

74. C. D. Livingston and G. J. Barton, *Int. J. Peptide Protein Res.*, in press.
75. A. Nicholls, R. Bharadwaj, B. Honig, *Biophys. J.* **64**, 166 (1991).
76. PTP domains were aligned by means of the multiple sequence alignment program PILEUP in the GCG software package of the Genetics Computer Group, Inc. (Madison, WI) using default parameters. The sequence identity between PTP1B and each of the other PTPs in their core catalytic domain, amino acids 31–280 of PTP1B, was calculated with the BESTFIT program. The mean overall identity to PTP1B is 40 percent. A measurement of the identity at each amino acid position in the multiple sequence alignment was obtained using the PLOTSIMILARITY program with the "identity" restriction applied. The identity score at each position is the average of the scores resulting from comparison of all possible pairs of sequences with an identical pair receiving a score of 1 and a non-identical pair receiving score 0. Identity scores of 1.0 reveal invariant amino acids and scores between 0.8 and 1.0 are defined here as "highly conserved." Each "highly conserved" amino acid is found in 24 or more of these aligned PTP domains. The data base accession numbers of the PTPs and their identities to PTP/B were hPTP1B (A33897, PIR, 100%), hTCPTP (P17706, SwissProt, 75%), DPTP61F (L11251, GenBank, 56%), hMEG1 (M68941, GenBank, 42%), hPTPH1 (P26045, SwissProt, 41%), hMEG2 (M83738, GenBank, 41%), hPTP-PEST (A45496, PIR, 38%), mPTP-PEP (M90388, GenBank, 39%), hSH-PTP1 (P29350, SwissProt, 40%), hSH-PTP2 (A46210, PIR, 40%), DdPTPa (L15420, GenBank, 47%), DdPTP1 (A44267, PIR, 35%), SpPYP1 (P27574, SwissProt, 34%), SpPYP2 (A45030, PIR, 37%), SpPYP3 (X69994, GenBank, 37%), ScPTP1 (P25044, SwissProt, 35%), hCD45-D1 (P08575, SwissProt, 38%), hPTP α -D1 (P18433, SwissProt, 38%), hLAR-D1 (Y00815, GenBank, 43%), hPTP δ -D1 (P23468, SwissProt, 41%), rPTP σ -D1 (L19180, GenBank, 43%), DLAR-D1 (M27700, GenBank, 38%), hPTP μ (P28827, SwissProt, 38%), mPTP κ -D1 (L10106, GenBank, 40%), hPTP ζ -D1 (P23471, SwissProt, 39%), DPTP (P16620, SwissProt, 33%), hPTP β (P23467, SwissProt, 42%).
77. Supported by the Robertson Research Fund (D.B.), NIH grant CA53840 (N.K.T.), and a Pew Scholar award in the Biomedical Sciences (N.K.T.), a Cancer Research Institute and F. M. Kirby Foundation Fellowship (A.J.F.). We thank J. W. Pflugrath for advice and discussions; X. Cheng, W. Furey, and S. Gambelin for discussions; B. Stillman and J. D. Watson for encouragement and support; E. Westbrook, R. Alkire, S. L. Ginell, L. J. Keefe, and J. Lazaiz for assistance and for access to beamline X8C at the National Synchrotron Light Source, Brookhaven National Laboratory; G. J. Barton for access to unpublished data; and H. Charbonneau, J. Horton, J. W. Pflugrath, and B. Stillman for comments on the manuscript. The atomic coordinates have been submitted to the Brookhaven protein data base.

25 January 1994; accepted 24 February 1994

in the catalytic domain of class 1 synthetase—exemplified by the crystal structures of the tyrosyl- (2), methionyl- (3) and glutamyl- (4) tRNA synthetases) and an antiparallel fold in the catalytic domain of class 2 synthetases as seen in the crystal structures of the seryl- (5) and aspartyl- (6) tRNA synthetases. Apart from the catalytic domain, aminoacyl-tRNA synthetases contain other domains, often as COOH- or NH₂-terminal extensions. These putative tRNA binding domains show more idiosyncratic sequence and structural variability than the catalytic domains although, within subclasses, such domains can be homologous (7).

Of particular interest are the determinants of the specific recognition between a synthetase and its cognate tRNAs. For many systems, tRNA identity elements have been located (8); that is, nucleotides that are crucial to the specific recognition and whose transplantation into other tRNAs can change their amino acid identity. In most cases, the identity elements include some or all of the anticodon nucleotides (the exceptions being the tRNAs for serine, alanine, and leucine in *Escherichia coli*). However other nucleotides, often in the acceptor stem, can be important for identity as well. The two published crystal structures of synthetase-tRNA complexes, the glutamine system (4, 9) from *E. coli* (class 1) and the aspartic acid system (6, 10) from yeast (class 2) reveal examples of specific anticodon recognition. In both cases an additional domain is responsible, but the fold of this domain and the distortion induced in the anticodon stem and loop of the tRNA are different (9, 10). Similarly, the mode of entry of the 3' end of the tRNA into the active site is not the same in the two systems (6) with the net result, presumably, being the correct positioning of the 2' OH (class 1) or 3' OH (class 2) of the terminal ribose for receiving the amino acid (1).

In view of the idiosyncracies of each synthetase system, it is necessary to obtain detailed structural information about other tRNA-synthetase complexes in order to identify common and distinct features. The seryl-tRNA synthetase, an α_2 dimeric class 2 synthetase, warrants inspection for several reasons. First, because of the variety of serine codons (from two distinct codon groups), it is not surprising that the anticodon is not recognized by the synthetase (11–14); this is consistent with the existence of a serine amber suppressor tRNA and the opal suppressor tRNA^{SerCys}, both of which are specifically charged by seryl-tRNA synthetase. Second, most tRNA^{Ser} moieties are peculiar in having a long variable arm of up to 20 nucleotides, a feature shared in prokaryotic systems only by tRNA^{Leu} and tRNA^{Tyr}. To our knowledge,

The 2.9 Å Crystal Structure of *T. thermophilus* Seryl-tRNA Synthetase Complexed with tRNA^{Ser}

Valérie Biou, Anna Yaremchuk, Michael Tukalo, Stephen Cusack*

The crystal structure of *Thermus thermophilus* seryl-transfer RNA synthetase, a class 2 aminoacyl-tRNA synthetase, complexed with a single tRNA^{Ser} molecule was solved at 2.9 Å resolution. The structure revealed how insertion of conserved base G20b from the D loop into the core of the tRNA determines the orientation of the long variable arm, which is a characteristic feature of most serine specific tRNAs. On tRNA binding, the antiparallel coiled-coil domain of one subunit of the synthetase makes contacts with the variable arm and T Ψ C loop of the tRNA and directs the acceptor stem of the tRNA into the active site of the other subunit. Specificity depends principally on recognition of the shape of tRNA^{Ser} through backbone contacts and secondarily on sequence specific interactions.

Aminoacyl-tRNA synthetases perform an essential step in assuring the accuracy of protein synthesis by specifically ligating amino acids to their cognate tRNA isoacceptors. This occurs in a two-step reaction in which the amino acid is first activated by adenosine triphosphate (ATP) to form the aminoacyl-adenylate intermediate and is subsequently transferred to the 3' terminal ribose of the tRNA. Considerable progress

has been made in understanding the structural basis of specificity by application of sequence analyses, site-directed mutagenesis of tRNAs and synthetases, chemical probing of synthetase-tRNA complexes and x-ray crystallography. On the basis of short conserved primary sequence motifs the 20 aminoacyl-tRNA synthetases have been partitioned into two classes, each containing 10 members (1). This has largely resolved the problem of the apparent wide diversity of synthetase primary and quaternary structures. From the known crystal structures of five synthetases it is now clear that this partition corresponds in structural terms to the existence of a Rossmann fold

The authors are at the European Molecular Biology Laboratory, Grenoble Outstation, c/o ILL, 156X, 38042 Grenoble Cedex, France. A. Yaremchuk and M. Tukalo are on leave from the Institute of Molecular Biology and Genetics, Kiev, Ukraine.

*To whom correspondence should be addressed.

no atomic structure of a long variable arm tRNA has yet been determined, although various models have been proposed (15, 16). In *E. coli*, a minimal length and the correct orientation of this arm is crucial for aminoacylation by seryl-tRNA synthetase, although the sequence seems to be less critical (12–14). Other identity elements of *E. coli* tRNA^{Ser} have been located in the acceptor stem and D stem (C11–G24) (11–13). Third, the crystal structures of the uncomplexed seryl-tRNA synthetase from *E. coli* (5) and *T. thermophilus* (17) have both been determined at 2.5 Å resolution. The *T. thermophilus* enzyme has 421 residues per subunit with a primary sequence identity of 34 percent to the *E. coli* enzyme and the three-dimensional structures are similar (17). The catalytic domain is based on a seven-stranded antiparallel β sheet with two connecting helices; a similar fold was subsequently found in the aspartyl-tRNA synthetase (6). In addition, seryl-tRNA synthetase has an NH₂-terminal domain comprising an exceptionally long, solvent exposed, antiparallel coiled-coil (the helical arm).

We describe the crystal structure at 2.9 Å of a complex between seryl-tRNA synthetase and tRNA^{Ser}(GGA) from *T. thermophilus* refined to a crystallographic *R* factor of 19.4 percent (Fig. 1). The stoichiometry of tRNA to synthetase dimer in this complex is 1:1. However it is known from both solution studies and from crystal structures of different crystal forms of the seryl-

tRNA synthetase-tRNA^{Ser} complex from both *T. thermophilus* (18) and *E. coli* (19) that the synthetase can simultaneously bind two tRNA molecules. About one-third of the single tRNA is not visible in the electron density map. The unobserved regions include the anticodon stem and loop and the variable arm loop, neither of which are in contact with the synthetase nor make crystal contacts. However, the 3' and 5' ends of the tRNA are also not visible so that no details are revealed about the interaction of the end of the acceptor stem with the active site of the enzyme. Nevertheless, the structure yields considerable information on the interactions of the antiparallel coiled-coil of the synthetase with the tRNA and on the tertiary interactions within the tRNA core that determine the orientation of the tRNA long variable arm. The main conclusions can be summarized as follows. (i) The tRNA binds across the two subunits of the dimer; (ii) upon tRNA binding the otherwise flexible, coiled-coil domain of the synthetase is stabilized in a particular orientation and curves between the TΨC loop and the long variable arm of the tRNA; (iii) the synthetase makes several backbone contacts but few base-specific interactions; (iv) contacts with the tRNA long variable arm backbone extend out to the sixth base pair, explaining the need for a minimum length of the arm, but allowing longer arms (as, for instance, in tRNA^{Ser}_{Cys}) to be accommodated; and (v) the bases 20a and 20b inserted into the D loop in the tRNA^{Ser} have unusual roles in the core of the tRNA. In particular, the base of G20b is stacked against the first base pair of the long variable arm and thus determines the orientation of the variable arm. These results are discussed below in the light of known features of the recognition between tRNA^{Ser} and seryl-tRNA synthetase. We also show the binding site within the complex of AMPPCP, an ATP analog with a methylene group between the β- and γ-

phosphates, as determined from crystallographic measurements at 3.1 Å resolution.

Structure determination. Several crystal forms of the complex between wild-type seryl-tRNA synthetase and tRNA^{Ser} from *T. thermophilus* have been obtained (18, 20). The highest resolution diffraction is given by crystal form 4, which is of space-group *P*2₁2₁2₁ and unit cell dimensions *a* = 124.5 Å, *b* = 128.9 Å, *c* = 121.2 Å. A data set with 41,675 independent reflections, 6.7 percent *R*_{merge} and 95 percent complete to 2.9 Å was obtained as described (18). A second data set was obtained from three crystals that had been soaked with 10 mM AMPPCP; it comprises 31,042 independent measurements to 3.1 Å (85 percent complete) with *R*_{merge} of 12.0 percent (25.2 percent between 3.1 and 3.2 Å). Data integration was performed with MOSFLM (21) and subsequent data processing was done with the CCP4 package (22).

The structure of uncomplexed seryl-tRNA synthetase from *T. thermophilus* (17) was used as a starting model for molecular replacement as described (18). An electron density map calculated with phases from the synthetase alone [SIGMAA (23)], showed good extra density for about half of a tRNA molecule. It also showed that the asymmetric unit contains one protein dimer bearing one single tRNA, the other, symmetrically related, potential tRNA site on the synthetase being blocked by crystal contacts. The protein model was improved by rigid-body refinement and by positional energy minimization by means of XPLOR (24), which resulted in a reduction of the crystallographic *R* factor from 38.6 percent to 31 percent. A model of tRNA^{Ser} (24a) with the nucleotide sequence of *T. thermophilus* tRNA₁^{Ser} (25) was then placed in the electron density (at this stage, the sequence of the tRNA₂^{Ser} actually present in the crystals was unknown). The orientations of the different parts of the tRNA were improved by rigid body minimization and sub-

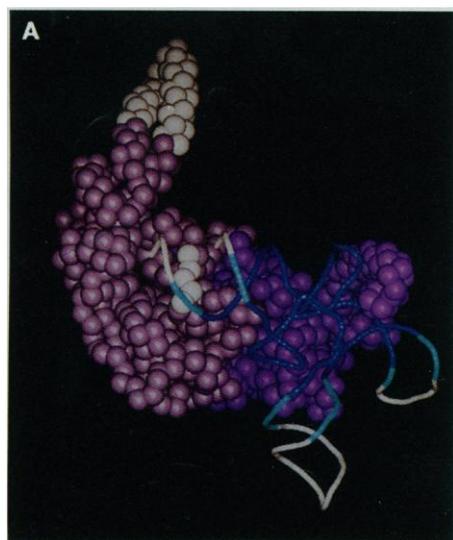


Fig. 1. The seryl-tRNA synthetase-tRNA^{Ser} complex represented by spheres at the C_α positions for the protein and the phosphate backbone for the tRNA. Monomer 1 (residues 1 to 421) is represented in plum with unobserved parts in white, and monomer 2 (residues 501 to 921) is in purple. Well-defined, partially defined, and unobserved regions of the tRNA are, respectively, in blue, cyan, and white (see Fig. 2).

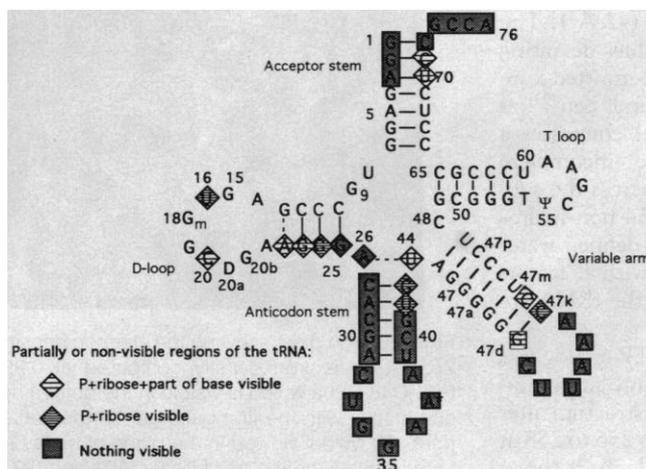
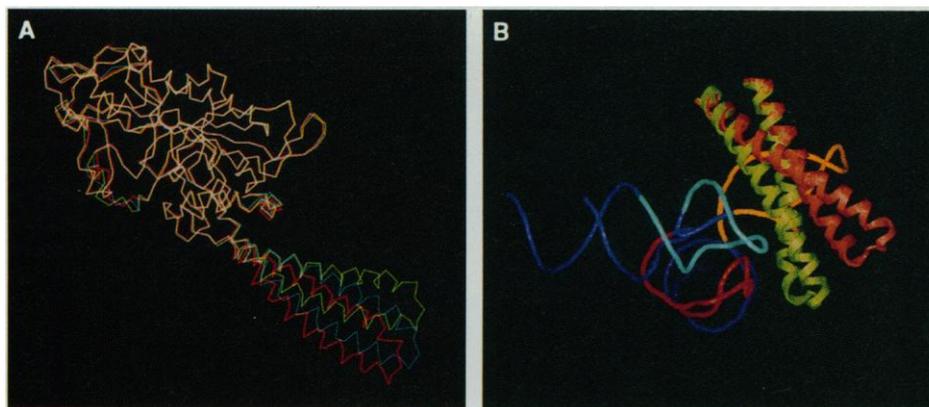


Fig. 2. Cloverleaf representation of the nucleotide sequence of *T. thermophilus* tRNA₂^{Ser}(GGA). The degree of definition of each nucleotide in the final electron density map is indicated. Nucleotides constrained to 2'-endo were G-7, G-9, G-18, G-19, C-20, G-20b, C-48, and A-58, U-60; all others are 3'-endo.

Fig. 3. (A) Superposition of C α traces of monomer 2 of the complex (green) and the two monomers of the uncomplexed seryl-tRNA synthetase (orange and blue) showing the variable orientation of the NH₂-terminal antiparallel coiled-coil domain. (B) Diagram illustrating in green the curvature of the helical arm of the synthetase around the tRNA in the complex compared to its position in one of the subunits of the uncomplexed synthetase (orange). The D loop (red), variable loop (yellow), and T Ψ C loop (cyan) are shown.



sequently by a series of refinement cycles involving energy minimization with XPLOR, map calculations with Fourier coefficients derived from SIGMAA (23), and model building with the graphics program O (26). Toward the end of the refinement, the sequence of the tRNA_{2^{Ser}} became known (25); it has 94 nucleotides and a GGA anticodon (Fig. 2). Ribose conformations were constrained to either 3'-endo or 2'-endo with dihedral angle constraints (Fig. 2), but no constraints were used for base planarity. Occupancies of some of the backbone atoms of the tRNA on the borders of the well-defined regions were fixed at 0.7 or 0.5 since difference maps showed the position of these atoms, but with an occupancy of 1, the B factors were unrealistically high (>100 Å²).

The R factor is 19.4 percent for 39,713 reflections >0.5 σ between 2.9 and 10 Å. The root-mean-square (rms) deviations from ideality for bond lengths and angles are, respectively, 0.015 Å and 3.3° for the protein and 0.014 Å and 3.4° for the regions of the tRNA with occupancy 1.0. The average B factors for backbone (side chain) atoms of monomer 1 and monomer 2 are, respectively, 20 Å² (26 Å²) and 14 Å² (21 Å²), suggesting that the binding of the tRNA mainly to monomer 2 significantly stabilizes the latter. The average B factors for backbone (bases) of the fully occupied parts of the tRNA are 50 Å² (42 Å²). The only residues with phi-psi values deviating significantly from those permitted are equivalent residues Leu³⁵³ and Leu⁸⁵³ as noted (17). The final model comprises a synthetase dimer with one incomplete tRNA molecule interacting across the subunits (Fig. 1). It contains 7754 non-hydrogen atoms including 44 well-defined water molecules that stably refine with B factors <30 Å². The visible parts of the tRNA are indicated in Fig. 2.

The structure with the ATP analog was refined by one round of energy minimization, starting with the above structure after manual adjustment of residues 256 to 258 at the beginning of the motif 2 loop. It shows

one AMPPCP molecule in each active site as described below.

Protein structure. The synthetase subunit whose active site receives the 3' end of the tRNA is referred to as monomer 1 (residues 1 to 421); the coiled-coil domain of monomer 2 (residues 501 to 921) is in contact with the tRNA variable arm and D

and T Ψ C loops (Fig. 1). Monomer 2 has a relatively ordered active site, because (i) electron density into which an adenosine molecule can be fitted is found at the ATP site (see below) and (ii) the motif 2 loop (residues 760 to 766) is stabilized by crystal contacts with the external side of the coiled-coil domain of a symmetry-related

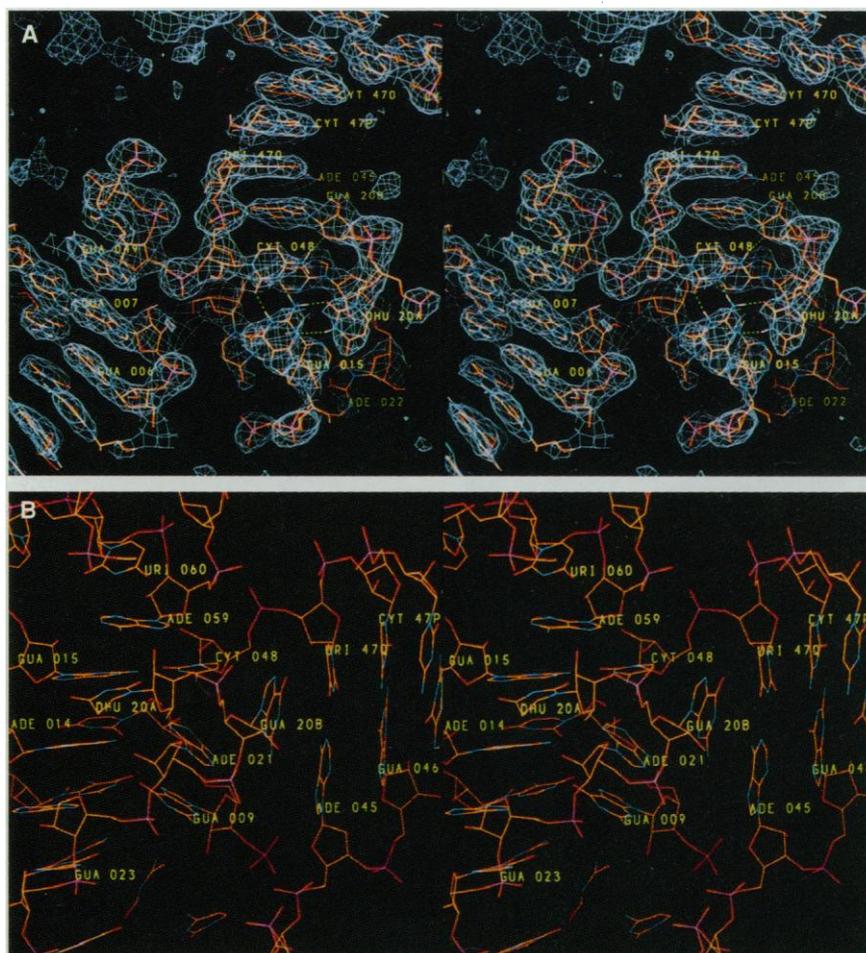


Fig. 4. Stereo pairs showing the electron density and model in the core region of the tRNA. (A) Final $2F_o - F_c$ electron density contoured at 0.95 σ and viewed perpendicular to the base triple G15-C48-D20a which is visible in the bottom right of the figure. G20b is perpendicular to this triple and stacks with the first base pair of the variable arm (A45-U47q) which extends toward the top right. (B) Model viewed in the plane of triple G15-C48-D20a with the variable arm stem extending to the right. Note the tilt of bases A21 and G9 and the insertion of base G20b.

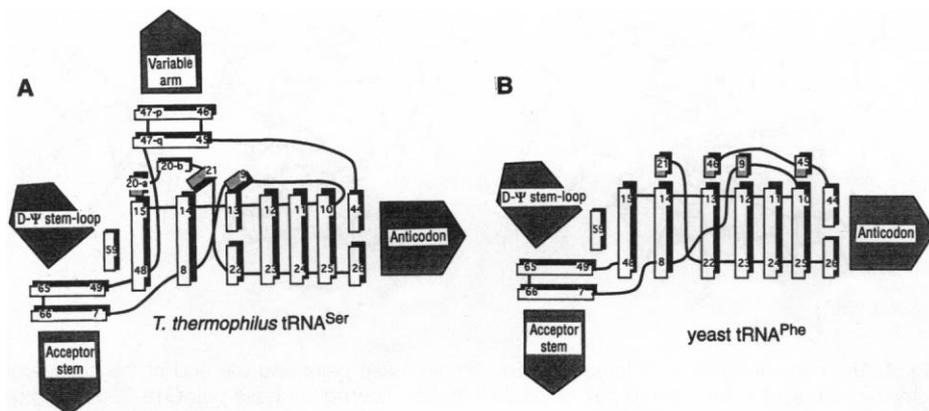


Fig. 5. Comparison of the core structures of (A) tRNA^{Ser} from *T. thermophilus* and (B) yeast tRNA^{Phe}.

molecule. In the active site of monomer 1 there is no interpretable density that could be attributed to the 3' end of the tRNA, and the motif 2 loop (260 to 264) is disordered. Two-thirds of the coiled-coil domain of monomer 1 (residues 36 to 86) are absent from the electron density (although there is room in the crystal for the remainder), consistent with the fact that this arm is flexible in the absence of stabilizing intermolecular contacts.

The seryl-tRNA synthetase structure in the complex is close to that of the uncomplexed synthetase (17), apart from the helical arm domain which has an altered orientation and curvature (Fig. 3A). Pairwise superpositions of the four catalytic domains (that is, residues 100 to 421, excluding the flexible motif 2 loop 258 to 268), two from the complex and two from the native enzyme, show that the rms deviation between the α carbons is about 0.5 Å in each case. These superpositions indi-

cate that the helical arm of monomer 2 in the complex has rotated by, respectively, 23° (19.9 Å displacement at the tip) and 17° (14.7 Å displacement at the tip) compared to monomers 1 and 2 of the uncomplexed synthetase. The new conformation of the arm permits it to fit inside the cleft between the D loop and the variable arm (Fig. 3B). The short helices at the NH₂-terminal move in concert with the arm (Fig. 3A). Flexibility of the helical arm has previously been observed in crystals of the uncomplexed *E. coli* enzyme, where different orientations have been found in a structure determined at 100 K and in a different crystal form (27). The position of the arm in the *E. coli* synthetase-tRNA complex

(19) is very similar to that found here in the *T. thermophilus* complex, showing that the presence of the tRNA locks the arm into a well-defined orientation.

The tRNA structure. Prokaryotic tRNA^{Ser} secondary structures have two features that distinguish them from most other tRNAs: (i) the D loop has a deletion at position 17 and a double insertion of 20a-20b and (ii) there is a long variable arm, which here contains 20 nucleotides (labeled 45, 46, 47, and 47a to 47q) and seven base pairs (Fig. 2). The tRNA^{Ser} tertiary structure observed in the complex (Fig. 4) provides an understanding of the structural consequences of these features by comparison with known tRNA structures such as tRNA^{Phe} (28) and tRNA^{Asp} (29).

The tertiary interactions in the core of the tRNA^{Ser} differ from those observed previously (Figs. 5 and 6). In tRNA^{Phe} and tRNA^{Asp}, the core includes four parallel stacked planes; three of them consist of a base triplet (Fig. 5B) and can be described by the notation (for tRNA^{Phe}) [G15-C48], A21-[U8-A14], 7MG46-[G22-C13] and A9-[A23-U12], where the square brackets indicate the principal base pair (not necessarily Watson-Crick), with the third base participating through additional hydrogen bonding. The equivalent region in tRNA^{Ser} (Fig. 5A) is reorganized to accommodate the insertion 20a-20b from the D loop and can be described by the notation D20a-[G15-C48], A21-[U8-A14], G9-[A22-

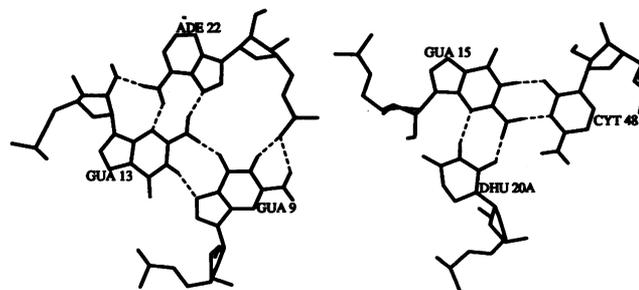


Fig. 6. Two of the base triplets occurring in the core of tRNA^{Ser}. The G9 is tilted out of the plane of G13 and A22 (Fig. 5A). Hydrogen bonds are dotted.

Table 1. Polar tRNA-protein interactions <3.5 Å visible in the complex.

Monomer 1			
Ser-156	N	C-67	O2P
156	OG	C-68	O1P
156	OG	C-68	O2P
Arg-267	NE	U-69	O2P
Monomer 2			
Lys-542	NZ	G-47b	O3'
542	NZ	G-47c	O2P
Gln-545	OE1	G-47a	N2
545	NE2	C-47n	O2
Gln-548	OE1	C-47o	O2'
548	NE2	C-47p	O2P
Thr-549	OG1	C-47n	O2'
Arg-551	NH2	G-57	O2P
Asn-552	ND2	C-47o	O2P
552	ND2	C-47p	O1P
Ala-555	O	G-19	N2
Arg-588	NH2	G-53	O2P
Arg-695	NH2	C-66	O3'
695	NH2	C-67	O2P
Arg-865	NH1	G-46	O2P
865	NH2	G-46	O1P
865	NH2	G-46	O2P

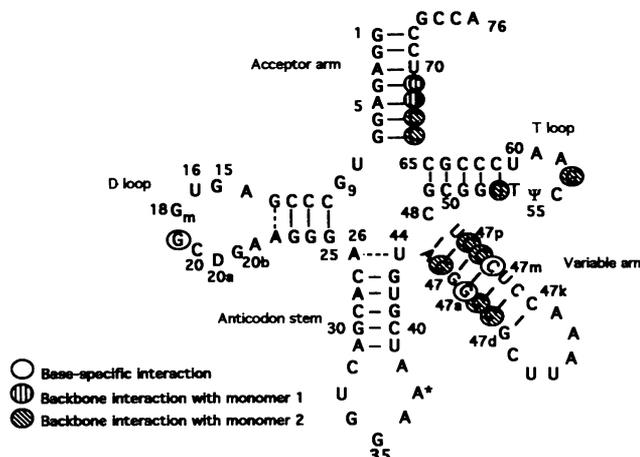


Fig. 7. Cloverleaf representation of the tRNA, showing the regions of interaction with the synthetase.

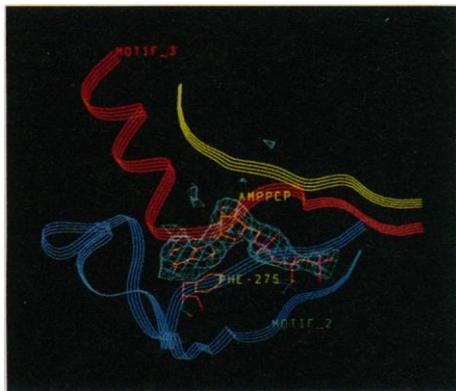


Fig. 10. Omit map showing difference electron density for the AMPPCP molecule in the active site of monomer 2 contoured at 3.5σ . Also shown is the stacking interaction with Phe²⁷⁵, a residue of motif 2 (blue) and the positions of motif 3 (red) and a fourth important β strand in the active site (green), from which a conserved glutamic acid residue (Glu³⁴⁵) interacts with the ribose.

with the variable arm stem, and Arg⁵⁵¹ and Arg⁵⁸⁸ with the T Ψ C loop. Another exposed loop region of the protein (residues 860 to 868), which is highly mobile in the absence of tRNA, contributes an interaction via Arg⁸⁶⁵ to the variable stem backbone. At the base of the acceptor stem, the 3' strand, between phosphates 67 and 68, straddles the two subunits interacting with both Ser¹⁵⁶ (monomer 1) and Arg⁶⁹⁵ (monomer 2) at the dimer interface. Only about half of these cited amino acid residues are conserved between *E. coli* and *T. thermophilus* seryl-tRNA synthetases and only Asn⁵⁵² (which interacts with the phosphate of the second nucleotide preceding 48, in this case C47p) is absolutely conserved among the four known seryl-tRNA synthetase sequences (Fig. 9). Most of the interactions with the tRNA are clustered in the residue range 542 to 555 (7 of 14 interact) on the first helix (H3) of the arm of the synthetase (Fig. 9). This correlates with the fact that comparison between the four known seryl-tRNA synthetase sequences shows that the first long helix (H3) is more conserved than the second long helix (H4) (Fig. 9).

ATP-binding site. The difference electron density map for the AMPPCP molecule bound in the active site of monomer 1 (Fig. 10) shows that the adenosine moiety makes a number of interactions with residues particularly from motif 2; one of them is a stacking interaction with Phe²⁷⁵ which is highly conserved among all class 2 synthetases (1, 7). Arg²⁵⁶, another highly conserved motif 2 residue, interacts with the α -phosphate and Glu³⁴⁵, from another β strand, with the ribose 3' hydroxyl. The triphosphate is in an extended conforma-

tion aligned along the β strands of motif 2 and 3, although recent results strongly suggest that in the active conformation of the enzyme bound Mg²⁺-ATP, the triphosphate is in a bent conformation (30). A more detailed presentation of the interactions of the AMP moiety with the active site derived from higher resolution data has been described by Belrhali *et al.* (30). Binding of the ATP analog in the active site did not significantly improve the visibility of the 3' end of the tRNA although the orientation of the acceptor stem helix was changed slightly.

Comparison with previous biochemical results. The *T. thermophilus* complex structure explains a number of previous results on the interaction of tRNA^{Ser} with seryl-tRNA synthetase. The iodine-phosphorothioate method was used to establish that for the *E. coli* complex (31) there was phosphate protection on the 3' side of the acceptor stem (P67-P70), the 5' side of the T Ψ C-loop (P52-P53) and the 5' side of the variable stem (P46), with the strongest protection at P68 and P52. In the crystal structure of *T. thermophilus* complex, phosphate contacts were observed in all these regions, showing that the mode of association must be similar to that in *E. coli*. However several additional contacts are observed in the crystal structure (notably on the 3' side of the variable stem) that were not detected in the solution experiments. In contrast, no specific contact with P52 was observed in the crystal. Whether these detailed differences reflect differences between solution and crystal conditions or between *E. coli* and *T. thermophilus* complexes remains to be seen (32). As noted above, phosphate-contacting polar residues are not necessarily conserved between *E. coli* and *T. thermophilus* and therefore slightly different contacts, although in the same regions of the tRNA, may be expected.

The long variable arm of the tRNA^{Ser} is of major importance for recognition by the synthetase and we can ask what distinguishes the tRNA^{Ser} variable arm from the other long variable arms in tRNA^{Leu} and tRNA^{Tyr} (often referred to as class II tRNAs, not to be confused with class 2 synthetases). A conserved feature in prokaryotic tRNA^{Ser} isoacceptors is the base-pairing of 45 to 48-1 (that is, 48 minus 1, in this case 47q) such that there are no unpaired bases at the base of the variable arm stem. In addition prokaryotic tRNA^{Ser} variable arms can have six or more base pairs (eight in tRNA^{Ser(Cys)}). The number of unpaired bases at the beginning of the variable arm stem differs among other long variable arm tRNAs and also correlates with the number of inserted nucleotides in the D loop (12, 33). The tRNA^{Leu} isoacceptors have one unpaired nucleotide at the 3' end of the variable stem (that is, 45 pairs with 48-2), and tRNA^{Tyr}

has two (that is, 45 pairs with 48-3). These differences may influence the orientation of the variable arm as well as the backbone conformation at the base of the arm and thus could be very important for discrimination between long variable arm tRNAs. Interestingly, tRNA^{Ser} from yeast (and all eukaryotic tRNA^{Ser} isoacceptors) is similar to tRNA^{Leu} from *E. coli* in having one unpaired base in the variable stem. This could partially explain the absence of cross-recognition of yeast tRNA^{Ser} by seryl-tRNA synthetase from *E. coli* (although the acceptor stem is also altered from *E. coli* tRNA^{Ser}).

Himeno *et al.* (12) took into account the above considerations when attempting to convert the identity of an *E. coli* tRNA^{Tyr} to tRNA^{Ser} using T7 transcripts and in vitro aminoacylation assays. Insertion of two nucleotides into the variable stem of the tRNA^{Tyr} between positions 44 and 45 (thus making its base-pairing similar to that of tRNA^{Ser}) immediately introduced serine charging activity, although at a low level. However, deletion of the two unpaired bases showed no detectable serine charging activity presumably because the arm is too short (only three base pairs and a three nucleotide loop). The replacement of U9 in the tRNA^{Tyr} by G9 as in tRNA^{Ser} makes the V_{max}/K_m four times higher. This can be attributed to the important role of G9 in the core of tRNA^{Ser} (Fig. 6); clearly a U in this position could not establish the same interactions. An optimal identity switch is obtained by a simultaneous replacement of A20b by G20b (as in tRNA^{Ser}) and an insertion of a four-nucleotide variable loop by insertion of G47:1 (equivalent to G47c). As discussed above, residue 20b directs the orientation of the variable stem, but since its base makes no specific interactions, the reason for the strong evolutionary preference for a G in this position in tRNA^{Ser} isoacceptors is unclear.

The results of Normanly *et al.* (11, 13) on the identity change from tRNA^{Leu} (UAG) into a specific serine suppressor tRNA using in vivo assays, are less easily explainable by the present structure. In the original identity switch experiment, the D loop was reconstructed as in tRNA^{Ser} (11) but later it was found that this was in fact detrimental to the suppressor efficiency, and not necessary for serine specificity (13). In addition, no modification of the tRNA^{Leu} variable arm was required to give the mutant LSM6 the same biological activity as the wild-type suppressor tRNA^{Ser}. In contrast, the base pair C11-G24 (and in particular the base C11) was shown to play a major role in serine specificity. In the *T. thermophilus* structure, although nucleotide G24 is not very well defined, it is clear that the base pair C11-G24 is not in contact with the synthetase. The importance of this

pair can therefore only be explained by an unknown structural effect, or that it is a negative determinant preventing mischarging of tRNA^{Ser} by, for example, leucyl-tRNA synthetase.

Recognition and specificity. The seryl-tRNA synthetase-tRNA^{Ser} complex structure shows that the protein principally contacts the tRNA backbone and thus recognizes the tRNA shape rather than its specific nucleotide sequence. Both distinctive features of the system, the synthetase coiled-coil domain and the tRNA^{Ser} long variable arm play major roles in this recognition. Comparison of the conformation of the protein in the presence and absence of tRNA suggests that the recognition process involves conformational adaptation of the two molecules, in particular the protein helical arm, to maximize contacts between them. Only with a tRNA possessing a long variable arm in the correct orientation will sufficient interactions be made for the synthetase helical arm to stably hold the tRNA in the required orientation for entry of its 3' end into the active site of the other subunit of the synthetase. In this model of the recognition process, sequence-specific recognition in the acceptor stem region (not visible in our structure) would be a second step in the discrimination against non-cognate tRNAs, after shape recognition. The importance of the orientation of the long variable arm for the recognition of tRNA^{Ser} and tRNA^{SerCys} by seryl-tRNA synthetase in the human system has been demonstrated (34).

The question remains as to why significant regions of the tRNA are disordered in this complex? This is not too surprising for the anticodon stem and the end of the variable arm which do not make any intermolecular contacts. The most obvious reason for the absence of electron density in the acceptor stem would be degradation of the tRNA. However, gel analysis of the tRNA recovered from well-washed crystals shows the tRNA to be intact (17). The mother liquor, however, can contain, in addition to intact tRNA, two major degradation products. Sequence analysis of end-labeled tRNA shows that this corresponds to a cleavage between U47g and A47h in the variable arm loop; spontaneous cleavage of pyrimidine-p-A sequences in tRNA loops is well known (35). The presence of this cut in a position not directly involved in synthetase-tRNA interactions would not easily explain the disordering of the active site region of the complex. An alternative explanation could be that an ordered active site only occurs after the formation of the bound seryl adenylate, that is, at the moment of aminoacylation; this would be a

further means of discrimination against mischarging. In this context we note that in the yeast aspartyl-tRNA synthetase complex, in the absence of other substrates, the 3' terminal adenosine partially occupies the ATP site in one subunit and its "correct" position in the other subunit, with a consequent small difference in the two acceptor stem helix conformations (10). If two or more such structures were averaged in our form 4 crystals this could explain the lack of interpretable electron density. However measurements at 2.8 Å resolution on form 4 crystals co-crystallized with an inhibitor, a stable sulfamoyl analog of seryl-adenylate (30), clearly show the inhibitor in the active site, but only a slight improvement in the ordering of the 3' end of the tRNA (36). In contrast, a recent 3.4 Å resolution structure determined from a new trigonal crystal form of the synthetase-tRNA complex (called form 5) obtained with the same batch of tRNA₂^{Ser} (GGA) as used in crystals of form 4, shows a much better ordered active site with the 3' end of the tRNA visible (36). In form 5 crystals there is, as in form 4, only one tRNA per synthetase dimer. This suggests that this stoichiometry might have a particular biological significance and not be a crystallization artifact. In the form 5 structure, most synthetase-tRNA contacts are as we describe for form 4; however, the acceptor stem helix is slightly reoriented and the 3' end of the tRNA is visible in close proximity to the adenylate analog with which the crystals were grown. We therefore conclude that in form 4 crystals something, perhaps crystal contacts, is preventing the formation of a "productive" conformation of the acceptor end of the tRNA, even in the presence of the seryl-adenylate analog.

REFERENCES AND NOTES

- G. Eriani, M. Delarue, O. Poch, J. Gangloff, D. Moras, *Nature* **347**, 203 (1990); D. Moras, *Trends Biochem. Sci.* **17**, 159 (1992).
- P. Brick, T. N. Bhat, D. M. Blow, *J. Mol. Biol.* **208**, 83 (1989).
- S. Brunie, C. Zelwer, J.-L. Risler, *ibid.* **216**, 411 (1990).
- M. A. Rould, J. J. Perona, D. Söll, T. A. Steitz, *Science* **246**, 1135 (1989).
- S. Cusack, C. Berthet-Colominas, M. Härtlein, N. Nassar, R. Leberman, *Nature* **347**, 249 (1990).
- M. Ruff *et al.*, *Science* **252**, 1682 (1991).
- S. Cusack, M. Härtlein, R. Leberman, *Nucleic Acids Res.* **19**, 3489 (1991).
- M. E. Saks, J. R. Sampson, J. Abelson, *Science*, in press; L. D. Schulman, *Prog. Nucleic Acid Res. Mol. Biol.* **41**, 23 (1991); W. H. McClain, *J. Mol. Biol.* **234**, 257 (1993).
- M. A. Rould, J. J. Perona, T. A. Steitz, *Nature* **352**, 213 (1991).
- J. Caverelli, B. Rees, M. Ruff, J.-C. Thierry, D. Moras, *ibid.* **362**, 181 (1993).
- J. Normanly, R. C. Ogden, S. J. Horwath, J. Abelson, *ibid.* **321**, 213 (1986); G. Sundharadas, J. Katze, D. Söll, W. Konigsberg, P. Lengyel, *Proc. Natl. Acad. Sci. U.S.A.* **61**, 693 (1968).
- H. Himeno, T. Hasegawa, T. Ueda, K. Watanabe, M. Shimizu, *Nucleic Acids Res.* **18**, 6815 (1990).
- J. Normanly, T. Olick, J. Abelson, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5680 (1992).
- J. R. Sampson and M. E. Saks, *Nucleic Acids Res.* **21**, 4467 (1993).
- T. Brennan and M. Sundaralingam, *ibid.* **3**, 3235 (1976).
- A. C. Dock-Bregeon, E. Westhof, R. Giegé, D. Moras, *J. Mol. Biol.* **206**, 707 (1989).
- M. Fujinaga, C. Berthet-Colominas, A. D. Yaremchuk, M. A. Tukalo, S. Cusack, *ibid.* **234**, 222 (1993).
- A. D. Yaremchuk, M. A. Tukalo, I. Krikliiviy, N. Malchenko, V. Biou, C. Berthet-Colominas, S. Cusack, *FEBS Lett.* **310**, 157 (1992b).
- S. Price, S. Cusack, F. Borel, C. Berthet-Colominas, R. Leberman, *ibid.* **324**, 167 (1993).
- A. D. Yaremchuk *et al.*, *J. Mol. Biol.* **224**, 519 (1992).
- R. Fourme *et al.*, *Rev. Sci. Instr.* **63**, 982 (1992).
- A. G. W. Leslie, Joint CCP4 and ESR-EACBM Newsletter on Protein Crystallography, No. 26 (April 1992), Daresbury Laboratory, Warrington WA4 4AD, England.
- CCP4, programme package Daresbury Laboratory, Warrington WA4 4AD, England.
- R. J. Read, *Acta Crystallogr.* **A42**, 140 (1986).
- A. T. Brünger, X-PLOR Version 3.1 (Yale Univ. Press, New Haven, 1993); *Acta Crystallogr.* **A46**, 46 (1990).
- Built by E. Westhof (IBMC, Strasbourg). Coordinates used for analysis here have been deposited at Brookhaven under 1SER, 1SES, and 1SET.
- Primary sequences of tRNA₁^{Ser}(GCU) and tRNA₂^{Ser}(GGA) from *T. thermophilus* have been determined by Z. Petrusenko and M. A. Tukalo (unpublished results). The sequence of a tRNA^{Ser} gene from *T. thermophilus* with 93 nucleotides and a GGA anticodon has been published (A. Venegas, *T. thermophilus* HB8 tRNA^{Ser} gene, EMBL data bank entry X07394, 1988). Our sequence of tRNA₂^{Ser}(GGA), verified by both gene and tRNA sequencing, has 94 nucleotides and differs from the previous sequence only in having an extra guanine in the long variable arm (that is, entry X07394 lacks G47d in Fig. 2).
- A. T. Jones, J.-Y. Zou, S. W. Cowan, M. Kjeldgaard, *Acta Crystallogr.* **A47**, 110 (1991).
- N. Nassar, Ph.D. thesis, University of Grenoble (1992); T. Thüne, thesis, Technical University of Munich (1993).
- S. R. Holbrook, J. L. Sussman, R. W. Warrant, S. H. Kim, *J. Mol. Biol.* **123**, 631 (1978).
- E. Westhof, P. Dumas, D. Moras, *ibid.* **184**, 119 (1985).
- H. Belrhali *et al.*, *Science* **263**, 1432 (1994).
- D. Schatz, R. Leberman, F. Eckstein, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 6132 (1991).
- Preliminary results on phosphate protection with ethylnitrosourea as probe on the complex of *T. thermophilus* seryl-tRNA synthetase and tRNA^{Ser}(GGA) are in good agreement with the crystal structure and confirm the differences with the *E. coli* system (Z. Petrusenko and M. A. Tukalo, unpublished results).
- Y. Komine, T. Adachi, H. Inokuchi, H. Pzeki, *J. Mol. Biol.* **212**, 579 (1990).
- X.-Q. Wu and H. J. Gross, *Nucleic Acids Res.* **21**, 5589 (1993).
- A. C. Dock-Bregeon and D. Moras, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 113 (1987).
- A. D. Yaremchuk, M. A. Tukalo, S. Cusack, unpublished results.
- We thank J.-P. Benoit, R. Fourme, and C. Berthet-Colominas for assistance in the diffraction measurements at LLURE and A. Jones for advice about the use of "O". Supported in part by NATO (collaborative research grant no. 920692).

27 September 1993; accepted 24 January 1994