L-Arginine recognition by yeast arginyl-tRNA synthetase

Jean Cavarelli¹, Bénédicte Delagoutte, Gilbert Eriani², Jean Gangloff² and Dino Moras

UPR 9004 Biologie Structurale, Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/ULP, BP 163, 67404 Illkirch Cedex and ²UPR 9002 Structure des Macromolécules Biologiques et Mécanismes de Reconnaissance, Institut de Biologie Moléculaire et Cellulaire du CNRS, 15 rue René Descartes, 67084 Strasbourg Cedex, France

¹Corresponding author
e-mail: cava@igbmc.u-strasbg.fr

The crystal structure of arginyl-tRNA synthetase (ArgRS) from *Saccharomyces cerevisiae*, a class I aminoacyl-tRNA synthetase (aaRS), with L-arginine bound to the active site has been solved at 2.75 Å resolution and refined to a crystallographic *R*-factor of 19.7%. ArgRS is composed predominantly of α -helices and can be divided into five domains, including the class I-specific active site. The N-terminal domain shows striking similarity to some completely unrelated proteins and defines a module which should participate in specific tRNA recognition. The C-terminal domain, which is the putative anticodon-binding module, displays an all- α -helix fold highly similar to that of *Escherichia coli* methionyl-tRNA synthetase. While ArgRS requires tRNA^{Arg} for the first step of the aminoacylation reaction, the results show that its presence is not a prerequisite for L-arginine binding. All H-bond-forming capability of L-arginine is used by the protein for the specific recognition. The guanidinium group forms two salt bridge interactions with two acidic residues, and one H-bond with a tyrosine residue; these three residues are strictly conserved in all ArgRS sequences. This tyrosine is also conserved in other class I aaRS active sites but plays several functional roles. The ArgRS structure allows the definition of a new framework for sequence alignments and subclass definition in class I aaRSs.

Keywords: aminoacylation/arginyl-tRNA synthetase/crystal structure/X-ray diffraction

Introduction

Aminoacyl-tRNA synthetases (aaRSs) catalyse the esterification of an amino acid to the 3'-terminal adenosine of a tRNA. Each aaRS is specific for one amino acid and several cognate isoacceptor tRNA species. Aminoacylation of a tRNA with its cognate amino acid is achieved by a two-step reaction. The first step produces a stable enzyme-bound intermediate, called aminoacyl-adenylate, from ATP and amino acid. The second step is the transfer of the aminoacyl moiety to one of the hydroxyl groups of the

3'-terminal adenosine of the tRNA. AaRSs can be divided into two classes of 10 members each, which correspond to two architectures of the active site core characterized by conserved amino acid residues (Eriani *et al.*, 1990). The class I aaRSs, whose active sites contain a canonical dinucleotide-binding fold [four- or five-stranded parallel β -sheet, Rossmann fold (Rossmann *et al.*, 1974)], display two signature amino acid sequences 'HIGH' (Barker and Winter, 1982) and 'KMSKS' (Webster *et al.*, 1984). Class II aaRSs are built around an antiparallel β -sheet partly closed by helices (Cusack *et al.*, 1990; Ruff *et al.*, 1991) and contain three characteristic homologous motifs. The topology of each active site is the structural constraint governing the conformation of the ATP molecule and conferring a regiospecificity to the second step of the reaction. Class I aaRSs aminoacylate the 2'-OH of the 3'-terminal adenosine of the tRNA, while class II aaRSs aminoacylate the 3'-OH.

Crystallographic studies of 12 aaRSs revealed the textbook modularity of these enzymes, whereby various insertions and additional domains are appended to the active site. Most of these domains play a functional role either in tRNA recognition or in proofreading mechanisms. The diversity of the appended domains outside the catalytic core, which range from all α -helical folds to all β -barrel structures, allow specialized patterns of tRNA recognition that are characteristic of each aminoacylation system (for a recent review, see Francklyn *et al.*, 1997).

Amino acid or aminoacyl-adenylate recognition has been examined in structural detail for several class I and II aaRSs: TyrRS (Brick *et al.*, 1989), TrpRS (Doublié *et al.*, 1995), GlnRS (Rath *et al.*, 1998), AspRS (Cavarelli *et al.*, 1994; Poterszman *et al.*, 1994), LysRS (Onesti *et al.*, 1995), HisRS (Arnez *et al.*, 1995), SerRS (Belrhali *et al.*, 1995) and, recently, IleRS (Nureki *et al.*, 1998). The large amount of information collected over the past years has produced a unified vision of the aaRSs world, also revealing the complexity of the fine tuning necessary for the aminoacylation reaction.

ArgRS deserves a special place among aaRSs. Like GlnRS and GluRS, ArgRS requires its cognate tRNA for the first step of the aminoacylation reaction, the amino acid activation (Mehler and Mitra, 1967; Mitra and Smith, 1969; Kern and Lapointe, 1980). In spite of extensive investigations performed by several groups, no definitive answer has been found favouring one of the two proposed mechanisms accounting for this peculiar behaviour: either a two-step mechanism, where tRNA binding is a prerequisite for adenylate formation, or a mechanism where ATP, arginine and tRNA co-react in a concerted way leading to a one-step aminoacylation of tRNA (Mehler and Mitra, 1967; Loftfield, 1972). However, results of kinetic investigations using non-chargeable tRNA^{Arg} argue in favour of the classical two-step mechanism (Fersht *et al.*, 1978; Gerlo

Table I. Crystallographic data

	Native	PCMBs ^a	K ₂ Hg(CN) ₄ ^a
Data collection			
Wavelength (Å)	1.1024	1.1024	1.1024
Resolution (Å)	30–2.75	50–3.95	50–3.45
No. of observations	168 917	69 530	56 364
No. of unique reflections	27 890	9528	13 386
Completeness (%)	99.4	99.0	92.8
2.75–2.85 Å ($F/\sigma > 3$)	80.0		
R_{sym}^b (%) (all range)	4.2	5.5	4.8
2.75–2.85 Å			
(12 392 obs/2660 reflections)	18.0		
R_{ano}^c (%) (all range)		2.6	
Heavy atom refinement and phasing			
Heavy atom concentration (mM)		1	2
Soaking time (h)		24	48
No. of sites per molecule		3	3
R_{iso}^d (%)		20.4	24.8
Phasing power ^e (centric–acentric)		1.22–1.82	0.82–1.35
R_{cullis}^f (centric–acentric)		0.63–0.69	0.76–0.76
Refinement statistics			
Refinement program	CNS		
Resolution (Å)	20–2.75		
No. of reflections	27 116		
No. of non-hydrogen proteins atoms	4883		
No. of ordered water molecules	280		
$R_{\text{model}}^g, R_{\text{free}}^h$ (%)	19.7, 26.9		
Overall G -factor ⁱ	0.29		
Average B -factors (Å ²)			
all non-hydrogen atoms	53.3		
main chain, side chain, water molecules	51.1, 56.1, 48.8		
R.m.s. deviations from ideal geometry ^j bonds (Å), angles (°), dihedral (°)	0.009, 1.52, 19.1		
Ramachandran plot quality			
Residues in core, allowed, generous regions (%)	84.7, 13.3, 0.40		
Coordinates errors from cross-validated Luzzati plots (Å)	0.33		

^aHeavy atoms abbreviations: PCMBs (*p*-chloromercuribenzenesulfonate), K₂Hg(CN)₄ (potassium tetracyano mercurate); anomalous data has been collected and used for PCMBs.

^b $R_{\text{sym}} = \sum_i \sum_j |I_{h,i} - I_{h,j}| / \sum_i \sum_j I_{h,i}$, where $I_{h,i}$ is the *i*-th observed intensity of reflection *h* and $\langle I_h \rangle$ is the average intensity for this unique reflection.

^c $R_{\text{ano}} = \sum_h | \langle I_{h^+} \rangle - \langle I_{h^-} \rangle | / \sum_h \langle I_{h^+} \rangle + \langle I_{h^-} \rangle$, where $\langle I_{h^+} \rangle$, $\langle I_{h^-} \rangle$ correspond to average intensities of each Friedel pair for reflection *h*.

^d $R_{\text{iso}} = \sum_h |F_{\text{der}} - F_{\text{nat}}| / \sum_h F_{\text{nat}}$, where F_{nat} and F_{der} are the native and derivative structure factor amplitudes.

^ePhasing power = (r.m.s. heavy atom structure factor)/(phase integrated lack of closure) (statistics of MLPHARE).

^f $R_{\text{cullis}} = (\text{phase integrated lack of closure}) / \text{r.m.s. isomorphous difference}$ (statistics of MLPHARE).

^g $R_{\text{model}} = \sum_h |F_{\text{obs}} - F_{\text{calc}}| / \sum_h F_{\text{obs}}$, where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively.

^hA small fraction (7.3%) of reflections with no cut-off between 20 and 2.75 Å (2030 out of 27 116 unique reflections) was excluded from the refinement and used for monitoring the course of the refinement.

ⁱOverall 'normality' as calculated by PROCHECK.

^jR.m.s. deviations were calculated using Engh and Huber parameters.

et al., 1982). Yeast ArgRS also displays an original feature in its ability to aminoacylate, in addition to its cognate tRNA, a non-cognate molecule, namely a yeast tRNA^{Asp} deprived of post-transcriptional modifications (Ebel *et al.*, 1973; Gangloff *et al.*, 1973). Here we describe the first step towards structural comprehension of this peculiar system, the crystal structure of ArgRS from the yeast *Saccharomyces cerevisiae* with L-arginine bound to the active site.

Results

Structure determination

Yeast ArgRS is a monomeric protein of 607 residues (mol. wt 69.5 kDa). The enzyme was expressed at high levels from an overproducing *Escherichia coli* strain. Protein purification involved two chromatographic steps: an anion exchanger followed by a hydrophobic interaction column. The structure of ArgRS was solved using crystals grown from ammonium sulfate at pH 7.0 in the presence of

L-arginine and ATP. Phases to 3.8 Å resolution were obtained by the multiple isomorphous replacement method and anomalous scattering from two heavy atom derivatives (Table I). The 3.8 Å map was improved by solvent flattening, solvent flipping and histogram matching, and then the phases were extended to 3.5 Å. The modified map allowed chain tracing for most parts of the polypeptide chain. The final model has been refined at 2.75 Å resolution to a crystallographic *R*-factor of 19.7% ($R_{\text{free}} = 26.9\%$) with good stereochemistry (see Table I for refinement statistics). The crystallographic asymmetric unit contains one molecule of ArgRS, the L-arginine substrate and 280 water molecules. The first four residues at the N-terminus of ArgRS are not visible in the electron density map and are presumed to be disordered.

Anatomy of class I ArgRS

The structure of ArgRS can be divided schematically into five domains: a catalytic domain that contains the class I

Fig. 1. The structure of cytoplasmic ArgRS from the yeast *S.cerevisiae*. (A) Schematic drawing (overall dimension $85 \times 65 \times 40$ Å) (drawn with SETOR, Evans, 1993). (B) Topology diagram of the secondary structure elements. Add-1 is in orange, the catalytic domain in red, Ins-1 in green, Ins-2 (CP1) in blue and Add-2 in yellow.

Fig. 2. Sequence alignment for ArgRSs from different organisms: yRrs = cytoplasmic *S.cerevisiae*, mRrs = mitochondrial *S.cerevisiae*, jRrs = *Methanococcus jannaschii*, hRrs = *Homo sapiens*, eRrs = *E.coli*. The secondary structure, determined with PROMOTIF (Hutchinson and Thornton, 1996), is also displayed for yRrs. Elements of secondary structure are indicated by: H for α -helix, S for strand of β -sheet, ξ for 3_{10} helix. Yeast numbering is shown above the sequence of yRrs. The colour code defined in Figure 1 is also used here. The residues of the class I signature sequences are highlighted in red and the residues involved in L-arginine binding are highlighted in pink. This sequence alignment was extracted from an alignment built with the program CLUSTAL W (Thompson et al., 1994) using 29 sequences of ArgRS from different organisms. This figure was produced with the program ALSCRIPT (Barton, 1993).

active site, to which four structurally defined domains are appended (Figure 1). Two of them, called additional domain 1 and 2 (Add-1 and Add-2) are attached respectively at the N- and C-terminal sides of the active site. Two domains (Ins-1 and Ins-2) are inserted into the catalytic core. The structure is composed predominantly

of α -helices, with 55% of the residues adopting a helical conformation and 9% forming β -strands. An alignment of ArgRS sequences from different organisms is shown in Figure 2.

The active site of ArgRS (colour coded red in Figures 1 and 2), which forms the scaffold of the Rossmann fold,

Fig. 3. Superposition of the C α atoms of the class I aaRSs onto ArgRS. (A) GlnRS (Rould *et al.*, 1989), (B) GluRS (Nureki *et al.*, 1995), (C) MetRS (Brunie *et al.*, 1990) and (D) TrpRS (Doublé *et al.*, 1995). ArgRS is traced in red, GlnRS in blue, GluRS in cyan, MetRS in green and TrpRS in black. Superimposing the six class I aaRSs structures gives r.m.s. deviations between corresponding C α atoms of 2.56 Å (GluRS/ArgRS over 182 residues), 2.61 Å (GlnRS/ArgRS over 175 residues), 2.47 Å (TyrRS/ArgRS over 119 residues), 2.64 Å (TrpRS/ArgRS over 129 residues) and 2.53 Å (ArgRS/ecMetRS over 222 residues). The structure superpositions have been made with the program SARF (Alexandrov, 1996). This figure was drawn with MOLSCRIPT (Kraulis, 1991).

is composed of two halves assembled from three peptides; the first half includes residues 143–194 and residues 266–293, while the second half includes residues 345–410. The first signature sequence of class I aaRSs (His159–Ala160–Gly161–His162) belongs to the first half of the nucleotide-binding fold, while the second motif, ‘MSK’ (Met408–Ser409–Thr410), belongs to the second half. Superimposing the five class I aaRSs for which coordinates are available gives root-mean-square (r.m.s.) deviations between corresponding C α atoms of ~2.5 Å. The number

of spatially equivalent residues ranges from 119 (ArgRS/TyrRS) to 222 (ArgRS/MetRS). This reveals a high structural similarity between ArgRS and MetRS and also significant similarities with GluRS and GlnRS which extend beyond the anticipated correspondence of their respective nucleotide-binding domains (Figures 3 and 4).

Additional domain 1

ArgRS is first characterized by its N-terminal domain (Add-1, 142 residues, colour coded brown in Figures 1

Fig. 4. Schematic representation of the six class I aaRSs highlighting the positions of the Rossmann fold moieties (RF1, RF2 colour coded red). The anticodon-binding module whose fold is similar to that of MetRS and ArgRS is colour coded in yellow. Another inserted peptide, called connecting peptide 2 (CP2), has been identified by sequence analysis in five class I aaRSs (LeuRS, ValRS, IleRS, CysRS and MetRS) (Starzyk *et al.*, 1987; Burbaum and Schimmel, 1991). The CP2 domain of MetRS is highlighted in grey. In ArgRS, as in GluRS, GlnRS, TyrRS and TrpRS, there is no insertion peptide corresponding to CP2.

and 2). Its topology, a $\beta\alpha\beta$ single split motif (S2–H4–S3–S4) to which another strand (S1) and three helices (H1–H2–H3) are added, is unique among the known aaRSs. An alignment of 29 sequences of ArgRS (five archaea, seven eukaryotes and 17 eubacteria; data not shown) shows that this extension is conserved among all species. The core of this domain, encompassing the β -sheet and two helices (H2 and H4), is strikingly similar in topology to several others proteins: (i) two Zn-binding proteins, muramoyl-pentapeptide carboxypeptidase from *Streptomyces albus* G (PDB code 1lbu) (Dideberg *et al.*, 1982) and the N-terminal signal domain of Sonic hedgehog (PDB code 1vhh) (Tanaka Hall *et al.*, 1995); (ii) the oligomerization and L-arginine-binding domain of the arginine repressor of *E. coli* (PDB code 1xxa) (Van Duyne *et al.*, 1996); (iii) a nucleotidyl transferase, the palm domain of human DNA polymerase β (PDB code 1bpy) (Sawaya *et al.*, 1994); and (iv) the C-terminal domain of glyceraldehyde-3-phosphate dehydrogenase (GAPDH, PDB code 1cer). The structural similarities were found by the programs SARF (Alexandrov, 1996) and DALI (Holm and Sander, 1994). Superimposition of the different structures gives r.m.s. deviations between corresponding C α atoms of 2.8 Å (ArgRS/1lbu over 79 residues), 2.2 Å (ArgRS/1bpy over 69 residues), 2.1 Å (ArgRS/1xxa over 48 residues), 2.1 Å (ArgRS/1vhh over 74 residues) and 2.75 Å (ArgRS/1cer over 75 residues). The origin of the structural similarity between these proteins is unclear. Due to the diversity of the functions involved, thermodynamic stability may be the common root. The single split motif is a recurrent protein structure and has been found as a building block for several domains (Orengo and Thornton, 1993). However, the common topology between the N-terminal domain of ArgRS and the palm domain of DNA polymerase β indicates that the face of the β -sheet is a platform which can be used for interactions with nucleic acids. A model for tRNA docking on ArgRS suggests that Add-1 of ArgRS will certainly be involved

in tRNA recognition, interacting on the D loop side of the tRNA (see Discussion).

Insertion domain 1

ArgRS possesses an insertion domain in the first half of the Rossmann fold between the second strand (S6) and the second helix (H11) (Ins-1, residues 195–265, colour coded green in Figures 1 and 2). This connection, which is a short loop in GluRS, GluRS, MetRS and TrpRS, and a slightly longer one in TyrRS (14 residues), is 80 residues long in ArgRS and contains four helices. This domain closes one side of the active site. The angle between the axes of two consecutive helices is $\sim 120^\circ$. Therefore, helix H11, which is part of the nucleotide-binding fold, has a different orientation with respect to the β -sheet and is much longer in ArgRS (17 residues).

Insertion domain 2

Domain Ins-2 of ArgRS, which ranges from residues 294 to 344 (colour coded blue in Figure 1), corresponds to the so-called connecting peptide 1 (CP1), which in each class I aaRS links the two halves of the Rossmann fold and is specific to each system (Starzyk *et al.*, 1987; Burbaum and Schimmel, 1991). In ArgRS, CP1 contains a helix (H12), a two-stranded antiparallel β -sheet (S8 and S9) and a strand S10, and sits on the top of the second half of the nucleotide-binding fold.

Though relatively short in size in TyrRS and TrpRS (25 residues, mainly two short helices), CP1 contains >100 residues in GlnRS (residues 100–210) and is folded into a five-stranded antiparallel β -sheet flanked by three helices. This domain has been called acceptor-binding domain in GlnRS because it has been shown to stabilize the hairpin conformation of the 3' end of the tRNA^{Gln}.

Despite an overall structure which is specific for each system, striking similarities between CP1s of GluRS, GlnRS, MetRS and ArgRS can be pointed out. These enzymes share two common substructures: (i) a helix–loop–strand–loop–helix motif (from helix H11 to helix H12) and (ii) a two-stranded antiparallel β -sheet (S8 and S9). The first common substructure is very similar in all four enzymes; GluRS possesses an extra long loop (residues 71–80 of GluRS). This common feature has already been mentioned for GlnRS, GluRS and MetRS (Landès *et al.*, 1995); it closes the back of the active site and may play a structural role in stabilizing the overall structure.

For the common β -sheet substructure, system specificity is expressed in the lengths and conformations of the connection between helix H12 and the first strand of the β -sheet (S8), the peptide linking the two strands (S8 and S9) and the peptide linking strand S9 to helix H13. In GlnRS, this last peptide, which spatially overlaps Ins-1 of ArgRS, has been shown to provide a binding pocket for the looped out base C74 of tRNA^{Gln}.

Anticodon-binding module

ArgRS possesses an α -helical C-terminal domain (Add-2, residues 411–607, colour coded yellow in Figures 1 and 2) which contains 10 helices and one strand (S13). By spatial analogy with other aaRSs, this module should be the putative anticodon-binding domain. Add-1 is spatially

adjacent to Add-2, which could imply that both modules will cooperate in specific tRNA recognition.

Helices H20 and H21 can be considered as a single helix kinked in the middle by a 3_{10} helix turn. Helices H18, H20, H22 and H23 constitute a four-helix bundle which has a topology similar to the C-terminal domain of MetRS (Brunie *et al.*, 1990). Superimposition of the corresponding C α atoms of the two domains (113 residues) yields r.m.s. deviations of 2.42 Å (ArgRS, residues 455–600; MetRS, residues 340–491). System specificity is expressed in the length of the helices. H19 is only present in ArgRS whereas MetRS has an extra C-terminal extension peptide.

Helices H17 (residues 475–479) and H18 (residues 486–500) present a striking structural feature. Their axes align with an interaxial angle of 0.8° , so they could make a single helix (which is observed in MetRS). However, a small contiguous peptide of six amino acids (Ser480–Phe481–Glu482–Gly483–Asp484–Thr485) joins these two helices and creates a protruding Ω loop at the surface of the protein. This exposed Ω loop, located just after the tRNA-anchoring platform (see below), may be used as a molecular switch by ArgRS for specific recognition. *In vivo* experiments have shown that a mutation of Gly483 to a serine is lethal for cell growth, which gives support for the functional role of this residue (Eriani *et al.*, in preparation).

The similar structures and functions of the anticodon-binding domains of ArgRS and MetRS (Figure 3C) suggest that they may have evolved from the same ancestor. The original subclass definition of class I aaRSs should, therefore, be revised. Our results would place ArgRS and MetRS in the same subgroup and confirm the subgroup classification proposed by Landès *et al.* (1995). Sequence alignments based on structure superposition of the two anticodon-binding domains highlight amino acid residues which are involved mainly in the packing of the helices. Detailed sequence comparisons between MetRS and ArgRS will be published elsewhere (in preparation).

Helices H21, H22 and H23 interact underneath the floor of the active site with helix H6, which carries the 'HIGH' signature motif. H6 seems to be a crucial node of communication between the different modules of ArgRS. It interacts (i) with helix H11 anchoring domain Ins-1, (ii) with helix H5 linking domain Add-1 to the active site, and also (iii) with helix H16. One can therefore easily imagine how any structural events caused by substrate binding can be communicated all over the structure via those long helices (H23 is 37.3 Å long, H22 is 32.0 Å long, H16 is 30.3 Å long, H6 is 30.7 Å long and H11 is 29.8 Å long).

L-Arginine binding and recognition

A well-ordered arginine molecule was found bound to the active site (Figure 5A). L-Arginine binds at the C-terminal end of the β -strands of the Rossmann fold in a crevice formed between the two symmetrical halves of the fold

Fig. 5. L-Arginine-binding site. (A) Final $(2F_o - F_c)$ cross-validated σ -weighted omit map (resolution limits 20–2.75 Å, all data used, contour level 1.2σ). (B) Interaction between the active site and L-arginine. The two histidines of the HIGH motif are also shown. (C) L-Arginine recognition: interactions between the substrate and the protein. The figure was drawn with SETOR (Evans, 1993).

(Figure 5B). Arginine recognition involves amino acid residues from strands S5 and S11, helices H6 and H13 and the loop between S5 and H6. Arginine is bound specifically to the protein by a network of H-bonds and salt bridge interactions (Figure 5C). All H-bonding capability of the substrate is used by the protein for the specific recognition. The α -amino group of the arginine molecule forms H-bonds to the amide oxygen of Asn153 and the main chain carbonyl of Ser151, while the α -carboxylate interacts with the amide nitrogens of Asn153 and Gln375. Note that strand S5 presents a β -bulge at residues 149–150. This peculiar conformation may be important for the correct recognition of the α -ammonium and α -carboxy groups of the L-arginine substrate by Ser151 and Asn153. The two histidine residues of the conserved signature of class I aaRSs also interact with the carboxylate atoms of L-arginine, either by a direct H-bond (His162) or by a water-mediated interaction (His159). The guanidinium moiety forms two salt bridge interactions with two carboxylates residues, Glu148 on one side, and Asp351 on the other. The phenolic oxygen of Tyr347 plays a dual role, accepting an H-bond from the η -nitrogen atoms of the substrate and donating one to the amide oxygen of Gln375. This residue links the specificity of side chain recognition of the substrate to the correct positioning of the main chain atoms. An alignment of 29 ArgRS sequences shows that Asp351, Asn153 and Tyr347 are strictly conserved, while Glu148 is sometimes replaced by an aspartate. The mode of binding of L-arginine can thus be generalized to all ArgRSs. Like GlnRS and GluRS, ArgRS requires its cognate tRNA for the first step of the aminoacylation reaction, the amino acid activation. Our structure shows that tRNA^{Arg} binding is not a prerequisite for L-arginine binding.

Despite our best efforts, no crystal has been obtained either for the ArgRS alone, or for ArgRS in the presence of ATP alone. Once the initial conditions of crystal growth have been established in the presence of ATP and L-arginine, we have been able to reproduce crystals in the presence of L-arginine alone. Therefore, only L-arginine is essential for crystal growth. One may postulate that L-arginine binding produces a conformational change in different loops around the active site which will promote intermolecular contacts necessary for crystal packing. One cannot exclude that ArgRS may also require the binding of tRNA to form the ATP-binding site, as seen in the GlnRS–tRNA^{Gln} structure, where the terminal adenosine of tRNA^{Gln} interacts with the α -phosphate of ATP.

ATP-binding site

While ArgRS crystals grow in the presence of L-arginine and ATP, no ATP molecule is visible in the electron density map. Packing effects can explain this absence. Previous results on class I aaRSs have shown that the second signature motif 'KMSKS' is part of a mobile loop that is involved in the stabilization of the transition state, the second lysine forming a salt bridge interaction with the α -phosphate of ATP (Brick *et al.*, 1989). In our crystal form, this loop (G407–M408–S409–T410–R411) is involved in a strong crystal packing interaction with the domain Add-1 of another molecule. The C-terminal side of the Rossmann fold, together with strands S13 and S14, presents an open platform on which the domain I of

a crystallographically equivalent molecule sits, interacting on the helix side. The 'KMSKS' loop is therefore located below helices H2 and H3 of the other molecule. Several H-bond interactions anchor the two molecules together with an intimate salt bridge between Glu294 and Lys79 of the crystallographically equivalent molecule. The crevice which should receive the AMP moiety is accessible, but there is no room to accommodate the β - and γ -phosphates.

Discussion

Amino acid recognition by class I aaRSs

A structural comparison of the active sites of class I aaRSs reveals that each enzyme has its own solution for the binding of the main chain atoms of the amino acid substrate. There are no strictly conserved residues of the protein dedicated for this purpose. The main chain atoms of residues at the C-terminal end of strands S5 or S11 are also involved in the recognition of the common peptidic part of the substrate. The local conformation of these strands is different in each enzyme, and irregularities such as β -bulges are often found. The system-specific conformation of S5 or the loop between S5 and H6 may explain the presence of at least one proline residue but at a different position in all class I aaRSs.

The variability within class I enzymes is apparent in the orientation of the peptidic atoms of the amino acid substrate, which is different in ArgRS compared with TyrRS and TrpRS. While the α -amino group of L-arginine points toward strand S5, those of Tyr and Trp are rotated 90° away and are directed toward helix H13. An orientation of the peptidic portion of the substrate similar to that found in ArgRS seems to be present in the structure of the complex between GlnRS–tRNA^{Gln} and a glutaminyl-adenylate analogue (Figure 3 in Rath *et al.*, 1998).

Only two residues are conserved in the amino acid-binding crevice of class I aaRSs, a tyrosine residue (Tyr347 in ArgRS) and an acidic residue (Asp191 in ArgRS). Tyr347 is special since it is strictly conserved in GlnRS, GluRS, TyrRS, TrpRS and ArgRS. In ArgRS, it cooperates in the recognition of the η -nitrogen atom of the substrate. In TyrRS (Tyr169) and TrpRS (Tyr125), this residue makes a H-bond to the α -amino group. In the recent crystal structure of GlnRS–tRNA^{Gln} in the presence of a stable glutaminyl-adenylate analogue (Rath *et al.*, 1998), this tyrosine residue (Tyr211) together with a water molecule has been shown to be responsible for the specific recognition of the two amide hydrogens of the substrate glutamine. By analogy, this tyrosine is likely to be in a key position for the recognition of L-glutamic acid by GluRS. It is astonishing that such a highly conserved residue may play several roles in different aaRSs.

The only other invariant residue, an acidic residue (Asp191 in ArgRS), is highly conserved in class I aaRSs (all but TyrRS and TrpRS) at the C-terminal end of strand S5 (Landès *et al.*, 1995). Our structure shows that this residue is not involved in substrate binding but has a structural role. It stabilizes the interactions of strand S5 with strand S6 and the surrounding secondary structures.

ATP-binding site

Crystal structures of class I aaRSs and extensive solution experiments linked the two signature peptides 'HIGH'

and 'KMSKS' to the ATP-binding site (for a recent review, see Meinnel *et al.*, 1995). The absolutely invariant glycine of the first signature sequence forms a flat platform at the N-terminus of helix H6 on which the adenine base of the ATP sits. The two histidines and the lysine have been shown to be involved in the stabilization of the first transition state to the aminoacyl-adenylate formation. In the GlnRS-tRNA^{Gln} complex, the two histidines make H-bonds with the negatively charged ATP phosphates. The ArgRS structure illustrates the dual functional role played by these two histidines; the first (His162) forms a direct H-bond with the carboxylate moiety of the L-arginine substrate and the second interacts through a water molecule. The structures of GlnRS, TyrRS and TrpRS have shown that the adenylate is formed with little displacement of the AMP and amino acid moieties, compared with their initial binding conformations. Model building suggests (data not shown) that His159 of ArgRS can easily interact with the α -phosphate of ATP in the initial conformation.

Based on sequence alignment of class I aaRSs, the precise location of this second signature motif 'KMSKS' in ArgRS was a subject of controversy (Landès *et al.*, 1995). Our structure clearly identifies G407-M408-S409-T410-R411 as the correct motif where the essential lysine is not present. Only 10 out of 29 ArgRS sequences known to date (data not shown) display a lysine residue at the third position of this signature sequence. The variability of the 29 ArgRS sequences at this location can be described by the following pattern M(F,I,L)S(K,R)T(G,K)R K(R,S,A,T,E)G(A). A single lysine residue is always present, either in position 2, 3 or 5. This lysine can play the role of the second lysine of the canonical 'KMSKS'. Position 4 is always occupied by an arginine residue, which is also present at a similar position in GlnRS and GluRS. In these three structures, this residue is directed toward the solvent and does not seem to be implicated directly in initial substrate binding (tRNA or ATP).

tRNA recognition and binding

The structural similarities between ArgRS and GlnRS mentioned above allow a model to be built by docking a tRNA molecule to the ArgRS structure, based on the crystal structure of the GlnRS-tRNA^{Gln} complex. The resulting model shows that (i) the amino acid acceptor stem is clamped on one side by Ins-1, and by the second half of the Rossmann fold on the other, and that (ii) domains Add-1 and Add-2 together recognize the anticodon stem and arm of tRNA^{Arg}.

A tRNA-anchoring platform

ArgRS, GlnRS, GluRS and MetRS share a common motif made of two strands (S13 and S14), located after the second half of the Rossmann fold, which are linked by a left-handed crossover connection. This connection is vital for class I aaRSs because it brings the two signature motifs on the same side on the active site. Neither TrpRS nor TyrRS has the second strand. This feature has already been mentioned for GlnRS, GluRS and MetRS (Landès *et al.*, 1995). This motif is involved in the anchoring of the tRNA molecule to the synthetase platform, as seen in the crystal structure of the GlnRS-tRNA^{Gln} complex. It interacts with the inside of the L-corner of tRNA^{Gln} at the junction between the two helical domains. This feature

clearly suggests a similar tRNA positioning by these four enzymes.

An RNA-binding domain at the N-terminus of ArgRS

Our model for tRNA docking on ArgRS indicates that domain Add-1 of ArgRS will most probably be involved in tRNA recognition, interacting with the D loop side of the tRNA. Add-1, particularly the exposed face of the β -sheet, possesses the characteristics which have been pointed out in many nucleic acid-binding proteins: an exposed aromatic or aliphatic residue (Phe109 and Leu70) which may be involved in van der Waals and hydrophobic interactions; positively charged residues (Arg66, Arg75 and Lys103) which may interact with the sugar-phosphate backbone; and polar side chains (Asn62, Asn106 and Gln111) which may be involved in direct or water-mediated interactions with the nucleic acid (for a review, see Arnez and Cavarelli, 1997).

Structural homology between Add-1 of ArgRS and the C-terminal domain of GAPDH was only revealed by the program SARF (Alexandrov, 1996). A monomer of GAPDH can be divided schematically into two domains, an N-terminal domain which contains the Rossmann fold (residues 1-140) and binds the cofactor nicotinamide adenine dinucleotide, and a C-terminal domain (residues 149-331) built around a twisted six-stranded antiparallel β -sheet. Structural superposition reveals that residues 155-177 and residues 237-317 of GAPDH have the same topology as Add-1 of ArgRS (r.m.s. deviation over 67 C α residues of 2.8 Å). This correlation gives a bona fide class I aaRS structure to monomeric GAPDH: a nucleotide-binding site and a putative RNA-binding domain. This may explain the observation that GAPDH was characterized as a tRNA-binding protein which may participate in tRNA export (Singh and Green, 1993).

The spatial arrangement of the secondary structural elements of Add-1, but with different connectivities, is also similar to that of two other RNA-binding proteins: the small nuclear ribonucleoprotein U1A protein (Nagai *et al.*, 1990; Oubridge *et al.*, 1994) and ribosomal protein S6 (Lindahl *et al.*, 1994).

Anticodon binding

According to our model, the protein-tRNA interface would involve residues from helices H21 and H22, the loops surrounding those two helices and the exposed face of the antiparallel β -sheet of the Add-1 domain. Add-1 would contribute to the recognition of the D loop and the anticodon arm. This is supported by the distribution of the electrostatic potential on the solvent-accessible surface of ArgRS, which is predominantly positive for domain Add-1 (Figure 6). It also agrees with solution studies which have shown that A20 and C35 are the nucleotides responsible for the arginine identity in *E. coli* tRNA^{Arg} (McClain and Foss, 1988; Schulman and Pelka, 1989; McClain *et al.*, 1990; Tamura *et al.*, 1992; Saks *et al.*, 1998).

Superposing the Rossmann folds of ArgRS, GlnRS and GluRS clearly emphasizes differences in the relative position in space of the anticodon-binding domain of ArgRS relative to the two other enzymes. Looking in a direction parallel to the plane of the Rossmann fold and

Fig. 6. Docking model of tRNA binding to ArgRS. The molecular surface of the protein showing the electrostatic potential was calculated with GRASP (Nicholls and Honig, 1991). Negatively charged regions are in red and positively charged areas in blue. The tRNA is rendered as a green coil. The orientation of the ArgRS molecule corresponds to that of Figure 1A.

perpendicular to the strand direction, with the C-termini of the strands on the right hand side, the anticodon-binding domains of GluRS and GlnRS are located on the right hand side, while the anticodon-binding domain of ArgRS is located on the left hand side (Figure 3). The two positions are separated by ~ 38 Å. The structural feature determining these alternative locations seems to be related to the orientation of helix H17–H18 relative to strand S14. This different positioning can be accommodated easily by a rotation of the tRNA along an axis parallel to the helical axis of the acceptor arm of the tRNA.

It is remarkable that MetRS, with a related anticodon-binding domain, has not retained a domain similar to Add-1 for helping in tRNA selection. The extra lock used by ArgRS may also be correlated to the numerous anticodons recognized by ArgRS (six codons) in comparison with MetRS (one codon). The three anticodon bases of tRNA^{Met} are identity determinants for the aminoacylation reaction by MetRS, which means that MetRS strongly discriminates its cognate tRNA at the anticodon level (Muramatsu *et al.*, 1988; Schulman and Pelka, 1988; Meinnel *et al.*, 1991a,b). In contrast, arginine identity at the anticodon level depends mainly on nucleotide C35. Several residues exposed at the surface of the helices seem to be good candidates for nucleic acid recognition and will be probed by site-directed mutagenesis. The crucial role of helices H17–H18 and H22 in tRNA recognition has already been probed by *in vivo* selection of enzyme mutations lethal for cell growth; seven out of 29 such mutations map to those two helices. A detailed structural analysis of these mutations will be published elsewhere.

Materials and methods

Gene expression and protein purification

The gene encoding yeast cytoplasmic ArgRS was isolated from genomic DNA by PCR amplification. The two oligonucleotides used for the experiment were designed in order to clone the gene (termed *RRS1*) into the *Nco*I site of the pTrc99 vector behind the strong *trc* promoter (Aman *et al.*, 1988). This construct led to the expression of a non-fusion protein with the authentic amino acid sequence. The *E. coli* strain TBI [*F*[−] *ara* Δ(*lac-proAB*) *hsdR* (*rk*− *mk*+) *rpsL*(*Str*) (φ80, *dlac*Δ(*lacZ*)M15] was transformed by the resulting vector pTrc99-RRS1. The moderate level of *RRS1* expression was improved by co-transformation with pSBETa, a vector carrying the *E. coli argU* gene (Schenk *et al.*, 1995). This gene encodes the rare tRNA^{Arg} decoding the AGA and AGG arginine codons rarely used in *E. coli*, but used at much higher frequencies in eukaryotic coding sequences (72% of the arginine codons are AGG or AGA codons in *RRS1*). The presence of the 'helper' tRNA in the *E. coli* cells increases the level of expression of *RRS1* and the proportion of soluble ArgRS. Gene expression and protein purification then followed a protocol already published (Sissler *et al.*, 1997).

Crystallization and data collection

Initial crystallization conditions were found by screening different precipitating agents for ArgRS alone and in the presence of the other substrates (L-arginine and ATP). Crystals initially were only obtained in the presence of L-arginine and ATP. The best crystals were grown at 17°C by the hanging drop vapour diffusion method against a reservoir containing 2.45 M ammonium sulfate pH 7, 100 mM Tris–HCl buffer pH 7.0. For the final setup, 4 μl of reservoir solution were mixed with 4 μl of the protein solution. This protein solution contains 13.5 mg/ml ArgRS, 5 mM L-arginine, 5 mM ATP, 10 mM MgSO₄ in 50 mM Tris–HCl buffer pH 7. Crystals grow after a few days. They belong to space group P4₃2₁2 with unit cell constants $a = b = 100.347$ Å, $c = 204.34$ Å and with one molecule of ArgRS in the asymmetric unit, resulting in a solvent content of 63%. X-ray diffraction data were collected at 100 K on a 30 cm Mar-Research image plate at the station BW7B of the EMBL Hamburg outstation at DESY ($\lambda = 1.1024$ Å). The data collection statistics are presented in Table I. Each data set was collected from a single frozen crystal. The crystals were prepared for cryocooling by transferring them first into solutions of mother liquor containing 20% glycerol for 2 min and then plunged into liquid ethane; they were inserted just before data collection into a nitrogen gas flow at 100 K. The native data set between 30 and 2.75 Å was obtained from 215 frames of 0.5° oscillations, split into two runs in order to collect a complete data set at low and high resolution (125 frames with a crystal–image plate distance of 350 mm, and 90 frames with a crystal–image plate distance of 450 mm). The native data set was 99.4% complete in the resolution range 30–2.75 Å. Two heavy atom derivatives, PCMBs (*p*-chloromercuribenzenesulfonate) and K₂Hg(CN)₄ (potassium tetracyano mercurate) were collected and used for the phasing process. For PCMBs, the anomalous signal was measured by collecting two sets of 100 frames (0.5° oscillation) separated by a rotation angle of 180°. Data were processed with the programs DENZO and SCALEPACK (Otwinowski and Minor, 1997).

Phasing

Heavy atom-binding sites were determined by difference Patterson and difference Fourier maps. The two derivatives have three heavy atom sites which correspond to the three cysteine residues of yeast ArgRS. Refinement of the heavy atom parameters was performed by the maximum likelihood approach as coded in the program MLPHARE (CCP4, 1994). The calculated phases gave an overall figure of merit of 0.51 for data between 30 and 3.75 Å resolution. The anomalous signal of the PCMBs derivative allowed the determination of the correct space group (P4₃2₁2 versus P4₁2₁2) and the correct hand of the heavy atom derivatives. An initial 3.8 Å map was improved by solvent flattening, using a solvent content of 60%, solvent flipping and histogram matching using the programs SOLOMON (CCP4, 1994) and DM (Cowtan, 1994). The calculated phases were then extended to 3.5 Å resolution using SOLOMON. All crystallographic calculations were carried out with the CCP4 package (CCP4, 1994).

Model building and refinement

The modified 3.5 Å MIRAS map allowed chain tracing for most parts of the polypeptide chain, and a polyaniline model of the ArgRS molecule was built using the program O (Jones *et al.*, 1991). Phases obtained by combining the MIRAS phases with those derived from the polyaniline

model using the program SIGMAA (Read, 1986) were then modified at 3.5 Å by solvent flattening. This phase modification process combined with model building was repeated several times. The amino acid sequence could be fitted for most of the chain to give an initial model which included 603 residues. The model was refined with the program CNS (Brünger *et al.*, 1998), using the Engh and Huber stereochemical parameters (Engh and Huber, 1991). Initially, only 3.2 Å resolution data were included for torsion angle molecular dynamics refinement using a cross-validated maximum likelihood crystallographic target. Inclusion of the bulk solvent correction, as implemented in CNS, allowed the use of all low resolution collected data. The crystallographic *R*-factor for the starting model was 38.8%. A random sample containing 7.5% of the data was excluded from the refinement and used for monitoring the course of the refinement (Brünger, 1992). Torsion angle refinement at 3.2 Å was followed by rounds of model building in cross-validated σ -weighted maps with coefficient ($2F_o - F_c$) and ($3F_o - 2F_c$) (Kleywegt and Jones, 1994; Kleywegt and Brünger, 1996) and by torsion angle refinement in which the resolution of the data was increased gradually. In the latest stage, cartesian coordinate refinement was followed by individual *B*-factor refinement. All rebuilding and graphics operations were done with O and related Uppsala programs. At every stage, models resulting from a refinement round were subjected to critical quality analyses, using the programs O, OOPS (Kleywegt and Jones, 1996), PROCHECK (Laskowski *et al.*, 1993) and WHATIF (Vriend, 1990).

The refined model

The current model contains one ArgRS molecule of 603 residues, one L-arginine substrate and 280 water molecules. The quality of the refined structure was assessed using the Biotech validation suite for protein structures (Wodak *et al.*, 1995). The crystallographic *R*-factor is 19.7% using all reflections between 20 and 2.75 Å with no sigma cut off ($R_{\text{free}} = 26.9\%$). The stereochemistry of the model was inspected by PROCHECK (see Table I for detailed analysis). The average *B*-factor of the model is 53.3 Å² except for two surface loops (residues 235–242 and residues 252–261) where *B*-factors exceed 100 Å². This is in agreement with the overall *B*-factor determined by Wilson plot on the collected data ($B = 43$ Å²). The high *B*-factor may be explained by the high solvent content in the crystal and the few packing interactions between the molecules in the crystal lattice. The atomic coordinates have been deposited at the Brookhaven Protein Data Bank.

Acknowledgements

We thank Professor Axel T. Brünger (Yale University) for allowing us to use a pre-release version of the CNS software system and Dr John G. Arnez for careful reading of the manuscript. We thank the staff of the EMBL Outstation at DESY for use of their synchrotron instrumentation (beam line BW7B). This work was supported by grants from CNRS and by EEC contracts.

References

- Alexandrov, N.N. (1996) Sarfing the PDB. *Protein Eng.*, **9**, 727–732.
- Aman, E., Ochs, B. and Abel, K.J. (1988) Tightly regulated *tac* promoter vectors useful for the expression of unfused and fused proteins in *E. coli*. *Gene*, **69**, 301–305.
- Arnez, J.G. and Cavarelli, J. (1997) Structures of RNA-binding proteins. *Q. Rev. Biophys.*, **30**, 195–240.
- Arnez, J.G., Harris, D.C., Mitschler, A., Rees, B., Francklyn, C.S. and Moras, D. (1995) Crystal structure of histidyl-tRNA synthetase from *Escherichia coli* complexed with histidyl-adenylate. *EMBO J.*, **14**, 4143–4155.
- Barker, D.G. and Winter, G. (1982) Conserved cysteine and histidine residues in the structures of the tyrosyl- and methionyl-tRNA synthetases. *FEBS Lett.*, **145**, 191–193.
- Barton, G.J. (1993) ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.*, **6**, 37–40.
- Belrhali, H., Yaremchuk, A., Tukalo, M., Berthet-Colominas, C., Rasmussen, B., Bosecke, P., Diat, O. and Cusack, S. (1995) The structural basis for seryl-adenylate and Ap4A synthesis by seryl-tRNA synthetase. *Structure*, **3**, 341–352.
- Brick, P., Bhat, T.N. and Blow, D.M. (1989) Structure of tyrosyl-tRNA synthetase refined at 2.3 Å resolution. Interaction of the enzyme with the tyrosyl adenylate intermediate. *J. Mol. Biol.*, **208**, 83–98.
- Brünger, A.T. (1992) Free-*R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–474.
- Brünger, A.T. *et al.* (1998) Crystallography and NMR system (CNS): a new software system for macromolecular structure determination. *Acta Crystallogr.*, **D**, in press.
- Brunie, S., Zelwer, C. and Risler, J.-L. (1990) Crystallographic study at 2.5 Å resolution of the interaction of methionyl-tRNA synthetase from *Escherichia coli* with ATP. *J. Mol. Biol.*, **216**, 411–424.
- Burbaum, J.J. and Schimmel, P. (1991) Structural relationships and the classification of aminoacyl-tRNA synthetases. *J. Biol. Chem.*, **266**, 16965–16968.
- Cavarelli, J. *et al.* (1994) The active site of yeast aspartyl-tRNA synthetase: structural and functional aspects of the aminoacylation reaction. *EMBO J.*, **13**, 327–337.
- CCP4 (1994) Collaborative Computational Project No. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.
- Cowan, K. (1994) dm: an automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, **31**, 34–38.
- Cusack, S., Berthet-Colominas, C., Härtlein, M., Nassar, N. and Leberman, R. (1990) A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase. *Nature*, **347**, 249–255.
- Dideberg, O., Charlier, P., Dive, G., Joris, B., Frère, J.M. and Ghuyens, J.M. (1982) Structure of a Zn²⁺ containing D-alanyl-D-alanine cleaving carboxypeptidase at 2.5 Å resolution. *Nature*, **299**, 469–470.
- Doublé, S., Bricogne, G., Gilmore, C. and Carter, C.J. (1995) Tryptophanyl-tRNA synthetase crystal structure reveals an unexpected homology to tyrosyl-tRNA synthetase. *Structure*, **3**, 17–31.
- Ebel, J.P., Giegé, R., Bonnet, J., Kern, D., Befort, N., Bollack, C., Fasiolo, F., Gangloff, J. and Dirheimer, G. (1973) Factors determining the specificity of the tRNA aminoacylation reaction. *Biochimie*, **55**, 547–557.
- Engh, R.A. and Huber, R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr.*, **A47**, 392–400.
- Eriani, G., Delarue, M., Poch, O., Gangloff, J. and Moras, D. (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature*, **347**, 203–206.
- Evans, S.V. (1993) SETOR: hardware lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graphics*, **11**, 134–138.
- Fersht, A.R., Gangloff, J. and Dirheimer, G. (1978) Reaction pathway and rate-determining step in the aminoacylation of tRNA^{Arg} catalyzed by the arginyl-tRNA synthetase from yeast. *Biochemistry*, **17**, 3740–3746.
- Francklyn, C., Musier-Forsyth, K. and Martinis, S.A. (1997) Aminoacyl-tRNA synthetases in biology and disease: new evidence for structural and functional diversity in an ancient family of enzymes. *RNA*, **3**, 954–960.
- Gangloff, J., Ebel, J.P. and Dirheimer, G. (1973) Isolation of a complex between yeast arginyl-tRNA synthetase and yeast tRNA^{Asp} and mischarging of tRNA^{Asp} with arginine. *Int. Res. Commun. System*, **12**, 3–8.
- Gerlo, E., Freist, W. and Charlier, J. (1982) Arginyl-tRNA synthetase from *Escherichia coli* K12: specificity with regard to ATP analogs and their magnesium complexes. *Hoppe Seyler's Z. Physiol. Chem.*, **363**, 365–373.
- Holm, L. and Sander, C. (1994) The FSSP database of structurally aligned protein fold families. *Nucleic Acids Res.*, **22**, 3600–3609.
- Hutchinson, E.G. and Thornton, J.M. (1996) PROMOTIF—a program to identify and analyze structural motifs in proteins. *Protein Sci.*, **5**, 212–220.
- Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.
- Kern, D. and Lapointe, J. (1980) The catalytic mechanism of glutamyl-tRNA synthetase of *Escherichia coli*. Evidence for a two-step aminoacylation pathway and study of the reactivity of the intermediate complex. *Eur. J. Biochem.*, **106**, 137–150.
- Kleywegt, G.J. and Brünger, A.T. (1996) Checking your imagination. Applications of the free *R* value. *Structure*, **4**, 897–904.
- Kleywegt, G.J. and Jones, T.A. (1994) Crystal structures of cellular retinoic acid binding proteins I and II in complex with all-*trans*-retinoic acid and a synthetic retinoid. *Structure*, **2**, 1241–1258.
- Kleywegt, G.J. and Jones, T.A. (1996) Efficient rebuilding of protein structures. *Acta Crystallogr.*, **D52**, 829–832.

- Kraulis, P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
- Landès, C., Perona, J.J., Brunie, S., Rould, M.A., Zelwer, C., Steitz, T.A. and Risler, J.L. (1995) A structure-based multiple sequence alignment of all class I aminoacyl-tRNA synthetases. *Biochimie*, **77**, 194–203.
- Laskowski, R.A., Mac Arthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structure. *J. Appl. Crystallogr.*, **26**, 283–291.
- Lindahl, M. et al. (1994) Crystal structure of the ribosomal protein S6 from *Thermus thermophilus*. *EMBO J.*, **13**, 1249–1254.
- Lofffield, R.B. (1972) The mechanism of aminoacylation of transfer RNA. *Prog. Nucleic Acid Res. Mol. Biol.*, **12**, 87–128.
- McClain, W.H. and Foss, K. (1988) Changing the acceptor identity of a transfer RNA by altering nucleotides in a 'variable pocket'. *Science*, **241**, 1804–1807.
- McClain, W.H., Foss, K., Jenkins, R.A. and Schneider, J. (1990) Nucleotides that determine *Escherichia coli* tRNA^{Arg} and tRNA^{Lys} acceptor identities revealed by analyses of mutant opal and amber suppressor tRNAs. *Proc. Natl Acad. Sci. USA*, **87**, 9260–9264.
- Mehler, A.H. and Mitra, S.K. (1967) The activation of arginyl transfer ribonucleic acid synthetase by transfer ribonucleic acid. *J. Biol. Chem.*, **242**, 5495–5499.
- Meinell, T., Mechulam, Y., Blanquet, S. and Fayat, G. (1991a) Binding of the anticodon domain of tRNA^{Met} to *Escherichia coli* methionyl-tRNA synthetase. *J. Mol. Biol.*, **220**, 205–208.
- Meinell, T., Mechulam, Y., LeCorre, D., Panvert, M., Blanquet, S. and Fayat, G. (1991b) Selection of suppressor methionyl-tRNA synthetases: mapping the tRNA anticodon binding site. *Proc. Natl Acad. Sci. USA*, **88**, 291–295.
- Meinell, T., Mechulam, Y. and Blanquet, S. (1995) Aminoacyl-tRNA synthetase: occurrence, structure and function. In Soll, D. and RajBhandry, U.L. (eds), *tRNA: Structure, Biosynthesis and Function*. ASM Press, Washington, DC, pp. 251–291.
- Mitra, S.K. and Smith, C.J. (1969) Absolute requirement for transfer RNA in the activation of arginine by arginyl transfer RNA synthetase of yeast. *Biochim. Biophys. Acta*, **190**, 222–224.
- Muramatsu, T., Nishikawa, K., Nemoto, F., Kuchino, Y., Nishimura, S., Miyazawa, T. and Yokoyama, S. (1988) Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. *Nature*, **336**, 179–181.
- Nagai, K., Oubridge, C., Jessen, T.H., Li, J.L. and Evans, P.R. (1990) Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature*, **348**, 515–520.
- Nicholls, A. and Honig, B. (1991) A rapid finite difference algorithm utilizing successive over-relaxation to solve the Poisson–Boltzmann equation. *J. Comp. Chem.*, **12**, 435–445.
- Nureki, O., Vassilyev, D.G., Katayanagi, K., Shimizu, T., Sekine, S., Kigawa, T., Miyazawa, T., Yokoyama, S. and Morikawa, K. (1995) Architectures of class defining and specific domains of glutamyl-tRNA synthetase. *Science*, **267**, 1958–1965.
- Nureki, O. et al. (1998) Enzyme structure with two catalytic sites for double-sieve selection of substrate. *Science*, **280**, 578–582.
- Onesti, S., Miller, A.D. and Brick, P. (1995) The crystal structure of the lysyl-tRNA synthetase (LysU) from *Escherichia coli*. *Structure*, **3**, 163–176.
- Orengo, C.A. and Thornton, J.M. (1993) Alpha plus beta folds revisited: some favoured motifs. *Structure*, **15**, 105–120.
- Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
- Oubridge, C., Nobutoshi, I., Evans, P.R., Teo, C.H. and Nagai, K. (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, **372**, 432–438.
- Poterszman, A., Delarue, M., Thierry, J.-C. and Moras, D. (1994) Synthesis and recognition of aspartyl-adenylate by *Thermus thermophilus* aspartyl-tRNA synthetase. *J. Mol. Biol.*, **244**, 158–167.
- Rath, V.L., Silvian, L.F., Beijer, B., Sproat, B.S. and Steitz, T.A. (1998) How glutamyl-tRNA synthetase selects glutamine. *Structure*, **6**, 439–449.
- Read, R.J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr.*, **A42**, 140–149.
- Rossmann, M.G., Moras, D. and Olsen, K.W. (1974) Chemical and biological evolution of a nucleotide-binding protein. *Nature*, **250**, 194–199.
- Rould, M.A., Perona, J.J., Söll, D. and Steitz, T.A. (1989) Structure of *E. coli* glutamyl-tRNA synthetase complexed with tRNA^{Gln} and ATP at 2.8 Å resolution. *Science*, **246**, 1135–1142.
- Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J.C. and Moras, D. (1991) Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA^{Asp}. *Science*, **252**, 1682–1689.
- Saks, M.E., Sampson, J.R. and Abelson, J. (1998) Evolution of a transfer RNA gene through a point mutation in the anticodon. *Science*, **279**, 1665–1670.
- Sawaya, M.R., Pelletier, H., Kumar, A., Wilson, S.H. and Kraut, J. (1994) Crystal structure of rat DNA polymerase β: evidence for a common polymerase mechanism. *Science*, **264**, 1930–1935.
- Schenk, P.M., Bauman, S., Mattes, R. and Steinbib, H.H. (1995) Improved high-level expression system for eukaryotic genes in *Escherichia coli* using T7 RNA polymerase and rare Arg tRNAs. *BioTechniques*, **19**, 196–200.
- Schulman, L.H. and Pelka, H. (1988) Anticodon switching changes the identity of methionine and valine transfer RNAs. *Science*, **242**, 765–768.
- Schulman, L.H. and Pelka, H. (1989) The anticodon contains a major element of the identity of arginine transfer RNAs. *Science*, **246**, 1595–1597.
- Singh, R. and Green, M.R. (1993) Sequence-specific binding of transfer RNA by glyceraldehyde-3-phosphate dehydrogenase. *Science*, **259**, 365–368.
- Sissler, M., Eriani, G., Martin, F., Giegé, R. and Florentz, C. (1997) Mirror image alternative interaction patterns of the same tRNA with either class I arginyl-tRNA synthetase or class II aspartyl-tRNA synthetase. *Nucleic Acids Res.*, **25**, 4899–4906.
- Starzyk, R.M., Webster, T.A. and Schimmel, P. (1987) Evidence for dispensable sequences inserted into a nucleotide fold. *Science*, **237**, 1614–1618.
- Tamura, K., Himeno, H., Asahara, H., Hasegawa, T. and Shimizu, M. (1992) *In vitro* study of *E. coli* tRNA^{Arg} and tRNA^{Lys} identity elements. *Nucleic Acids Res.*, **20**, 2335–2339.
- Tanaka, Hall, T.M., Porter, J.A., Beachy, P.A. and Leahy, D.J. (1995) A potential catalytic site revealed by the 1.7 Å crystal structure of the amino-terminal signalling domain of Sonic hedgehog. *Nature*, **378**, 212–216.
- Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
- Van Duyn, G.D., Ghosh, G., Maas, W.K. and Sigler, P.B. (1996) Structure of the oligomerization and L-arginine binding domain of the arginine repressor of *Escherichia coli*. *J. Mol. Biol.*, **256**, 377–391.
- Vriend, G. (1990) WHATIF: a molecular modelling and drug design program. *J. Mol. Graphics*, **8**, 52–56.
- Webster, T., Tsai, H., Kula, H.T.M., Mackie, G.A. and Schimmel, P. (1984) Specific sequence homology and three-dimensional structure of an aminoacyl transfer RNA synthetase. *Science*, **226**, 1315–1317.
- Wodak, S.J., Pontius, J., Vaguine, A. and Richelle, J. (1995) Validating protein structures. From consistency checking to quality assessment. In *Proceedings of the CCP4 Daresbury Study Weekend: Making the Most of Your Model*. SERC Daresbury Laboratory, Warrington, UK, pp. 41–51.

Received June 16, 1998; revised and accepted July 13, 1998