A role for [Fe$_4$S$_4$] clusters in tRNA recognition—a theoretical study

Martin T. Stiebritz[*]

Laboratorium für Physikalische Chemie, ETH Zurich, Vladimir-Prelog-Weg 2, 8093 Zurich, Switzerland

ABSTRACT

Over the past several years, structural studies have led to the unexpected discovery of iron–sulfur clusters in enzymes that are involved in DNA replication/repair and protein biosynthesis. Although these clusters are generally well-studied cofactors, their significance in the new contexts often remains elusive. One fascinating example is a tryptophanyl-tRNA synthetase from the thermophilic bacterium *Thermotoga maritima*, *Tm*TrpRS, that has recently been structurally characterized. It represents an unprecedented connection among a primordial iron–sulfur cofactor, RNA and protein biosynthesis. Here, a possible role of the [Fe$_4$S$_4$] cluster in tRNA anticodon-loop recognition is investigated by means of density functional theory and comparison with the structure of a human tryptophanyl-tRNA synthetase/tRNA complex. It turns out that a cluster-coordinating cysteine residue, R224, and polar main chain atoms form a characteristic structural motif for recognizing a putative 5′ cytosine or 5′ 2-thiocytosine moiety in the anticodon loop of the tRNA molecule. This motif provides not only affinity but also specificity by creating a structural and energetical penalty for the binding of other bases, such as uracil.

INTRODUCTION

Iron–sulfur clusters play a variety of important roles in the biochemistry of the cell. For a long time they have been known to facilitate electron transport, as for example in ferredoxin-type proteins (1,2), but also enzymatic conversion reactions, with the tricarbonic acid cycle-enzyme aconitase being a famous example (3–5). The significant progress of structural biology over the last two decades that culminated in structural genomics initiatives has shown that iron–sulfur clusters are far more widespread than previously thought and appear at prominent positions in crucial cellular processes such as protein biosynthesis (6–8) and DNA replication and repair without, however, directly

unveiling their functional role (9–11). It could be demonstrated for example that the [Fe$_4$S$_4$] cluster harbored by the MutY protein (*Escherichia coli*) facilitates electron transfer steps that occur during DNA damage recognition and repair (12–15). Frequently, however, the exact functions of the iron–sulfur clusters remain elusive. This is the case for the recently structurally resolved DNA Helicase XPD (10,11) (Xeroderma pigmentosum D in human cells or Ding in *E. coli*) and Trp-tRNA synthetase (TrpRS) from the thermophilic bacterium *Thermotoga maritima* (*Tm*TrpRS). According to current knowledge, the emergence of iron–sulfur clusters appears to be connected exclusively with bacterial life, an idea that is supported by the fact that in eukaryotic cells, intact mitochondria are required for cluster biosynthesis (16,17). This observation, together with the seemingly self-sustained existence of chemosynthetic bacteria and archaea at hydrothermal vents along the mid-oceanic ridges, provides support for an iron–sulfur-based origin of life as proposed by Wächtershäuser (18). We might never be able to uncover fully the processes that took place billions of years ago and eventually resulted in the first life forms, but the occurrence of an iron–sulfur cluster in a tRNA synthetase nevertheless provides a fascinating and suggestive extant connection among an ancient cofactor, nucleic acid-based information decoding and protein biosynthesis. Whether the cluster in this particular enzyme is really a reminiscence of early life on this planet or just the result of a secondary adaptation process to the extreme environment, its host organism lives in, cannot be easily answered. In any case, the assignment of a functional role in the process of aminoacylation is a critical first step in addressing this question.

The crystal structure of *Tm*TrpRS, recently solved by Han *et al.* (7), consists of a catalytic domain with canonical Rossmann fold and a tRNA^{Trp} anticodon-binding (TAB) domain that harbors an [Fe$_4$S$_4$] cluster (see Figure 1). Structural superpositioning with human TrpRS (*Hs*TrpRS) in a complex with bovine tRNA (*Bt*-tRNA) led the authors to speculate about a possible involvement of the cubane in anticodon recognition. Here, the properties of the peculiar Fe–S cluster are studied by means of density functional theory (DFT), and the interaction with its direct protein environment is analyzed. Due to the lack of structural informa-

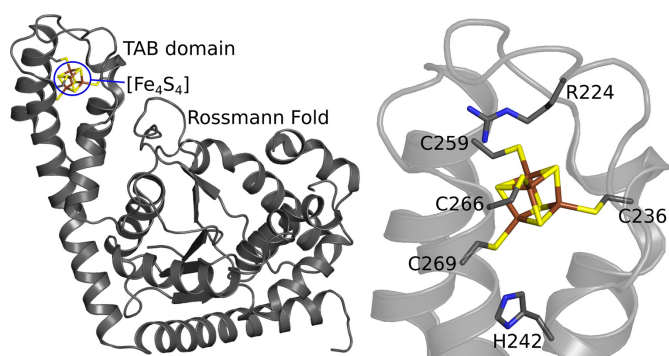[*]Tel: +41 44 633 38 66; Fax: +41 44 633 15 94; Email: martin.stiebritz@phys.chem.ethz.ch

Figure 1. Tryptophanyl-tRNA synthetase from *Thermotoga maritima* (*Tm*TrpRS). Left panel: Overall fold of one monomer of the homodimeric enzyme, showing the embedding of the [Fe$_4$S$_4$] cluster in the tRNA anticodon-binding (TAB) domain. Right panel: Detailed view of the Fe–S cubane with the neighboring residues R224 and H242. The models shown were generated from the crystal structure by Han *et al.* (7) (PDB entry 2G36). Element color code—gray: C, blue: N, yellow: S, brown: Fe.

tion about how *Tm*TrpRS binds to its cognate tRNA^{Trp}, the *Hs*TrpRS/*Bt*-trRNA complex is used to derive a minimal model for studying the recognition of a plausible 5′ base in the anticodon by the TAB domain of *Tm*TrpRS. It turns out that a characteristic structural motif formed by the cubane and side and main chain atoms of the enzyme could provide a specific interaction site suitable for tRNA base recognition.

MATERIALS AND METHODS

All-electron calculations were carried out with the DFT programs provided by the TURBOMOLE suite (19) on model structures of the tryptophanyl-tRNA synthetases from *Thermotoga maritima* (PDB entry 2G36 (7)), *Aeropyrum pernix* (PDB entry 3A05 (8)) and *Homo sapiens* (PDB entry 2DR2 (20)). Amino acids were included by saturating their carboxy-termini with NH-CH$_3$ groups and their N-termini with acetyl groups. The coordinates for the capping groups were taken from the preceding or following residues in the amino acid sequence. During structural optimizations the C, O and N atoms of the main chain were kept fixed. All other atoms, including C$_\alpha$ atoms, were allowed to relax freely. H atoms were added with the MolProbity server (21). As starting geometries for the bases, cytosine, 2-thiocytosine and uracil the coordinates of 5′ cytosine in the anticodon of *Bt*-tRNA^{Trp} (complex structure 2DR2 (20)) were chosen. For 2-thiocytosine and uracil, atoms were edited accordingly. Structural alignments were generated with Pymol (22). In the case of the complex structures of *Tm*TrpRS and *Ap*TrpRS, steric clashes between the enzyme models and the ligands were resolved by force field pre-optimization, using the force field MMFF94 (23) as implemented in Avogadro (24). For quantum mechanical structural optimizations, DFT was applied with the TPSS (Tao, Perdew, Staroverov, and Scuseria) exchange-correlation functionals (25). A triple-ζ def2-TZVP basis set (26), referred to as TZVP throughout the paper, was chosen and, optionally, empirical dispersion corrections (DFT-D3 (27), labeled by TPSS-D3 in the manuscript) were considered as provided by the Turbomole package (19). As shown
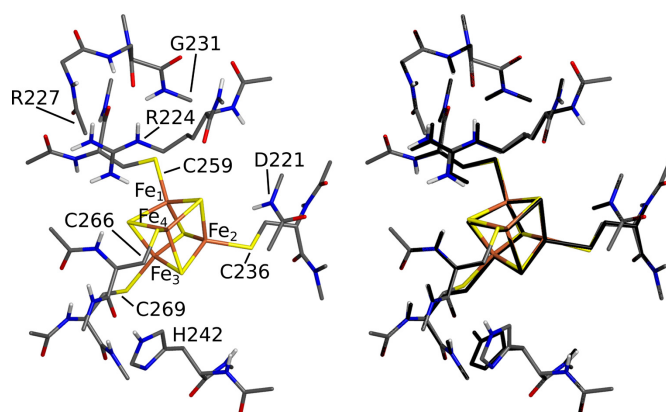

Figure 2. Molecular model of the TAB domain of *Tm*TrpRS considered in this study. Left panel: Model extracted from the *Tm*TrpRS crystal structure (PDB entry 2G36). Cysteine residues C236, C259, C266 and C269 were included together with their main chain atoms, as well as R224 and H242, and the main chain atoms between the C$_\alpha$ atoms of R227 and G231. Only polar hydrogen atoms are depicted. Hydrogen atoms/protons were added by using the MolProbity Server (21). Right panel: Overlay of the starting model (black) shown in the left panel and the structure optimized with TPSS/TZVP for the ^{2}[Fe$_4$S$_4$]$^{1+}$ state of the cubane cluster (colored). Element color code—white: H, red: O, gray: C, blue: N, yellow: S, brown: Fe.

previously, the combination of the TPSS functional with a TZVP basis is well suited to describe the active centers of iron–sulfur proteins (28–30). To accelerate calculations, advantage was taken of the resolution-of-the-identity approximation (RI) with the corresponding auxiliary basis sets (31). In order to account for screening effects of the protein matrix and the solvent, the polarizable continuum solvent model, COSMO (32), was considered, with a dielectric constant of ε = 4. Homology models of TrpRS enzymes were generated with the SWISS-MODEL web service (33–38). Electron density maps (not shown in this paper) were calculated for the crystal structures 2G36, 3A05 and 3JXE with the crystallography package Phenix (39) using the published structure factors. Images of molecular structures were created with Pymol (22).

RESULTS AND DISCUSSION

Minimal model and cluster properties

In the crystal structure of *Tm*TrpRS by Han *et al.* (7), the Fe–S cluster located in the putative anticodon-binding (TAB) domain appears as a regular [Fe$_4$S$_4$] cubane (see Figure 1). For the theoretical analysis of its properties pursued here, a model was chosen that comprises the cluster with the four coordinating cysteine residues and all side chain and main chain atoms that are in close contact with the cubane moiety and that could play an additional role in base recognition (see below). The *Tm*TrpRS model, which is depicted in Figure 2 (left panel), is large enough to investigate all relevant questions addressed in this work while still being computationally tractable with DFT.

As of now, there have been no experimental data available that would help to assign the redox and magnetic states of the cluster. For this reason, a paramagnetic ^{2}[Fe$_4$S$_4$]$^{1+}$ and a diamagnetic ^{1}[Fe$_4$S$_4$]$^{2+}$ state, which are commonly encoun-
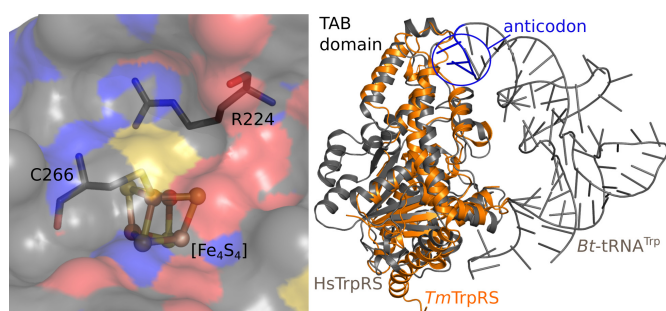
Figure 3. *Tm*TrpRS TAB domain surface and potential tRNA interaction. Left panel: Solvent-excluded surface of the TAB domain of *Tm*TrpRS near the [Fe$_4$S$_4$] cluster that shows that the S atom of C266 is directly solvent-exposed and forms a potential binding groove together with the residue R224. Right panel: Superposition of *Tm*TrpRS (orange) with the structure (gray) of the complex of human TrpRS (*Hs*TrpRS) and bovine tRNA (*Bt*-tRNA^{Trp}).
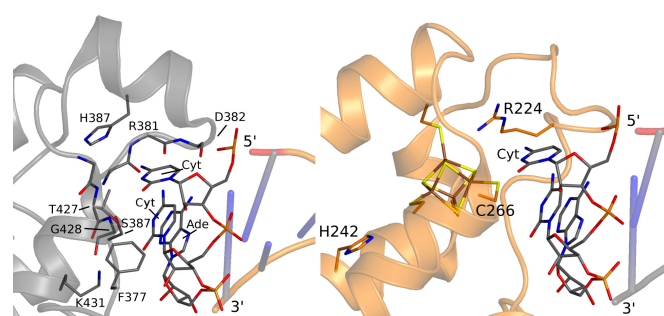

Figure 4. Interaction between tRNA and TrpRS. Left panel: Detailed view of the complex structure *Hs*TrpRS/*Bt*-tRNA^{Trp} of Shen *et al.* (20) (PDB entry 2DR2) that shows the interaction of the anticodon bases with key residues and main chain atoms of the synthetase. The three bases of the anticodon, 5′-CCA-3′, are explicitly labeled (Cyt, Cyt, Ade). Right panel: Detailed view of the hybrid structure of *Tm*TrpRS and *Bt*-tRNA that was obtained by superposition of *Tm*TrpRS (PDB entry 2G36) and *Hs*TrpRS (PDB entry 2DR2). The 5′ cytosine (Cyt) is located between the Fe-S cubane and R224. Because the current study focuses on this cytosine, the other two bases are not labeled explicitly for the sake of clarity. Element color code—red: O, gray/orange: C, blue: N, yellow: S, brown: Fe.

tered in other proteins harboring Fe–S cubanes, were considered. For an approximate description of the diamagnetic state, broken-symmetry (BS) DFT was applied (40–43).

As can be seen from the crystal structure, the cubane is deeply embedded in the TAB domain, with only the sulfur atom of C266 being directly solvent-accessible (Figure 3, left panel). In addition to the coordinating residues C236, C259, C266 and C269, the cluster is in direct contact with R224, which presumably forms an H-bond with the sulfur of the cubane facet bordered by C259 and C266 that faces away from the protein surface. As the solvent-excluded surface shows, the guanidinium group of R224 is partially exposed to the solvent (see Figure 3, left panel). The second sulfur atom of the aforementioned cubane facet is within H-bonding distance of the main chain NH group of D221 (see Figure 2, left panel). H242, being electrically neutral according to the prediction by the program package Mol-Probity (21), forms an H-bond with the sulfur atom of C269 (see Figure 2, left panel). The main chain atoms from R227 to G231 of *Tm*TrpRS were included in the model because they provide a framework to hold R224 in place. Indeed, structural optimization with TPSS/TZVP resulted in only small deviations with respect to the crystallographically determined atom positions (see Figure 2, right panel). This holds true for both oxidation states considered.

In order to analyze the electronic structure of the optimized models, the number of unpaired electrons on selected cluster atoms, $N_{\alpha, i} - N_{\beta, i}$, was calculated as well as the partial charges obtained by Mulliken population analysis (see Table 1). Additionally, optimizations of just the cluster with its coordinating cysteine residues were performed to determine whether both properties are affected by the presence of the amino acids in direct contact with the cubane, R224 and H244. As can be seen from Table 1, the spin polarization of the iron atoms of the Fe–S cluster is almost unchanged when the isolated cubane and the *Tm*TrpRS model are compared. A reduced spin polarization on Fe$_4$ in the reduced oxidation state can also be found in both cases and is therefore not caused by the closeness of R224 or H244, and rather due to the fact that the number of electrons is uneven for the ^{2}[Fe$_4$S$_4$]$^{1+}$ state as opposed to the anti-ferromagnetic singlet, ^{1}[Fe$_4$S$_4$]$^{2+}$. On the S atoms of C266 and C269, only

small residual spin polarization could be detected. Whereas the residues R224 and H244 do not affect spin polarization, they do apparently somewhat alter the charge distribution in the cluster. As a result, the Fe atoms Fe$_3$ and Fe$_4$, belonging to the facet that is almost collinear with the axis formed by the imidazole and the guanidinium group of H244 and R224, are more positively charged, and the iron atoms Fe$_1$ and Fe$_2$, of the opposite facet, are more negatively charged than the corresponding atoms of the free cubane (see Table 1). Interestingly, the charges on the two S atoms considered are almost unchanged by the presence of R224 and H244.

Comparison of *Hs*TrpRS and *Tm*TrpRS TAB domains

Currently, only one crystal structure of a TrpRS/tRNA complex has been solved (at 3.1 Å resolution) and consists of human *Hs*TrpRS and bovine *Bt*-tRNA^{Trp} (20). Although *Tm*TrpRS and *Hs*TrpRS share only 13.4 % identical residues, both proteins feature a high degree of structural conservation, especially in the anticodon-binding (TAB) domain (see Figure 3, right panel, 3.6 Å r.m.s.d. overall). This similarity allowed Han *et al.* to speculate about possible protein/tRNA interactions for the bacterial enzyme for which no complex structure is available to date (7). The location of the [Fe$_4$S$_4$] cluster in the center of the TAB domain could indeed allow for direct interaction with the anticodon loop of a tRNA molecule and suggests a possible role in anticodon recognition. How tRNA proofreading can in general be accomplished by aminoacyl-tRNA synthetases is exemplified by the *Hs*TrpRS/*Bt*-tRNA^{Trp} complex structure (20). *Hs*TrpRS recognizes tRNA^{Trp} by its discriminator adenine in the tRNA acceptor stem as well as by specific interactions with tRNA base moieties in the anticodon loop (20) (see Figure 3, right panel).

Figure 4 shows the interplay among the three bases of the anticodon-loop, 5′-CCA-3′, of *Bt*-Trp-tRNA and *Hs*TrpRS as present in the crystal structure of the complex. The adenine is in H-bonding contact with D382 and possibly has

Table 1. Number of unpaired electrons ($N_{\alpha, i} - N_{\beta, i}$) on atom i, derived from the difference of alpha and beta spin densities ($\rho_{\alpha, i} - \rho_{\beta, i}$), and Mulliken partial charges for the Fe atoms of the [Fe$_4$S$_4$] cluster and S atoms of C266 (S$_1$) and C269 (S$_2$) of *Tm*TrpRS

| | $N_{\alpha, i} - N_{\beta, i}$ | | | | | |
	Fe$_1$	Fe$_2$	Fe$_3$	Fe$_4$	S$_1$	S$_2$
[Fe$_4$S$_4$]$^{1+}$	3.2	−3.0	3.2	−2.7	−0.1	0.1
[Fe$_4$S$_4$]$^{2+}$	3.1	3.1	−3.1	−3.1	−0.1	−0.2
*Tm*TrpRS–[Fe$_4$S$_4$]$^{1+}$	3.1	−3.0	3.2	−2.7	−0.1	0.1
*Tm*TrpRS–[Fe$_4$S$_4$]$^{2+}$	3.1	3.0	−3.0	−3.0	− 0.1	−0.2
*Tm*TrpRS(H244A)–[Fe$_4$S$_4$]1+	3.1	−3.0	3.2	−2.7	−0.1	0.1
*Tm*TrpRS(H244A)–[Fe$_4$S$_4$]$^{2+}$	3.1	3.0	−3.1	−3.0		
	Partial charges					
	Fe$_1$	Fe$_2$	Fe$_3$	Fe$_4$	S$_1$	S$_2$
[Fe$_4$S$_4$]$^{1+}$	0.43	0.39	0.43	0.38	−0.58	−0.58
[Fe$_4$S$_4$]$^{2+}$	0.40	0.37	0.40	0.38	−0.49	−0.45
*Tm*TrpRS–[Fe$_4$S$_4$]$^{1+}$	0.42	0.36	0.45	0.42	−0.56	−0.57
*Tm*TrpRS–[Fe$_4$S$_4$]$^{2+}$	0.38	0.34	0.41	0.43	−0.47	−0.46
*Tm*TrpRS(H244A)–[Fe$_4$S$_4$]$^{1+}$	0.42	0.36	0.41	0.42	−0.56	−0.55
*Tm*TrpRS(H244A)–[Fe$_4$S$_4$]$^{2+}$	0.37	0.35	0.38	0.43	−0.48	−0.42

[Fe$_4$S$_4$] refers to just the cubane with coordinating cysteine residues, *Tm*TrpRS assigns the *Tm*TrpRS model studied here and *Tm*TrpRS(H244A) refers to the same model where H244 was replaced by alanine. The structures were optimized with TPSS-D3/TZVP. Without explicit dispersion interactions the numbers are almost unchanged (see Supplementary Table S1), except for the H244A mutant where no BS solution could be converged. For each compound, different BS solutions were tested and the one corresponding to the lowest energy is given. The energetical deviation, however, never exceeded 2 kcal/mol.

π-interactions with F377; the cytosine following in 5′ direction forms H-bonds with the side chains of S378, K431, T427, and with the main chain oxygen atoms of G380 and R381. The 5′ cytosine is in H-bonding contact with the main chain NH-groups of T427 and G428, and its NH$_2$ group seemingly interacts with H387 (see Figure 4, left panel).

In order to identify possible interactions that would facilitate tRNA recognition in *Tm*TrpRS, the latter was superimposed with the complex structure of the human tRNA-synthetase and bovine tRNA^{Trp} using the structural alignment algorithm as implemented in Pymol (final r.m.s.d. 3.6 Å) (22). The superposition of *Tm*TrpRS and *Hs*TrpRS shows that in the given setting, a direct interaction between the tRNA anticodon loop and the [Fe$_4$S$_4$] cluster would be possible, with the 5′ cytosine pointing directly to it (see Figure 4, right panel). Apparently, this interaction could be supported by R224, which interacts with the Fe–S cluster, as discussed above (see Figure 4, right panel). Remarkably, the structural motif formed by the cubane moiety and the guanidinium group of R224 somewhat resembles the situation in *Hs*TrpRS, where H387 is most probably involved in anticodon recognition (see Figure 4, left panel). The superposition locates R224 at almost the same position as H387 in *Hs*TrpRS. Of course, structural alignment of the complex structure and *Tm*TrpRS cannot provide accurate atom coordinates, and due to sequence variations and deviations of main chain coordinates, clashes occur between the bovine tRNA molecule and the bacterial enzyme.

Base recognition and binding affinity

The superposition of *Tm*TrpRS and *Hs*TrpRS predicts that the [Fe$_4$S$_4$] cluster predominantly interacts with the 5′ cytosine of the anticodon. Therefore, the quantum mechanical study of tRNA anticodon recognition by *Tm*TrpRS presented here focuses on the binding of the 5′ base in order to render the computational problem tractable. One

unknown in this context is created by the fact that tRNA base modifications have not yet been characterized in *Thermotoga maritima*. Hence, the precise modes of interaction between *Tm*TrpRS and its cognate tRNA might differ from those studied here. However, unmodified cytosine as it appears in the anticodon loop of bovine tRNA^{Trp} is clearly the structurally simplest option and shall serve as an example to study how the peculiar Fe–S cluster could be involved in base recognition. Along with cytosine, 2-thiocytosine is considered as a known base modification in tRNA molecules in order to account for variations in the anticodon. Deciphering the correct 3′ base in a mRNA codon (corresponding to the 5′ base in the anticodon of the tRNA) is most important in the case of tryptophan as it is the only amino acid besides methionine that is represented by a single triplet in the genetic code (5′-UGG-3′); a wobble base in this position would therefore interfere with translational fidelity. As a consequence, tryptophanyl tRNA-synthetases have to recognize the correct 5′ base of the anticodon with great accuracy.

Cytosine and 2-thiocytosine, which are considered in this study, are represented by the base moieties together with their associated sugar residues. In order to avoid unrealistic rearrangements during structural optimization due to isolated charges and polar groups that would be compensated in the native enzyme, the hydroxy and bridging phosphate groups of the sugar molecules were removed. Only the 2′ O-group was included and neutralized by a proton (see Figure 5). Note that, for the sake of simplicity, the 5′ base models considered here are referred to by the names of their nucleobases (cytosine, 2-thiocytosine, uracil) consistently and not by the naming convention for the corresponding nucleosides (cytidine, 2-thiocytidine, uridine). This prevents frequent switching between nomenclatures.

The strength and specificity of the interaction between the 5′ base of the anticodon loop and the *Tm*TrpRS model was determined by calculating the interaction energy by
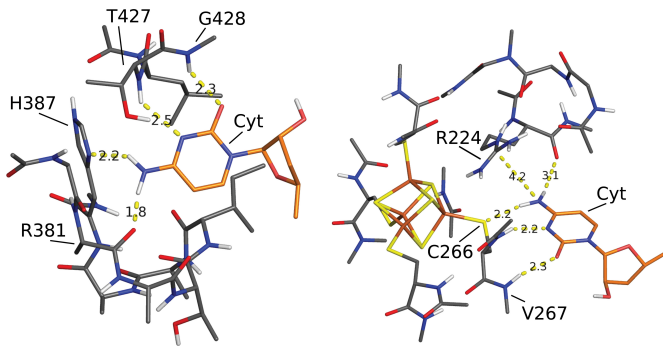
Figure 5. Interaction of *Hs*TrpRS and *Tm*TrpRS models with 5′ cytosine. Left panel: TPSS/TZVP optimized structure of the *Hs*TrpRS model and cytosine. Dispersion interactions were not considered explicitly. Right panel: TPSS/TZVP optimized structure of the *Tm*TrpRS model and cytosine. The oxidation state of the cubane was $[Fe_4S_4]^{1+}$. Dispersion interactions were not considered explicitly. H244 is not shown explicitly for the sake of better visibility of cytosine binding. Distances are given in angstroms. Element color code—white: H, red: O, gray/orange: C, blue: N, yellow: S, brown: Fe.

means of DFT. In order to judge whether the energies obtained are indicative of a significant interaction, analogous calculations were performed for the human *Hs*TrpRS model. DFT typically underestimates dispersion interactions and has therefore been complemented (*ad hoc*) by empirical correction terms to alleviate this shortcoming (44,45). Therefore, results are given with and without the explicit consideration of dispersion interactions throughout this paper.

At first, in order to resolve structural clashes between the ligand atoms (cytosine/2-thiocytosine) and the *Tm*TrpRS model that were caused by the simplistic structural-alignment approach, a constrained molecular mechanics optimization was performed, where only the bases and the attached sugar moieties were allowed to relax structurally. The pre-optimized coordinates were then used for full quantum mechanical optimizations.

The optimization of the *Hs*TrpRS/cytosine model corroborates and refines the information provided by the X-ray structure published by Shen *et al.* and shows that the 5′ cytosine is held in place by interactions with main chain atoms and H387 (see Figure 5, left panel). Specifically, the carbonyl group of cytosine interacts with the peptide-NH of G428. The imine-N of the pyrimidine ring is in H-bonding contact with the peptide-NH of T427 and the amino group of cytosine interacts with the peptide carbonyl group of R381 and with the imidazole side chain of H387. Evaluating the energetic contribution to tRNA recognition and binding reveals a rather weak interaction of −7 kcal/mol if dispersion interactions are not explicitly considered in the DFT approach (see Table 2). Dispersion interactions may play a role in non-specifically stabilizing the base in its position, for their consideration leads to a much stronger binding energy of −32 kcal/mol.

As the results in Table 2 show, cytosine would also bind to the bacterial enzyme *Tm*TrpRS, and with −8 kcal/mol even slightly more exothermic than to *Hs*TrpRS. Apparently, all polar groups of the cytosine molecule are in contact with atoms of the bacterial enzyme (see Figure 5, right panel).
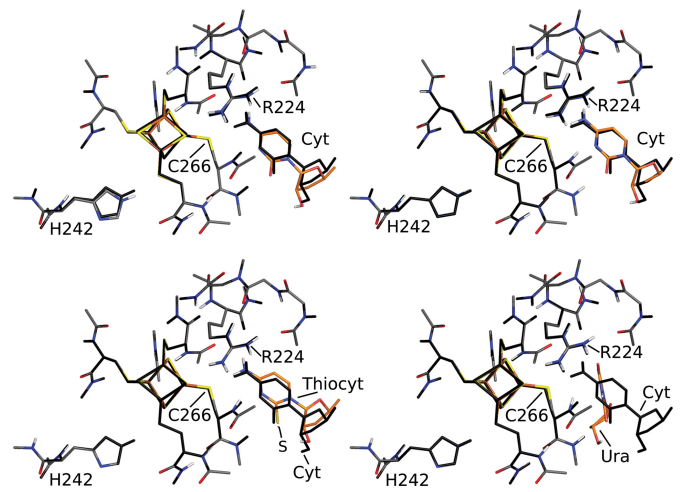
Figure 6. Strength and specificity of *Tm*TrpRS/anticodon base interactions. All structures were optimized with TPSS/TZVP or TPSS-D3/TZVP. To demonstrate structural changes, the complex *Tm*TrpRS/cytosine in oxidation state $[Fe_4S_4]^{1+}$ and optimized with TPSS/TZVP is always depicted in black. Upper left panel: *Tm*TrpRS/cytosine in oxidation state $[Fe_4S_4]^{2+}$, Cyt: cytosine. Upper right panel: *Tm*TrpRS/cytosine in oxidation state $[Fe_4S_4]^{1+}$ calculated with the explicit consideration of dispersion interactions (TZVP-D3/TZVP), Cyt: cytosine. Lower left panel: *Tm*TrpRS/2-thiocytosine in oxidation state $[Fe_4S_4]^{1+}$, Cyt: cytosine, Thiocyt: 2-thiocytosine. Lower right panel: *Tm*TrpRS/uracil in oxidation state $[Fe_4S_4]^{1+}$, Cyt: cytosine, Ura: uracil. Element color code—white: H, red: O, gray/orange: C, blue: N, yellow: S, brown: Fe.

Figure 7. Distribution of the Fe–S cluster sequence motif in TrpRS proteins among representatives of the bacterial and archaeal classes. The multiple sequence alignment was generated with Clustal Omega (52,53). The organisms *Thermotoga maritima* and *Aeropyrum pernix*, the TrpRS of which were studied quantum mechanically here, are highlighted in bold face. The cysteine residues of the consensus sequence 5′−C−X_6−C−X_2−C−3′ are highlighted and correspond to C259, C266 and C269 in *Tm*TrpRS.

The carbonyl group of cytosine forms an H-bond with the peptide-NH of V267. The imine-N of the pyrimidine ring interacts with the peptide-NH of C266, whereas the cluster-coordinating S atom of this residue is in H-bonding contact with the NH_2 group of the cytosine moiety. The N atom of the amino group apparently also interacts with the guanidinium group of R224 which, in turn, is held in position due to its interaction with the cubane cluster and surrounding peptide groups (Figure 5).

Table 2. Energies for the binding of cytosine, 2-thiocytosine and uracil models to TrpRS models from *Homo sapiens* (*Hs*TrpRS), *Thermotoga maritima* (*Tm*TrpRS) and *Aeropyrum pernix* (*Ap*TrpRS)

TrpRS model	Oxidation state	Ligand	Interaction Energy	
			—	Dispersion Corrections +
*Hs*TrpRS	—	cytosine	−6.5	−32.3
*Tm*TrpRS	$[Fe_4S_4]^{1+/2+}$	cytosine	−7.8/−7.9	−20.6/−20.6
*Tm*TrpRS(H244A)	$[Fe_4S_4]^{1+/2+}$	cytosine	−7.9/−8.0	−20.4/−20.0
*Tm*TrpRS	$[Fe_4S_4]^{1+/2+}$	2-thiocytosine	−9.5/−8.9	−21.7/−20.8
*Tm*TrpRS	$[Fe_4S_4]^{1+/2+}$	uracil	+2.0/+5.1	−5.6/−5.7
*Ap*TrpRS	$[Fe_4S_4]^{1+/2+}$	cytosine	−5.1/−4.7	−18.6/−16.7

Energies are given in kcal/mol. Structural optimizations were performed with TPSS/TZVP, with and without explicit dispersion corrections (TPSS-D3). For the archaeal and bacterial systems, ligand coordinates were derived by structural superposition with the *Hs*TrpRS/*Bt*-tRNA^{Trp} complex (PDB entry 2DR2). Due to sterical clashes, the ligand molecules were subsequently relaxed by force field optimization prior to quantum mechanical structural optimizations. For *Tm*TrpRS and *Ap*TrpRS the first number refers to the $[Fe_4S_4]^{1+}$ and the second one to the $[Fe_4S_4]^{2+}$ oxidation state.

Interestingly, the redox state of the Fe–S cluster has almost no effect on binding strength. The interaction energies are virtually identical for both forms considered, $[Fe_4S_4]^{1+}$ and $[Fe_4S_4]^{2+}$ (see Table 2). Structurally, a change in the redox state leads only to small deviations of the cluster atoms and to only a slight movement of the cytosine ligand (see Figure 6, upper left panel). As in the case of *Hs*TrpRS, dispersion interactions contribute significantly, but apparently non-specifically, to the interaction, because their explicit inclusion results only in minor changes of the ligand and *Tm*TrpRS model coordinates (see Figure 6, upper right panel). Replacing cytosine by 2-thiocytosine does not change this picture and the structural characteristics of the interaction remain the same. With approximately −9 kcal/mol, the interaction with the enzyme model is even slightly stronger than in the the case of cytosine (see Table 2). The cytosine/2-thiocytosine exchange does not alter the binding mode of the base moiety with respect to the *Tm*TrpRS model and therefore the binding partners for cytosine and 2-thiocytosine are the same (see Figure 6, lower left panel). Analogous results were obtained with explicit dispersion corrections (see Supplementary Figure S1). Due to the bigger van der Waals radius of sulfur as compared to oxygen, the distance between the thiocarbonyl group of 2-thiocytosine and the NH-group of V267 is increased compared to the situation with cytosine. This results in a slight shift of the 2-thiocytosine molecule (see Figure 6, lower left panel). With about −10 kcal/mol, the background generated by dispersion interactions remains almost unchanged (see Table 2).

Taken together, these data strongly suggest that the structural motif given by the Fe–S cluster and its surroundings provides a specific pattern of interaction partners for recognizing a cytosine or 2-thiocytosine 5′ base in the anticodon loop of a tRNA^{Trp} molecule.

Specificity of base recognition

As shown in the previous section, cytosine or 2-thiocytosine show clear affinities to *Tm*TrpRS that are comparable to cytosine binding in the *Hs*TrpRS/*Bt*-Trp-tRNA complex. This observation, however, directly raises the question of specificity, i.e. is the structural motif given by the $[Fe_4S_4]$ cluster and its embedding suited to discriminate between cognate and non-cognate bases? Given the structural characteristics of the bacterial enzyme, the binding of a purine base at the same position as cytosine or 2-thiocytosine seems sterically too demanding. Therefore, in an independent set of calculations, uracil was studied in order to determine whether or not its binding would be energetically sufficiently disfavored to allow for the specific recognition of cytosine/2-thiocytosine.

The results in Table 2 demonstrate that uracil shows no affinity for the *Tm*TrpRS model if no dispersion interactions are considered. For both oxidation states, the binding energetics is endothermic. As in the case of cytosine/2-thiocytosine, dispersion interactions contribute exothermically on the order of −10 kcal/mol. For the given model sizes, i.e. the number of interacting atoms, this amount apparently represents the non-specific interaction between the base moiety and the enzyme. The structurally optimized coordinates that are depicted in Figure 6 (lower right panel) show that fewer favorable and specific interactions between uracil and the bacterial synthetase (in comparison to cytosine/2-thiocytosine) could be identified and the optimization resulted in a reorientation of uracil with respect to cytosine due to sterical clashes, which explains the clear energetic penalty for the binding of uracil over cytosine/2-thiocytosine to *Tm*TrpRS. With explicit dispersion corrections, a similar displacement of the uracil ligand could be observed (see Supplementary Figure S1).

These data show that the putative binding motif formed by the Fe–S cubane and its direct protein environment could indeed serve as a pattern for the selective recognition of cytosine/2-thiocytosine in the 5′ position of the tRNA anticodon loop, as it provides both affinity and specificity of binding.

Significance and structural conservation of the R-FeS-H motif

Given that the Fe–S cluster in the TAB domain of *Tm*TrpRS with its characteristic embedding provides a specific and selective binding platform for cytosine/2-thiocytosine and could therefore be important for tRNA recognition and decoding, as demonstrated above, the next question becomes whether the complete structural unit, that could be characterized as an arginine/Fe–S cluster/histidine (R-FeS-H)

Figure 8. Conservation of residues in the assumed cluster environments of the TrpRS proteins shown in Figure 8. The multiple sequence alignment was generated with Clustal Omega (52,53). The organisms *Thermotoga maritima* and *Aeropyrum pernix*, the TrpRS of which were studied quantum mechanically here, are highlighted in bold face. The sequence positions explicitly considered in *Tm*TrpRS and *Ap*TrpRS are highlighted and correspond to R224 in *Tm*TrpRS, Q293 in *Ap*TrpRS and H244 in *Tm*TrpRS.
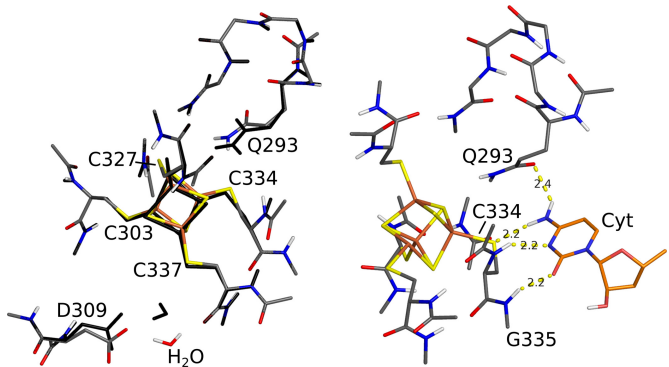

Figure 9. Potential binding of cytosine to *Ap*TrpRS. Left panel: Superposition of the starting model (black) extracted from the crystal structure of *Ap*TrpRS (PDB entry 3A04) with the TPSS/TZVP structure (color). Right panel: TPSS/TZVP structure of the *Ap*TrpRS/cytosine complex. For better visibility of cytosine binding, D309 and the crystal water are not shown. Distances are given in angstroms. Element color code—white: H, red: O, gray/orange: C, blue: N, yellow: S, brown: Fe.

motif, is conserved and thus strictly required for the assumed function.

Phylogenetic analysis done by taking TrpRS genes of representatives from all bacterial and archaeal classes into account shows that both R224 and H242 are conserved in bacteria which presumably contain an Fe–S cluster in their TrpRS enzymes, according to the occurrence of the cysteine-rich consensus sequence, $5'-C-X_6-C-X_2-C-3'$ (see Figure 7). In order to determine whether or not H242 is required for base recognition, a *Tm*TrpRS model was generated that carries an H242A mutation and for which cytosine binding energies were determined by DFT calculations. Interestingly, despite being in H-bond contact with the S atom of C269, the substitution of H242 does not change the partial charges or the spin polarization on the cluster atoms significantly (see Table 1), in both oxidation states considered. Only the

partial charge on the S atom of C269 becomes more positive due to the missing H-bond otherwise formed with the imidazole side chain. This seems to be compensated for by a decrease of positive charge on the iron atom Fe_3 of the cluster (compare Figure 2). Moreover, binding energy and binding mode for cytosine are almost unchanged in the H242A mutant model (see Table 2 and Supplementary Figure S1).

Therefore, the remarkably strong conservation of H242 in the bacterial TrpRS proteins phylogenetically considered here seems not to be directly related to enzyme/base interaction and could be due to its involvement in other processes such as cluster bio-synthesis and assembly/insertion.

Phylogenetic analysis shows that H242 is not conserved in archaeal representatives of the TrpRS enzyme class (see Figure 8). Notably, only one crystal structure of an archaeal TrpRS carrying a cubane cluster—from the organism *Aeropyrum pernix* K1—has been published to date (*Ap*TrpRS) (8) (PDB entry 3A04). The TAB domain region with its Fe–S cluster is shown in Figure 9 (left panel). The cluster is again coordinated by an amino acid that is surface-exposed, as in *Tm*TrpRS, but it is a glutamine residue (Q293) instead of an arginine. Similarly, H242 that is conserved in bacterial species is missing in *Ap*TrpRS and replaced by an aspartate, D309. Due to this exchange, the carboxyl group of D309 is not in direct contact with the cluster. According to the refinement of Tsuchiya *et al.* (8), a crystal water is located between D309 and the S atom of C337 (see Figure 9, left panel). Because of this apparent hydrogen bonding network, D309 and the water molecule were included in a minimal model of *Ap*TrpRS (Figure 9, left panel). In order to study the binding of cytosine, Q293 (which according to structural alignment adopts the place of R224 in *Tm*TrpRS) was considered for model construction. Furthermore, main chain atoms of the direct protein environment of the cubane and Q293 were included.

The binding energetics documented in Table 2 show that cytosine also binds to the model of the archaeal enzyme *Ap*TrpRS, albeit with a somewhat lower affinity (around −5 kcal/mol). The binding mode of the cytosine moiety is comparable to that which was observed in *Tm*TrpRS, with hydrogen bonds between the NH_2 group of cytosine and the S atom of a cluster-coordinating cysteine (C337) (see Figure 9, right panel). The same NH_2 group also interacts with the carbonyl-O atom of the amide group of Q293. The 3′-N of the pyrimidine ring forms a hydrogen acceptor for the peptide NH group of C337; and the pyrimidine's carbonyl-O interacts with the peptide NH group of G335. The interaction between a 5′ cytosine and *Ap*TrpRS calculated here is corroborated by an experimental study of Tsuchiya and Hasegawa, who investigated the binding of tRNA molecules that were obtained by an *in vitro* transcription approach, to overexpressed *Ap*TrpRS. They were able to demonstrate that mutating the cytosine residues in the tryptophan anticodon, $5'-CCA-3'$, severely impedes aminoacylation, whereas adenine mutants show wild-type activity (8). The potential hydrogen bonding network formed by C337, D309, and the crystal water was not conserved during structural optimization and seems not to be important for cluster function. Rather, the water molecule changed position and formed a hydrogen bond with the peptide carbonyl-O of C337 (see Figure 9, left panel).

Apparently, recognition of a 5′ base could be accomplished by different choices of amino acid side chains around an Fe–S cluster. As can be seen from the alignment in Figure 8, not all bacterial TrpRS analyzed here feature an arginine residue at the sequence position corresponding to R224 in *Tm*TrpRS. In the enzymes from *Chlorobium ferrooxidans* and *Aquifex aeolicus*, arginine is replaced by lysine. Due to comparable side chain lengths of arginine and lysine, it seems very likely that lysine can fulfill the role of R224 in *Tm*TrpRS by interacting with the NH_2 group of cytosine. In the archaeal candidates, this sequence position does not show any signs of conservation. Instead, Q293 serves the function of recognizing the 5′ base in *Aeropyrum pernix* as discussed above. Interestingly, Q293 is replaced by histidine in *Methanopyrus kandleri* and *Methanocaldococcus jannaschii*.

Secondary adaptation or remnant of early life?

Iron–sulfur clusters can be considered an early invention in the emergence of life. Iron–sulfur proteins are generally highly conserved and appear at crucial positions in the metabolism of cells. The iron–sulfur world hypothesis even claims that the evolution of a metabolism preceded the invention of a genetic code (18). On the other hand, RNA molecules have always been considered attractive candidates for early life because they can at the same time possess enzymatic activity, i.e. as ribozymes, and carry genetic information. An iron–sulfur enzyme like *Tm*TrpRS that is involved in deciphering nucleic-acid encoded information and protein biosynthesis, therefore could echo in a present-day organism a possible convergence of iron–sulfur and ribonucleic acid-based origins of life. As tempting as such an explanation might be, it must be noted that the occurrence of iron–sulfur clusters in *Tm*TrpRS or *Ap*TrpRS could be a mere secondary adaptation to extreme environmental conditions.

It is difficult to answer this question by phylogeny based on a single gene, because the high degree of lateral gene transfer inside and between the kingdoms *bacteria* and *archaea* (46–48) prevents us from clearly determining whether a TrpRS gene with an iron–sulfur cluster binding domain represents the chronologically older variant. The alignment shown here does not provide an exhaustive overview of bacterial and archaeal TrpRS gene distribution. However, it does reveal a number of intriguing details. Interestingly, the distribution of putative Fe–S clusters is not linked to certain bacterial classes. For example, in the class gamma-proteobacteria, the prototypical enterobacterium *E. coli* carries a TrpRS lacking the characteristic cysteine-rich consensus sequence indicative of a $[Fe_4S_4]$ cluster, $5'-C-X_6-C-X_2-C-3'$, whereas in the order *Legionellales* and *Pseudomonadales*, TrpRS that most probably contain the cluster occur. Examples are the medically relevant species *Legionella pneumophila* and *Pseudomonas stutzeri* that clearly do not belong to the group of thermophilic bacteria (see Figure 7). In order to prevent false positive results, comparative modeling was applied to confirm that the sequences are in accordance with the protein fold of an Fe–S cluster-containing TAB domain (see Supplementary Figure S2 and Table S2). Given the subtle structural interplay in
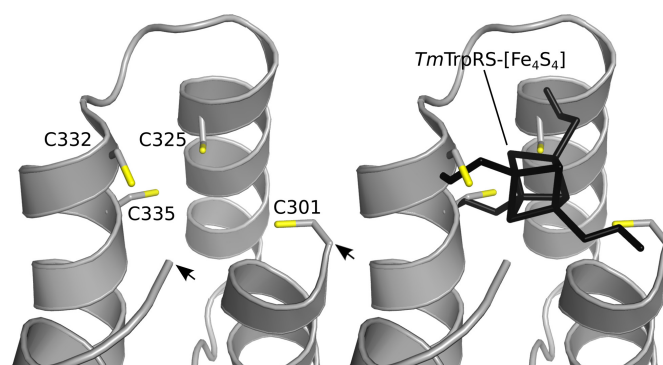

Figure 10. A missing $[Fe_4S_4]$ cluster in TrpRS from *Pyrococcus horikoshii*? Left panel: Putative TAB domain of *Ph*TrpRS with the four free cysteine residues highlighted (PDB entry 3JXE (51)). The black arrows assign the strand break in the crystal structure. Right panel: $[Fe_4S_4]$ cluster and co-ordinating cysteine side chains (black) from the *Tm*TrpRS structure 2G36 positioned in the TAB domain of *Ph*TrpRS by superimposing the crystal structures of *Ph*TrpRS and *Tm*TrpRS.

the recognition of tRNA by TrpRS, an Fe–S cluster in the TAB domain seems plausible under extreme environmental conditions due to the structural rigidity the cubane scaffold provides. However, the occurrence of Fe–S cluster-free TrpRS in thermophilic bacteria like *Thermus thermophilus* or archaea like *Sulfolobus solfataricus* (see Figure 7) shows that this motif is not strictly required for translational fidelity.

A curiosity in this respect is the parasitic or symbiotic relationship between the archaebacteriae *Nanoarchaeum equitans* and *Ignicoccus hospitalis* (49). The extremophile organism *N. equitans* lives near hydrothermal vents. It requires the presence of a host organism because its reduced genome lacks several enzymes required to sustain essential pathways. Interestingly, a Trp-tRNA-synthetase is present (like all gene products required for replication, transcription and translation (49,50)) but does not contain a single cysteine residue and is therefore unlikely to contain an Fe–S cluster. In contrast, potential host organisms existing in the same habitat, like *Ignicoccus hospitalis*, carry an Fe–S cluster consensus sequence (see Figure 7). The strict colocalization of both organisms therefore further calls into question the claim that Fe–S clusters represent a mere adaptation for Trp-tRNA aminoacylation under extreme conditions.

Clearly, sequence homology alone has to be considered carefully. This becomes especially apparent in the example of *Pyrococcus horikoshii* whose TrpRS protein (*Ph*TrpRS) features an Fe–S cluster consensus sequence (see Figure 7). Surprisingly, a crystal structure of the synthetase that has been solved (51) does not show an iron–sulfur cluster. However, closer inspection of the TAB domain region and comparison with the *Tm*TrpRS structure shows that in *Ph*TrpRS, four cysteine residues are located in the same way as in *Tm*TrpRS and should therefore be suited to accommodate and coordinate an Fe–S cubane (see Figure 10). The *Ph*TrpRS crystal structure features a strand break and the quality of the electron density in the TAB domain is rather low (not shown); this might be due to an increased flexibility induced by the free four cysteine residues. An explanation for the missing $[Fe_4S_4]$ cluster could be that the overexpression system (*E. coli* strain BL21) that was used for isolating

*Ph*TrpRS does not provide the proper *P. horikoshii*-specific Fe–S cluster maturation and insertion machinery and that the native protein indeed carries a cubane cluster. According to the phylogenetic and structural analysis provided by Dong *et al.*, the origin of TrpRS lies in the kingdom archaea and was acquired subsequently by bacterial ancestors of extant organisms (51). A convergence of iron–sulfur and RNA-based life exemplified by Fe–S cluster-containing TrpRS proteins seems therefore plausible.

Interestingly, an Fe–S cluster consensus motif is also present in the TrpRS of the bacterium *Desulforudis audaxviator* that was discovered in 2008 in a South African gold mine at 2.8-km depth and lives in complete autarky in this isolated ecosystem (54). The high sequence homology with *Tm*TrpRS (52% identical residues) allows for reliable comparative modeling and shows that an Fe–S cluster should be present even in *Da*TrpRS (see Supplementary Figure S2 and Table S2).

CONCLUSION

Attributing a role to a well-known structural unit that appears in an unprecedented context can be difficult. As shown here, computational chemistry can complement biochemical experiments in the interpretation of experimental findings. Theoretical modeling of iron–sulfur proteins is challenging because transition metal complexes require a quantum mechanical description that results in significant computational effort, limiting thereby the model sizes that can be treated. Nevertheless, the data provided here clearly support the idea that iron–sulfur clusters in prokaryotic and archaeal TrpRS proteins are able to provide affinity and specificity for deciphering the 5′ base in the anticodon loop of their cognate tRNA^{Trp} molecules by providing a stable structural pattern that assures affinity to the base and at the same time specificity by producing an energetic and structural penalty against the binding of non-cognate base moieties. The models considered here suggest that it is mostly the thiolate-S atom of a cluster-coordinating cysteine that directly interacts with the base. Clearly, an $[Fe_4S_4]$ cluster is perfectly suited to provide this functionality in a structurally restrained setting that is robust against structural disturbances and changes in the redox conditions of the cell. In comparison to the typical role Fe–S clusters play in electron transport, this function seems not to be of major importance for the specific recognition of anticodon bases, as no significant differences could be observed for the two redox states analyzed.

The occurrence of $[Fe_4S_4]$ cluster-free TrpRS representatives in thermophilic organisms demonstrates that the problem of anticodon recognition can be solved even under extreme conditions without additional cofactors, and the question remains why the energetically demanding process of cluster formation is maintained in extant organisms. It might very well be that iron–sulfur clusters in the context of tRNA decoding are remnants of an early convergence of RNA and iron–sulfur-based life forms in a setting that provided a plethora of metal sulfides, as do hydrothermal vents.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Tagawa,K. and Arnon,D.I. (1962) Ferredoxins as electron carriers in photosynthesis and in the biological production and consumption of hydrogen gas. *Nature*, **195**, 537–543.
2. Valentine,R.C., Jackson,R.L. and Wolfe,R.S. (1962) Role of ferredoxin in hydrogen metabolism of Micrococcus lactilyticus. *Biochem. Biophys. Res. Commun.*, **7**, 453–456.
3. Martius,C. (1937) *Z. Physiol. Chem.*, **247**, 104–110.
4. Martius,C. (1939) *Z. Physiol. Chem.*, **257**, 29–42.
5. Speyer,J.F. and Dickman,S.R. (1956) On the mechanism of action of aconitase. *J. Biol. Chem.*, **220**, 193–208.
6. Sunita,S., Tkaczuk,K.L., Purta,E., Kasprzak,J.M., Douthwaite,S., Bujnicki,J.M. and Sivaraman,J. (2008) Crystal structure of the Escherichia coli 23S rRNA:m5C methyltransferase RlmI (YccW) reveals evolutionary links between RNA modification enzymes. *J. Mol. Biol.*, **383**, 652–666.
7. Han,G.W., Yang,X.-L., McMullan,D., Chong,Y.E., Krishna,S.S., Rife,C.L., Weekes,D., Brittain,S.M., Abdubek,P., Ambing,E. *et al.* (2010) Structure of a tryptophanyl-tRNA synthetase containing an iron-sulfur cluster. Acta Cryst. F, **66**(Pt 10), 1326–1334.
8. Tsuchiya,W. and Hasegawa,T. (2009) Molecular recognition of tryptohan tRNA by tryptophanyl-tRNA synthetase from Aeropyrum pernix K1. *J. Biochem.*, **145**, 635–641.
9. Fromme,J.C. and Verdine,G.L. (2003) Structure of a trapped endonuclease III-DNA covalent intermediate. *EMBO J.*, **22**, 3461–3471.
10. Liu,H., Rudolf,J., Johnson,K.A., McMahon,S.A., Oke,M., Carter,L., McRobbie,A.M., Brown,S.E., Naismith,J.H. and White,M.F. (2008) Structure of the DNA repair helicase XPD. *Cell*, **133**, 801–812.
11. Wolski,S.C., Kuper,J., Hänzelmann,P., Truglio,J.J., Croteau,D.L., Van Houten,B. and Kisker,C. (2008) Crystal structure of the FeS cluster-containing nucleotide excision repair helicase XPD. *PLoS Biol.*, **6**, e149.
12. Boon,E.M., Livingston,A.L., Chmiel,N.H., David,S.S. and Barton,J.K. (2003) DNA-mediated charge transport for DNA repair. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 12543–12547.
13. Ren,B., Duan,X. and Ding,H. (2009) Redox control of the DNA damage-inducible protein DinG helicase activity via its iron-sulfur cluster. *J. Biol. Chem.* **284**, 4829–4835.
14. Yeeles,J.T.P., Cammack,R. and Dillingham,M.S. (2009) An iron-sulfur cluster is essential for the binding of broken DNA by AddAB-type helicase-nucleases. *J. Biol. Chem.*, **284**, 7746–7755.
15. Engstrom,L.M., Partington,O.A. and David,S.S. (2012) An iron-sulfur cluster loop motif in the Archaeoglobus fulgidus uracil-DNA glycosylase mediates efficient uracil recognition and removal. *Biochemistry (Mosc.)*, **51**, 5187–5197.
16. Rouault,T.A. (2012) Biogenesis of iron-sulfur clusters in mammalian cells: new insights and relevance to human disease. *Dis. Model. Mech.*, **5**, 155–164.
17. Stehling,O. and Lill,R. (2013) The role of mitochondria in cellular iron-sulfur protein biogenesis: mechanisms, connected processes, and diseases. *Cold Spring Harb. Perspect. Biol.*, **5**, a011312.
18. Wächtershäuser,G. (1992) Groundworks for an evolutionary biochemistry: The iron-sulphur world. *Prog. Biophys. Mol. Biol.*, **58**, 85–201.
19. Ahlrichs,R., Bär,M., Häser,M., Horn,H. and Kölmel,C. (1989) *Chem. Phys. Lett.*, **162**, 165–169.
20. Shen,N., Guo,L., Yang,B., Jin,Y. and Ding,J. (2006) Structure of human tryptophanyl-tRNA synthetase in complex with tRNATrp reveals the molecular basis of tRNA recognition and specificity. *Nucleic Acids Res.*, **34**, 3246–3258.
21. Chen,V.B., Arendall,W.B., Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S. and Richardson,D.C.(2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta. Cryst. D*, **66**, 12–21.

22. The PyMOL Molecular Graphics System, Version 1.6.0.0, Schrödinger, LCC.

23. Halgren,T.A. (1996) Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comp. Chem.*, **17**, 490–519.

24. Hanwell,M.D., Curtis,D.E., Lonie,D.C., Vandermeersch,T., Zurek,E. and Hutchison,G.R. (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.*, **4**, 17.

25. Tao,J., Perdew,J.P., Staroverov,V.N. and Scuseria,G.E. (2003) Climbing the density functional ladder: nonempirical meta-generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.*, **91**, 146401.

26. Weigend,F. and Ahlrichs,R. (2005) Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.*, **7**, 3297–3305.

27. Grimme,S., Antony,J., Ehrlich,S. and Krieg,H. (2010) A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.*, **132**, 154104.

28. Stiebritz,M.T. and Reiher,M. (2009) Theoretical study of dioxygen induced inhibition of [FeFe]-hydrogenase. *Inorg. Chem.*, **48**, 7127–7140.

29. Stiebritz,M.T., Finkelmann,A.R. and Reiher,M. (2011) Oxygen coordination to the active site of Hmd in relation to [FeFe] hydrogenase. *Eur. J. Inorg. Chem.*, **2011** 1163–1171.

30. Bruska,M.K., Stiebritz,M.T. and Reiher,M. (2011) Regioselectivity of H cluster oxidation. *J. Am. Chem. Soc.*, **133**, 20588–20603.

31. Eichkorn,K., Weigend,F., Treutler,O. and Ahlrichs,R. (1997) Auxiliary basis sets for main row atoms and transition metals and their use to approximate coulomb potentials. *Theor. Chem. Acc.*, **97**, 119–124.

32. Klamt,A and Schüürmann,G. (1993) COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2* 799–805.

33. Peitsch,M.C. (1995) Protein modeling by E-mail. *Nat. Biotech.*, **13**, 658–660.

34. Guex,N. and Peitsch,M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.

35. Schwede,T., Kopp,J., Guex,N. and Peitsch,M.C. (2003) SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.*, **31**, 3381–3385.

36. Arnold,K., Bordoli,L., Kopp,J. and Schwede,T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, **22**, 195–201.

37. Kiefer,F., Arnold,K., Künzli,M., Bordoli,L. and Schwede,T. (2009) The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.*, **37**, D387–D392.

38. Benkert,P., Biasini,M. and Schwede,T. (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, **27**, 343–350.

39. Adams,P.D., Afonine,P.V., Bunkóczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.-W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst. D*, **66**, 213–221.

40. Noodleman,L. and Norman,J.G. (1979) The Xα valence bond theory of weak electronic coupling. Application to the low-lying states of $Mo_2Cl_8^{4-}$. *J. Chem. Phys.*, **70**, 4903–4906.

41. Noodleman,L. (1981) Valence bond description of antiferromagnetic coupling in transition metal dimers. *J. Chem. Phys.*, **74**, 5737–5743.

42. Noodleman,L., Post,D. and Baerends,E. (1982) Symmetry breaking and ionization from symmetry equivalent inner shells and lone pairs in Xα theory. *Chem. Phys.*, **64**, 159–166.

43. Noodleman,L. and Davidson,E.R. (1986) Ligand spin polarization and antiferromagnetic coupling in transition metal dimers. *Chem. Phys.*, **109**, 131–143.

44. Antony,J. and Grimme,S. (2006) Density functional theory including dispersion corrections for intermolecular interactions in a large benchmark set of biologically relevant molecules. *Phys. Chem. Chem. Phys.*, **8**, 5287–5293.

45. Grimme,S. (2006) Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.*, **27**, 1787–1799.

46. Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A. *et al.* (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima. *Nature*, **399**, 323–329.

47. Nesbo,C.L., L'Haridon,S., Stetter,K.O. and Doolittle,W.F. (2001) Phylogenetic analyses of two "archaeal" genes in thermotoga maritima reveal multiple transfers between archaea and bacteria. *Mol. Biol. Evol.*, **18**, 362–375.

48. DeBoy,R.T., Mongodin,E.F., Emerson,J.B. and Nelson,K.E. (2006) Chromosome evolution in the Thermotogales: large-scale inversions and strain diversification of CRISPR sequences. *J. Bacteriol.*, **188**, 2364–2374.

49. Huber,H., Hohn,M.J., Rachel,R., Fuchs,T., Wimmer,V.C. and Stetter,K.O. (2002) A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature*, **417**, 63–67.

50. Das,S., Paul,S., Bag,S.K. and Dutta,C. (2006) Analysis of Nanoarchaeum equitans genome and proteome composition: indications for hyperthermophilic and parasitic adaptation. *BMC Genomics*, **7**, 186.

51. Dong,X., Zhou,M., Zhong,C., Yang,B., Shen,N. and Ding,J. (2010) Crystal structure of Pyrococcus horikoshii tryptophanyl-tRNA synthetase and structure-based phylogenetic analysis suggest an archaeal origin of tryptophanyl-tRNA synthetase. *Nucleic Acids Res.*, **38**, 1401–1412.

52. Goujon,M., McWilliam,H., Li,W., Valentin,F., Squizzato,S., Paern,J. and Lopez,R. (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.*, **38**, W695–W699.

53. Sievers,F., Wilm,A., Dineen,D., Gibson,T.J., Karplus,K., Li,W., Lopez,R., McWilliam,H., Remmert,M. and Söding,J. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, **7**, 539.

54. Chivian,D. (2008) Environmental genomics reveals a single-species ecosystem deep within Earth. *Science*, **322**, 275–278.