

Crystal Structures of the Human and Fungal Cytosolic Leucyl-tRNA Synthetase Editing Domains: A Structural Basis for the Rational Design of Antifungal Benzoxaboroles

Elena Seiradake¹, Weimin Mao², Vincent Hernandez², Stephen J. Baker², Jacob J. Plattner², M.R.K. Alley² and Stephen Cusack^{1*}

¹European Molecular Biology Laboratory, Grenoble Outstation
6 rue Jules Horowitz, BP 181,
38042 Grenoble Cedex 9, France

²Anacor Pharmaceuticals
Incorporated, 1020 East
Meadow Circle, Palo Alto,
CA 94303, USA

Received 17 February 2009;
received in revised form
5 April 2009;
accepted 20 April 2009
Available online
6 May 2009

Leucyl-tRNA synthetase (LeuRS) specifically links leucine to the 3' end of tRNA^{leu} isoacceptors. The overall accuracy of the two-step aminoacylation reaction is enhanced by an editing domain that hydrolyzes mischarged tRNAs, notably ile-tRNA^{leu}. We present crystal structures of the editing domain from two eukaryotic cytosolic LeuRS: human and fungal pathogen *Candida albicans*. In comparison with previous structures of the editing domain from bacterial and archeal kingdoms, these structures show that the LeuRS editing domain has a conserved structural core containing the active site for hydrolysis, with distinct bacterial, archeal, or eukaryotic specific peripheral insertions. It was recently shown that the benzoxaborole antifungal compound AN2690 (5-fluoro-1,3-dihydro-1-hydroxy-1,2-benzoxaborole) inhibits LeuRS by forming a covalent adduct with the 3' adenosine of tRNA^{leu} at the editing site, thus locking the enzyme in an inactive conformation. To provide a structural basis for enhancing the specificity of these benzoxaborole antifungals, we determined the structure at 2.2 Å resolution of the *C. albicans* editing domain in complex with a related compound, AN3018 (6-(ethylamino)-5-fluorobenzo[c][1,2]oxaborol-1(3*H*)-ol), using AMP as a surrogate for the 3' adenosine of tRNA^{leu}. The interactions between the AN3018-AMP adduct and *C. albicans* LeuRS are similar to those previously observed for bacterial LeuRS with the AN2690 adduct, with an additional hydrogen bond to the extra ethylamine group. However, compared to bacteria, eukaryotic cytosolic LeuRS editing domains contain an extra helix that closes over the active site, largely burying the adduct and providing additional direct and water-mediated contacts. Small differences between the human domain and the fungal domain could be exploited to enhance fungal specificity.

© 2009 Elsevier Ltd. All rights reserved.

Keywords: leucyl-tRNA synthetase; editing; *Candida albicans*; benzoxaborole; X-ray crystallography

Edited by J. Doudna

*Corresponding author. E-mail address: cusack@embl.fr.

Present address: E. Seiradake, Division of Structural Biology, Wellcome Trust Center for Human Genetics, University of Oxford, Roosevelt Drive, Headington OX3 7BN, UK.

Abbreviations used: LeuRS, leucyl-tRNA synthetase; aaRS, aminoacyl-tRNA synthetase; IleRS, isoleucyl-tRNA synthetase; ValRS, valyl-tRNA synthetase; CP1, connective peptide 1; PDB, Protein Data Bank; I2ae, insertion 2; I3ae, insertion 3; I4ae, insertion 4; ESI, electrospray ionization; PEG, polyethylene glycol; ESRF, European Synchrotron Radiation Facility; TEV, tobacco etch virus.

Introduction

Aminoacyl-tRNA synthetases (aaRS) are evolutionary ancient enzymes that play an essential role in protein synthesis.¹ They catalyze the specific attachment of amino acids to their cognate tRNAs in a two-step reaction: activation of amino acid with ATP to form enzyme-bound aminoacyl-AMP (with release of pyrophosphate), and transfer of amino acid moiety to cognate tRNA, releasing AMP and charged tRNA. There are two distinct classes of aaRS (classes I and II) characterized by very different

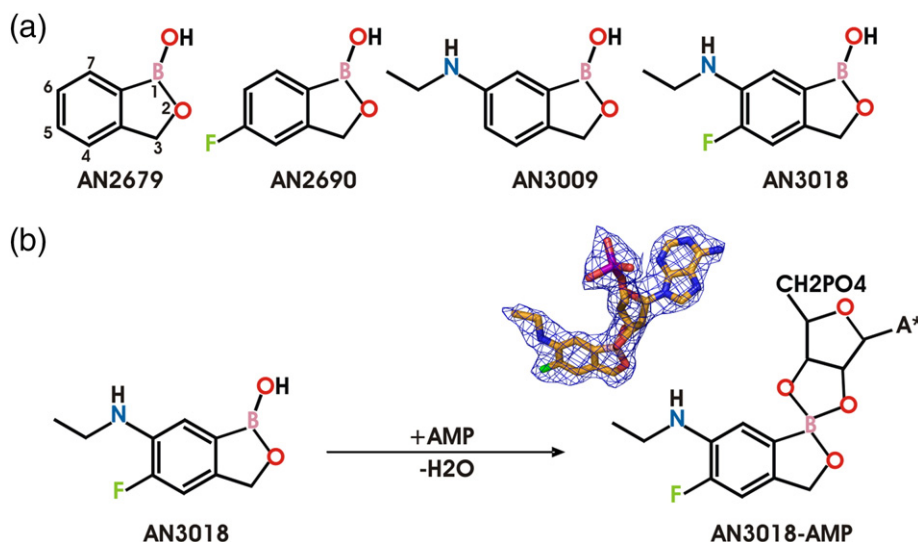


Fig. 1. (a) Schematic representations of benzoxaborole compounds AN2679, AN2690, AN3009, and AN3018. Noncarbon atom colors are as follows: fluorine (green), boron (pink), phosphorus (purple), oxygen (red), and nitrogen (blue). (b) In the presence of AMP, AN3018 is converted within the active site of CP1 into the AN3018-AMP adduct. Electron density, contoured at 1σ , is shown in blue and was calculated with phases from the refined protein/water model before inclusion of the adduct in the model [brown sticks; noncarbon atom colors as in (a)]. In the accompanying schematic diagram, A* represents adenosine moiety.

conserved core catalytic domains.² During evolution, additional less conserved domains and extensions have been added, resulting in the architectures of extant aaRS appearing extremely diverse.

To achieve protein synthesis that is sufficiently accurate for cell viability, the accurate transfer of amino acids to their cognate tRNAs must have an error rate of less than 1:10,000. To increase fidelity, many aaRS that have to discriminate between chemically similar amino acids have acquired a proofreading (editing) active site in an additional editing domain to hydrolyze mischarged tRNAs. The combined aminoacylation and editing active sites enhance overall accuracy by deploying complementary binding specificities, a mechanism that has been likened to a “double sieve.”³ Thus, for instance, isoleucyl-tRNA synthetase (IleRS) accommodates the smaller valine in the isoleucine-binding pocket of the aminoacylation active site, but compensates by sterically excluding the larger cognate isoleucine from its editing site, which thus only hydrolyzes mischarged val-tRNA^{Ile}. The so-called posttransfer editing activity of this kind exists in three closely related class Ia enzymes: leucyl-tRNA synthetase (LeuRS), IleRS, and valyl-tRNA synthetase (ValRS) (reviewed by Jakubowski and Goldman⁴), as well as class II prolyl-tRNA synthetase,⁵ threonyl-tRNA synthetase,⁶ alanyl-tRNA synthetase,⁷ and phenylalanyl-tRNA synthetase.⁸ The class Ia editing enzymes possess a conserved inserted domain of about 200 amino acids

[sometimes referred to as connective peptide 1 (CP1)] that are responsible for posttransfer editing. CP1 is connected to the bulk of the synthetase by a flexible β -ribbon and undergoes significant rotation during the aminoacylation/editing cycle.⁹ Although the class Ia editing domain must have appeared early in evolution before the branching of the universal phylogenetic tree, there are two distinct insertion points for CP1 within the primary sequence of extant class Ia synthetases. One is shared by all IleRS, ValRS, and eukaryotic/archaeal LeuRS; the other is found only in bacterial LeuRS.¹⁰ Interestingly, crystallographic analyses show that, despite having the same insertion point, the CP1 domain of archeal LeuRS is rotated by $\sim 180^\circ$ compared to IleRS and ValRS, giving it the same orientation as that of the CP1 domain of bacterial LeuRS, which has a different insertion point.^{10–12} In the case of LeuRS, interactions of peripheral regions of CP1 with the bulk of the synthetase are likely to be important during the aminoacylation/editing cycle^{13,14} and may depend on kingdom-specific insertions depending on the point of insertion of the domain.

The structural basis for the recognition and hydrolysis of mischarged tRNA^{Leu} by LeuRS CP1 has been revealed by crystal structures of LeuRS from the bacteria *Thermus thermophilus* in complex with a nonhydrolyzable posttransfer editing substrate¹⁵ and tRNA^{Leu} in the editing conformation,⁹ as well as of *Escherichia coli* LeuRS CP1 domain alone, in

Fig. 2. (a) Distance tree based on CP1 structural homology as calculated with programme SHP²⁴. (b) Structure based sequence alignment for class Ia synthetase CP1 domains. The colors used in the ribbon diagrams correspond to those of the bars above the sequence alignment. Red, conserved core elements; gray, eukaryotic insertion 1 (I1e); yellow, archeal/eukaryotic I2ae; green, archeal/eukaryotic I3ae; blue, archeal/eukaryotic I4ae. Light green bars below the sequences indicate conserved motifs. Stars mark strictly conserved residues discussed in the text. C.a., *C. albicans*; H.s., *Homo sapiens*; P.h., *Pyrococcus horikoshii* (PDB entry 1WK8); E.c., *E. coli* (PDB entry 2AJH); T.t., *T. thermophilus* (PDB entry 1H3N).

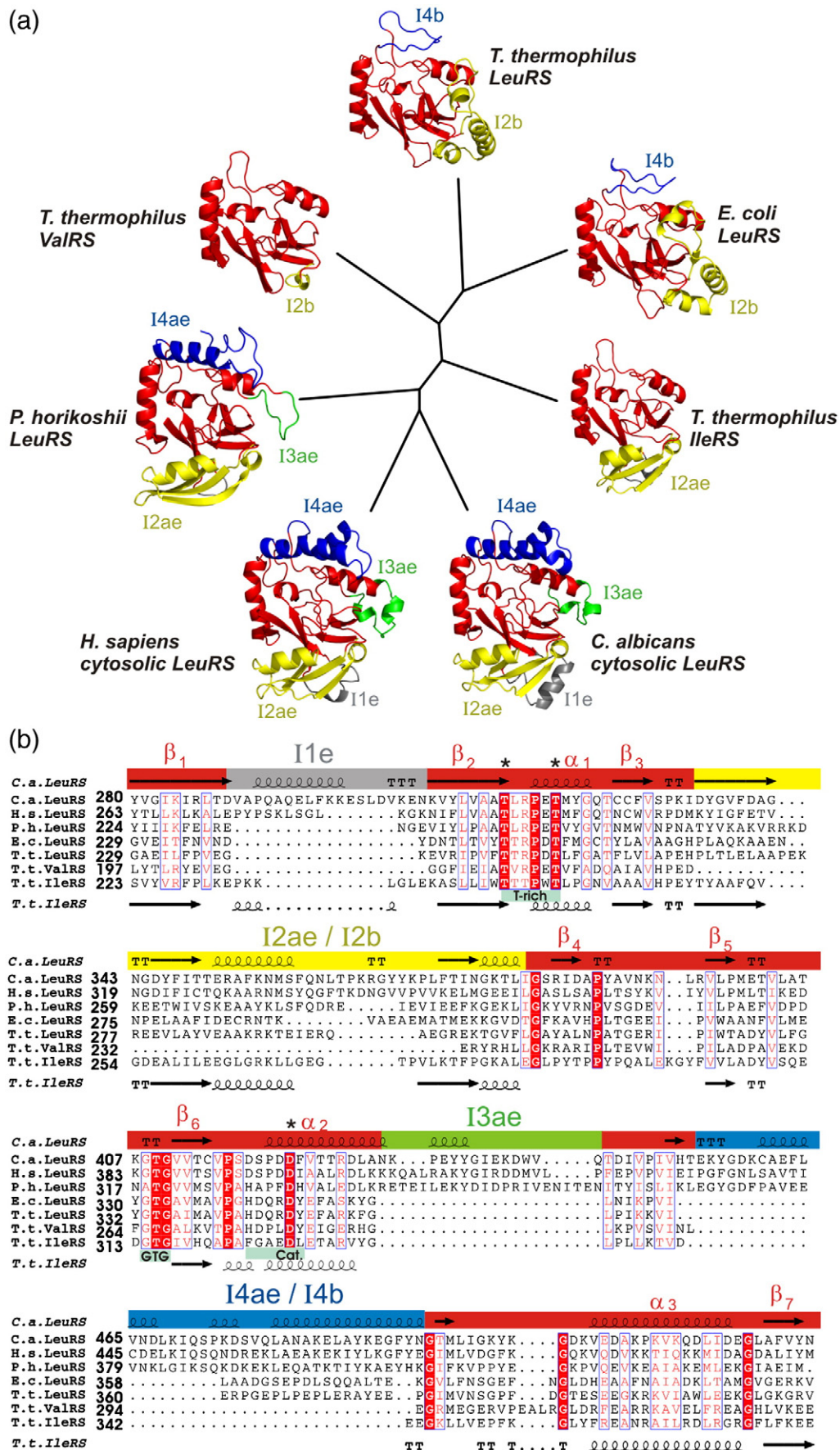


Fig. 2 (legend on previous page)

complex with the noncognate amino acids isoleucine and methionine.¹⁶ These structures, together with mutagenesis and functional studies, reveal that posttransfer editing depends critically on a strictly conserved aspartic acid (also conserved in ValRS and IleRS editing domains) that interacts with the α -amino group of the noncognate amino acid and positions the substrate for hydrolysis by a water molecule.¹⁵ The side chain of the noncognate amino acid (e.g., isoleucine, methionine, or norvaline) is accommodated in a rigid pocket that excludes cognate leucine because its γ -branched side chain sterically clashes with a strictly conserved threonine (Thr252 in *T. thermophilus* LeuRS).^{15,17} Mutation of this threonine allows manipulation of the amino acid specificity of the editing activity,^{18,19} for example, for protein engineering applications.²⁰

Because of their importance as essential players in the translation of the genetic code, aaRS have long been considered as valid targets for antibiotic drugs, and a number of clinically interesting anti-synthetase compounds have recently been identified.²¹ These include a family of recently discovered boron-based benzoxaborole compounds that have broad-spectrum antifungal activity (Fig. 1a). One of these, AN2690 (5-fluoro-1,3-dihydro-1-hydroxy-1,2-benzoxaborole), is in clinical trials as a topical treatment for onychomycosis.^{22,23} AN2690 covalently binds, via its boron atom, to the 2' and 3' hydroxyl groups of the 3'-terminal ribose of tRNA^{leu} in the editing active site of LeuRS. By irreversibly (on the timescale of normal turnover) locking the bound tRNA in its posttransfer editing conformation onto the protein, the enzyme is inactivated and protein synthesis is interrupted.²³ An important current research goal is improving the potency of benzoxaborole com-

pounds, as well as enhancing specificity towards particular pathogen classes (e.g., fungi, bacteria, or eukaryotic parasites). To this end, we solved the crystal structure of the cytosolic LeuRS CP1 from the pathogenic fungus *C. albicans* in its apo form and in complex with an AMP adduct of another benzoxaborole, AN3018 (6-(ethylamino)-5-fluorobenzo[*c*] [1,2]oxaborol-1(3*H*)-ol) (Fig. 1a). In addition, we solved the crystal structure of human cytosolic LeuRS CP1 domain. Together, these results will facilitate the design of new antifungals with increased potency. We compare these first structures of eukaryotic LeuRS CP1s to the equivalent domains of known structures from bacteria and archaea, revealing kingdom-specific peripheral insertions around a highly conserved catalytic core.

Results

The crystal structure of the human cytosolic LeuRS CP1 domain was determined at 3.25 Å resolution using selenomethionine-labeled protein (Fig. 2). The domain crystallized in space group *P*6₅ with two copies in the asymmetric unit. The two copies form a tight face-to-face dimer, with the two N-termini crossing over to bind in the active site of the opposite molecule (Fig. 3). The N-terminus includes a (seleno) methionine that was artificially introduced during cloning. Interestingly, this selenomethionine binds in the noncognate-amino-acid-binding pocket of the editing domain as observed, for example, in the complex of the *E. coli* LeuRS editing domain with methionine [Protein Data Bank (PDB) entry 2AJH].¹⁶ This competition likely explains the fact that although human CP1 domain was crystallized in the presence

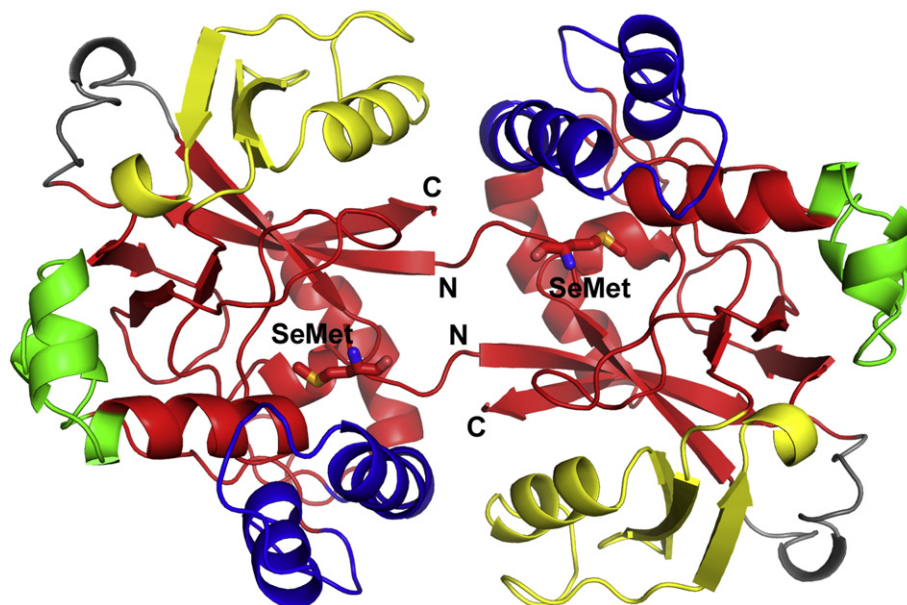


Fig. 3. Ribbon diagram showing the crystallographic homodimer of human cytosolic LeuRS CP1. Colors are as in Fig. 2. The N-terminus of each chain extends into the active site of the neighboring molecule, where an extra selenomethionine remaining after TEV protease cleavage (shown in stick representation in red) binds close to the noncognate-amino-acid-binding pocket.

of the RNA trinucleotide Cyt-Cyt-Ade and a benzoxaborole, which were expected to form an adduct in the active site, no density could be attributed to these ligands. Similar face-to-face dimers were present in other crystal forms of this construct (data not shown). Because the active site is inaccessible in the dimer, making this construct unsuitable for ligand binding studies, we expressed an alternative construct lacking the four N-terminal residues involved in artificial dimerization. However, upon expression, this construct was insoluble and not pursued further. The fungal (*C. albicans*) cytosolic LeuRS CP1 domain crystallized in space group $P2_1$ (four chains per asymmetric unit) and its structure were determined at 2.9 Å resolution (Fig. 2). This domain did not form tight dimers. Crystallographic details for all structures are summarized in Table 1.

The human and fungal CP1 domains have a sequence identity of 43%, and their structures are very similar: 221 residues (88%) of the human and *Candida* apo structures can be structurally aligned with an RMSD of 1.22 Å on C^α positions. Like all LeuRS CP1 domains examined to date, they consist of a highly conserved core comprising about 120 residues with additional variable inserted elements (Fig. 2). The core residues form a seven-stranded β -sheet and three α -helices. To facilitate description, we number the core secondary structure elements according to their relative positions in the primary sequence: β_1 - β_2 - α_1 - β_3 - β_4 - β_5 - β_6 - α_2 - α_3 - β_7 (Fig. 2b). Conserved active-site features include the “threonine-rich region,” which contains two strictly conserved threonines—one at the end of strand β_2 ,

which is important for hydrolysis mechanism (T316 in *C. albicans*, T297 in humans, and T247 in *T. thermophilus* and *E. coli*), and one in helix α_1 for amino acid side-chain discrimination (T321 in *C. albicans*, T302 in humans, and T252 in *T. thermophilus* and *E. coli*) (both marked with a star in the alignment in Fig. 2b); the “GTG region” (in between strands β_5 and β_6), which is a flexible loop that is important for binding adenine 76 of the tRNA; and a strictly conserved aspartic acid in helix α_2 (D422 in *C. albicans*, D398 in humans, D347 in *T. thermophilus*, and D345 in *E. coli*; marked with a star in Fig. 2b) involved in binding the α -amino group of the noncognate amino acid and in positioning the substrate for catalysis.¹⁵

In addition to the conserved core, eukaryotic cytosolic LeuRS CP1 domains contain four major insertions (Fig. 2). Eukarya-specific insertion 1 (denoted insertion 1e) occurs between core strands β_1 and β_2 . It comprises 13 residues in humans and 20 residues in *C. albicans*, where it forms a three-turn α -helix (Fig. 2). Insertion 2 (I2ae) is a large insertion in eukaryotic and archeal LeuRS. It is a 48-residue β - β - α - β element that is inserted between the core strands β_3 and β_4 . In eukaryotes, it provides both direct and water-mediated hydrogen bonds to the active-site “GTG region,” presumably adding stability to this important region. Interestingly, I2ae exists also in bacterial IleRS CP1, which groups closer to eukaryotic/archeal LeuRS than to bacterial LeuRS in the structure-based phylogenetic tree (Fig. 2). Insertion 3 (I3ae) is a small region that is present in eukaryotes and archaea. It is inserted after core helix

Table 1. Crystallographic data collection and refinement statistics

	Human LeuRS CP1	<i>C. albicans</i> LeuRS CP1	<i>C. albicans</i> LeuRS CP1 in complex with AN3018-AMP
Protein sample	Selenomethionine derivative	Native	Native
Space group	$P6_5$	$P2_1$	$P2_1, 2_1$
Cell dimensions a, b, c (Å)	94.5, 94.5, 148.8	101.1, 43.0, 122.2	43.0, 58.0, 97.6
Cell dimensions α, β, γ (°)	90, 120, 90 $\beta = 120$	90, 107.6, 90 $\beta = 107.6$	90, 90, 90
Resolution range (Å) (last shell) (Å)	29.2–3.25 (3.25–3.30)	28.8–2.90 (3.02–2.90)	49.9–2.20 (2.20–2.26)
Beamline	ESRF ID14-4	ESRF ID23-1	ESRF ID14-1
Wavelength (Å)	0.9795	1.0723	0.934
Detector	ADSC Quantum Q315r	ADSC Quantum Q315r	ADSC Quantum Q210
Completeness (last shell) (%)	99.8 (99.5)	94.8 (71.5)	99.7 (99.5)
R_{merge} (last shell)	0.149 (0.661)	0.149 (0.399)	0.161 (54.8)
$I/\sigma I$ (last shell)	7.46 (1.98)	8.55 (2.92)	8.16 (2.26)
R -factor (last shell)	0.202 (0.300)	0.242 (0.292)	0.185 (0.288)
R_{free} (last shell)	0.240 (0.250)	0.301 (0.394)	0.262 (0.383)
Number of protein (non-H) atoms	3922 (two chains)	7760 (four chains)	1974 (one chain)
Number of water molecules	0	0	247
Number of ligand atoms	0	0	36
Number of reflections used in refinement	11,198	20,554	12,300
Fraction of reflections used for R_{free} calculation	0.05	0.05	0.05
Solvent content (%)	63.2	43.5	41.1
Mean B -value (Å ²)	59.4	28.4	23.0
Ramachandran plot MOLPROBITY			
Favored (%)	92.4	96.6	97.6
Allowed (%)	99.8	99.8	100
RMSD from ideal values			
Bond distances (Å)	0.010	0.007	0.009
Angles (°)	1.318	1.023	1.153

α_2 and packs against core strands β_4 and β_5 . Insertion 4 (I4ae) precedes core helix α_3 , and its C-terminal-long helix forms part of the active site. This helix has relatively high *B*-values in the apo form of *C. albicans* cytosolic LeuRS CP1, suggesting that it is not rigidly fixed. In the ligand-bound structures of *C. albicans* LeuRS CP1 (see the text below), this helix is better ordered and closes over the active site, where it binds the ligand.

The benzoxaborole AN2690 is a small boron-based molecule that forms a long-lived adduct with tRNA^{leu} (or AMP) within the active site of LeuRS CP1 domain, thus inhibiting enzyme turnover and stopping protein synthesis²³ (Fig. 1). Previous crystallographic studies using bacterial LeuRS from *T. thermophilus* have shown that the AN2690–AMP adduct contacts many highly conserved residues, suggesting that the binding mode is likely to be similar in all LeuRS CP1 domains. In order to help design better antifungal benzoxaboroles by improving their antifungal potency, we solved the crystal structure of a CP1 domain from a fungal cytosolic LeuRS, that of *C. albicans*, to 2.2 Å resolution in complex with AMP and AN3018 (Fig. 4). The complex crystallized in space group *P*₂₁₂₁ with one protein chain per asymmetric unit. AN3018 and its des-fluoro derivative AN3009 (6-(ethylamino)benzo[*c*][1,2]oxaborol-1(3*H*)-ol) inhibit *C. albicans* cytosolic LeuRS with IC₅₀ values of 38 and 4.3 μM, respectively. This is in comparison with the parental

benzoxaborole AN2679, which has an IC₅₀ of 5.7 μM. Although the ethylamine substitution at position 6 on the core benzoxaborole structure (Fig. 1) does not add additional potency, it does permit further exploration of the volume and interactions available for optimizing inhibitor design.

The *C. albicans* crystal structure confirms that adduct binding to the LeuRS CP1 editing domain is similar in eukaryotes and bacteria (Figs. 4 and 5). Conserved and structurally similar interactions are shown in Fig. 5a and b. Notably, a strictly conserved threonine (T316 in *C. albicans* and T247 in *T. thermophilus*), which normally would bind to the amino acid backbone oxygen of mischarged tRNA^{leu}, binds to the oxygen of AN3018 benzoxaborole rings. Further conserved interactions involve backbone atoms of nonconserved residues, such as the main-chain oxygens of D418 (*C. albicans*) and H343 (*T. thermophilus*), which both contact the AN2690–AN3018 fluorine atom. Fixed waters bridge between the AN2690–AN3018 adduct and the protein active sites (one direct bridge in *T. thermophilus* and two direct bridges in *C. albicans*). Notably, a cavity between the AN3018 adduct and the long helix of insertion I4ae in *C. albicans* is filled by a network of fixed waters (Fig. 5a and c). Compared to AN2690, the extra ethylamine group of AN3108 forms two additional contacts with active-site residues (Fig. 5d). These are a hydrogen bond with the backbone oxygen of D418 via the ethylamine nitrogen atom

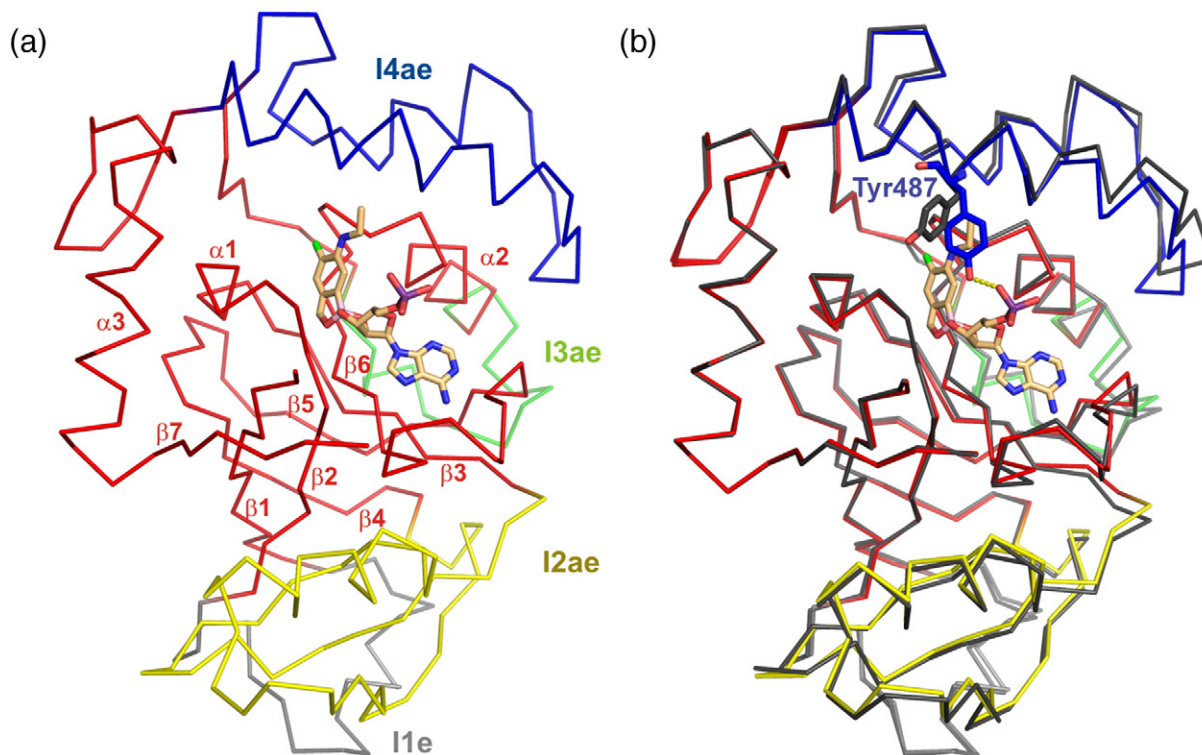


Fig. 4. (a) The structure of *C. albicans* cytosolic LeuRS CP1 in complex with AN3018–AMP at 2.2 Å resolution. The protein backbone is shown with colors as in Fig. 2. AN3018–AMP is depicted as sticks with colors as in Fig. 1a. (b) The structure of *C. albicans* cytosolic LeuRS CP1+AN3018–AMP as in (a), superposed onto the corresponding native (unliganded) structure shown in black. Tyrosine 487, shown in stick representation, changes rotamer upon AN3018–AMP adduct binding.

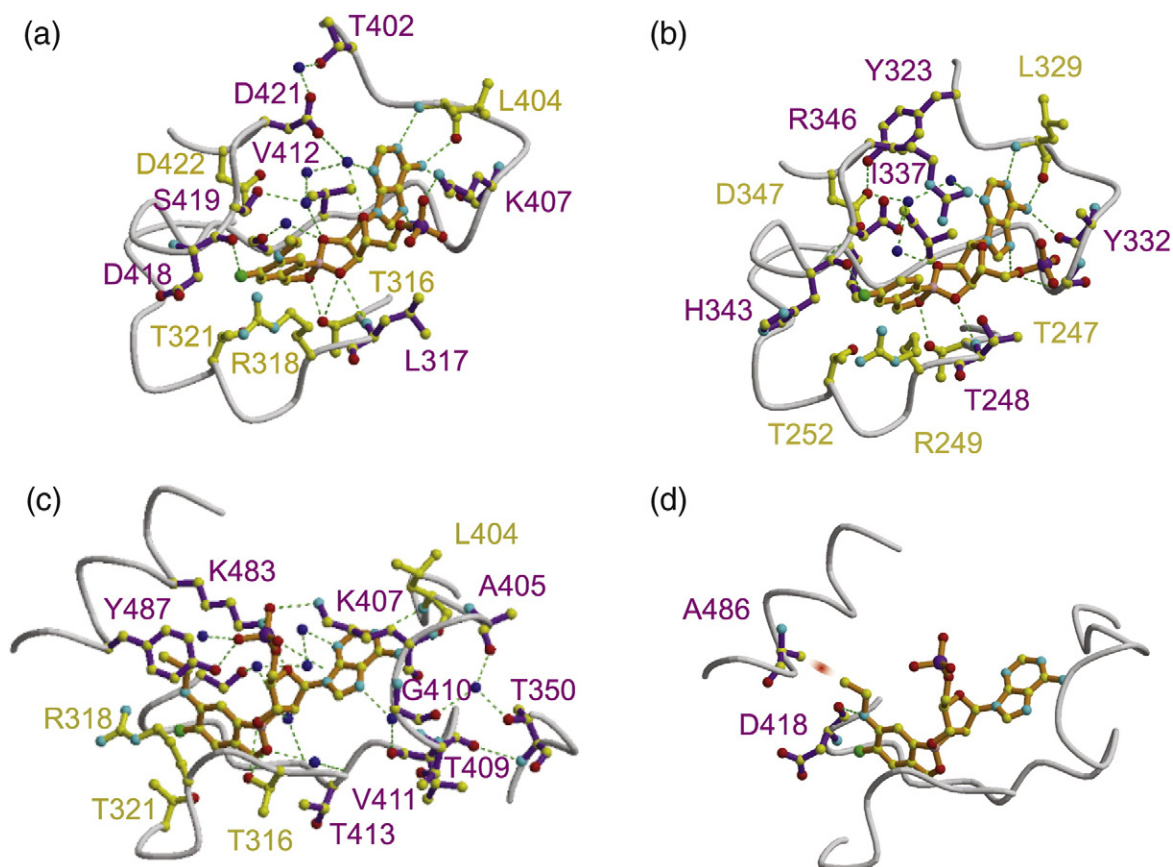


Fig. 5. (a) Selected active-site residues of *C. albicans* cytosolic LeuRS CP1 in complex with AN3018–AMP shown in stick representation. Residues that are conserved in the bacterium *T. thermophilus* are shown in yellow; nonconserved residues are shown in purple. Putative hydrogen bonds are shown as dotted lines. (b) Selected active-site residues of *T. thermophilus* LeuRS CP1 in complex with AN2690–AMP shown in stick representation. The orientation of the active site corresponds to that of the *C. albicans* structure in (a). Conserved residues are shown in yellow; nonconserved residues are shown in purple. Putative hydrogen bonds are shown as dotted lines. (c) As in (a), but showing a different view of the *C. albicans* cytosolic LeuRS CP1 active site and highlighting other interactions. Colors as in (a). (d) As in (c); however, only the interactions provided by the AN3018-specific ethylamine group are shown. Hydrophobic contact between the ligand and alanine 486 is indicated as a fuzzy red cloud.

and a hydrophobic interaction of the terminal CH₃ group with the side chain of A486.

Discussion

The three class Ia aaRS—LeuRS, IleRS, and ValRS—each contain structurally similar editing domains (CP1) that specifically recognize and hydrolyze mischarged tRNAs. In this work we present crystal structures of eukaryotic cytosolic LeuRS CP1 domains from *C. albicans* and humans. Together with previously determined structures from bacterial¹⁰ and archaeal LeuRS,¹¹ this completes the structural characterization of LeuRS editing domains from representatives of each kingdom. The core structure of the editing domains from all IleRS, ValRS, and LeuRS is highly conserved (Fig. 2), reflecting the common mode of posttransfer editing substrate binding and the common mechanism of hydrolysis. However, peripheral insertions vary considerably, depending on the enzyme and the kingdom. ValRS has the most compact CP1 domain

consisting of the minimum conserved core together with a very short insertion I2b (Fig. 2). Indeed, the structure-based phylogenetic tree shown in Fig. 2 shows ValRS CP1 between two main branches: that of the bacterial LeuRS CP1 domains, which contain two insertions I2b and I4b, and that of the IleRS and archaeal/eukaryotic cytosolic LeuRS CP1 domains, which both contain a quite distinct insertion I2ae. The eukaryotic and archaeal LeuRS CP1 domains also contain the unique insertions I3ae and I4ae (Fig. 2).

These observations suggest the following evolutionary scenario, starting from the idea that an ancestral class Ia synthetase that did not discriminate between isoleucine, valine, and leucine existed. This enzyme likely acquired at some stage a free-standing CP1-like domain that may have resembled the minimal ValRS CP1 seen today. It is indeed possible that a ValRS-type editing activity first arose to eliminate the mischarged polar amino acid threonine from otherwise hydrophobic positions in proteins. Subsequent gene duplication and divergence would eventually give rise to the first IleRS (addition of the I2ae insertion) and then archaeal/

eukaryotic-type LeuRS (addition of unique LeuRS insertions), both of which have the I2ae insertion and both of which have the editing domain inserted in the same position as in ValRS.¹⁰ Editing in eubacterial LeuRS must have evolved by a different pathway, as the bacterial CP1 is inserted in a different position into the body of the enzyme¹⁰ and the domain has a distinct I2b insertion. The two possibilities are as follows: firstly, on a second occasion, a free-standing CP1-like domain was inserted into a duplicated ancestral enzyme, but at a different point in the primary sequence (and this LeuRS variant survived in eubacteria), or, secondly, through some intragene rearrangement, the point of insertion of the CP1 domain in LeuRS was switched at some stage in the common ancestor of all bacteria. Distinguishing these possibilities or other more complicated scenarios would require a much more detailed analysis of the comparative phylogeny of LeuRS, IleRS, and ValRS in the three kingdoms, taking into account the probable different evolutionary history of the editing domain compared with the catalytic anti-codon binding and other tRNA-binding domains (e.g., the C-terminal domain), as well as the coevolution with tRNA, particularly the acquisition of the long variable arm of tRNA^{leu}. This analysis is beyond the scope of the current work, although some aspects have been previously discussed.^{11,25} In particular, the structure of the archeal LeuRS showed that, despite having the same point of insertion as ValRS and IleRS (and despite being distinct from bacterial LeuRS), its CP1 rotational orientation is similar to that of bacterial LeuRS, being 180° different from that of ValRS and IleRS.¹¹ The bacteria-like orientation of archeal LeuRS CP1 is thought to be favored by the fact that the protruding C-terminal loop of the helix in insertion I2ae and the entire insertion I4ae both mask regions involved in anchoring CP1 in the alternative IleRS/ValRS-like orientation.¹¹ These two features are present also in the human and fungal cytosolic LeuRS CP1, suggesting an archeal-like orientation of the editing domain in full-length eukaryotic LeuRS proteins.

LeuRS is the target of a new family of benzoxaborole antifungals.²⁶ One member of this family, AN2690 (Fig. 1), is being developed as a topical treatment for onychomycosis.²⁶ Crystallographic studies using bacterial LeuRS showed that AN2690 covalently binds in the noncognate-amino-acid-binding pocket of the CP1 domain, with its boron atom covalently linked to both hydroxyl groups of the 3'-terminal ribose of the tRNA. This AN2690-tRNA adduct locks the tRNA on the enzyme in the editing conformation, thereby inhibiting the enzyme, resulting in the shutdown of protein synthesis.²³ The crystallographic studies also showed that AMP can act as a surrogate for tRNA and also forms a tightly binding adduct with AN2690.²³ In order to improve the potency of AN2690 as an antifungal agent and to initiate the design of other species-specific benzoxaborole drugs, we solved the crystal structure of a target eukaryotic cytosolic

LeuRS CP1 domain, that of the pathogenic fungus *C. albicans*, to 2.2 Å resolution in complex with AMP and AN3018 (Figs. 4 and 5). AN3018 is a derivative of AN2690 that also inhibits LeuRS and fungal growth. AN3018-AMP binds to *C. albicans* cytosolic LeuRS CP1 in essentially the same way that AN2690-AMP binds to bacterial LeuRS, but provides additional interactions due to its extra ethylamine substitution at position 6 (Fig. 5).

An analysis of spontaneously emerged mutations in the presence of AN2690 and subsequent site-specific mutagenesis in the yeast *Saccharomyces cerevisiae* cytosolic LeuRS CP1 active site revealed point mutations that provide resistance to AN2690.^{23,27} The residues involved correspond to T316, L317, R318, T321, C328, G408, K407, S419, and D421 in *C. albicans*. They either contact AN3018-AMP directly or indirectly via water, such as K407 or S419, or play a role in maintaining active-site integrity (for instance, D421). Interestingly, mutations in two nonconserved residues (corresponding to T350 and T413 in *C. albicans*) do not provide resistance to AN2690, but still affect editing activity.²⁷ T413 is part of the active site, and T350 is located on a loop in insertion I2ae, which stabilizes the active-site GTG loop (Figs. 1b and 5).

The eukaryotic cytosolic LeuRS editing active site contains a specific long second helix insertion, I4ae. Upon binding AN3018-AMP, this helix moves 2.9 Å towards the active site, closing over the ligand and providing additional contacts (Fig. 4). The cavity resulting from the I4ae lid-like helix is large enough to accommodate AN3018-AMP plus a network of fixed water molecules within the active site. These bound water molecules bridge between the ligand and the protein, mainly along the helix of I4ae. This difference between bacteria and eukaryotes makes it particularly interesting for the development of novel antibacterial benzoxaboroles. These could exploit the fact that the bacterial LeuRS CP1 active site is more open than the cytosolic eukaryotic one and is incompatible with eukaryotic I4ae long helix residues.

AN3018 differs from AN2690 in that it contains an additional ethylamine group (Fig. 1). This additional group points into the active-site hydrophobic pocket and forms two additional contacts compared to AN2690-AMP. These are a hydrogen bond to the backbone oxygen of D418 via the nitrogen atom and hydrophobic contacts to the side chain of A486. A486 is one of the two residues that differ between the active sites of humans and fungi: A486 and L451 are both replaced by bulkier isoleucines in humans, thus reducing the size of the hydrophobic pocket. To rationally design a better fungal-specific compound, one could add additional groups to the inward-pointing regions of the AN3018 moiety, in expectation that these will reduce affinity for the human enzyme while still fitting into the fungal homologue.

To conclude, we have presented the first eukaryotic LeuRS CP1 crystal structures to date, which give insight into fundamental questions concerning LeuRS evolution. Moreover, we showed how a benzoxaborole antifungal agent (AN3018) binds to

its target, the fungal LeuRS cytosolic CP1. With the apo forms of the corresponding domains in humans and fungi, we provide key information for the rational optimization of these novel antibiotics.

Materials and Methods

Synthesis of AN3018

5-Fluorobenzo[*c*][1,2]oxaborol-1(3*H*)-ol²² (20 g, 132 mmol) was added to 75 ml of fuming nitric acid at -50°C . The resulting suspension was stirred for 30 min then warmed briefly to -10°C before being added to crushed ice. A light-yellow precipitate was filtered and air dried, then recrystallized from EtOAc to obtain 16 g of 5-fluoro-6-nitrobenzo[*c*][1,2]oxaborol-1(3*H*)-ol as a white solid with a melting point at $162\text{--}167^{\circ}\text{C}$. To a solution of 5-fluoro-6-nitrobenzo[*c*][1,2]oxaborol 1(3*H*)-ol (2 g, 10.2 mmol) in methanol (50 ml) was added 1 ml of Raney nickel (50% slurry in H_2O). The resulting mixture was stirred, then 3 ml of hydrazine monohydrate was added dropwise in portions over 2 h (gas evolution). After the reaction has been completed, the mixture was filtered through Celite and evaporated to a light-brown solid. This material was dissolved in a minimum amount of EtOAc and filtered to obtain 1.1 g of 6-amino-5-fluorobenzo[*c*][1,2]oxaborol-1(3*H*)-ol as a white solid with a melting point of $142\text{--}145^{\circ}\text{C}$. To a solution of 6-amino-5-fluorobenzo[*c*][1,2]oxaborol-1(3*H*)-ol (250 mg, 1.5 mmol) in acetonitrile (25 ml) were added sodium bicarbonate (280 mg, 3.3 mmol) and iodoethane (146 μl , 1.8 mmol). The resulting mixture was stirred overnight at room temperature and monitored by thin-layer chromatography. As an incomplete reaction was observed, iodoethane was subsequently titrated until all starting material had been consumed. The reaction was then quenched with water and extracted with EtOAc. The product was purified on a silica column eluting with 1:2 EtOAc/HXN to yield 82 mg of a tan solid of AN3018 with a melting point of $116\text{--}121^{\circ}\text{C}$. The ^1H NMR assignments at 300 MHz in $\text{DMSO-}d_6$ are as follows: δ 8.96 (s, 1H), 6.97–7.08 (m, 2H), 5.25 (bs, 1H), 4.81 (s, 2H), 3.07–3.11 (q, 2H), 1.14–1.19 (t, 1H). Mass spectrometry electrospray ionization (ESI) ($-$) 194 [M–H].

Synthesis of AN3009

This compound, a yellow semisolid, was synthesized using the same route as described for AN3018 starting from the commercially available benzo[*c*][1,2]oxaborol-1(3*H*)-ol. The ^1H NMR assignments at 300 MHz in $\text{DMSO-}d_6$ are as follows: δ 8.91 (s, 1H), 7.05–7.08 (d, 1H), 6.84 (s, 1H), 6.68–6.71 (d, 1H), 5.44 (bs, 1H), 4.81 (s, 2H), 2.96–3.05 (q, 2H), 1.12–1.17 (t, 1H). Mass spectrometry ESI ($-$) 176 [M–H] ESI(+) 178 [M+H].

Cloning, expression, and biochemical inhibition of *C. albicans* cytosolic LeuRS

The entire coding sequence of cytosolic LeuRS from *Candida albicans* ATCC 90028 was amplified using the primers GAGAGCTAGCCATCATCATCATCAT-CACATGAGTGGTCTGTTACTTTTGAAAAGAC and AGAGCTCGAGTTATTCGACATTTTAATAAAGATAC-CAGCG and cloned into the NheI and XhoI sites of pET21b. The 5' primer introduced an N-terminal His-tag, which was used to purify the protein. Benzoxaborole inhibitors were

assayed by preincubation with *C. albicans* cytosolic LeuRS (0.1 nM) in the presence of 16 μM baker's yeast tRNA (Roche), 50 mM Hepes–KOH (pH 7.6), 5 mM MgCl_2 , 60 mM KCl, 0.02% (wt/vol) bovine serum albumin, 1 mM DTT, and 33 μM [^{14}C]leucine (Perkin-Elmer; specific activity, 11.1 GBq/mmol) at 30°C . After 20 min, 4 mM ATP was added to each well in a 96-well nitrocellulose membrane filter plate (Millipore Multiscreen HTS, MSHAN4B50) to start the reaction. The reaction was stopped after 15 min by the addition of 10% (wt/vol) trichloroacetic acid to each well. Each well was then washed three times with 100 μl of 5% trichloroacetic acid. Filter plates were then dried under a heat lamp, and the precipitated [^{14}C]leucine tRNA^{leu} was quantified by liquid scintillation counting using a Wallac MicroBeta Trilux model 1450 counter. One-third serial dilutions of the inhibitors were used, and each dilution was performed in triplicate. Percent inhibition was used by Prism to determine IC_{50} values.

Production and crystallization of *C. albicans* cytosolic LeuRS CP1

A DNA fragment coding for *C. albicans* cytosolic LeuRS CP1 (residues 280–530) was amplified using primers GAGCCATGGGTCCACAAGAATATGTTGG and GAACCTCGAGGTAAACAAAAGCAAGACCTTCATC and cloned into the NcoI–XhoI sites of pET21d (Novagen). This fragment does not contain any CUG codon that would be read as leucine by *E. coli*, rather than serine as in *C. albicans*, due its uniquely altered genetic code.²⁸ The resulting construct contained a C-terminal 6 \times histidine tag and was expressed in the soluble fraction of *E. coli* strain 10G (a derivative of DH10B) at 37°C . The cells were resuspended in lysis buffer [20 mM Tris–HCl (pH 7.5), 300 mM NaCl, and 20 mM imidazole; Boehringer Complete ethylenediaminetetraacetic-acid-free protease inhibitor cocktail] and lysed by sonication. The cell lysate was centrifuged for 30 min at 25,000g, and the supernatant was loaded on a Ni Sepharose column (Fast Flow; GE Healthcare). Protein bound to the resin was washed with wash buffer 1 [20 mM Tris–HCl (pH 7.5), 1 M NaCl, and 20 mM imidazole] and wash buffer 2 [20 mM Tris–HCl (pH 7.5), 200 mM NaCl, and 70 mM imidazole], and eluted with elution buffer [20 mM Tris–HCl (pH 7.5), 200 mM NaCl, and 400 mM imidazole]. The eluted fractions were concentrated to 1–2 mg/ml and loaded onto a Superdex200 column (GE Healthcare). The running buffer contained 200 mM NaCl, 20 mM Tris (pH 7.5), and 5 mM DTT. The main peak fractions were collected and concentrated to 6–8 mg/ml. Protein crystals of the native protein grew in hanging drops containing 2 μl of protein solution and 2 μl of well solution [0.2 M ammonium acetate, 0.01 M magnesium acetate tetrahydrate, 0.05 M sodium cacodylate trihydrate (pH 6.5), and 30% wt/vol polyethylene glycol (PEG) 8000] and were frozen in a cryoprotectant solution containing 25% glycerol. Complexes with AN3018–AMP were produced by adding AMP and AN3018 up to final concentrations of 20 and 1.5 mM, respectively. Crystals grew in hanging drops of 2 μl of protein solution and 2 μl of well solution [0.2 M sodium acetate, 0.1 M sodium cacodylate (pH 6.5), and 30% PEG 8000] and were frozen with a cryoprotectant solution containing 25% glycerol.

C. albicans cytosolic LeuRS CP1 structure determination

Data of native apo and inhibitor–AMP adduct-bound crystals were collected at European Synchrotron Radiation Facility (ESRF) beamlines ID23-1 and ID14-EH4, respec-

tively, following a strategy determined by MOSFLM²⁹ and integrated using XDS.³⁰ The structure was solved by the SAD method with SHELXD³¹ and SHARP³² using selenomethionine-labeled protein cocrystallized with the AN2690–AMP adduct (data not shown). The resulting phases were used for automatic model building in ARP/WARP.³³ The AN3018–AMP complex (space group $P2_12_12_1$; one molecule per asymmetric unit) and the apo structure (space group $P2_1$; four molecules per asymmetric unit) were subsequently solved by molecular replacement. Coot³⁴ was used for manual model editing, REFMAC³⁵ was used for refinement, and PROCHECK³⁶ and MOLPROBITY³⁷ were used for model validation. Crystallographic details are summarized in Table 1. For the apo structure, TLS refinement (one group per chain) and strict noncrystallographic symmetry were applied to the four chains, except where there were clear deviations, notably in residues 357–369 and 465–481. Chain B is the most complete, and chains A and B have lower overall B -factors (22.5 and 24.5 Å², respectively) compared to chains C and D (32.8 and 34.0 Å², respectively).

Production and crystallization of human cytosolic LeuRS CP1

DNA coding for residues 260–511 corresponding to the human cytosolic LeuRS CP1 was generated from a full-length cDNA using PCR and cloned into the NcoI–EcoRI sites of the expression vector pPROEX HTb coding for an additional N-terminal 6× histidine tag and a tobacco etch virus (TEV) protease site. The protein was expressed at 37 °C in the soluble fraction of *E. coli* strain BL21 Star(DE3) cells (Life Technologies) and purified on a Ni Sepharose column as described above for the *C. albicans* homologue. To cleave off the His-tag, the proteins were incubated with 1% His-tagged TEV protease at 10 °C overnight. Imidazole was removed by dialysis against lysis buffer. Uncleaved proteins and TEV protease were removed by binding to a second Ni Sepharose column. Cleaved protein was concentrated to 0.5 mg/ml and loaded onto a Superdex200 column. The running buffer contained 300 mM NaCl and 20 mM Tris (pH 7.5). Peak fractions were concentrated to 2 mg/ml and submitted to extensive crystallization trials using a Cartesian PixSys 4200 robot.

Hexagonal bipyramidal crystals containing PEG, ammonium sulfate, or other salts grew under numerous conditions at various pH values. Most crystals tested indexed with hexagonal lattices, with related cell dimensions of either $a=93.77$ Å, $b=93.77$ Å, and $c=299.10$ Å (maximum resolution of 4.7 Å; four molecules per asymmetric unit) or $a=93.02$ Å, $b=93.02$ Å, and $c=140.95$ Å (maximum resolution of 6 Å; two molecules per asymmetric unit). One crystal that had been frozen after slowly increasing the PEG concentration in the mother liquor by 15–35% gave a data set at 3.3 Å resolution, but could not be reproduced. In this case, the cell dimensions were $a=140.53$ Å, $b=91.78$ Å, $c=160.19$ Å, and $\beta=92.75$ and belonged to space group $P2_1$, with an estimated 12 molecules in the asymmetric unit. None of the above crystal forms could be solved by molecular replacement using the *C. albicans* editing domain. Therefore, a second construct (residues 260–509) was expressed and purified as previously described, and similar hexagonal crystals diffracting only to low resolution again grew in numerous conditions. The best data set and that which allowed independent structure solution was obtained from a selenomethionine-labeled sample of the second construct and cocrystallized with 5 mM short synthetic RNA (CCA, mimicking the 3' end of the tRNA)

and 7 mM of another oxaborole compound (AN2962) of the same family. This crystal was grown in sitting drops of 0.1 µl of protein solution and 0.1 µl of well solution [0.2 M sodium citrate tribasic dihydrate (pH 8.3) and 20% wt/vol PEG 3350]; had cell dimensions of $a=94.54$ Å, $b=94.54$ Å, and $c=148.83$ Å; belonged to space group $P6_5$ (with two molecules in the asymmetric unit); and diffracted to 3.2 Å resolution.

Human cytosolic LeuRS CP1 structure solution

The structure of the selenomethionine-labeled domain was solved with SAD, as described above. Twenty of 22 Se positions were found corresponding to two molecules in the asymmetric unit, including the extra methionine introduced by the cloning procedure [the terminal sequences of the construct are GAMG(260)PQ...YME (509), with italics indicating extra residues]. Model building and refinement were performed as described above. TLS refinement (one group per molecule) and strict noncrystallographic symmetry restraints were used. Subsequent analysis by molecular replacement showed that the $P2_1$ crystal form has six such dimers in the asymmetric unit. Crystallographic details are summarized in Table 1.

Model analysis and figures

Ribbon diagrams were made with PyMOL³⁸ and MOLSCRIPT.³⁹ Structural alignments were performed with the Protein Structure Comparison Service (secondary-structure matching) at the European Bioinformatics Institute.⁴⁰

Accession numbers

Coordinates and structure factors have been deposited in the PDB with accession number 2WFD for the human cytosolic LeuRS editing domain and accession numbers 2WFE and 2WFG for the apo- and AN3018–AMP-bound forms of the *C. albicans* cytosolic LeuRS editing domain, respectively.

Acknowledgements

We thank the beamline staff of the ESRF and the European Molecular Biology Laboratory for help with X-ray data collection, and platforms of the Partnership for Structural Biology for access. E.S. was supported by an “E-STAR” fellowship funded by the EU FP6 Marie Curie Host fellowship for Early Stage Research Training under contract no. MEST-CT-2004-504640. Part of this work was funded by a National Institutes of Health grant awarded to J.J.P. (R01 DE16835).

References

1. Woese, C. R., Olsen, G. J., Ibba, M. & Soll, D. (2000). Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol. Mol. Biol. Rev.* **64**, 202–236.

2. Cusack, S. (1995). Eleven down and nine to go. *Nat. Struct. Biol.* **2**, 824–831.
3. Fersht, A. R. (1977). Editing mechanisms in protein synthesis. Rejection of valine by the isoleucyl-tRNA synthetase. *Biochemistry*, **16**, 1025–1030.
4. Jakubowski, H. & Goldman, E. (1992). Editing of errors in selection of amino acids for protein synthesis. *Microbiol. Rev.* **56**, 412–429.
5. Beuning, P. J. & Musier-Forsyth, K. (2000). Hydrolytic editing by a class II aminoacyl-tRNA synthetase. *Proc. Natl Acad. Sci. USA*, **97**, 8916–8920.
6. Dock-Bregeon, A., Sankaranarayanan, R., Romby, P., Caillet, J., Springer, M., Rees, B. *et al.* (2000). Transfer RNA-mediated editing in threonyl-tRNA synthetase. The class II solution to the double discrimination problem. *Cell*, **103**, 877–884.
7. Beebe, K., Ribas De Pouplana, L. & Schimmel, P. (2003). Elucidation of tRNA-dependent editing by a class II tRNA synthetase and significance for cell viability. *EMBO J.* **22**, 668–675.
8. Roy, H., Ling, J., Irnov, M. & Ibba, M. (2004). Post-transfer editing *in vitro* and *in vivo* by the beta subunit of phenylalanyl-tRNA synthetase. *EMBO J.* **23**, 4639–4648.
9. Tukalo, M., Yaremchuk, A., Fukunaga, R., Yokoyama, S. & Cusack, S. (2005). The crystal structure of leucyl-tRNA synthetase complexed with tRNA^{leu} in the post-transfer-editing conformation. *Nat. Struct. Mol. Biol.* **12**, 923–930.
10. Cusack, S., Yaremchuk, A. & Tukalo, M. (2000). The 2 Å crystal structure of leucyl-tRNA synthetase and its complex with a leucyl-adenylate analogue. *EMBO J.* **19**, 2351–2361.
11. Fukunaga, R. & Yokoyama, S. (2005). Crystal structure of leucyl-tRNA synthetase from the archaeon *Pyrococcus horikoshii* reveals a novel editing domain orientation. *J. Mol. Biol.* **346**, 57–71.
12. Fukunaga, R. & Yokoyama, S. (2005). Aminoacylation complex structures of leucyl-tRNA synthetase and tRNA^{leu} reveal two modes of discriminator-base recognition. *Nat. Struct. Mol. Biol.* **12**, 915–922.
13. Li, T., Guo, N., Xia, X., Wang, E. D. & Wang, Y. L. (1999). The peptide bond between E292-A293 of *Escherichia coli* leucyl-tRNA synthetase is essential for its activity. *Biochemistry*, **38**, 13063–13069.
14. Williams, A. M. & Martinis, S. A. (2006). Mutational unmasking of a tRNA-dependent pathway for preventing genetic code ambiguity. *Proc. Natl Acad. Sci. USA*, **103**, 3586–3591.
15. Lincecum, T. L., Jr., Tukalo, M., Yaremchuk, A., Mursinna, R. S., Williams, A. M., Sproat, B. S. *et al.* (2003). Structural and mechanistic basis of pre- and posttransfer editing by leucyl-tRNA synthetase. *Mol. Cell*, **11**, 951–963.
16. Liu, Y., Liao, J., Zhu, B., Wang, E. D. & Ding, J. (2006). Crystal structures of the editing domain of *Escherichia coli* leucyl-tRNA synthetase and its complexes with Met and Ile reveal a lock-and-key mechanism for amino acid discrimination. *Biochem. J.* **394**, 399–407.
17. Mursinna, R. S., Lincecum, T. L., Jr. & Martinis, S. A. (2001). A conserved threonine within *Escherichia coli* leucyl-tRNA synthetase prevents hydrolytic editing of leucyl-tRNA^{leu}. *Biochemistry*, **40**, 5376–5381.
18. Mursinna, R. S. & Martinis, S. A. (2002). Rational design to block amino acid editing of a tRNA synthetase. *J. Am. Chem. Soc.* **124**, 7286–7287.
19. Tang, Y. & Tirrell, D. A. (2002). Attenuation of the editing activity of the *Escherichia coli* leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins *in vivo*. *Biochemistry*, **41**, 10635–10645.
20. Wu, N., Deiters, A., Cropp, T. A., King, D. & Schultz, P. G. (2004). A genetically encoded photocaged amino acid. *J. Am. Chem. Soc.* **126**, 14306–14307.
21. Ochsner, U. A., Sun, X., Jarvis, T., Critchley, I. & Janjic, N. (2007). Aminoacyl-tRNA synthetases: essential and still promising targets for new anti-infective agents. *Expert Opin. Invest. Drugs*, **16**, 573–593.
22. Baker, S. J., Zhang, Y. K., Akama, T., Lau, A., Zhou, H., Hernandez, V. *et al.* (2006). Discovery of a new boron-containing antifungal agent, 5-fluoro-1,3-dihydro-1-hydroxy-2,1-benzoxaborole (AN2690), for the potential treatment of onychomycosis. *J. Med. Chem.* **49**, 4447–4450.
23. Rock, F. L., Mao, W., Yaremchuk, A., Tukalo, M., Crepin, T., Zhou, H. *et al.* (2007). An antifungal agent inhibits an aminoacyl-tRNA synthetase by trapping tRNA in the editing site. *Science*, **316**, 1759–1761.
24. Bamford, D. H., Grimes, J. M. & Stuart, D. I. (2005). What does structure tell us about virus evolution? *Curr. Opin. Struct. Biol.* **15**, 655–663.
25. Dohm, J. C., Vingron, M. & Staub, E. (2006). Horizontal gene transfer in aminoacyl-tRNA synthetases including leucine-specific subtypes. *J. Mol. Evol.* **63**, 437–447.
26. Alley, M. R., Baker, S. J., Beutner, K. R. & Plattner, J. (2007). Recent progress on the topical therapy of onychomycosis. *Expert Opin. Invest. Drugs*, **16**, 157–167.
27. Yao, P., Zhou, X. L., He, R., Xue, M. Q., Zheng, Y. G., Wang, Y. F. & Wang, E. D. (2008). Unique residues crucial for optimal editing in yeast cytoplasmic leucyl-tRNA synthetase are revealed by using a novel knockout yeast strain. *J. Biol. Chem.* **283**, 22591–22600.
28. Santos, M. A., Ueda, T., Watanabe, K. & Tuite, M. F. (1997). The non-standard genetic code of *Candida* spp.: an evolving genetic code or a novel mechanism for adaptation? *Mol. Microbiol.* **26**, 423–431.
29. Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Jt. CCP4+ESF-EAMCB Newsl. Protein Crystallogr.* **26**.
30. Kabsch, W. (1988). Evaluation of single-crystal X-ray diffraction from a position-sensitive detector. *J. Appl. Crystallogr.* **21**, 916–924.
31. Schneider, T. R. & Sheldrick, G. M. (2002). Substructure solution with SHELXD. *Acta Crystallogr. Sect. D*, **58**, 1772–1779.
32. Fortelle de La, E. & Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 590–620.
33. Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458–463.
34. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D*, **60**, 2126–2132.
35. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D*, **53**, 240–255.
36. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
37. Lovell, S. C., Davis, I. W., Arendall, W. B., III, de

- Bakker, P. I., Word, J. M., Prisant, M. G. *et al.* (2003). Structure validation by C α geometry: phi,psi and Cbeta deviation. *Proteins*, **50**, 437–450.
38. DeLano, W. L. (2002). *The PyMOL User's Manual* DeLano Scientific, Palo Alto, CA, USA.
39. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950.
40. Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. Sect. D*, **60**, 2256–2268.