defined based on sequence analysis: TruA, TruB, RsuA and RluA [1,2]. The only common conserved sequence feature between these families is a short motif [2] containing an aspartate residue essential for the catalytic activity [3,4]. Structural studies of members in the TruA, TruB, RsuA and RluA families have revealed that even though the enzymes have low sequence similarity, their catalytic domains have similar overall structure and also identical location of the conserved aspartate residue in the active site. This implies a common ancestry of the pseudouridine-synthase families [5–11].

The mechanism of the pseudouridine-synthase catalysed isomerisation is debated but evidence is accumulating that the reaction proceeds through a covalent intermediate formed between an aspartate residue and C6 of U, allowing a 180° rotation of the pyrimidine, and the subsequent re-ligation resulting in a C–C glycosidic bond [3]. Some, but not all, pseudouridine synthases are inhibited by 5-fluorouridine, rather than U, at the site of modification. The fluorine interferes with the re-ligation step, causing the enzyme to be trapped with a covalent adduct at the catalytic Asp [3,12,13]. Co-crystal structures of TruB and stem-loop RNA 5-fluorouridine-based inhibitors have been determined [6,8], but unfortunately none of these structures display a covalent intermediate. Recent investigations show that TruB is not able to form a covalent adduct with 5-fluorouridine in solution [13], thus, additional co-crystal structures with pseudouridine synthases that have been successfully inhibited by 5-fluorouridine are needed to shed light on the mechanism.

Recently, a new pseudouridine-synthase family, with no sequence homology to known pseudouridine synthases, was identified. A yeast representative, *pus7p*, was found to modify position U35 in *Saccharomyces cerevisiae* U2 snRNA, U13 in yeast tRNA and U35 in pre-tRNA^{Tyr} [14,15]. The presence of ψ35 in U2 snRNA is important for high splicing efficiency in *S. cerevisiae* in that it re-sculptures the spliceosomal branch site to enable the first step of splicing [16]. Independently of the yeast work, the search for the pseudouridine synthase responsible for the modification of U13 in tRNA^{Glu} in *E. coli* revealed a *pus7p* orthologue [17]. The gene, *ygbO*, encoding a polypeptide of 39 kDa, was renamed TruD, representing the fourth tRNA-modifying pseudouridine synthase to be characterised. Following the convention to name each pseudouridine-synthase family after the first *E. coli* enzyme to be cloned, this new family of pseudouridine synthases is referred to as the TruD family. In a database search, 58 representatives of the TruD family were detected of which 24 are from Eubacteria, 19 from Archea and 16 from Eukarya [17]. Previously known RNA-binding motifs were not represented among the six motifs identified, indicating that this class of proteins should contain novel RNA-binding sequences [17]. Mutational studies of the only completely conserved aspartate (Asp80 in *E. coli* and Asp256 in *S. cerevisiae*) showed that it is essential for in vitro activity [15,17]. Here, we describe the crystal structure of *E. coli* TruD that represents the first 3D structure from this

* Corresponding author. Fax: +46-8-5537-8358.
E-mail address: hallberg@dbb.su.se (B.M. Hallberg).

0014-5793/$22.00 © 2004 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.febslet.2004.03.085

Table 1
Data collection, phasing and refinement statistics

	Se-Met Peak	Se-Met Highres
Cell constants *a*, *b*, *c*[a] (Å)	63.56, 108.5, 112.2	63.18, 108.1, 112.5
Wavelength (Å)	0.97970	0.93219
Resolution (Å)	35–2.55 (2.64–2.55)	35–2.0 (2.1–2.0)
Unique/observed reflections	25 511/91 839	52 753/467 504
Completeness (%)	97.4 (80.0)	99.8 (100.0)
$\langle I/\sigma(I)\rangle$	18.3 (3.16)	15.9 (5.35)
R_{sym}	9.6 (31.3)	7.2 (36.0)
Anomalous signal-to-noise ratio[b]	1.76 (1.14)	–
CC (All/Weak)[c]	33/17.6	–
R_{work}/R_{free}[d] (%)	–	19.5/24.3
RMSD bond lengths (Å)/angles (°)	–	0.020/2.0
Ramachandran plot outliers[e] (%)	–	2.0

[a] Space group $P2_12_12_1$.
[b] $\langle|F^+ - F^-|/\sigma(F^+ - F^-)\rangle$ as calculated by XPREP (Bruker AXS).
[c] Correlation coefficient as a measure for the agreement between E_o^2 and E_c^2; expressed as a percentage for the best solution in SHELXD [21].
[d] 3.1% of reflections excluded from refinement.
[e] Percentage of residues that fall outside core regions of the Ramachandran plot [34].

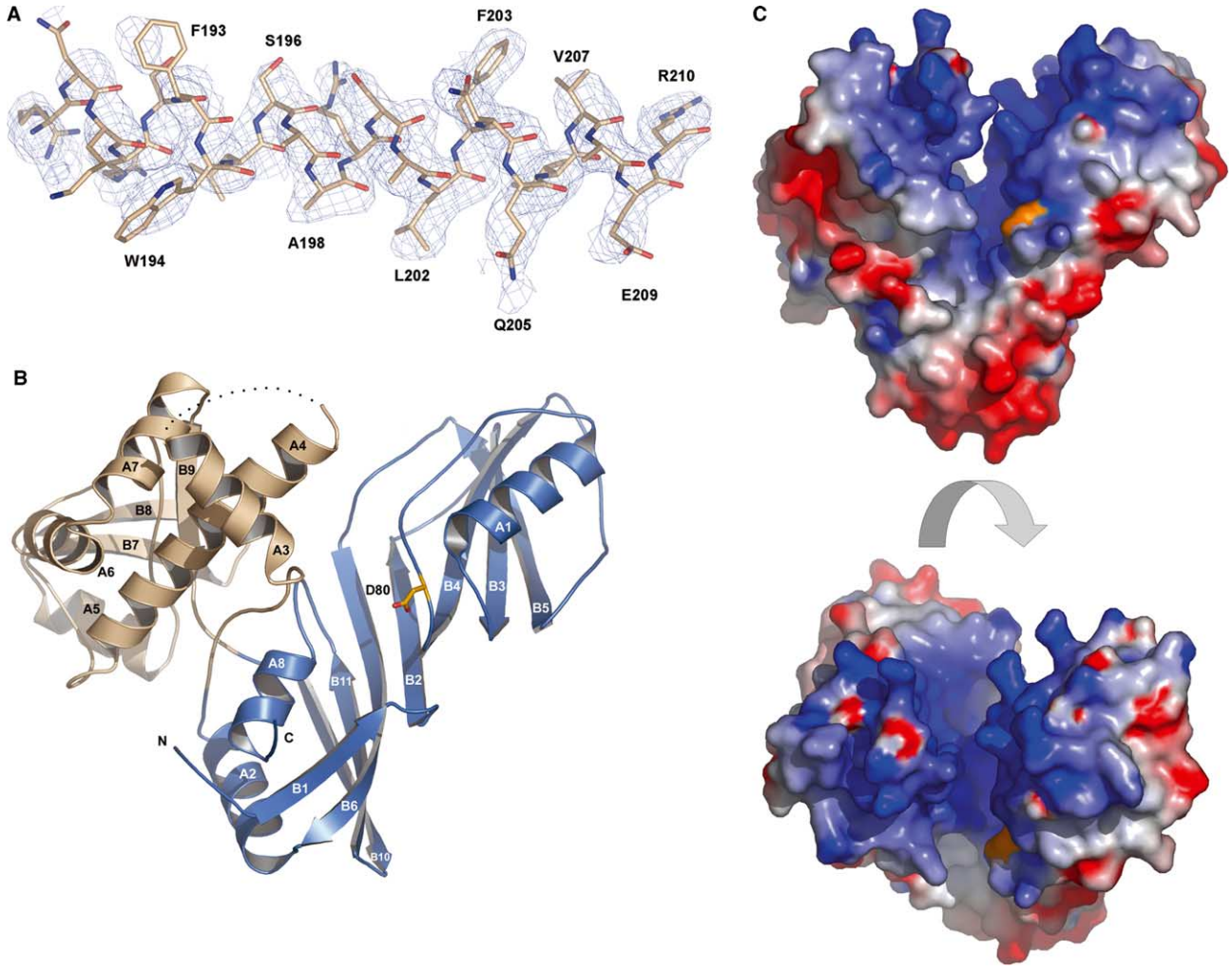

Fig. 1. (A) Unbiased electron density from RESOLVE [23] contoured at 1σ around parts of motif 5 in the insertion domain. (B) Ribbon diagram of *E. coli* TruD. The catalytic domain (residues 1–155 and 304–341) is in blue and the insertion domain (residues 156–303) is in beige. The proposed catalytic aspartate is shown in stick representation. The loop between residues 180 and 187 is not present in the crystallographic model and is represented by a dotted curve in the picture. Molecule A was used in this figure. (C) A molecular surface representation of TruD coloured by the local electrostatic potential (blue, +8 kT; red −5 kT). The surface of the catalytic Asp80 is shown in orange. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

recently discovered, and widely distributed, family of pseudouridine synthases.

2. Materials and methods

2.1. Structure determination

TruD was expressed, purified and crystallised as described elsewhere [18] with the modification that the methionine pathway inhibition method [19] was used to obtain selenomethionine-substituted protein. A single-wavelength anomalous dispersion data set was collected at beamline PSF-BL1 (BESSY, Berlin, Germany) (Table 1). Diffraction data were processed and scaled with HKL2000 using an empirical absorption correction to maximise the anomalous signal [20]. After data preparation with XPREP (Bruker AXS), seven out of eight possible selenium sites were identified with SHELXD [21]. The identified selenium positions were used to calculate phases with the program SOLVE [22]. Density modification using RESOLVE [23] produced an easily interpreted electron density map (Fig. 1A).

A higher resolution data set (Table 1) was collected for automatic model building in ARP/wARP [24] and was processed with the XDS package [25]. The resulting crystallographic model was refined using REFMAC5 [26] with alternating rebuilding with O [27]. Only one region (residues 180–187) was poorly resolved and did not allow model building. A solvent model was built automatically in the ARP/wARP package. Coordinates and structure factors have been deposited in the PDB (ID code: 1SZW).

Rigid-body docking calculations of the complex between tRNA from 1J2B [28] and TruD were performed with the program HEX [29] by a global search using a ligand, receptor and twist range of 180°, 180° and 360°, respectively.

3. Results and discussion

3.1. Overall structure

TruD is a V-shaped molecule (Fig. 1B) with two structurally distinct domains: a catalytic domain ($\sim 20 \times 26 \times 53$ Å^3; residues 1–155 and 304–341) and an insertion domain ($\sim 22 \times 26 \times 45$ Å^3; residues 156–303). A hinge provides numerous inter-domain interactions at the bottom of the V and an extended, quite narrow cleft between the two domains. The rendering of an electrostatic surface on TruD exposes a rift with strong positive charge between the two domains and positive surfaces at the top of the V (Fig. 1C). These surfaces are the most probable sites of tRNA interaction. A structural alignment of TruD-related pseudouridine-synthase sequences (Fig. 2) reveals that the eukaryotic members have two insertions: one N-terminal, and one after strand B2. These insertions may encode domains necessary for recognition of substrates exclusive to eukaryotes. The six motifs identified previously [17] are all found along the cleft between the two domains in the structure. The two molecules constituting the crystal asymmetric unit have slightly different inter-domain

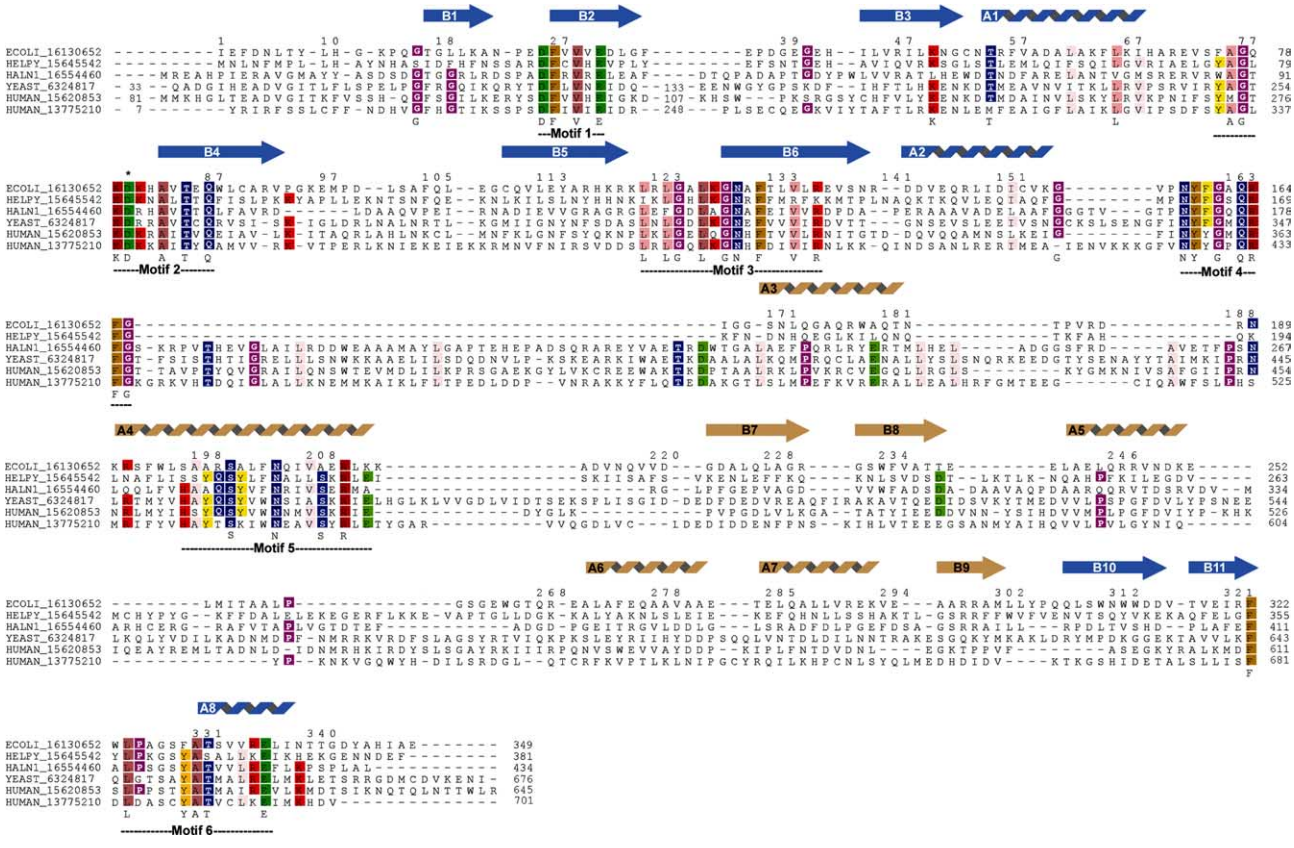

Fig. 2. Structure-based multiple sequence alignment of members in the TruD family of pseudouridine synthases. Each homologue is denoted with its Swiss-Prot organism code (ECOLI, *E. coli*; HELPY, *Helicobacter pylori*; HALN1, *Halobacterium* sp. *NRC-1*; YEAST, *S. cerevisiae*; HUMAN, *Homo sapiens*) followed by its *gi* number. Secondary-structure elements are shown above and consensus residues (at 80% identity) below the alignment. Sequence conservation is shown according to Zappo colour coding (i.e., aliphatic, pink; aromatic, orange; positive, red; negative, green; hydrophilic, blue; conformationally special, magenta). Conserved sequence motifs are denoted below the alignment. Insertions in the eukaryotic sequences are marked by the number of excluded residues. The catalytic aspartate is denoted with an asterisk. Secondary-structure elements are denoted as in Fig. 1B. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

orientation, presumably due to different packing environment. This is manifested in a more closed overall V-shape in one of the molecules, highlighting significant plasticity of the TruD molecule. Furthermore, during the review process of this manuscript, a crystal structure of *E. coli* TruD was reported elsewhere [30], independently of our work. Although similar in shape, the other TruD model appears to have a more closed overall V-shape compared with the structure reported here. This stresses further the plasticity of the TruD molecule.

3.2. The catalytic domain of TruD reveals structural conservation in the absence of sequence identity

Although there is no significant sequence homology, the catalytic domain is dominated by a large curved β-sheet featuring the tandem ferredoxin-like fold found in the catalytic domains of the previously characterised pseudouridine synthases. A DALI [31] search with the catalytic domain as a search probe returns TruA as the structurally most similar protein (Fig. 3A), followed by the catalytic subunits of TruB and RsuA. An optimised superposition of the available pseudouridine-synthase structures to TruD (Table 2) with the program O [27] reveals a sequence identity of the structural superposition approaching the limit of random alignment [32]. However, the structural and functional similarity supports a divergent evolution from a common ancestry.

The proposed catalytic residue Asp80 [17] is the only structurally conserved residue in the active site. However, two structural features are retained, highlighting their importance for recognition and catalysis. The residue Phe131 is conserved in the TruD family, while it is a conserved tyrosine in the other pseudouridine-synthase families. In the modelled TruB-TSL precursor complex [6], this residue (Tyr76 in TruB) makes a stacking interaction with the base to be modified (U55), thereby orienting it for catalysis. Similarly, a conserved leucine residue (Leu200, TruB numbering) is involved in stacking with U55 in TruB and superimposes well with a phenylalanine (Phe27) conserved in the TruD family (Fig. 3B). A feature exclusive to TruB and TruD is Tyr179 (TruB numbering) that is proposed to stabilise a reaction intermediate [6]. The corresponding residue (Phe329) in the TruD family is either a Phe or a Tyr (Fig. 2). Similar to other pseudouridine synthases, a catalytic pocket that can accept the flipped-out U to be modified is formed in TruD. Although substrate recognition is idiosyncratic in pseudouridine synthases, the catalytic pocket and the overall structural conservation of the active site indicates that pseudouridine synthases have a common catalytic reaction mechanism. The loop that contains motif 2 (Fig. 2) adopts different conformations in the A and B molecule of the asymmetric unit. Interestingly, the displacement is particularly pronounced for Asp80 (4.6 Å on Cα).

3.3. The insertion domain has a new fold

The insertion domain features a mixed α/β structure dominated by an extended 24-residue α-helix. A search with DALI [31] revealed no significant structural similarity to any entry in the PDB. Thus, the structure of this domain constitutes a new protein fold, and considering the strong positive charge present on the side facing the catalytic domain, the inserted domain is most likely involved in RNA binding (Fig. 1C). Sequence conservation in this domain is localised to the side of the long α-helix that faces the cleft between the domains (motif 5). A basic local alignment search tool [33] search for short, almost perfect matches using motif 5 as a search template revealed high local sequence similarity with a part of the *Arabidopsis thaliana* RecQ DNA-helicase (gi:12320889), implying that the insertion domain may have a relative involved in DNA interaction.

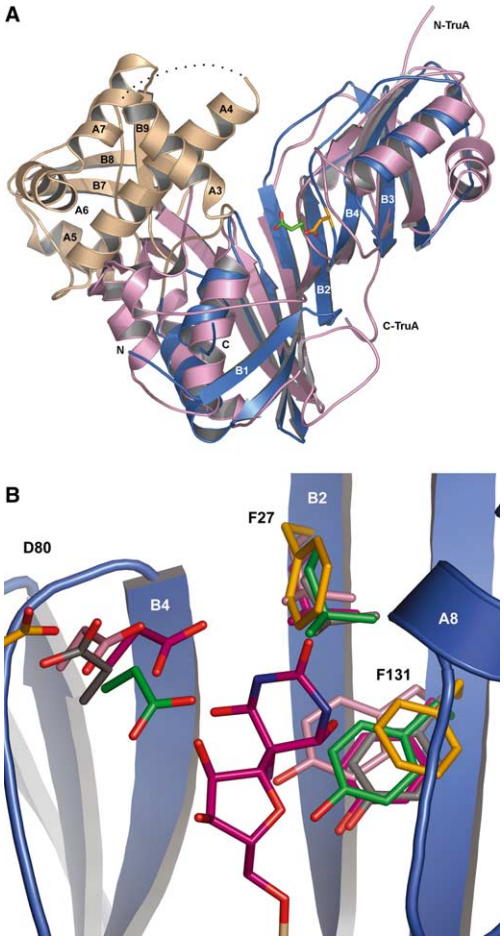

Fig. 3. (A) Superposition of TruD and TruA. TruD has the same colouring scheme as in Fig. 1(B). TruA is coloured in pink with its catalytic aspartate (D60) in stick representation (green). Secondary-structure elements are denoted as in Fig. 1B. Molecule A of TruD was used in this figure. (B) Superposition of universally conserved structural features in the active-site region, also including the 5-fluoro-6-hydroxy-pseudouracil from TruB (1K8W). Colour scheme: TruD ribbon in blue, side chains in gold; RsuA (1KSK) in grey; TruA (1DJO) in green; TruB (1K8W) in magenta (phosphate in beige); RluD (1PRZ) in pink. Secondary-structure elements are denoted as in Fig. 1B. Molecule A of TruD was used in this figure. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The domain is inserted in one of the ferredoxin-like units in the catalytic domain, suggesting that the insertion occurred after the gene duplication event presumed to have generated the tandem ferredoxin-like fold of the catalytic domains of pseudouridine synthases [9]. The structure of TruD is the first example of a structurally characterised pseudouridine synthase that has an external domain as an insertion in the primary structure of the catalytic domain.

3.4. Interaction with tRNA

The nucleotide modified by TruD, U13, resides in the D-loop that is part of the structural core of tRNA. This is the

Table 2
Optimised structural superimposition[a] of pseudouridine synthases and the sequence identity based on the structural alignment

	TruA (1DJ0)	TruB (1K8W)	RsuA (1KSK)	RluD (1PRZ)[b]	TruD[c]
TruA	–	15%	10%	10%	8%
TruB	1.7 Å; 107 Cα	–	23%	19%	8%
RsuA	2.0 Å; 106 Cα	1.5 Å; 110 Cα	–	21%	11%
RluD	2.3 Å; 91 Cα	1.9 Å; 115 Cα	1.7 Å; 105 Cα	–	12%
TruD[2]	2.1 Å; 118 Cα	2.2 Å; 122 Cα	2.2 Å; 92 Cα	2.1 Å; 91 Cα	–

[a] The structures were searched in the program O [27] for the longest possible matching structural fragments of at least three residues for which each atom pair could be aligned within a distance of 3.8 Å.
[b] 1QYU is very similar to 1PRZ and is therefore not used in this comparison.
[c] Molecule A was used in the structural superimposition.

only known pseudouridine in the D-loop of *E. coli* tRNAs. Any post-transcriptional modification of residues in the D-loop would need a partial unfolding of the L-shaped structure of tRNA. This hypothesis was corroborated by the recent co-crystal structure of an archeosine tRNA–guanine transglycosylase [28], which exchanges G15 in archeal D-loops to a 7-cyano-7 deazaguanine by transglycosylation. In this structure, the canonical core and the D-loop of the bound tRNA are unfolded. The D-stem protrudes to form a lip giving the tRNA a λ-form, thus enabling the enzyme the access necessary to exchange G15. Following initial, unsatisfactory, attempts to dock the relatively wide classical L-form of tRNA into the narrow cleft between the two domains in TruD, the λ-form of tRNA from the archeosine tRNA–guanine transglycosylase structure (PDB entry 1J2B; [28]) was docked to TruD. The dominant solutions clustered in a docking mode where the single-stranded D-loop penetrates into the rift between the two domains (Fig. 4). The interaction surface of the complex is extensive (3900 Å^2), and has a high degree of charge and shape complementarity. Interestingly, the only disordered part of the TruD structure described here (residues 180–187 in the insertion domain) would be able to interact with the exposed core of the tRNA molecule so as to stabilise the unfolded D-stem, a role similar to that of the β18–β19 hairpin in the archeosine tRNA–guanine transglycosylase [28]. In this model, the function of the insertion domain would be to recognise and stabilise the λ-form of tRNA. Given the extensive flexibility of the hinge region, as shown by the discrepancy between the two molecules constituting the crystal asymmetric unit, we cannot exclude the possibility that the cleft between the domains can widen enough to allow entry of the L-form of tRNA. However, in that case, the problem with accessing U13 would still remain.

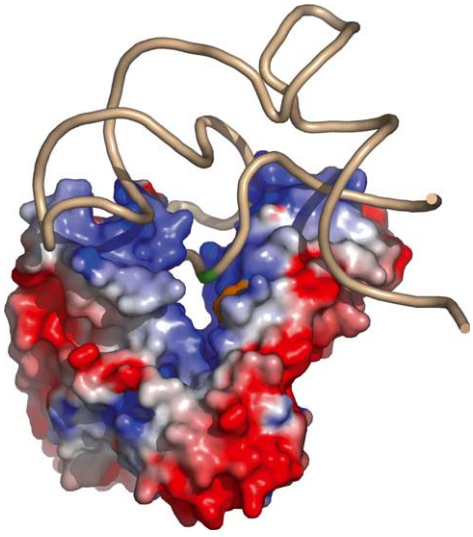

Fig. 4. A rigid-docking model of TruD with the λ-form of tRNA from the structure of archeosine tRNA–guanine transglycosylase (PDB entry 1J2B; [28]). The surface of the catalytic Asp80 is shown in orange, whereas the site to be modified (U13) is coloured green on the λ-tRNA phosphate backbone ribbon. The molecular surface is coloured by its local electrostatic potential (blue, +8 kT; red −5 kT). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.5. Conclusions and future prospects

The crystal structure of TruD is the first from a new and widely distributed family of pseudouridine synthases with representatives from all kingdoms of life. The structure reveals two domains: one that has the same overall fold and location of the active site as the catalytic domains observed in other pseudouridine synthases, and a smaller domain that displays a novel fold inserted into the primary structure of the catalytic domain. It provides a structural basis for further biochemical, biophysical and genetic studies on members of this family. Future challenges include investigation of co-crystal structures with full-length tRNA to understand how recognition and catalysis is achieved at a molecular level.

Acknowledgements: We thank the staff (Uwe Müller and Martin Fieber-Erdmann) at PSF-BESSY (Berlin, Germany) for support during data collection. This study was supported by the Swedish Research Council, the Wallenberg Consortium North and the EU framework V Network SPINE.

References

[1] Gustafsson, C., Reid, R., Greene, P.J. and Santi, D.V. (1996) Nucleic Acids Res. 24, 3756–3762.
[2] Koonin, E.V. (1996) Nucleic Acids Res. 24, 2411–2415.
[3] Huang, L., Pookanjanatavip, M., Gu, X. and Santi, D.V. (1998) Biochemistry 37, 344–351.
[4] Ramamurthy, V., Swann, S.L., Paulson, J.L., Spedaliere, C.J. and Mueller, E.G. (1999) J. Biol. Chem. 274, 22225–22230.
[5] Foster, P.G., Huang, L., Santi, D.V. and Stroud, R.M. (2000) Nat. Struct. Biol. 7, 23–27.
[6] Hoang, C. and Ferré-D'Amaré, A.R. (2001) Cell 107, 929–939.
[7] Mueller, E.G. (2002) Nat. Struct. Biol. 9, 320–322.
[8] Pan, H., Agarwalla, S., Moustaka, D.T., Finer-Moore, J. and Stroud, R.M. (2003) Proc. Natl. Acad. Sci. USA 100, 12648–12653.
[9] Sivaraman, J., Sauve, V., Larocque, R., Stura, E.A., Schrag, J.D., Cygler, M. and Matte, A. (2002) Nat. Struct. Biol. 9, 353–358.
[10] Sivaraman, J., Iannuzzi, P., Cygler, M. and Matte, A. (2004) J. Mol. Biol. 335, 87–101.

[11] del Campo, M., Ofengand, J. and Malhotra, A. (2004) RNA 10, 231–239.
[12] Samuelsson, T. (1991) Nucleic Acids Res. 19, 6139–6144.
[13] Spedaliere, C.T. and Mueller, E.G. (2004) RNA 10, 192–199.
[14] Ma, X., Zhao, X. and Yu, Y.T. (2003) EMBO J. 22, 1889–1897.
[15] Behm-Ansmant, I., Urban, A., Ma, X., Yu, Y.T., Motorin, Y. and Branlant, C. (2003) RNA 9, 1371–1382.
[16] Newby, M.I. and Greenbaum, N.L. (2002) Nat. Struct. Biol. 9, 958–965.
[17] Kaya, Y. and Ofengand, J. (2003) RNA 9, 711–721.
[18] Ericsson, U.B., Andersson, M.E., Engvall, B., Nordlund, P. and Hallberg, B.M. (2004) Acta Crystallogr. D 60, 775–776.
[19] Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L. and Clardy, J. (1993) J. Mol. Biol. 229, 105–124.
[20] Otwinowski, Z., Borek, D., Majewski, W. and Minor, W. (2003) Acta Crystallogr. A 59, 228–234.
[21] Schneider, T.R. and Sheldrick, G.M. (2002) Acta Crystallogr. D 58, 1772–1779.
[22] Terwilliger, T.C. and Berendzen, J. (1999) Acta Crystallogr. D 55, 849–861.
[23] Terwilliger, T.C. (2002) Acta Crystallogr. D 58, 1937–1940.
[24] Perrakis, A., Morris, R. and Lamzin, V.S. (1999) Nat. Struct. Biol. 6, 458–463.
[25] Kabsch, W. (1993) J. Appl. Cryst. 26, 795–800.
[26] Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Acta Crystallogr. D 53, 240–255.
[27] Jones, T.A., Zou, J.-Y., Cowan, S.W. and Kjeldgaard, M. (1991) Acta Crystallogr. A 47, 110–119.
[28] Ishitani, R., Nureki, O., Nameki, N., Okada, N., Nishimura, S. and Yokoyama, S. (2003) Cell 113, 383–394.
[29] Ritchie, D.W. and Kemp, G.J.L. (2000) Proteins 39, 178–194.
[30] Kaya, Y., del Campo, M., Ofengand, J. and Malhutra, A. (2004) J. Biol. Chem. (in press).
[31] Holm, L. and Sander, C. (1998) Proteins 33, 88–96.
[32] Rost, B. (1997) Fold. Des. 2, S19–S22.
[33] Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[34] Kleywegt, G.J. and Jones, T.A. (1996) Structure 4, 1395–1400.