The EMBO Journal vol.15 no.17 pp.4749–4758, 1996

Structure of crystalline *Escherichia coli* methionyl-tRNA$_f^{Met}$ formyltransferase: comparison with glycinamide ribonucleotide formyltransferase

Emmanuelle Schmitt[1], Sylvain Blanquet and Yves Mechulam

Laboratoire de Biochimie, Unité de Recherche Associée no.1970 du Centre National de la Recherche Scientifique, Ecole Polytechnique, F-91128 Palaiseau cedex, France

[1]Corresponding author

Formylation of the methionyl moiety esterified to the 3′ end of tRNA$_f^{Met}$ is a key step in the targeting of initiator tRNA towards the translation start machinery in prokaryotes. Accordingly, the presence of methionyl-tRNA$_f^{Met}$ formyltransferase (FMT), the enzyme responsible for this formylation, is necessary for the normal growth of *Escherichia coli*. The present work describes the structure of crystalline *E.coli* FMT at 2.0 Å resolution. The protein has an N-terminal domain containing a Rossmann fold. This domain closely resembles that of the glycinamide ribonucleotide formyltransferase (GARF), an enzyme which, like FMT, uses *N*-10 formyltetrahydrofolate as formyl donor. However, FMT can be distinguished from GARF by a flexible loop inserted within its Rossmann fold. In addition, FMT possesses a C-terminal domain with a β-barrel reminiscent of an OB fold. This latter domain provides a positively charged side oriented towards the active site. Biochemical evidence is presented for the involvement of these two idiosyncratic regions (the flexible loop in the N-terminal domain, and the C-terminal domain) in the binding of the tRNA substrate.

Keywords: crystalline structure/formylation/tRNA/ translation initiation

Introduction

Among the various tertiary structures adopted by RNAs, that of tRNAs appears relatively invariant in the living world. Despite a constant shape, each tRNA is recognized with high specificity by its cognate aminoacyl-tRNA synthetase. A few nucleotides, called the tRNA identity elements, which in most cases are located in the anticodon loop and the acceptor stem, participate in this recognition. In turn, aminoacyl-tRNA synthetases are built up of several domains, one of which, the catalytic domain, binds the acceptor stem, while another one usually recognizes the anticodon loop.

Each tRNA species is also a specific target for many other enzymes and proteic factors, including ribosomes. The specificity of the corresponding interactions is the basis of the accuracy and efficiency of the translation process. For instance, the start of translation involves a specialized initiator tRNA which, after aminoacylation by methionyl-tRNA synthetase, has, in prokaryotes, to be modified further by the addition of a formyl group. The initiator formyl-Met-tRNA$_f^{Met}$ is then complexed by initiation factor 2 (IF2) to be recognized by IF3 on the 30S ribosomal subunit.

The transfer of a formyl group from *N*-10 formyltetrahydrofolate (FTHF) to the methionyl group esterified to the 3′ end of initiator tRNA$_f^{Met}$ is catalyzed by methionyl-tRNA$_f^{Met}$ formyltransferase (FMT). This reaction is essential for the orientation of the initiator tRNA towards the translation start machinery (Guillon *et al.*, 1993; Mangroo and RajBhandary, 1995). Obviously, formylase must be highly specific for its tRNA substrate and, in particular, must not sustain formylation of the elongator Met-tRNA$_m^{Met}$, a substrate of the ribosomal A site. The nucleotidic determinants specifying formylation of Met-tRNA$_f^{Met}$ are clustered at the top of the acceptor stem. Rather than the presence of a particular nucleotide sequence, it is the absence of a strong base pairing at position 1–72 of the acceptor stem which allows tRNA$_f^{Met}$ to be a substrate for *Escherichia coli* FMT (Lee *et al.*, 1991; Guillon *et al.*, 1992b). Previous work (Kahn *et al.*, 1980; Blanquet *et al.*, 1984) showed that, under low ionic strength conditions, formylase binds any tRNA with high affinity with a stoichiometry of nearly 10 formylase molecules per tRNA. However, under nearly physiological conditions (150 mM KCl and 10 mM $MgCl_2$), the enzyme distinguishes Met-tRNA$_f^{Met}$ among all other assayed elongator tRNAs, and the binding stoichiometry is close to 1:1. These properties led to the idea that formylase has the ability to melt the secondary structures of RNA molecules but that, under physiological conditions, this melting is easier with tRNA$_f^{Met}$ than with elongator tRNAs, because of the absence of base pairing at position 1–72 in the former. Such melting would be a prerequisite to allow the methionyl moiety attached to the acceptor end of initiator tRNA to reach the formylation site. More recently, this model was reinforced by NMR experiments indicating that several imino protons in the acceptor stem of the nucleic acid become solvent accessible upon formation of the 1:1 FMT–tRNA$_f^{Met}$ complex (Wallis *et al.*, 1995). As a first step in determining which structural motifs of the enzyme are responsible for the recognition of the acceptor stem of tRNA$_f^{Met}$ and to further our understanding of the mechanism of action of FMT, we have solved the three-dimensional structure of this protein which we had crystallized previously (Schmitt *et al.*, 1996).

Results and discussion

Structure solution

Trigonal crystals of FMT belonging to space group $P3_221$ and having cell dimensions of $a = b = 151.0$ Å, $c = 81.8$ Å were used (Schmitt *et al.*, 1996). The structure of

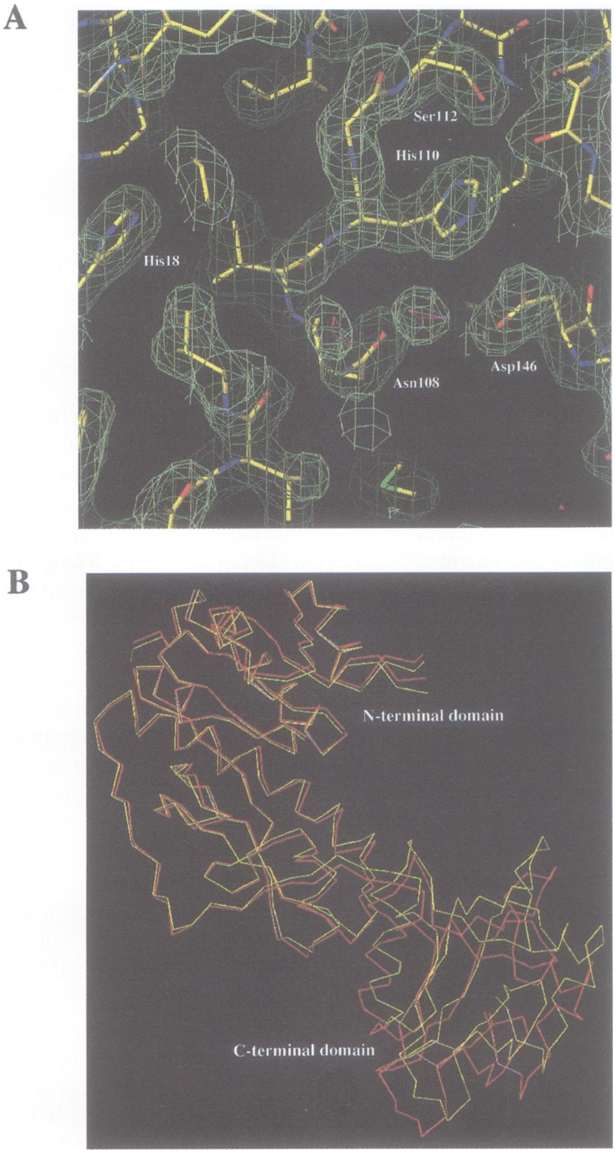

Fig. 1. (**A**) View of the 2.0 Å resolution '$2F_o$–F_c' map contoured at 1.3 standard deviation showing the active site region, with residues Asn108, His110 and Asp146. Two water molecules are also visible and drawn as purple triangles. The figure was drawn using the program O (Jones *et al.*, 1991). (**B**) Comparison of the Cα traces of the two formylase molecules in the asymmetric unit. The structures were superimposed after fitting the Cα atoms of the two N-terminal domains (residues 1–189).

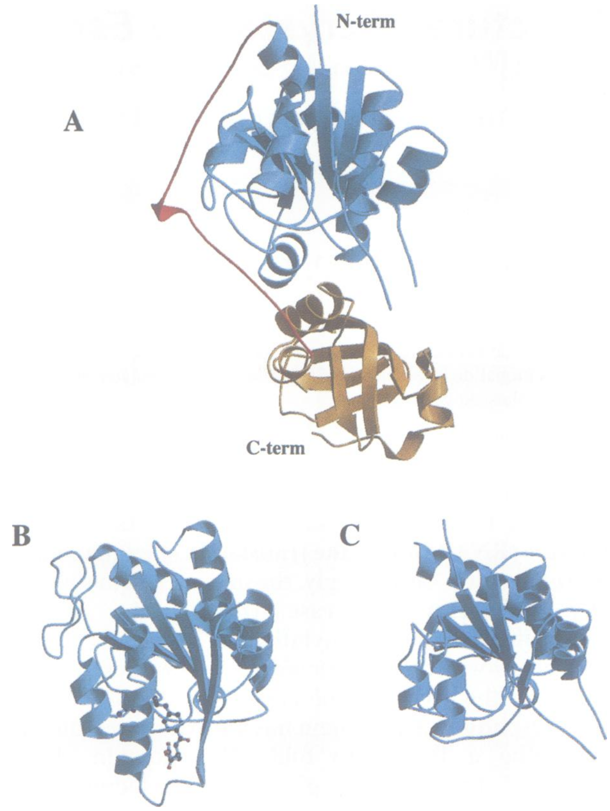

Fig. 2. (**A**) Ribbon representation of methionyl-tRNA$_f^{Met}$ formyltransferase from *E.coli.* The N-terminal domain (residues 1–189) is cyan colored, the linker domain (residues 190–208) red colored and the C-terminal domain yellow colored. (**B**) and (**C**) Comparison of the three-dimensional structure of the N-terminal domain (residues 1–189) of FMT (C) with that of the glycinamide ribonucleotide formyltransferase complexed to 5dTHF (B). The molecules are represented in the same orientation. The inhibitor 5dTHF is drawn in the ball-and-stick representation. The figure was generated using the programs MOLSCRIPT (Kraulis, 1991) and Raster3D (Bacon and Anderson, 1988).

FMT was solved by multiple isomorphous replacement (MIR) using four derivatives, followed by solvent flattening and 2-fold non-crystallographic averaging, since the asymmetric unit contains two molecules. The model finally encompasses 308 residues out of the 314 comprising the polypeptide. A total of 151 water molecules have been placed in the asymmetric unit. The crystallographic *R*-factor is 21.1% for 60 405 unique reflections between 8.0 Å and 2.0 Å (2 σ cut off). The free *R*-factor calculated on the remaining 6792 reflections, which have not been included in the refinement (Brunger, 1992), is 25.6%. The root mean square (r.m.s.) coordinate error was estimated to be 0.25 Å by the method of Luzatti (1952). A representative portion of the final electron density map is shown in Figure 1A. The model has a good geometry with an r.m.s. deviation from ideal geometry of 0.011 Å for bond lengths and of 2° for bond angles. All residues, except Met235, have φ and ψ angles within the allowed regions of the Ramachandran plot, with 91.2% in the most favored regions. The average temperature factor for all protein atoms is 28.4 Å^2.

Overall structure

The two molecules in the asymmetric unit are related by an improper non-crystallographic symmetry. The second monomer can be deduced from the first by a rotation of 98.7°. The refined model at 2.0 Å resolution contains all amino acids with the exception of residues 40–45 in the two monomers and of residue 1 in monomer 2. The enzyme, which behaves in solution as a monomer, is composed of two domains connected by a linker (Figure 2). The N-terminal domain is formed by residues 1–189 and exhibits a Rossmann-type nucleotide binding fold (Rossmann *et al.*, 1974). The linker is formed from residues 189–208. The C-terminal domain, encompassing residues 209–314, is mainly composed of β-strands. A schematic representation of the topology of the enzyme is shown in Figure 3.

The two monomers of FMT in the asymmetric unit are

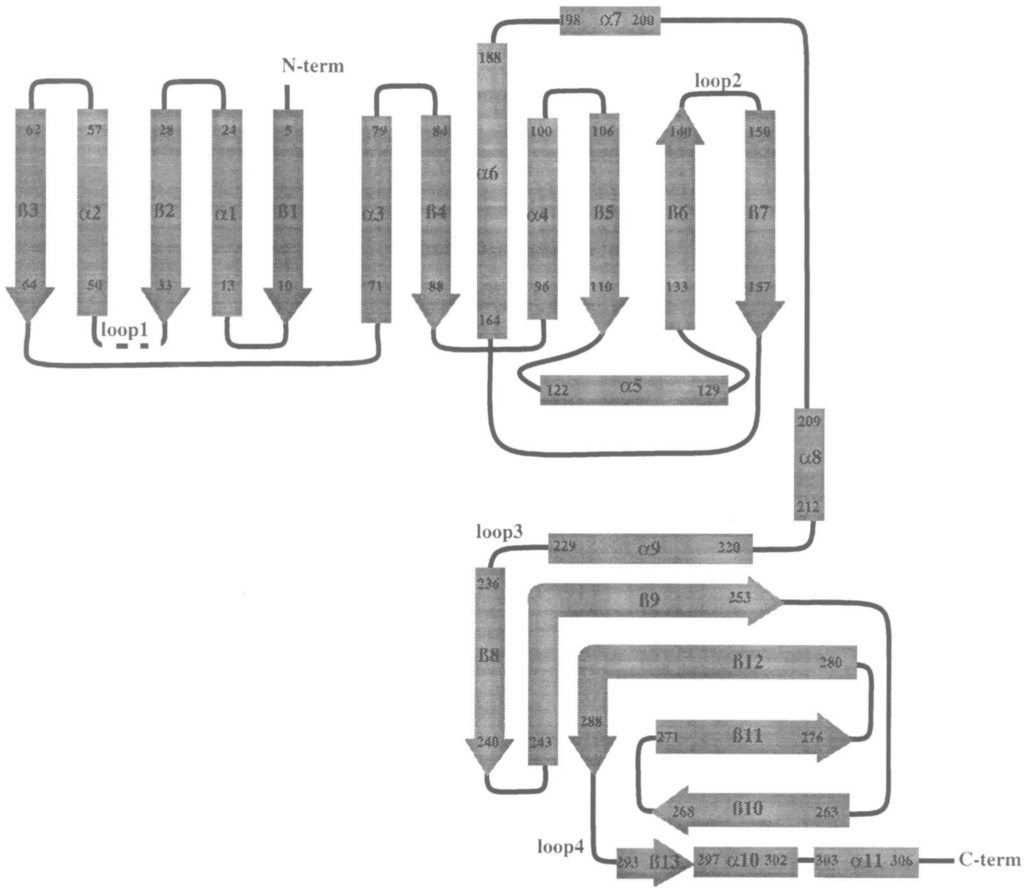

Fig. 3. Schematic representation of the topology of FMT. The β-strands are represented as arrows and the helices as rods. Secondary structure elements were assigned by using the PROCHECK program (Laskowski *et al.*, 1993). The numbering refers to the residues delineating each secondary structure element.

very similar according to the calculation of the r.m.s. difference between their atomic coordinates. However, the r.m.s. difference between the main chain atom positions of the two N-terminal domains (0.52 Å) or of the two C-terminal domains (0.58 Å) are smaller than the r.m.s. difference calculated from the whole molecules (1.12 Å). This discrepancy is therefore accounted for by a slight variation between the relative positions of the N- and C-terminal domains inside each of the two molecules (Figure 1B).

The N-terminal domain of FMT closely resembles glycinamide ribonucleotide formyltransferase

The N-terminal domain of FMT is made up of a seven stranded β-sheet surrounded on both sides by two α-helices (Figure 2). The first part of this nucleotide binding fold corresponds to three parallel β-strands connected by two cross-over α-helices (β1/α1/β2/α2/β3). The second part, joined to the first by helix α3, is made of two parallel strands (β4,β5) connected by a helix (α4), and is followed by a helix (α5) and two antiparallel strands (β6,β7). Therefore, the first five strands and β7 are parallel, and β6 is antiparallel (Figure 3). Downstream from the seven stranded β-sheet is a long α-helix (α6) of 25 residues located in the 3D structure on the same side of the β-sheet as α1 and α2.

Sequence homologies between FMT and other formyltetrahydrofolate binding enzymes have been reported previously (Guillon *et al.*, 1992a; Meinnel *et al.*, 1993a). One of these enzymes, the glycinamide ribonucleotide formyltransferase (GARF) from *E.coli* has been crystallized and its structure was solved, both free and complexed with substrate analogs (Almassy *et al.*, 1992; Chen *et al.*, 1992). The structures of the two enzymes, GARF and FMT, were superimposed. The best fit was obtained by comparison of the 189 N-terminal residues of FMT with the entire GARF molecule complexed with both an analog of FTHF (5-deaza-5,6,7,8-tetrahydrofolate or 5dTHF) and glycinamide ribonucleotide (GAR). Upon alignment of 167 pairs of atoms, a small r.m.s. deviation for Cα atoms of 1.46 Å was calculated. Such a value reflects a close similarity between all secondary structure elements of the compared domains (Figures 2 and 4). Despite this, the sequence alignment deduced from the structural alignment showed only 36 strict identities (19%) plus 36 conservative replacements (Figure 4). The highest score of sequence homology is obtained in the region of the '$_{112}$SLLP$_{115}$' motif, common to several enzymes capable of binding tetrahydrofolate derivatives. The loop carrying the SLLP motif is well ordered in the FMT structure as well as in the complexed GARF structure, whereas it is disordered in the uncomplexed GARF structure. In the case of FMT, there are strong electrostatic interactions between the linker peptide (residues 198, 202 and 204) and the C-terminal part of the 111–121 loop. This feature may account for a stabilization of the 111–121 loop in spite of the absence of ligand.

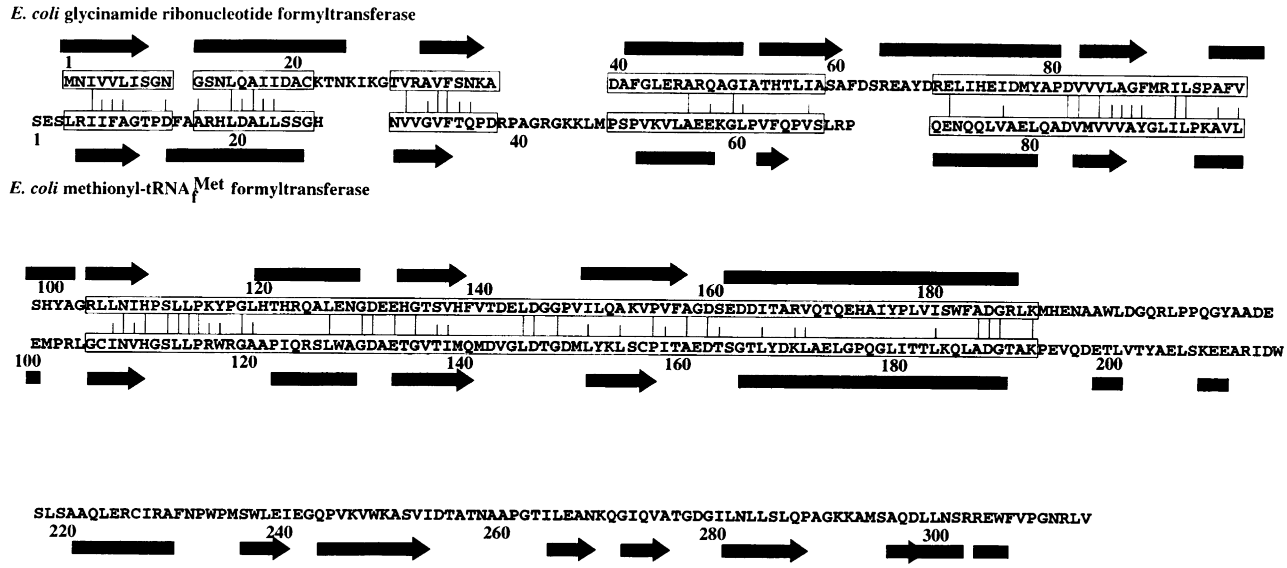

Fig. 4. Alignment of the amino acid sequence of glycinamide ribonucleotide formyltransferase from *E.coli* with that of methionyl-tRNA$_f^{Met}$ formyltransferase from *E.coli*, based on the structural homologies between the two enzymes. The structural alignment was performed by using the 'lsq' commands of the program O (Jones *et al.*, 1991). The secondary structure elements of each protein are schematized, with β-strands represented as arrows and helices as rods.

The quality of the superimposition of the structure of the complexed form of GARF on that of FMT allowed us to dock the analog of formyltetrahydrofolate within its putative active site. The deduced model of the FMT–5dTHF complex was energy minimized using the CHARMm program (Brooks *et al.*, 1983). Such a minimization resulted in almost no atom movement and showed that the above docking could be obtained without any steric clash. The docked inhibitor 5dTHF lies in a crevice at the carboxyl end of the β-sheet, surrounded by the β4 and β5 strands and the well-ordered loop 144–148 (Figure 5). In the native FMT structure, this crevice is occupied by several well-defined water molecules, which probably are displaced upon binding of the substrate (Figure 1A). This crevice is adjacent to the classical nucleotide binding site, usually located between the two halves of the Rossmann fold (i.e. between β1 and β4; Bränden, 1980) and which in the GARF enzyme is occupied by the GAR substrate. In FMT, as in the complexed GARF, the main chain carboxyl groups of residues 142 and 144 (according to the FMT numbering) are hydrogen bonded to the N3 group of 5dTHF. Interestingly, β4 and β5 strands as well as loop 144–148 carry residues conserved in the two enzymes. Among these residues, Asn108, His110 and Asp146 (according to the FMT numbering) previously were proposed to play a role in the GARF catalytic function (Inglese *et al.*, 1990; Almassy *et al.*, 1992). We noted that, in the structure of FMT, the side chains of these three residues adopt conformations identical to those which they have in the complexed GARF structure. In the case of FMT, they are in the close vicinity of the docked 5dTHF molecule (the γ oxygen of D146 is 4.0 Å from O4 of the bicyclic ring of 5dTHF). The idea of a functional role for these residues in the formylation reaction of FMT is reinforced by the occurrence of a network of electrostatic contacts involving Asp146, His110 and the SLLP loop (see Figure 1A). Finally, the benzoyl and bicyclic rings of the docked ligand lie in a pocket surrounded by

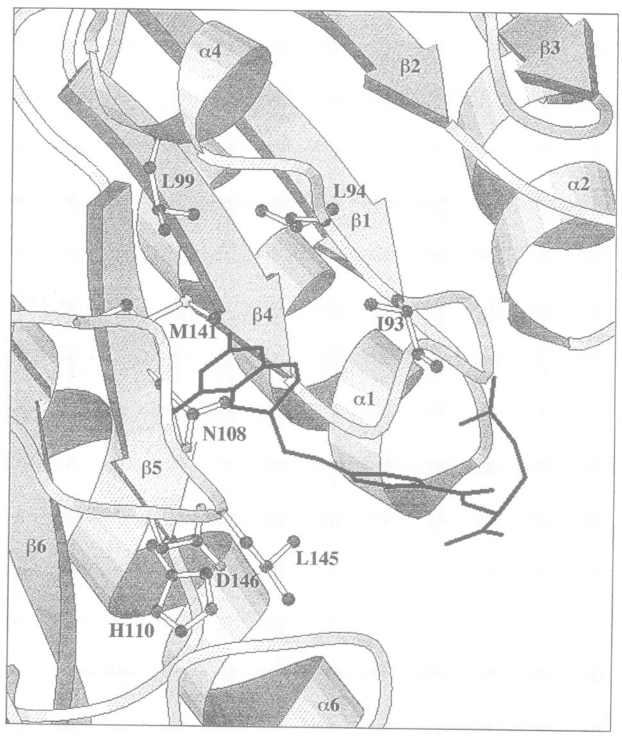

Fig. 5. Close up view of the putative FTHF binding crevice of FMT. Secondary structure elements are labeled. The docked 5dTHF molecule is drawn with solid lines, and the surrounding residues are in the ball-and-stick representation. The figure was generated using the program MOLSCRIPT (Kraulis, 1991).

hydrophobic residues (Ile93, Leu94, Leu99, Leu145, Met141). Among these residues, Met141 might form a sulfur aromatic interaction with the bicyclic ring of 5dTHF (Reid *et al.*, 1985).

On the other hand, comparison of the structures of the two enzymes reveals some differences. In the GARF complexed structure, the main chain NH groups of residues

11, 12 and 13 of the α1 helix are involved in the recognition of the GAR substrate. In addition, the γ oxygen of Ser12 and the δ-NH_2 of Asn13 interact with the oxygen atoms of the GAR phosphate group. The corresponding α1 helix region of FMT is not superimposable. Instead, the positioning of the α1 helix of FMT is shifted towards the N-terminus when compared with the case of the GARF enzyme. Therefore, the side chain of Phe14 of FMT occupies the place of the GAR binding site. This organization may be useful to avoid recognition of the GAR substrate by FMT *in vivo*. Moreover, residues belonging to the α1 helix of FMT, such as Phe14, may be involved in the specific binding of the 3′ end of Met-tRNA$_f^{Met}$. In the GARF enzyme, the long α6 helix also exposes residues involved in the binding of the GAR substrate. In the two enzyme structures, this helix is bent by a proline residue. In the case of FMT, this bend is found one turn of the helix before its corresponding location in GARF. Such a difference is likely to reflect the particular specificities of the two enzymes.

Another important difference between the two proteins is the insertion in FMT of a loop between strand β2 and helix α2 (residues 34–49, loop I in Figure 3). This solvent-exposed loop contains several basic residues and is disordered from residue 40 to residue 45. This loop, which is clearly located above the active site crevice, is likely to have functional importance, as shown by the following experiments. Upon limited trypsinolysis of FMT (1/1000 w/w of trypsin with respect to FMT) in a buffer containing 150 mM KCl and 7 mM $MgCl_2$, a single cleavage after residue Arg 42 occurred with complete loss of enzyme activity. The presence of saturating formylmethionyl-tRNA$_f^{Met}$ protected the loop against proteolysis. Indeed, after 40 min of trypsinolysis, <20% of the FMT was cleaved as compared with >80% in the absence of tRNA. Moreover, under otherwise identical conditions, the protection was less pronounced in the presence of non-aminoacylated tRNA$_f^{Met}$ and even less pronounced in the presence of tRNA$_m^{Met}$. It should be borne in mind that tRNA$_f^{Met}$ and tRNA$_m^{Met}$, although not esterified, can bind the enzyme with affinities ensuring saturation of the enzyme under the assayed conditions (Kahn *et al.*, 1980). Together, these data suggest a possible role for loop I in the specific recognition of the esterified 3′ end of the tRNA substrate.

The C-terminal domain participates in the binding of the tRNA molecule

FMT possesses a C-terminal domain (residues 209–314) distinguishing it from the GARF enzyme. This domain is the result of two orthogonal sheets made of five antiparallel strands. The global fold of the domain is that of a β-barrel, surrounded by α-helices, and has characteristics in common with 'the oligonucleotide binding fold' or 'OB fold' (Murzin, 1993) observed in verotoxin (Stein *et al.*, 1992). In agreement with the case of a standard OB fold, three layers of hydrophobic residues fill the inside cavity of the FMT C-terminal domain. Moreover, an α-helix (α9), perpendicular to the β-barrel axis, packs against one side of the barrel. The N-terminal domains of aspartyl- and lysyl-tRNA synthetases, which ensure binding of the anticodons of the cognate tRNAs, also show an OB fold (Figure 6, Ruff *et al.*, 1991; Commans *et al.*, 1995; Onesti *et al.*, 1995). The C-terminal domain of FMT could be superimposed on the N-terminal domain of LysRS. It yields an r.m.s. deviation of 2.2 Å for 38 pairs of compared Cα atoms.

Important topological differences between the canonical OB fold and the C-terminal domain of FMT must be underlined however. In particular, the relative arrangement of the secondary structure elements comprising FMT is different from that in the cases of verotoxin, LysRS or AspRS. For instance, the helix packed against the barrel is located before strand 1 (β8, Figure 3) of FMT, whereas it is inserted between the third and fourth strands of the barrel in a standard OB fold. Moreover, in FMT, strand β8 does not interact with strand β10. Therefore, the β-barrel of FMT is open on one side (located at the back in Figure 6B).

The above comparison, together with the idiosyncratic character of the C-terminal domain of FMT as compared with GARF, makes this C-terminal domain an obvious candidate for the interaction with tRNA$_f^{Met}$. To consider this possibility, the surface potential of FMT was calculated with the DelPhi program (Nicholls and Honig, 1991). Figure 6 clearly shows that FMT, although on the whole an acidic protein, concentrates a positive surface potential in the region of the C-terminal domain pointing towards the active site crevice. More precisely, two solvent-exposed loops (loops 3 and 4, see Figure 3) as well as helix α11, all rich in aromatic and basic residues, can be suspected to interact with the polyanionic tRNA molecule.

To probe further the RNA binding capacity of the FMT C-terminal domain, a polypeptide corresponding to residues 209–314 of FMT was produced by genetic engineering (see Materials and methods). This proteic domain (FCTER) was purified to homogeneity. At low ionic strength, in the absence of added KCl, the intrinsic fluorescence of this tryptophan-rich domain was quenched by 33% upon saturation by tRNA$_f^{Met}$ and by 57% upon saturation by tRNA$_m^{Met}$. FCTER bound tRNA$_f^{Met}$ and tRNA$_m^{Met}$ equally well, with dissociation constants of 0.7 ± 0.1 μM and 1.1 ± 0.1 μM, respectively. In each case, the stoichiometry was 5 ± 2 proteins per tRNA molecule. Unfortunately, in the presence of 150 mM KCl and 7 mM $MgCl_2$, too high dissociation constants (>10 μM) precluded reliable measurements. These experiments show that the C-terminal domain participates in the binding of tRNA and that, at low ionic strength, it binds almost indiscriminantly to either tRNA$_f^{Met}$ or tRNA$_m^{Met}$ with high affinity, in a manner identical to that of intact FMT (Kahn *et al.*, 1980; Blanquet *et al.*, 1984).

Conformational variability and the interface between the two domains

As noted above, the respective orientations of the N- and C-terminal domains in each of the two molecules comprising the asymmetric unit are different (Figure 1B). This is clearly shown in a plot of the distances between pairs of corresponding Cα atoms from the two molecules, after superimposition of the N-terminal domains (Figure 7). In order to evaluate the basis of this difference, the contacts between the two domains were first examined. The interface comprises the α5 helix and loop 159–164 of the N-terminal domain, and the α9 helix and the following loop (loop 3) of the C-terminal domain. The

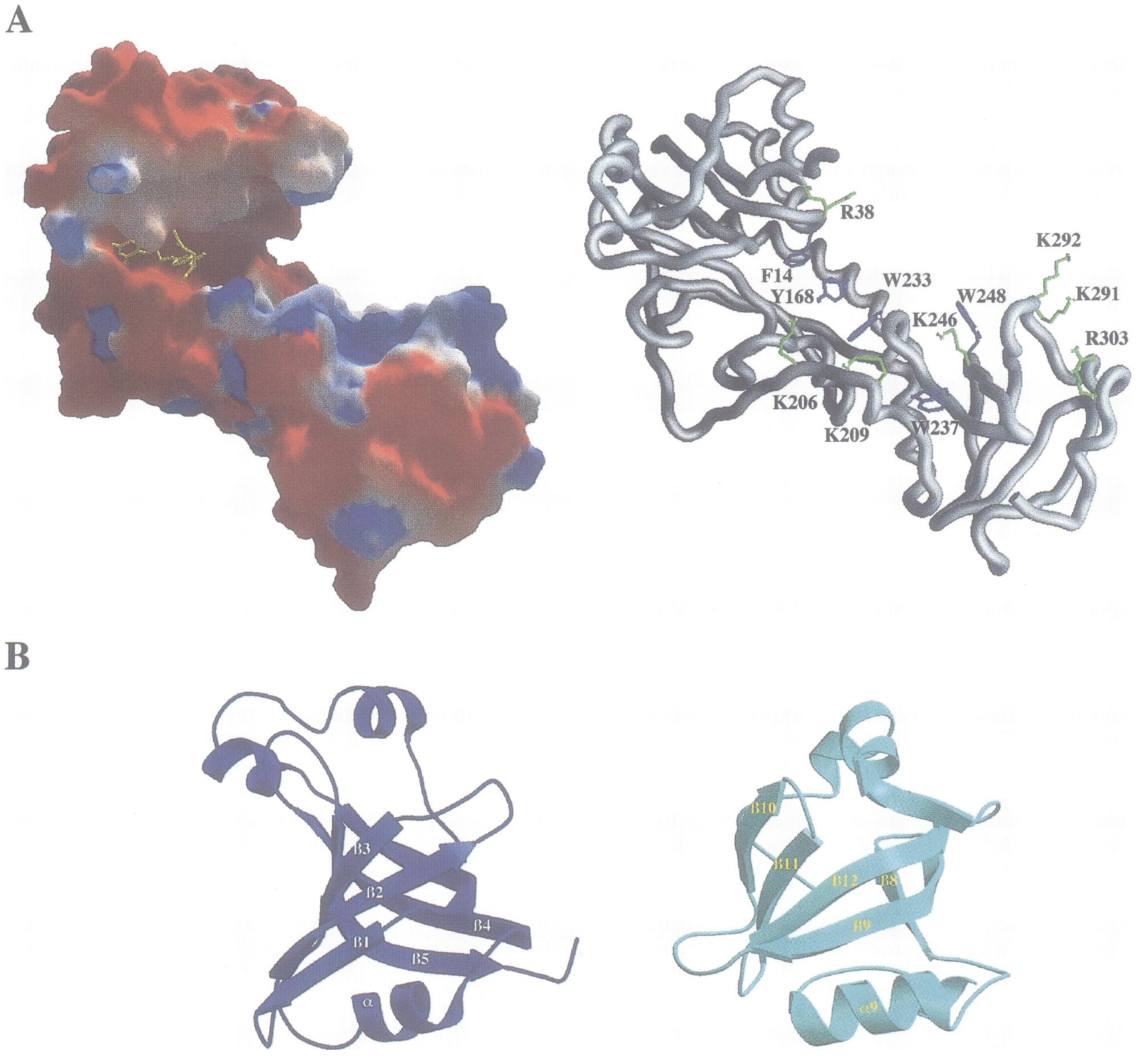

Fig. 6. (**A**) Left: molecular surface of the protein showing the electrostatic potentials calculated with DelPhi (Nicholls and Honig, 1991) and rendered with GRASP (Nicholls *et al.*, 1991). Negatively charged regions are colored in red and positively charged areas are in blue. The 5dTHF, docked within the active site crevice of FMT, is shown with yellow sticks. Right: schematic drawing of the Cα trace of the FMT molecule. Several basic (green) and aromatic (blue) side chains forming a positively charged channel leading to the active site are drawn with solid sticks.
(**B**) Comparative views of the β-barrel of *E.coli* lysyl-tRNA synthetase (left) and of the C-terminal domain of FMT (right). The figure was generated using the programs MOLSCRIPT (Kraulis, 1991) and Raster3D (Bacon and Anderson, 1988), using the co-ordinates of the LysRS N-terminal domain (Commans *et al.*, 1995). The molecules are represented with the same orientation. The aligment was made with help of the 'lsq' commands of the O program. An r.m.s. deviation of 2.2 Å was obtained for 38 pairs of Cα atoms. The correspondence between the secondary structures elements of the two proteins are: β1 of LysRS with β9 of FMT, β2 with β12, β3 with β11, β4 with β8 and α with α9. The numbering of the LysRS secondary structures elements is that of the standard OB fold as defined in Murzin (1993).

204–208 region and the α8 helix of the linker also participate in the packing. Numerous electrostatic interactions involve residues contributed by all these regions. For instance, hydrogen bonds between the side chains of Gln222 and Arg225 (α9) and the main chain carbonyl group of Trp128 (α5) link the antiparallel helices α5 and α9. Moreover, a salt bridge between the side chains of Arg125 (α5) and Glu211 (α8), as well as hydrophobic packing involving residues Leu127, Trp128, Leu207 and Phe230, contribute to the stacking between the two domains. In each of the two molecules of the asymmetric unit, this complex network of hydrophobic and electrostatic interactions appears identical. Accordingly, the conformational variation in the two molecules can be described by a bending of the C-terminal part of the linker along with the C-terminal domain and with the interface region in its entirety (Figure 1B). Indeed, the plot in Figure 7 shows that the peptide chains diverge from residue 205 onwards, with such a bending continuing along the interface (see regions 122–129, 159–164 and 220–229 in Figure 7).

One notable consequence of the distinct organizations of the two molecules in the asymmetric unit is the side chain conformation of Tyr168. In one enzyme molecule,

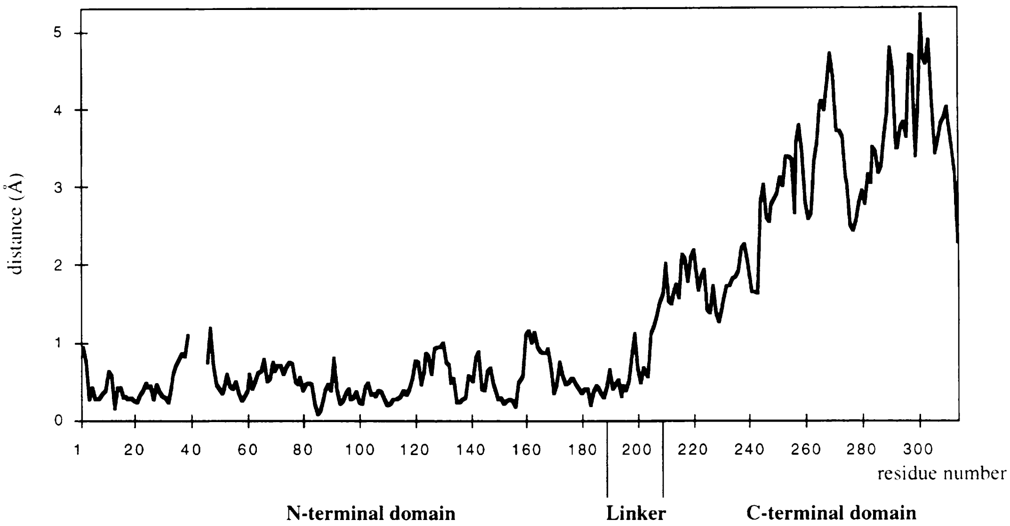

Fig. 7. Comparison of the two formylase molecules in the asymmetric unit. The N-terminal domains of the two molecules were superimposed (see Figure 1A), and the distance between pairs of corresponding Cα atoms was plotted versus the residue number.

Tyr168 points towards the C-terminal domain, whereas, in the other, the side chain hydroxyl is oriented in the direction of the active site and forms a hydrogen bond with the main chain nitrogen of Phe14, a residue belonging to the α1 helix and located at the border of the active site crevice. In addition, the N-terminal side of the long α6 helix, to which Tyr168 belongs, also has distinct positions in the two compared molecules. All these observations suggest that an equilibrium between the two different relative orientations of the N- and C-terminal domains might be relevant to the enzyme's mechanism.

Concluding remarks

The catalytic domain of FMT, built around a Rossmann fold, strikingly resembles the *E.coli* GARF. While these two enzymes are likely to have derived from a common ancestor, the FMT has acquired a supplementary module inserted within the active site domain, namely loop 1 which displays an electropositive character. Our data favor the idea that this idiosyncratic insertion, characteristic of all known FMT sequences (Figure 8), is involved in the recognition of methionylated initiator tRNA^{Met}. Additional structural adjustements may partially account for the different specificities of the two enzymes, FMT and GARF. The most remarkable difference lies at the N-terminus of the α1 helix of FMT, where we propose that residue Phe14 is involved in the binding of the 3′ end of the substrate. Notably, an aromatic residue is found systematically at this position in the sequence of an FMT, whatever its origin (Figure 8).

The main feature of FMT distinguishing it from GARF is the presence of a C-terminal domain. Sequence comparisons strongly suggest the occurrence of such a domain in all Met-tRNA formyltransferases (Figure 8). Such a modular organization of FMT is reminiscent of that of aminoacyl-tRNA synthetases, a family of tRNA binding enzymes. Indeed, these enzymes are systematically built with at least two functional domains, one corresponding to the catalytic center and another capable of recognizing a region of the tRNA molecule carrying idendity elements, often the anticodon. Notably, the C-terminal domain of

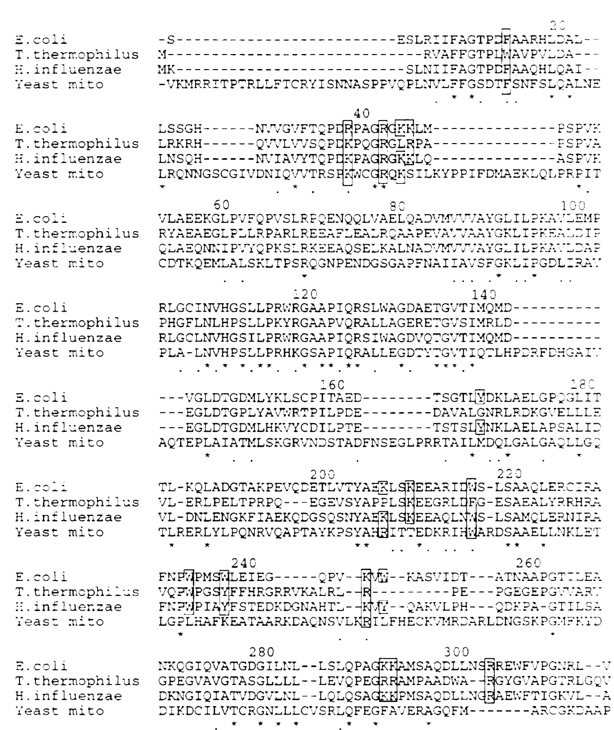

Fig. 8. Multiple sequence alignment of FMT polypeptides from various sources. The positions strictly conserved in the four compared sequences are marked with an asterisk below the sequence. Positions with conservative replacements are signaled by a point. The basic and aromatic residues forming a positively charged channel leading to the active site (see Figure 6 and text) are boxed. The alignment was performed with the CLUSTAL V program (Higgins and Sharp, 1989). The sequences are those of the enzymes from *E.coli* (Guillon *et al.*, 1992a), *Thermus thermophilus* (Meinnel and Blanquet, 1994), *Haemophilus influenzae* (Fleishman *et al.*, 1995) and *Saccharomyces cerevisiae* mitochondrion (Skala *et al.*, 1992).

FMT shares some characteristics of the well-known OB fold, present as an anticodon binding domain in at least two aminoacyl-tRNA synthetases (AspRS, Ruff *et al.*, 1991, LysRS, Commans *et al.*, 1995; Onesti *et al.*, 1995). However, the FMT domain exhibits marked differences, such as a peculiar topology of the β-sheets. These differ-

Table I. Native and heavy atom derivative data used in the structure determination of FMT

Data set	Native	PCMBS	PHMB	$KAuCl_4$	TMLA
Reagent concentration (mM)	–	5	2.5	1.7	17
Soaking time (days)	–	5	5	7	2
X-ray source	LURE W32	LURE W32	LURE W32	LURE W32	Lab.
Processing	MOSFLM	MOSFLM	MOSFLM	MOSFLM	XDS
Resolution used (Å)	30–2.0	30–3.1	12–2.7	8–2.7	30–4.0
Completeness (%)	96.6	97	95	98	85
Redundancy	3.2	2.5	2.4	2.7	2.2
R_{sym} (I) (%)[a]	5.4	4.5	4.8	3.8	3.2
ΔFiso (%)[b]	–	22.2	10.5	17.8	9.3
No. of heavy atom sites	–	5	7	6	6
R_{cullis}[c]	–	0.77	0.72	0.52	0.91
Phasing power[d]	–	1.4	1.5	2.2	0.8
Mean overall figure of merit	0.64 (30–2.7 Å)				

Each data set was collected with a single crystal. The derivatives are: PCMBS, sodium parachloromercuribenzylsulfonate; PHMB, sodium parahydroxymercuribenzoate; $KAuCl_4$, potassium tetrachloroaurate; TMLA, trimethyllead acetate.

$$^{a}R_{sym}(\mathrm{I}) = \frac{\sum_{hkl}\sum_{i} |\langle I_{hkl}\rangle - I_{hkl,i}|}{\sum_{hkl}\sum_{i} |I_{hkl}|}$$ where i is the number of reflections hkl.

$$^{b}\Delta\mathrm{Fiso} = \frac{\sum_{hkl} |Fph_{hkl} - Fp_{hkl}|}{\sum_{hkl} Fp_{hkl}}$$ where Fph are the structure factors of the derivative and Fp, those of the native crystal.

$$^{c}R_{cullis} = \frac{\sum_{hkl} |Fph - |\vec{Fp} + \vec{Fh}||}{\sum_{hkl} |Fph - Fp|}$$ for the centric terms only, where Fh are the structure factors of the heavy atom.

$$^{d}\text{Phasing power} = \left[\frac{\sum_{hkl} |Fh|^2}{\sum_{hkl} (|Fph|^2 - |Fph(calc)|^2)}\right]^{1/2}$$

ences render it unlikely that these folds could have derived from a common ancestor and rather argue in favor of convergent evolution.

The FMT β-barrel displays on its surface many basic and aromatic side chains. These side chains form a positively charged channel which might be involved in the orientation of the acceptor stem of the Met-$tRNA_f^{Met}$ substrate in the direction of the N-terminal domain. The linker region and a few residues of the N-terminal domain could complete this binding surface up to the catalytic crevice of the enzyme (Figure 6). Interestingly, the residues involved, as defined in Figure 6, are among the most conserved residues of prokaryotic methionyl-$tRNA_f^{Met}$ formyltransferases (Figure 8).

In relation to the above picture, the C-terminal domain would participate in a rather unspecific manner in the initial anchoring of the tRNA substrate and in the orientation of its esterified 3′ end with respect to the active center. Full expression of FMT specificity, possibly involving the melting of the C1A72 region of the acceptor stem of tRNA (Blanquet *et al.*, 1984; Kahn *et al.*, 1980; Wallis *et al.*, 1995; Guillon *et al.*, 1992b), would then be expressed by interactions with residues belonging to the linker and to the N-terminal domain. Residues of the α1 and α6 helices and of loop 1 could be particularly involved. Finally, the flexibility in the relative positions of the two domains can also be an important feature of the enzyme's mechanism of action. Further structural and biochemical investigations are required to assess all these ideas.

Materials and methods

Crystallization and data collection

FMT was purified from overproducing cells as previously described (Schmitt *et al.*, 1996). Suitable crystals for X-ray experimentation were obtained by using ammonium sulfate as precipitant and glycerol as additive (Schmitt *et al.*, 1996). Crystals obtained using these conditions were trigonal, space group $P3_221$, with unit cell dimensions $a = b = 151.0$ Å and $c = 81.8$ Å. For data collection and soaking, crystals were stabilized in 60% (w/v) ammonium sulfate, 5% (v/v) glycerol, 100 mM KCl, 10 mM KH_2PO_4 pH 7.3. Mercury derivatives were prepared by soaking native crystals in 5 mM sodium parachloromercuriphenylsulfonate (PCMBS), or in 2.5 mM sodium parahydroxymercuribenzoate (PHMB), for 5 days. Lead derivatives were obtained by soaking native crystals in 17 mM trimethyllead acetate (TMLA) for 2 days. Aurate crystals were obtained by soaking native crystals in 1.66 mM $KAuCl_4$ for 6 days (Table I). Data were collected at 0°C by using a synchrotronic source ($\lambda = 0.902$ Å) at the LURE (Orsay, France) on a MAR-Research phosphor image plate system (Hamburg, Germany), or on a rotating anode source (Siemens, Karlsruhe, Germany) with a Hi-star area detector (Siemens, Karlsruhe, Germany). Diffraction images were analyzed either with the MOSFLM program (A.G.W.Leslie, Laboratory of Molecular Biology, Daresbury, UK) or with the XDS program (Kabsch, 1988) and the data processed further using programs from the CCP4 package (Collaborative computational project No.4, 1994).

Heavy atom and phase determination

The positions of the mercury atoms of the PCMBS derivative were first identified by analysis of the Patterson difference map. Further, the position of the major site was used to calculate SIR (single isomorphous replacement) phases using the CCP4 program MLPHARE (Otwinowski, 1991). These phases were then used to compute difference Fourier maps. From these maps, the positions of minor sites of the mercury atoms of the PCMBS derivative as well as the positions of the heavy atoms in the other derivatives were determined iteratively. For each new site, the corresponding Patterson was checked systematically with the help of the PATGEN program (Chevrier, 1994). The analysis of all the derivatives allowed the determination of MIR phases by using the MLPHARE program. The mean figures of merit in the resolution range of 30–2.7 Å were 0.82 for the centric data and of 0.64 for the acentric data (Table I).

Electron density map averaging and model building

The determination of MIR phases allowed the calculation of MIR maps at 3.5, 3.0 and 2.7 Å resolution. The phases were then improved by solvent flattening and histogram matching using the DM program (Cowtan, 1994). In the resulting density maps, the secondary structures could be identified clearly, and model building was undertaken with the help of bones (calculated using MAPMAN, Kleijwegt and Jones, 1994) and of the O program (Jones *et al.*, 1991). The quality of the maps was such that most of the backbone of one molecule could be readily constructed. At this stage, the position of the second molecule in the asymmetric unit was identified unambiguously, and the non-crystallographic symmetry operators deduced. In a second step, the non-crystallographic symmetry operators were refined by a stage of rigid-body refinement with the program X-PLOR (Brunger *et al.*, 1987). Averaged density maps were then calculated using a mask derived from the first model. The quality of such maps was clearly improved. Hence, the amino acid sequence could be fitted unambiguously along the chain tracing. The initial model included 305 residues out of a total of 314 for one molecule.

Refinement

The model was first refined against the 8.0–2.7 Å native data using the program X-PLOR (Brunger *et al.*, 1987). The crystallographic *R*-factor of the starting model was 41.9%. A random sample containing 10% of the total data was excluded from the refinement and the agreement between the calculated and observed structure factors corresponding to these reflections (R_{free}) was used to monitor the course of the refinement procedure (Brunger, 1992). A round of positional refinement and simulated annealing was performed enforcing strict non-crystallographic 2-fold symmetry. This step lowered the *R*-factor to 0.345 and the R_{free} to 0.396. Further refinement by using the non-crystallographic symmetry constraints did not improve the R_{free} value. Therefore, refinement of the atomic positions and of the temperature factors was carried out independently for the two monomers of the asymmetric unit. During this process, resolution was gradually increased to 2.0 Å. The resulting *R*-factor was 21.4% and the R_{free} 25.6%. The stereochemistry and geometry were analyzed using the program PROCHECK (Laskowski *et al.*, 1993).

Mild proteolysis

Proteolysis of FMT (60 μM) was performed at 37°C in 0.1 M Tris, pH 7.5, 0.1 M KCl, 10 mM $MgCl_2$ and 1 mM 2-mercaptoethanol. Proteolysis was initiated by the addition of trypsin from bovine pancreas [Sigma; 1/1000 (w/w) with respect to FMT]. Samples (10 μl) were withdrawn at various times (from 1 to 60 min), and the trypsinolysis was quenched by the addition [2/1 (w/w) with respect to trypsin] of chicken egg white ovomucoid (Sigma). The reaction products were analyzed by SDS–PAGE on 12.5% (w/v) homogeneous gels (Phastsystem, Pharmacia). The cleavage site was identified by matrix-assisted laser desorption time-of-flight mass spectrometry (Fisons Instruments, UK). The experiments in the presence of saturating amounts (80 μM) of tRNA$_m^{Met}$, tRNA$_f^{Met}$ or formyl-Met-tRNA$_f^{Met}$, taking into account the corresponding dissociation constants (Kahn *et al.*, 1980), were performed under the same conditions (60 μM FMT). The efficiency of protection from trypsin attack conferred by the presence of the nucleic acids was estimated from SDS–PAGE analysis on 12.5% (w/v) homogeneous gels.

Expression and characterization of the C-terminal domain of FMT

In order to produce the C-terminal domain of FMT (residues 209–314), the corresponding DNA region of *fmt* (codons 208–314) and the downstream region, up to the next *Bam*HI site, was PCR amplified using pUC18Fatg as template (Schmitt *et al.*, 1996). The upstream nucleotide primer was designed so as to introduce an *Nde*I site (CATATG), with the ATG replacing codon 208. The resulting sub-gene was cloned between the *Nde*I and *Bam*HI sites of the pET3a vector (Rosenberg *et al.*, 1987). The resulting plasmid pET3aFCTER was then cleaved with *Xba*I and *Hin*dIII to yield a fragment carrying an open reading frame corresponding to the C-terminal domain under the control of the translation initiation signals and the transcription terminator of the bacteriophage T7 gene *10*. This fragment was then subcloned between the corresponding sites of the pUC18 vector to yield the pUC18FCTER plasmid. Taking into account the post-translational removal of the N-terminal methionine (Hirel *et al.*, 1989), this plasmid produces a polypeptide corresponding exactly to amino acids 209–314 of the FMT. This protein (FCTER) was overproduced in *E.coli* JM101Tr (Hirel *et al.*, 1989) extracts and purified to homogeneity by using two chromatographic steps, first on a Q-Hiload ion exchange column (10 cm×2.5 cm; Pharmacia) and second on a Superdex75 column (60 cm×1.6 cm; Pharmacia).

Variations of the intrinsic FCTER fluorescence upon titration with tRNA$_f^{Met}$ or tRNA$_m^{Met}$ [1500 pmol of methionine acceptance per A_{260}, as described by Meinnel *et al.* (1993b) and Meinnel and Blanquet (1995)] were followed by using a spectrofluorimeter (λ excitation = 295 nm, λ emission = 340 nm) at 25°C, in a buffer containing 20 mM Tris–HCl pH 7.6, 0.1 mM EDTA, 10 mM 2-mercaptoethanol (Blanquet *et al.*, 1973). Different titrations were recorded either in the presence or absence of 150 mM KCl and 7 mM $MgCl_2$. In the absence of increased ionic strength, the quenching of FCTER (2 μM) fluorescence observed upon titration with tRNAs (0–3 μM) could be fitted to the theoretical equation after correction for dilution and inner filter effect (Blanquet *et al.*, 1973). From this, an apparent dissociation constant and the number of binding sites could be derived. In the presence of 150 mM KCl and 7 mM $MgCl_2$, within the range of tRNA concentrations used (0–3 μM), after correction for inner filter effect, a slight decrease of intrinsic fluorescence was observed without evidence of saturation. This indicated a dissociation constant >10 μM.

Acknowledgements

We are much indebted to Dino Moras and all his colleagues at the 'Unité de Biologie structurale' (UPR CNRS no. 9004, Illkirch, France), and in particular to Jean Cavarelli, André Mitschler, Marc Ruff and Jean-Claude Thierry, for their kind hospitality during data collection and for sharing their experience of protein crystallography. Roger Fourme and Marc Schiltz (Laboratoire pour l'Utilisation du rayonnement Electromagnétique, UMR CNRS no. 130 Orsay, France) are gratefully acknowledged for assistance during data collection at the beamline DW32. We also thank Jean-Marie Schmitter for performing the mass spectrometry measurements and Laurent Maveyraud for helpful advice.

References

Almassy,R.J., Janson,C.A., Kan,C. and Hostomska,Z. (1992) Structures of apo and complexed *Escherichia coli* glycinamide ribonucleotide transformylase. *Proc. Natl Acad. Sci. USA*, **89**, 6114–6118.

Bacon,D.J. and Anderson,W.F. (1988) A fast algorithm for rendering space-filling molecule pictures. *J. Mol. Graphics*, **6**, 219–220.

Blanquet,S., Iwatsubo,M. and Waller,J.-P. (1973) The mechanism of action of methionyl-tRNA synthetase. 1. Fluorescence studies on tRNA^{Met} binding as a function of ligands, ions and pH. *Eur. J. Biochem.*, **36**, 213–226.

Blanquet,S., Dessen,P. and Kahn,D. (1984) Properties and specificity of methionyl-tRNA$_f^{Met}$ formyltransferase from *Escherichia coli*. *Methods Enzymol.*, **106**, 141–152.

Bränden,C.-I. (1980) Relation between structure and function of α/β-proteins. *Q. Rev. Biophys.*, **13**, 317–338.

Brooks,B., Bruccoleri,R., Olafson,B., States,D., Swaminathan,S. and Karplus,M. (1983) CHARMm: a program for macromolecular energy minimization and molecular dynamics calculations. *J. Comp. Chem.*, **4**, 187–217.

Brunger,A.T. (1992) Free-R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–474.

Brunger,A.T., Kuryian,J. and Karplus,M. (1987) Crystallographic R-factor refinement by molecular dynamics. *Science*, **235**, 458–460.

Chen,P., Schulze-Gahmen,U., Stura,E.A., Inglese,J., Johnson,L.D., Marolewski,A., Benkovic,S.J. and Wilson,I.A. (1992) Crystal structure of glycinamide ribonucleotide transformylase from *Escherichia coli* at 3.0 Å resolution. *J. Mol. Biol.*, **227**, 283–292.

Chevrier,B. (1994) PATGEN: an automatic program to generate theoretical Patterson peaks and to compare them with experimental Patterson peaks. *J. Appl. Crystallogr.*, **27**, 860–861.

Collaborative Computational Project No.4 (1994) The CCP4 suite: programs from protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.

Commans,S., Plateau,P., Blanquet,S. and Dardel,F. (1995) Solution structure of the anticodon binding domain of *Escherichia coli* lysyl-tRNA synthetase and studies of its interaction with $tRNA^{Lys}$. *J. Mol. Biol.*, **253**, 100–113.

Cowtan,K. (1994) 'dm': an automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, **31**, 34–38.

Fleishman,R.D. *et al.* (1995) Whole-genome random sequencing of *Haemophilus influenzae* Rd. *Science*, **269**, 496–512.

Guillon,J.M., Mechulam,Y., Schmitter,J.M., Blanquet,S. and Fayat,G. (1992a) Disruption of the gene for Met-$tRNA_f^{Met}$ formyltransferase severely impairs growth of *Escherichia coli*. *J. Bacteriol.*, **174**, 4294–4301.

Guillon,J.M., Meinnel,T., Mechulam,Y., Lazennec,C., Blanquet,S. and Fayat,S. (1992b) Nucleotides of tRNA governing the specificity of *Escherichia coli* methionyl-$tRNA_f^{Met}$ formyltransferase. *J. Mol. Biol.*, **224**, 359–367.

Guillon,J.M., Mechulam,Y., Blanquet,S. and Fayat,G. (1993) Importance of formylability and anticodon stem sequence to give $tRNA^{Met}$ an initiator identity in *Escherichia coli*. *J. Bacteriol.*, **175**, 4507–4514.

Higgins,D.G. and Sharp,P.M. (1989) Fast and sensitive multiple sequence alignments on a microcomputer. *Comput. Applic. Biosci.*, **5**, 151–153.

Hirel,P.H., Schmitter,J.-M., Dessen,P., Fayat,G. and Blanquet,S. (1989) Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proc. Natl Acad. Sci. USA*, **86**, 8247–8251.

Inglese,J., Smith,J.M. and Benkovic,S.J. (1990) Active-site mapping and site-specific mutagenesis of glycinamide ribonucleotide transformylase from *Escherichia coli*. *Biochemistry*, **29**, 6678–6687.

Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for the building of protein models in electron density maps and the location of errors in these models. *Acta Cystallogr.*, **A47**, 110–119.

Kabsch,W.J. (1988) Evaluation of single crystal X-ray diffraction data from a position sensitive detector. *J. Appl. Crystallogr.*, **21**, 916–924.

Kahn,D., Fromant,M., Fayat,G., Dessen,P. and Blanquet,S. (1980) Methionyl-transfer-RNA transformylase from *Escherichia coli*: purification and characterisation. *Eur. J. Biochem.*, **105**, 489–497.

Kleijwegt,G.J. and Jones,T.A. (1994) *Proceedings of the CCP4 Study Weekend: from First Map to Final Model*. SERC Daresbury Laboratory, Warrington, UK, pp. 59–66.

Kraulis,P. (1991) A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

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.

Lee,C.P., Seong,B.L. and RajBhandary,U.L. (1991) Structural and sequence elements important for recognition of *Escherichia coli* formylmethionine tRNA by methionyl-tRNA transformylase are clustered in the acceptor stem. *J. Biol. Chem.*, **266**, 18012–18017.

Luzatti,P.V. (1952) Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Crystallogr.*, **5**, 802–810.

Mangroo,D. and RajBhandary,U.L. (1995) Mutants of *Escherichia coli* initiator tRNA defective in initiation. *J. Biol. Chem.*, **270**, 12203–12209.

Meinnel,T. and Blanquet,S. (1994) Characterization of the *Thermus thermophilus* locus encoding peptide deformylase and methionyl-$tRNA_f^{Met}$ formyltransferase. *J. Bacteriol.*, **176**, 7387–7390.

Meinnel,T. and Blanquet,S. (1995) Maturation of pre-$tRNA_f^{Met}$ by *E.coli* RNase P is specified by a guanosine of the 5′ flanking sequence. *J. Biol. Chem.*, **270**, 15906–15914.

Meinnel,T., Mechulam,Y. and Blanquet,S. (1993a) Methionine as translation start signal: a review of the enzymes of the pathway in *Escherichia coli*. *Biochimie*, **75**, 1061–1075.

Meinnel,T., Mechulam,Y., Lazennec,C., Blanquet,S. and Fayat,G. (1993b) Critical role of the acceptor stem of $tRNAs^{Met}$ in their aminoacylation by *Escherichia coli* methionyl-tRNA synthetase. *J. Mol. Biol.*, **229**, 26–36.

Murzin,A.G. (1993) OB (oligonucleotide/oligosaccharide)-fold: common structural and functional solution for non-homologous sequences. *EMBO J.*, **12**, 861–867.

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.

Nicholls,A., Sharp,K.A. and Honig,B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–286.

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.

Otwinowski,Z. (1991) Maximum likelihood refinement of heavy atom parameters. In Wolf,W., Evans,P.R. and Leslie,A.G.W. (eds), *Isomorphous Replacement and Anomalous Scattering*. Science and Engineering Research Council, Warrington, UK, pp. 80–86.

Reid,K.S.C., Lindley,P.F. and Thornton,J.M. (1985) Sulphur-aromatic interactions in proteins. *FEBS Lett.*, **190**, 209–213.

Rosenberg,A., Lade,N., Chui,D., Lin,S., Dunn,J. and Studier,F. (1987) Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene*, **56**, 125–135.

Rossmann,M.G., Moras,D. and Olsen,K.W. (1974) Chemical and biological evolution of a nucleotide-binding protein. *Nature*, **250**, 194–199.

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.

Schmitt,E., Mechulam,Y., Ruff,M., Mitschler,A., Moras,D. and Blanquet,S. (1996) Crystallization and preliminary X-ray analysis of *Escherichia coli* methionyl-$tRNA_f^{Met}$ formyltransferase. *Proteins*, **25**, in press.

Skala,J., Van Dyck,L., Purnelle,B. and Goffeau,A. (1992) The sequence of a 8 kb segment on the left arm of chromosome II from *Saccharomyces cerevisiae* identifies five new open reading frames of unknown functions, two tRNA genes and two transposable elements. *Yeast*, **8**, 777–785.

Stein,P.E., Boodhoo,A., Tyrrell,G.J., Brunton,J.L. and Read,R.J. (1992) Crystal structure of the cell-binding B oligomer of verotoxin-1 from *Escherichia coli*. *Nature*, **355**, 748–750.

Wallis,N.G., Dardel,F. and Blanquet,S. (1995) Heteronuclear NMR studies of the interactions of the ^{15}N-labeled methionine-specific transfer RNAs with methionyl-tRNA transformylase. *Biochemistry*, **34**, 7668–7677.

Received on April 11, 1996; revised on May 13, 1996