Structure of the EMAPII domain of human aminoacyl-tRNA synthetase complex reveals evolutionary dimer mimicry

Louis Renault, Pierre Kerjan, Sebastiano Pasqualato, Julie Ménétrey, Jean-Charles Robinson, Shin-ichi Kawaguchi¹, Dmitry G. Vassilyev¹, Shigeyuki Yokoyama¹, Marc Mirande and Jacqueline Cherfils²

Laboratoire d'Enzymologie et Biochimie Structurales, CNRS, 1 Avenue de la Terrasse, 91198 Gif sur Yvette cedex, France and ¹Riken Harima Institute, 1-1-1, Kouto, Mikazuki-cho, Sayo-gun, Hyogo 679-5148, Japan

²Corresponding author
e-mail: cherfils@lebs.cnrs-gif.fr

The EMAPII (endothelial monocyte-activating polypeptide II) domain is a tRNA-binding domain associated with several aminoacyl-tRNA synthetases, which becomes an independent domain with inflammatory cytokine activity upon apoptotic cleavage from the p43 component of the multisynthetase complex. It comprises a domain that is highly homologous to bacterial tRNA-binding proteins (Trbp), followed by an extra domain without homology to known proteins. Trbps, which may represent ancient tRNA chaperones, form dimers and bind one tRNA per dimer. In contrast, EMAPII domains are monomers. Here we report the crystal structure at 1.14 Å of human EMAPII. The structure reveals that the Trbp-like domain, which forms an oligonucleotide-binding (OB) fold, is related by degenerate 2-fold symmetry to the extra-domain. The pseudo-axis coincides with the dyad axis of bacterial TtCsaA, a Trbp whose structure was solved recently. The interdomain interface in EMAPII mimics the intersubunit interface in TtCsaA, and may thus generate a novel OB-fold-based tRNA-binding site. The low sequence homology between the extra domain of EMAPII and either its own OB fold or that of Trbps suggests that dimer mimicry originated from convergent evolution rather than gene duplication.

Keywords: aminoacyl-tRNA synthetase complex/convergent evolution/crystal structure/cytokine/EMAPII

Introduction

Endothelial monocyte-activating polypeptide II (EMAPII) is a 22 kDa multifunctional polypeptide with dual activities as both a cytokine and a tRNA-binding domain, which was initially isolated from cultures of methylcholanthrene A-transformed fibrosarcoma cells (Kao *et al.*, 1992). *In vitro*, EMAPII stimulates chemotactic migration of polymorphonuclear leukocytes and mononuclear phagocytes, and induces expression of tissue factors on endothelial cells. Treatment of certain tumours with

EMAPII followed by systemic administration of tumour necrosis factor- α (TNF- α) resulted in acute thrombohaemorrhage and tumour regression (Kao *et al.*, 1994b; Marvin *et al.*, 1996). EMAPII is the maturation product of a 43 kDa precursor, proEMAPII. It has been proposed that its cleavage and release occurs during apoptosis, since EMAPII was observed in the culture medium of apoptotic cells (Knies *et al.*, 1998) and is cleaved *in vitro* from proEMAPII by an apoptotic protease, Caspase 7 (Behrendorf *et al.*, 2000). We recently identified proEMAPII as a protein involved in protein synthesis in mammals, the p43 component of the aminoacyl-tRNA synthetase complex, and showed that proEMAPII and EMAPII bind tRNA in a non-specific manner (Quevillon *et al.*, 1997). An EMAPII-like domain was also recovered from the C-terminal polypeptide extension of human TyrRS and was shown to have cytokine activities similar to those of EMAPII (Wakasugi and Schimmel, 1999b).

Aminoacyl-tRNA synthetases are ubiquitous enzymes that catalyse aminoacylation of tRNAs by their cognate amino acid (Schimmel and Söll, 1979). In higher eukaryotes, nine aminoacyl-tRNA synthetases (GluProRS, IleRS, LeuRS, MetRS, GlnRS, ArgRS, LysRS and AspRS) are associated with three auxiliary proteins (p43, p38 and p18) to form a multienzyme complex (Mirande, 1991) with a definite topology (Quevillon *et al.*, 1999). In this complex, p43 makes protein-protein interactions with p38, ArgRS and GlnRS. The hamster cDNA encoding p43 was isolated and shown to specify a 359-amino acid protein (Quevillon *et al.*, 1997). Its C-terminal moiety, starting from residue Asp193, was shown to be the hamster homologue of human and murine EMAPII isolated by Kao *et al.* (1994b). The hamster cDNA has a coding potential for 46 additional N-terminal residues as compared with human and murine cDNAs, suggesting that the latter lack a short segment in their 5'-upstream sequences. Interestingly, the EMAPII region of human p43 displays extensive similarities to the C-terminal domain of human TyrRS (54% sequence identity; Kleeman *et al.*, 1997), of nematode (57%) or plant MetRS (53%) (Kaminska *et al.*, 2000), and of Arc1p (52%), a cofactor of yeast MetRS and GluRS (Simos *et al.*, 1996). The finding that pro-EMAPII and EMAPII-like domains are invariably associated with various aminoacyl-tRNA synthetases suggested that this recurrent polypeptide might be primarily involved in protein synthesis. Accordingly, EMAPII-like domains have a general RNA-binding property and therefore might be involved in recruiting tRNA for aminoacylation (Quevillon *et al.*, 1997; Simos *et al.*, 1998; Chihade and Schimmel, 1999; Wang and Schimmel, 1999; Kaminska *et al.*, 2000). Also in agreement with a major role of EMAPII-like domains as cofactors for aminoacyl-tRNA synthetases is the finding that the N-terminal segment of EMAPII is also present in several bacterial MetRSs or

Table I. Statistics on data collection and phasing

	Crystal form 1				Crystal form 2
	Native	Derivative 1 ^a	Derivative 2 ^a	Derivative 2 ^a	
Data collection					
$P2_1$ unit cell parameters					
a (Å)	35.95	35.98	35.93	35.99	36.14
b (Å)	59.68	59.81	59.72	59.72	53.94
c (Å)	37.85	37.88	37.86	37.90	41.37
β (°)	114.82	115.15	115.02	115.04	111.92
wavelength (Å)	0.934	1.0009	1.0386	1.0009	0.934
resolution limits (Å)	30–1.14	30–1.64	30–1.69	30–1.69	30–1.90
reflections					
measured	269210	63936	67865	76366	119628
unique	52816	34168	31260	30882	11716
completeness (%) ^b	99.9 (99.1)	97.3 (95.0)	97.6 (94.8)	96.1 (92.7)	99.4 (96.5)
R_{symm} (%) ^{b,c}	6.2 (15.7)	6.2 (27.8)	7.1 (35.2)	6.2 (25.8)	4.2 (18.3)
I/σ^b	23.5 (6.4)	8.2 (2.5)	9.7 (3.1)	10.3 (3.2)	24.2 (6.0)
MIRAS phasing					
phasing power (ac/c) ^{d,e}		2.64/1.33	2.63/1.31	2.04/1.03	
R_{cullis} (ac/c) ^{e,f}		0.56/0.78	0.57/0.77	0.66/0.82	
anomalous R_{cullis} ^f		0.97	0.99	0.96	
overall figure of merit	0.521				
Refinement					
resolution (Å)	30–1.50				30–2.05
reflections for $R_{\text{cryst}}/R_{\text{free}}$	23301/1127				9020/947
R_{cryst} (%) ^g	22.1 (26.2)				22.1 (29.6)
R_{free} (%) ^g	23.0 (34.7)				27.2 (34.9)
No. of atoms/No. of water atoms	1276/250				1337/142
average B -factor (Å ²)	20.6				32.6
r.m.s.d. bond length (Å)	0.005				0.005
r.m.s.d. angles (°)	1.6				1.5

^aDerivatives 1 and 2 (see Materials and methods for crystal preparation) share the same three heavy-atom binding sites (Cys211, Cys229 and Cys284) with different occupancy. Crystal form 2 (see Materials and methods) shows heavy-atom binding sites at Cys161 and Cys284.

^bValues in parentheses correspond to the highest resolution shells: 1.16–1.14 Å for native, 1.68–1.64 Å for derivative 1, 1.72–1.69 Å for derivative 2 and 1.97–1.90 Å for crystal form 2.

^c $R_{\text{symm}} = \sum_h \sum_i |I_{h,i} - \langle I_h \rangle| / \sum_h \sum_i I_{h,i}$

^dPhasing power = $[\sum |F_H|^2] / \sum (F_{\text{PH,obs}} - F_{\text{PH,calc}})^2]^{1/2}$, where $F_{\text{PH,obs}}$ and $F_{\text{PH,calc}}$ are the observed and calculated structure factor amplitudes for the derivative, respectively.

^eac/c refers to acentric and centric data.

^f $R_{\text{cullis}} = \sum_h ||F_{\text{PH}} \pm F_{\text{pl}} - |F_H|| / \sum_h |F_{\text{PH}} \pm F_{\text{pl}}|$, where F_H is the calculated structure factor for heavy atom.

^gValues in parentheses correspond to highest resolution shells used in refinement: 1.59–1.50 for crystal form 1 and 2.18–2.05 for crystal form 2.

PheRSs, and forms an independent subunit in a novel family of bacterial tRNA-binding proteins, which includes Trbp111 (Morales *et al.*, 1999) and CsaA (Müller *et al.*, 2000; Kawaguchi *et al.*, 2001). In contrast with EMAPII, which is monomeric in solution (Quevillon *et al.*, 1997), Trbp111 and CsaA are dimers (Morales *et al.*, 1999; Müller *et al.*, 2000). Likewise, the Trbp-like C-terminal domain of *Escherichia coli* or *Thermus thermophilus* MetRSs is involved in their dimerization (Cassio and Waller, 1971; Kohda *et al.*, 1987; Mellot *et al.*, 1989). Another major difference between EMAPII and Trbp111 is their relative affinities for tRNAs. Whereas EMAPII binds tRNAs with an apparent K_D of 20 μM (Quevillon *et al.*, 1997), Trbp111 forms stable complexes with tRNAs (apparent K_D = 32 nM) (Morales *et al.*, 1999).

These studies have shown that the EMAPII domain of human p43 has the dual potential to bind tRNA and elicit cytokine activities. Because cleavage of EMAPII from the multisynthetase complex results in the loss of a cofactor for aminoacylation, p43 might identify with a molecular fuse that would trigger the irreversible cell growth/cell death transition. In order to uncover the structural determinants responsible for this functional duality, we

have undertaken the molecular characterization of human EMAPII. The crystal structure revealed an unexpected evolutionary relationship of the eukaryotic EMAP family to bacterial Trbp/CsaA proteins. In contrast to Trbp/CsaA proteins, which form dimers of oligonucleotide-binding (OB) fold subunits, EMAPII builds a monomeric pseudo-dimer where the dimer interface is mimicked by the interdomain interface. As compared with bona fide OB folds (Draper and Reynaldo, 1999), the RNA binding site in EMAPII should differ significantly, involving the interfacial region in addition to the apex of the $\beta 1$ – $\beta 2$ – $\beta 3$ sheet of the OB fold.

Results

Description of the EMAPII crystal structure

The crystal structure of the EMAPII domain of human p43 was determined to 1.14 Å resolution by the MIRAS method (crystal form 1), and in a second crystal form to 1.9 Å resolution by molecular replacement (crystal form 2) (Table I). The protein forms a compact structure, which can be described as two tightly associated domains of 100 and 60 amino acids, respectively (Figure 1). The

N-terminus domain (residues 148–252) forms an open β -barrel, characteristic of the OB fold, as predicted from its homology to the B2 domain of phenylalanyl-tRNA synthetase (PheRS) to which it is structurally very close [root mean square deviation (r.m.s.d.) = 1.2 Å on 92 C α s] (Mosyaka *et al.*, 1995; Goldgur *et al.*, 1997; Quevillon *et al.*, 1997). The OB fold displays two β -sheets, β 1– β 2– β 3 and β 1– β 4– β 5, which share the twisted β 1 strand and interact at almost a right angle with each other. The β -barrel constitutes the platform for intervening loops, several of which form short organized structures in EMAPII: small 3_{10} helices in the N-terminus and after β 3, and β -hairpins between β 4 and β 5 (β 1'– β 2') and following β 5 (β 3'– β 4'). The loops between β 1– β 2 and

Whereas the OB fold is common in RNA-binding proteins and in a number of other proteins unrelated to oligonucleotide binding (Murzin, 1993; Draper and Reynaldo, 1999), no protein with a significant similarity score to the C-terminal domain (C-domain, residues 253–312) was detected by the program DALI (<http://www2.embl-ebi.ac.uk/dali>). The C-domain is connected to the OB fold by a linker (residues 253–268), which wraps

Fig. 2. The internal pseudo-dyad in human EMAPII coincides with the 2-fold axis in bacterial CsaA. **(A)** Stereoview of EMAPII (in violet) superimposed onto itself (in orange), with superimposable regions shown as thick lines. The pseudo 2-fold axis is shown by an arrow. **(B)** The C-terminus domain of EMAPII (in yellow) is related by 2-fold symmetry to a subset of the OB fold (in dark violet). **(C)** Crystal structure of CsaA from *T.thermophilus* (Kawaguchi *et al.*, 2001). A subset of the N-terminus of the symmetrical subunit (in yellow) matches the C-terminus domain of EMAPII. Orientations in (A), (B) and (C) are as in Figure 1A. **(D)** Structure-based sequence alignment of the C-terminal domain of EMAPII with its N-terminus domain and with the symmetry-related subunit of CsaA. Helices and strands are on a coloured background, with helices boxed. The C-domain of EMAPII is superimposable as a continuous peptide. Dots indicate residues of EMAPII (N-terminal domain) or CsaA (symmetry-related subunit) that superimpose with the C-terminus domain within a 2.5 Å cutoff; numbers indicate the length of non-superposable intervening sequences.

around the N-terminus and strand $\beta 1$, and includes two turns of 3_{10} helix. The domain itself is made of six short β -strands, associated in an antiparallel four-stranded β -sheet (βe – βa – βb – βc) and a β -hairpin between the bent βc and βd strands. The two domains interact through a

tight hydrophobic interface, and display a groove at their interface region that accommodates the C-terminus. The EMAPII construct used in this study is fused to a Leu-Glu-His₆ tag in the C-terminus, which is disordered beyond Glu³¹⁴ in crystal form 1 and clearly visible in crystal

Fig. 3. The interfacial regions of EMAPII and CsaA. The orientations in (A), (B) and (C) are as in Figure 1A, and the secondary structures colour coded in (B) and (C) as in Figure 2. (A) Interdomain interface of EMAPII. The C-terminus region is in yellow. Trp271 adopts two different conformations in the two crystal forms (see Results). Lysines 267, 268 and 269 have no equivalent in CsaA, but the KKKxW stretch is highly conserved in EMAPII-like proteins. (B) Intersubunit interface in CsaA. The symmetry-related subunit is in yellow. Residues from the symmetry-related subunit are marked with an asterisk. (C) Electrostatic polarization of EMAPII, with positive potential in blue and negative potential in red. The positive potential is located at the domain interface and the $\beta'1$ - $\beta'2$ hairpin. This image is drawn with SwissPdbViewer (<http://www.expasy.ch/spdbv>).

form 2. In one crystal form, but not the other, the Leu side chain from the tag is buried in a slot created at the domain

interface by the displacement of Trp²⁷¹. This suggests that Trp²⁷¹ has the dynamics to build a binding site for a hydrophobic ligand (see below).

EMAPII forms a truncated dimer of OB folds

Purified EMAPII is a monomer (Quevillon *et al.*, 1997) and, although p43/proEMAPII is a dimer, its dimerization is sustained by the N-terminal domain that is cleaved during maturation of EMAPII (J.-C. Robinson and M. Mirande, unpublished data). In agreement with these observations, EMAPII is present as a monomer in the crystal, which indicates that it does not dimerize even at the submillimolar concentration used for crystallization. The OB fold domain of EMAPII has 30% sequence identity with a recently characterized family of bacterial tRNA-binding proteins, which includes Trbp111 (Morales *et al.*, 1999) and CsaA (Stover *et al.*, 2000). Unlike EMAPII, the Trbp111 polypeptide is a bare OB fold without an extra C-terminal domain and it associates into a stable dimer, which raised the issue that the EMAPII OB fold could have this ability as well (Morales *et al.*, 1999). Having ruled out that native EMAPII forms dimers, we investigated for internal symmetry as an alternative to dimerization and identified a 2-fold pseudo-symmetry that relates the extra domain to the OB fold (Figure 2A and B). Pseudo-symmetry superimposes most secondary structures of the C-domain onto a subset of the N-domain that encompasses the N-terminus helix, the $\beta1$ - $\beta4$ - $\beta5$ sheet and the $\beta'3$ - $\beta'4$ hairpin (r.m.s.d. = 1.4 Å on 37 C α s) but excludes the $\beta1$ - $\beta2$ - $\beta3$ β -sheet and the $\beta'1$ - $\beta'2$ hairpin. Surprisingly, except for the network of hydrophobic residues at the interdomain interface, there is almost no sequence homology between the two domains (Figure 2D). The C-domain superimposes onto the N-domain as a continuous polypeptide, with the short loops that connect its β -strands making short-cuts into long segments of the OB fold. Thus, EMAPII mimics a dimeric OB fold whose second subunit would have been trimmed from its first β -sheet.

Pseudo-symmetry axis in EMAPII coincides with bacterial Trbp/CsaA dyad axis and creates a conserved interface region

The structure of TtCsaA from the extreme thermophile *T. thermophilus*, a close homologue of *Aquifex aeolicus* Trbp111 and *Bacillus subtilis* CsaA, has been solved recently at 2 Å resolution (Kawaguchi *et al.*, 2001) (Figure 2C). The molecule associates as a dimer, which we compared with the monomeric structure of EMAPII. The OB fold of EMAPII closely resembles that of CsaA (r.m.s.d. = 1.1 Å on 97 C α s) and was chosen as a guide for superimposition. Remarkably, we find that the internal dyad in EMAPII coincides exactly with the 2-fold axis in the CsaA dimer and superimposes the C-domain onto the symmetry-related subunit of the CsaA dimer (r.m.s.d. 1.3 Å on 39 C α s). Regions of the second CsaA OB fold that superimpose on the C-domain of EMAPII match those identified in the OB fold of EMAPII by internal symmetry (Figure 2B and C), and, similarly, have little sequence homology with the C-domain except for the hydrophobic interface (Figure 2D). These findings provide a striking confirmation that, as identified from internal symmetry, EMAPII has the characteristics of a dimeric OB fold. As a

consequence, the face of the OB fold that forms the dimer interface in CsaA is buried by the C-domain in EMAPII, where it is unavailable for homologous dimerization. Thus, unlike CsaA and Trbp111, the EMAPII domain cannot form dimers, regardless of whether it has been cleaved from proEMAPII/p43 or not.

Besides these topological similarities, 2-fold symmetry also generates a shallow depression at the interdomain (EMAPII) and intersubunit (CsaA) interface that extends on both side of the C-terminus, which in CsaA is contributed by the symmetry-related subunit. It is lined on one side by the $\beta 5$ strands from the N-domain and capped by the $\beta 1$ – $\beta 2$ hairpin. Its other side is bordered by residues 260–275 from the linker in EMAPII, or the N-terminus helix of the symmetry-related subunit in CsaA. We find that this interface region has similarities in shape, organization and amino acid composition in the two proteins (Figure 3A and B). There is an exact correspondence of the C-terminal residue in both interfaces and of several exposed hydrophobic and basic residues in its vicinity. Thus, pseudo-symmetry in EMAPII mimics an interface region that is created by homodimerization in CsaA. We discuss below the implications of these findings for the quaternary organization of Trbp/Csa/EMAPII-related proteins and how this could be related to a conserved function.

Discussion

Convergent evolution of a pseudo-dimeric OB-fold-based domain

The crystal structure of the EMAPII domain of human p43, a component of the multisynthetase complex, has revealed that an internal pseudo-dyad relates the classical OB fold in the N-terminus to a novel domain in the C-terminus and mimics the 2-fold symmetry in TtCsaA, a bacterial tRNA-binding protein whose monomer is an independent OB fold (Kawaguchi *et al.*, 2001). Interdomain/intersubunit 2-fold symmetry generates an interfacial region with similarities in organization and chemical composition between the prokaryotic OB fold dimer and mammalian EMAPII monomer. The pseudo-symmetry of EMAPII and its relationship to CsaA, which were not expected from the sequences alone, raise the possibility that EMAPII may have evolved by domain fusion from an ancient dimeric tRNA-binding OB fold exemplified by the CsaA dimer. Dimers fused into a symmetrical monomer are illustrated by at least two sets of structures in the Protein Data Bank: the bacterial elongation factor EF-Ts, which is a monomer in *E. coli* (Kawashima *et al.*, 1996) and a dimer in *T. thermophilus* (Wang *et al.*, 1997); and monomeric dihaemic cytochrome *c4* from *Pseudomonas stutzeri* (Kadziola and Larsen, 1997), which mimics dimeric cytochrome *c552* from *Pseudomonas nautica* (Brown *et al.*, 1999). In both systems, there is significant sequence similarity of both domains to the parent dimer subunit, although we note that the similarity is higher for one domain than for the other. In addition, the monomer contributes two equivalent subsets to its cognate dimer. This is in contrast with EMAPII and CsaA in several aspects. First, whereas there is clear sequence similarity of TtCsaA to the N-terminal OB fold of EMAPII, its similarity to the C-domain is

barely detectable even from a structure-based sequence alignment (Figure 2D). Secondly, only one OB fold from the dimer remains intact in EMAPII, whereas the second has been widely trimmed. This would suggest that the C-terminal OB fold, but not the N-terminal OB fold, underwent accelerated divergence from the original OB fold following domain fusion, including the loss of several long loops. This hypothesis, which implies that there is little evolutionary pressure on the interfacial region, accounts poorly for the conserved organization of this region. An alternative hypothesis is that the topology of EMAPII is the result of convergent evolution, possibly from the fusion of an ancestral CsaA-like OB fold with an unrelated domain. This suggests that the interface region has been reconstituted by evolution in monomeric EMAPII to fulfil a common essential function.

Implications for tRNA binding

Two families of EMAPII-like OB folds have been identified to date. Bare domains are found in bacterial proteins of the Trbp111/CsaA family. Alternatively, the domain can be fused to an extra C-terminal domain, and be associated to aminoacyl-tRNA synthetases either *in trans*, as in p43/proEMAPII from different species (Quevillon *et al.*, 1997) and yeast Arc1p (Simos *et al.*, 1998), or *in cis*, as in several mammalian aminoacyl-tRNA synthetases (Kleeman *et al.*, 1997). Both families have the ability to bind tRNA (Simos *et al.*, 1996; Quevillon *et al.*, 1997; Morales *et al.*, 1999; Kawaguchi *et al.*, 2001). The bacterial Trbp111/CsaA family achieves this function as a dimer, which implies that two symmetry-related binding sites are potentially available. However, Trbp111 and CsaA bind only one tRNA (Morales *et al.*, 1999; Kawaguchi *et al.*, 2001), either because of steric or short-range electrostatic hindrance at the second binding site once the first tRNA is bound, or because the tRNA binding site might span both subunits (see below). The crystal structure of EMAPII shows that the region responsible for dimerization in TtCsaA is not available for dimer formation in EMAPII where it is masked by the C-terminal domain. Degenerate dimeric features are restored by pseudo-symmetry, but they do not duplicate either the OB fold or the interfacial region. Consequently, EMAPII cannot form dimers and is likely to carry a single tRNA binding site.

Our structural study clarifies the organization of tRNA-binding proteins that incorporate EMAPII-like OB folds. It shows that the C-terminus of the OB fold must not be fused to an extra-domain to allow protein dimerization, and that the pattern of hydrophobic residues that are conserved at the dimer interface in CsaA or the domain interface in EMAPII are a signature of Csa/Trbp/EMAPII-related OB folds that form actual or pseudo-dimeric interfaces (Figure 1B). Thus, yeast Arc1p, human TyrRS and MetRS from plants, which carry an extra-domain homologous to the C-terminal domain of EMAPII, are unlikely to form dimers from their EMAPII-like domain. A related case is that of MetRS from the spirochetes *Borrelia burgdorferi* and *Treponema pallidum*, which have a C-terminal domain appended to their CsaA-like OB fold. Although this extra domain has little sequence identity with the C-terminal domain of EMAPII (Figure 1B), the OB fold displays the conserved pattern of hydrophobic residues,

which suggests that it may form a domain interface but not a dimer. TtPheRS OB-fold is another related case, as its OB fold carries a C-terminus extension that prevents dimerization; however, its N- and C-terminus extensions fold into a domain that is unrelated to the C-terminus of EMAPII and does not display internal symmetry (Mosyak *et al.*, 1995; Goldgur *et al.*, 1997). Conversely, bacterial MetRSs that have CsaA-like OB folds devoid of an extra C-terminus domain, such as those of *E.coli* and *T.thermophilus* (Cassio and Waller, 1971; Kohda *et al.*, 1987; Mellot *et al.*, 1989), may use this domain as a dimerization unit.

The structure also reveals an extended similarity between the EMAPII OB-fold and that of bacterial PheRS and Trbp111, and the recurrence of the interfacial region in EMAPII and in the CsaA dimer. Thus, both the classical nucleotide binding pocket on the OB fold (reviewed in Draper and Reynaldo, 1999) and the distinctive interfacial region display evolutionarily conserved features, which point to a possible common strategy of this class of proteins in binding tRNA. Although OB folds have so far escaped the definition of an unequivocal RNA binding site, it has been proposed that the $\beta 1$ – $\beta 2$ – $\beta 3$ sheet forms a general scaffold whose specific properties are modulated by the loops grafted onto the β -barrel. This view has been extended by recent structural studies of RNA-binding proteins that carry extra-domains N- or C-terminal to their OB fold, including ribosomal L2 (Nakagawa *et al.*, 1999), the translation initiation factor eIF1A (Battiste *et al.*, 2000) and the transcription termination factor Rho (Briercheck *et al.*, 1998). These studies have mapped amino acids that are critical for RNA binding not only on the $\beta 1$ – $\beta 2$ – $\beta 3$ pocket, but also at the interface between the OB fold and the extra domain on its opposite side.

With this background, and bearing in mind the conservation of both the OB fold architecture and the interfacial region, we propose that the tRNA-binding site straddles both sides of the OB fold, including the interfacial region and the $\beta'1$ – $\beta'2$ pliers at the tip of the $\beta 1$ – $\beta 2$ – $\beta 3$ sheet. In support of this hypothesis, we find that EMAPII is strongly polarized, with a positively charged surface, as expected for long-range attraction for tRNA, spanning the $\beta'1$ – $\beta'2$ hairpin and the linker at the interfacial region (Figure 3C). Conversely, the opposite side of EMAPII develops a negative potential which rules out that this region binds tRNA. Remarkably, the positively charged region includes the contribution of a patch of residues, ²⁶⁷KKK ϕ WE²⁷² in EMAPII, which is the best-conserved signature of EMAPII-like extra domains so far (Figure 1B). Furthermore, by comparing our two crystal forms we find that Trp²⁷¹ in this region can be displaced by the flexible His-tag to accommodate a hydrophobic ligand, and may form a non-specific binding site for a tRNA base. Although this places the His-tag close to the potential tRNA-binding site, binding of tRNA is not hindered by the presence of the tag, as assessed with an EMAPII construct devoid of a His-tag (M.Mirande, unpublished gel retardation data), probably owing to its flexibility.

At the moment, we consider that these data are insufficient to dock tRNA onto EMAPII reliably, all the more so because the actual binding site of tRNA may also include the polybasic region N-terminal to EMAPII in

proEMAPII/p43. In support of this hypothesis, both domains of p43 bind tRNA, the affinity of p43 for tRNA being higher than that of its individual domains (M.Mirande, unpublished results). In addition, a polybasic region in the N-terminus of yeast Arc1p was also reported to synergize with its EMAPII-like domain for binding tRNA (Simos *et al.*, 1998). This suggests that the tRNA-binding site on p43 is bimodal, including the EMAPII domain and the N-terminal polybasic domain. Cleavage of the N-terminus polybasic region by caspases would produce an EMAPII domain with reduced tRNA-binding ability, in agreement with its functional switch. Trbp111 may also be expected to have a bimodal binding site (Morales *et al.*, 1999), which would have the interfacial region, but not the second binding site, in common with the EMAPII-like family. The crystal structure of EMAPII now provides rationale guidelines for subsequent experiments, which are the next stage towards the validation and refinement of this model.

Structural insight into the cytokine activity

EMAPII has first been identified as a polypeptide released from immunogenic tumour cells and capable of eliciting multiple biological responses, including inflammatory chemoattraction (Kao *et al.*, 1992) and antiangiogenic activity in tumour vascularization (Kao *et al.*, 1994b). It was subsequently demonstrated that EMAPII is a proteolytic product of p43 (also called proEMAPII), a non-enzymic component of the eukaryotic multisynthetase complex (Quevillon *et al.*, 1997), and that it can be cleaved by an apoptotic caspase *in vitro* (Behrendorf *et al.*, 2000) and probably *in vivo* (Knies *et al.*, 1998). The EMAPII-like domain of human TyrRS elicits analogous responses when cloned as an independent domain (Wakasugi and Schimmel, 1999b). Based on the sequence similarity of the N-terminus of EMAPII to the von Willebrand factor and the similar cytokine-like properties of the two proteins, N-terminal peptides were derived and shown to mimic partially the properties of native EMAPII (Kao *et al.*, 1994a; Wakasugi and Schimmel, 1999a). Our crystal structure of EMAPII does not display any resemblance to known cytokines, as assessed by DALI. However, it shows that the N-terminal residues of EMAPII (residues 148–150) extend into the solvent, in agreement with their accessibility as a caspase cleavage site in immature proEMAPII. The structure also provides the native architecture of the active peptide, whose shortest active construct, ¹⁵⁸RIGRIVT¹⁶⁴, corresponds to strand $\beta 1$ of the OB fold (Figure 1). It shows that residues 158, 159, 160 and 162 are buried in the protein core, where they form tight interactions at the interdomain interface. Since the replacement of positions 158 and 161 by other residues has no effect on activity (Kao *et al.*, 1994a), only Val¹⁶³ and Thr¹⁶⁴ are similarly accessible to protein interactions in the peptide and in native EMAPII, and may be involved in cytokine activity. A highly specific response is unlikely to depend solely on two residues, which questions the correlation between the activity of the peptide and its presence in EMAPII. Our crystal structure rules out that the peptide unmask upon conformational changes following caspase cleavage, since this would require major denaturation of EMAPII at the tightly associated pseudo-dimer interface. This implies that either the biological

effects of the peptide are serendipitous, or that a similar peptide is exposed to protein–protein interactions in another protein with cytokine properties, a plausible candidate being the von Willebrand factor. Novel peptides could, however, be designed from the crystal structure taking their structural accessibility in native EMAPII into account, which could be evaluated for potential anti-cancer activity.

Materials and methods

Crystallization of EMAPII

The C-terminal residues 147–312 of human p43, corresponding to mature EMAPII, were expressed in *E. coli* with a Leu-Glu-His₆ tag in the C-terminus and purified as described in Quevillon *et al.* (1997). Crystals were grown by the vapour diffusion method in hanging drops containing equal volumes of EMAPII at 11 mg/ml and of the reservoir (28–35% PEG3000 or PEG 6000, 100 mM Tris pH 7.5, 100–300 mM NaCl, 2 mM dithiothreitol). Crystals were stabilized in 37% PEG in the same buffer before flash-freezing in liquid ethane. Two related crystal forms diffracting to 1.14 and 1.9 Å, respectively, were obtained with these conditions, with the same space group but different cell parameters. Heavy-atom derivatives were obtained for crystal form 1 by incubating the protein prior to crystallization with a 3.2-fold excess of mercury acetate over the protein concentration and subsequent crystal soaking in 2 mM gold potassium cyanide (derivative 1) or in a cocktail of mercury acetate, methylmercuric chloride, Thimerosal and gold potassium cyanide, each at 0.5 mM (derivative 2) for 1 h (Table I). The second crystal form was identified in the course of the heavy-atom screen, and only a derivatized data set, obtained by soaking crystals in 2.5 mM methylmercuric chloride for 1 h, could be collected.

Native data sets were collected at the synchrotron facilities of LURE (Orsay, France) (beam line W32) and ESRF (Grenoble, France) (beam line ID14-1). Data sets for the derivatives were collected at the Hg L2 edge (1.0009 Å) and Au L2 edge (1.0386 Å) on beam line BM30 (CRG/ESRF). Reflections were reduced and scaled with Denzo and Scalepack (Otwinowski, 1993) or XDS (Kabsch, 1993). Both crystal forms contain one molecule per asymmetric unit. Statistics for data collection, structure determination and refinement are shown in Table I.

Structure determination and refinement

Heavy-atom sites for crystal form 1 were identified by SOLVE (Terwilliger and Berendzen, 1999). Multiple heavy-atom isomorphous replacement and anomalous scattering (MIRAS) phases were generated with MLphare (CCP4, 1994) to 1.64 Å. Phase improvement with solvent flattening and histogram matching was done in DM (CCP4, 1994). Experimental phases were improved and extended to 1.14 Å using the automated refinement procedures of wARP (Perrakis *et al.*, 1999), resulting in automated building of 85% of the structure with side chains matched. The model was completed using the graphics program O (Jones and Kjeldgaard, 1997) and refined to 1.5 Å with conventional refinement in CNS (Brunger *et al.*, 1998). Refinement to 1.14 Å resolution using SHELX (Sheldrick and Schneider, 1997) is in progress.

The crystal form 2 was solved by molecular replacement with AMORE (Navaza, 1994) using the first crystal form refined to 1.5 Å as a starting model, and refined using TURBO (A.Roussel, A.G.Inisan and C.Cambillau, TURBO, AFMB and BioGraphics, Marseille, France) and CNS. It is essentially identical to other one except for a conformational change at the C-terminal His-tag, which results in different cell parameters and crystal packing.

Structure analysis

Superpositions were done with TURBO, using a cutoff of 2.5 Å. Accessible surfaces areas were computed with ASA (author, A.Lesk). Figures were drawn with Molscript (Kraulis, 1991) and Raster3D (Merrit and Murphy, 1994). Sequence alignments were performed with Clustal X (Jeanmougin *et al.*, 1998). Coordinates have been deposited in the Protein Data Bank under accession codes 1FL0 and 1E7Z.

Acknowledgements

We thank the staff at the LURE and ESRF synchrotron facilities for making beamlines W32 (LURE), BM30 and ID14 (ESRF) available to us,

and V.Shalak and M.Kaminska (LEBS) for unpublished gel retardation data. L.R. is supported by a grant from the European Molecular Biology Organization (ALTF264-1999) and S.P. by a grant from Institut Curie. This work was supported by grants from the Association pour la Recherche contre le Cancer, the Ligue contre le Cancer and Physique-Chimie du Vivant (CNRS).

References

- Battiste,J.L., Pestova,T.V., Hellen,C.U. and Wagner,G. (2000) The eIF1A solution structure reveals a large RNA-binding surface important for scanning function. *Mol. Cell*, **5**, 109–119.
- Behrendorf,H.A., van de Craen,M., Knies,U.E., Vandenabeele,P. and Clauss,M. (2000) The endothelial monocyte-activating polypeptide II (EMAP II) is a substrate for caspase-7. *FEBS Lett.*, **466**, 143–147.
- Briercheck,D.M., Wood,T.C., Allison,T.J., Richardson,J.P. and Rule,G.S. (1998) The NMR structure of the RNA binding domain of *E. coli* rho factor suggests possible RNA–protein interactions. *Nature Struct. Biol.*, **5**, 393–399.
- Brown,K., Nurizzo,D., Besson,S., Shepard,W., Moura,J., Moura,I., Tegoni,M. and Cambillau,C. (1999) MAD structure of *Pseudomonas nautica* dimeric cytochrome *c*552 mimicks the *c*4 dihemic cytochrome domain association. *J. Mol. Biol.*, **289**, 1017–1028.
- Brunger,A.T. *et al.* (1998) Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.
- Cassio,D. and Waller,J.P. (1971) Modification of *E. coli* methionyl-tRNA synthetase by proteolytic cleavage and properties of the trypsin modified enzyme. *Eur. J. Biochem.*, **20**, 283–300.
- Cavarelli,J., Rees,B., Ruff,M., Thierry,J.C. and Moras,D. (1993) Yeast tRNA(Asp) recognition by its cognate class II aminoacyl-tRNA synthetase. *Nature*, **362**, 181–184.
- Chihade,J.W. and Schimmel,P. (1999) Assembly of a catalytic unit for RNA microhelix aminoacylation using nonspecific RNA binding domains. *Proc. Natl Acad. Sci. USA*, **96**, 12316–12321.
- Collaborative Computational Project No. 4 (1994) The CCP4 suite: program for protein crystallography. *Acta Crystallogr. D*, **50**, 760–763.
- Draper,D.E. and Reynaldo,L.P. (1999) RNA binding strategies of ribosomal proteins. *Nucleic Acids Res.*, **27**, 381–388.
- Goldgur,Y., Mosyak,L., Reshetnikova,L., Anikilova,V., Lavrik,O., Khodyreva,S. and Safro,M. (1997) The crystal structure of phenylalanyl-tRNA synthetase from *Thermus thermophilus* complexed with cognate tRNA^{Phe}. *Structure*, **5**, 59–68.
- Jeanmougin,F., Thompson,J.D., Gouy,M., Higgins,D.G. and Gibson,T.J. (1998) Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.*, **23**, 403–405.
- Jones,T.A. and Kjeldgaard,M. (1997) Electron-density map interpretation. *Methods Enzymol.*, **227**, 173–208.
- Kabsch,W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell content. *J. Appl. Crystallogr.*, **26**, 795–800.
- Kadziola,A. and Larsen,S. (1997) Crystal structure of the dihaem cytochrome *c*4 from *Pseudomonas stutzeri* determined at 2.2 Å resolution. *Structure*, **5**, 203–216.
- Kaminska,M., Deniziak,M., Kerjan,P., Barciszewski,J. and Mirande,M. (2000) A recurrent general RNA-binding domain appended to plant methionyl-tRNA synthetase acts as a *cis*-acting cofactor for aminoacylation. *EMBO J.*, **19**, 6908–6917.
- Kao,J. *et al.* (1992) Endothelial monocyte-activating polypeptide II. A novel tumor-derived polypeptide that activates host-response mechanisms. *J. Biol. Chem.*, **267**, 20239–20247.
- Kao,J. *et al.* (1994a) A peptide derived from the amino terminus of endothelial-monocyte-activating polypeptide II modulates mononuclear and polymorphonuclear leukocyte functions, defines an apparently novel cellular interaction site and induces an acute inflammatory response. *J. Biol. Chem.*, **269**, 9774–9782.
- Kao,J. *et al.* (1994b) Characterization of a novel tumor-derived cytokine. Endothelial monocyte activating polypeptide II. *J. Biol. Chem.*, **269**, 25106–25119.
- Kawaguchi,S.-i., Müller,J., Linde,D., Kuramitsu,S., Shibata,T., Inoue,Y., Vassilyev,D.G. and Yokoyama,S. (2001) The crystal structure of the tCsaA protein: an export-related chaperone from *Thermus thermophilus*. *EMBO J.*, **20**, 562–569.
- Kawashima,T., Berthet-Colominas,C., Wulff,M., Cusack,S. and

- Leberman,R. (1996) The structure of the *Escherichia coli* EF-Tu.EF-Ts complex at 2.5 Å resolution. *Nature*, **379**, 511–518.
- Kim,Y., Shin,J., Li,R., Cheong,C., Kim,K. and Kim,S. (2000) A novel anti-tumor cytokine contains an RNA binding motif present in aminoacyl-tRNA synthetases. *J. Biol. Chem.*, **275**, 27062–27068.
- Kleman,T.A., Wei,D.B., Simpson,K.L. and First,E.A. (1997) Human tyrosyl tRNA synthetase shares amino acid sequence homology with a putative cytokine. *J. Biol. Chem.*, **272**, 14420–14425.
- Knies,U.E., Behrendorf,H.A., Mitchell,C.A., Deutsch,U., Risau,W., Drexler,H.C. and Clauss,M. (1998) Regulation of endothelial monocyte-activating polypeptide II release by apoptosis. *Proc. Natl Acad. Sci. USA*, **95**, 12322–12327.
- Kohda,D., Yokoyama,S. and Miyazawa,T. (1987) Functions of isolated domains of methionyl-tRNA synthetase from an extreme thermophile, *Thermus thermophilus* HB8. *J. Biol. Chem.*, **262**, 558–563.
- Kraulis,P.J. (1991) Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
- Marvin,M.R. *et al.* (1996) A novel tumor-derived mediator that sensitizes cytokine-resistant tumors to tumor necrosis factor. *J. Surg. Res.*, **63**, 248–255.
- Mellot,P., Mechulam,Y., Le Corre,D., Blanquet,S. and Fayat,G. (1989) Identification of an amino acid region supporting specific methionyl-tRNA synthetase:tRNA recognition. *J. Mol. Biol.*, **208**, 429–443.
- Merrit,E.A. and Murphy,M.E.P. (1994) Raster3D Version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. D*, **50**, 869.
- Mirande,M. (1991) Aminoacyl-tRNA synthetase family from prokaryotes and eukaryotes: structural domains and their implications. *Prog. Nucleic Acid Res. Mol. Biol.*, **40**, 95–142.
- Morales,A.J., Swairjo,M.A. and Schimmel,P. (1999) Structure-specific tRNA-binding protein from the extreme thermophile *Aquifex aeolicus*. *EMBO J.*, **18**, 3475–3483.
- Mosyak,L., Reshetnikova,L., Goldgur,Y., Delarue,M. and Safo,M.G. (1995) Structure of phenylalanyl-tRNA synthetase from *Thermus thermophilus*. *Nature Struct. Biol.*, **2**, 537–547.
- Müller,J.P., Ozegowski,J., Vettermann,S., Swaving,J., Van Wely,K.H. and Driessen,A.J. (2000) Interaction of *Bacillus subtilis* CsaA with SecA and precursor proteins. *Biochem. J.*, **348**, 367–373.
- Murzin,A.G. (1993) OB (oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J.*, **12**, 861–867.
- Nakagawa,A., Nakashima,T., Taniguchi,M., Hosaka,H., Kimura,M. and Tanaka,I. (1999) The three-dimensional structure of the RNA-binding domain of ribosomal protein L2; a protein at the peptidyl transferase center of the ribosome. *EMBO J.*, **18**, 1459–1467.
- Navaza,J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A*, **50**, 157–163.
- Otwinowski,Z. (1993) Oscillation data reduction program. In Sawyer,L., Isaacs,N. and Bailey,S. (eds), *CCP4 Study Weekend: Data Collection and Processing*. Daresbury Laboratory, Warrington, UK, pp. 56–62.
- Perrakis,A., Morris,R. and Lamzin,V.S. (1999) Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.*, **6**, 458–463.
- Quevillon,S., Agou,F., Robinson,J.C. and Mirande,M. (1997) The p43 component of the mammalian multi-synthetase complex is likely to be the precursor of the endothelial monocyte-activating polypeptide II cytokine. *J. Biol. Chem.*, **272**, 32573–32579.
- Quevillon,S., Robinson,J.C., Berthonneau,E., Siatecka,M. and Mirande,M. (1999) Macromolecular assemblage of aminoacyl-tRNA synthetases: Identification of protein-protein interactions and characterization of a core protein. *J. Mol. Biol.*, **285**, 183–195.
- Schimmel,P. and Söll,D. (1979) Aminoacyl-tRNA synthetases: general features and recognition of transfer RNAs. *Annu. Rev. Biochem.*, **48**, 601–648.
- Sheldrick,G. and Schneider,T.R. (1997) SHELX: high resolution refinement. *Methods Enzymol.*, **277**, 319–343.
- Simos,G., Segref,A., Fasiolo,F., Hellmuth,K., Shevchenko,A., Mann,M. and Hurt,E.C. (1996) The yeast protein Arc1p binds to tRNA and functions as a cofactor for the methionyl- and glutamyl-tRNA synthetases. *EMBO J.*, **15**, 5437–5448.
- Simos,G., Sauer,A., Fasiolo,F. and Hurt,E.C. (1998) A conserved domain within Arc1p delivers tRNA to aminoacyl-tRNA synthetases. *Mol. Cell*, **1**, 235–242.
- Stover,C.K. *et al.* (2000) Complete genome sequence of *Pseudomonas aeruginosa* PAO1 and opportunistic pathogen. *Nature*, **406**, 959–964.
- Terwilliger,T.C. and Berendzen,J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr. D*, **55**, 849–861.
- Wakasugi,K. and Schimmel,P. (1999a) Highly differentiated motifs responsible for two cytokine activities of a split human tRNA synthetase. *J. Biol. Chem.*, **274**, 23155–23159.
- Wakasugi,K. and Schimmel,P. (1999b) Two distinct cytokines released from a human aminoacyl-tRNA synthetase. *Science*, **284**, 147–151.
- Wang,C.C. and Schimmel,P. (1999) Species barrier to RNA recognition overcome with nonspecific RNA binding domains. *J. Biol. Chem.*, **274**, 16508–16512.
- Wang,Y., Jiang,Y., Meyering-Voss,M., Sprinzl,M. and Sigler,P.B. (1997) Crystal structure of the EF-Tu.EF-Ts complex from *Thermus thermophilus*. *Nature Struct. Biol.*, **4**, 650–656.

Received September 19, 2000; revised November 21, 2000;
accepted December 8, 2000

Note added in proof

While this manuscript was being completed, the crystal structure of EMAPII was reported by another laboratory (Kim *et al.*, 2000).