

# Crystallization and preliminary X-ray diffraction analysis of three EGF domains of EMR2, a 7TM immune-system molecule

Rachel J. M. Abbott,<sup>a</sup> Vroni Knott,<sup>b</sup> Pietro Roversi,<sup>a</sup> Saskia Neudeck,<sup>a</sup> Petra Lukacik,<sup>a</sup> Penny A. Handford<sup>b</sup> and Susan M. Lea<sup>a\*</sup>

<sup>a</sup>Laboratory of Molecular Biophysics, Department of Biochemistry, South Parks Road, Oxford OX1 3QU, England, and <sup>b</sup>Department of Biochemistry, South Parks Road, Oxford OX1 3QU, England

Correspondence e-mail:  
susan.lea@biop.ox.ac.uk

Crystals of three epidermal growth-factor-like (EGF) domains of EMR2 (143 residues) have been grown. EMR2 is a member of the EGF-TM7 family of proteins. Different splice variants exist with between three and five consecutive EGF modules linked to a seven-span transmembrane G-protein-coupled receptor. Although its precise function is unknown, EMR2 is highly expressed in immune tissues and has been shown to weakly bind CD55, a complement-system regulator. Here, crystallization of EMR2 in the presence of Ca<sup>2+</sup>, Ba<sup>2+</sup> and Sr<sup>2+</sup> ions is reported. A complete data set has been collected from all three crystal types, all of which belong to space group *P*<sub>2</sub><sub>1</sub>. An anomalous Patterson map from the Ba<sup>2+</sup> crystal data reveals three Ba<sup>2+</sup> ions bound within the asymmetric unit.

Received 13 February 2004  
Accepted 4 March 2004

## 1. Introduction

Most cell-surface proteins are comprised of different combinations of a limited set of protein domains. These include small modules such as the epidermal growth-factor-like (EGF) and short consensus-repeat (SCR) domains, which are both less than 60 amino-acid residues long and frequently occur as multiple repeats. Structural studies (Rao *et al.*, 1995; Downing *et al.*, 1996; Casasnovas *et al.*, 1999) have shown that SCR and EGF modules fold to form all- $\beta$ -strand domains that are strengthened by two and three disulfide bonds, respectively.

The EGF module is widely distributed among extracellular proteins involved in adhesion, receptor–ligand interactions, extracellular matrix structure, determination of cell fate and blood coagulation (Campbell & Bork, 1993). A subset of EGF domains contains a Ca<sup>2+</sup>-binding (cb) consensus sequence (Rees *et al.*, 1988; Handford *et al.*, 1991; Mayhew *et al.*, 1992). The Ca<sup>2+</sup> ion is thought to be important for the orientation of cbEGF pairs by restricting the conformational flexibility of interdomain linkages, as shown for fibrillin-1 (Werner *et al.*, 2000). The solution structures of pairs of cbEGF domains from fibrillin-1 show the two domains in an extended rod-like conformation (Downing *et al.*, 1996; Smallridge *et al.*, 2003). Meanwhile, the crystal structure of three contiguous EGF domains, 4–6, from thrombomodulin (in complex with thrombin) shows a Y-shaped fragment. Thrombomodulin EGF domains 4 and 6 fold similarly to other EGF-like domains with the usual (1–3, 2–4, 5–6) disulfide pairing, while the central 5 domain has an unusual (1–2, 3–4, 5–6) disulfide

pairing and diverges strongly from archetypal EGF structures.

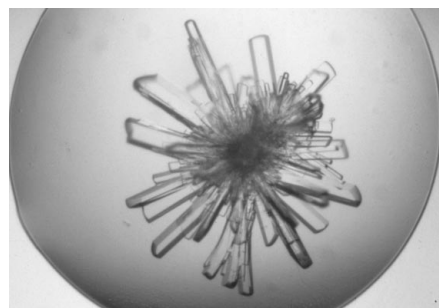
SCR domains are also known as complement protein (CP) modules; they are frequently found as components of complement-system proteins (Wiles *et al.*, 1997). Indeed, many of the complement regulatory proteins (*e.g.* CD46 and CD55) and complement receptors (*e.g.* CD21 and CD35) consist entirely of repeated SCR domains. However, SCR domains have been identified in over 50 distinct mammalian proteins. Multiple SCR domains occur in tandem, separated by three to five amino acids, and two to four contiguous SCR modules are believed to be required to form a ligand-binding site. The solution structures of individual (Norman *et al.*, 1991; Barlow *et al.*, 1992) and pairs (Barlow *et al.*, 1993; Wiles *et al.*, 1997) of SCR modules exist as well as the crystal structure of the four SCR domains of CD55 (Lukacik *et al.*, 2004), Vaccinia virus complement-control protein (Murthy *et al.*, 2001) and  $\beta$ 2-glycoprotein I (Schwarzenbacher *et al.*, 1999). These reveal that consecutive SCR modules form elongated structures.

EMR2 and CD97 are members of the EGF-TM7 family of proteins with between three and five consecutive EGF modules linked to a seven-span transmembrane class B G-protein-coupled receptor (GPCR) *via* a mucin-like spacer domain (Hamann *et al.*, 1995; Lin *et al.*, 2000). The EGF domains of EMR2 and CD97 are highly homologous, with 95–100% sequence identity in corresponding EGF domains. Of the 236 amino-acid residues in the five domain isoforms, only two residues in domain 1, one residue in domain 2 and three residues in domain 3 differ between EMR2 and CD97. Domains 4 and 5 are 100% identical.

Significant sequence homology between EMR2 and CD97 also extends to the mucin-like stalk region (46%), the 7TM region (45%) and multiple potential N- and O-glycosylation sites within the extracellular domain. EGF domains 2–5 in both EMR2 and CD97 are of the Ca<sup>2+</sup>-binding type.

Both EMR2 and CD97 molecules are highly expressed in immune tissues (Lin *et al.*, 2000). EMR2 is restricted to monocytes/macrophages and granulocytes, while CD97 is ubiquitously expressed in most cell types. Notably, CD97 is rapidly up-regulated in activated T and B cells. Their function is largely unknown. Based on unique hybrid structures and restricted expression patterns, it is thought that they may play a role in the immune system by interacting with either cell-surface or extracellular matrix proteins. CD55 (or decay-accelerating factor, DAF) has been shown to be a biological ligand of CD97 (Hamann *et al.*, 1996), demonstrating that these molecules do have the potential to act as cell-adhesion molecules. They are also likely to have a signalling role, given that class B GPCRs are known to signal *via* G-protein activation in response to peptide-ligand binding. However, G-protein-mediated signalling by EGF-TM7s still awaits functional demonstration.

The biological ligand of CD97, CD55, consists of four N-terminal SCR domains that are attached to many mammalian cell types *via* a heavily O-glycosylated spacer that is linked to a C-terminal glycosylphosphatidylinositol (GPI) anchor (Lublin & Atkinson, 1989). It is important for complement regulation (Brodbeck *et al.*, 2000) and cell signalling (Hamann *et al.*, 1996). CD97–CD55 binding is mediated exclusively by protein–protein interactions of the EGF and SCR modules located at the N-terminus of each protein (Lin *et al.*, 2001). The interaction is Ca<sup>2+</sup>-dependent and the CD97 splice variant with the highest affinity comprises three EGF domains: 1, 2 and 5



**Figure 1**  
Ba<sup>2+</sup> crystals of EGF domains 1, 2 and 5 from EMR2 (300 µm in the longest direction).

**Table 1**  
Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell.

Data set	EMR2 native (Ca <sup>2+</sup> )	EMR2 Ba <sup>2+</sup>	EMR2 Sr <sup>2+</sup>
ESRF beamline	ID14-2 ESRF	BM14 ESRF	ID14-2 ESRF
Wavelength (Å)	0.9340	1.3776	0.9340
Space group	P2 <sub>1</sub>	P2 <sub>1</sub>	P2 <sub>1</sub>
Unit-cell parameters (Å, °)	<i>a</i> = 35.6, <i>b</i> = 61.5, <i>c</i> = 81.0, β = 84.2	<i>a</i> = 35.1, <i>b</i> = 61.4, <i>c</i> = 90.9, β = 90.0	<i>a</i> = 35.4, <i>b</i> = 61.6, <i>c</i> = 91.2, β = 90.2
<i>d</i> <sub>max</sub> – <i>d</i> <sub>min</sub> (Å)	29.6–2.6 (2.7–2.6)	45.6–2.6 (2.7–2.6)	36.8–2.5 (2.7–2.5)
No. unique reflections	10840	12551	13148
Redundancy	3.6 (3.6)	7.1 (5.2)	2.5 (2.5)
Completeness (%)	100 (100)	100 (100)	96 (96)
Anomalous completeness (%)	N/A	98 (90)	N/A
<i>R</i> <sub>merge</sub> † (%)	10 (30)	4 (11)	11 (33)
<i>R</i> <sub>anom</sub> ‡ (%)	N/A	4 (8)	N/A
<i>I</i> /σ( <i>I</i> )	4.0 (2.4)	12.5 (5.8)	3.6 (2.1)

† *R*<sub>merge</sub> = ∑|*I*<sub>*j*</sub> – ⟨*I*⟩|/∑*I*<sub>*j*</sub>, where *I*<sub>*j*</sub> is the intensity of an individual reflection and ⟨*I*⟩ is the average intensity of that reflection. ‡ *R*<sub>anom</sub> is defined in the same way to *R*<sub>sym</sub>, except the summation is over anomalous pairs.

(143 residues in total, compared with the 236 residues of the five-domain isoform). EMR2 also binds CD55, but with a *K*<sub>d</sub> at least one order of magnitude weaker than that of CD97. This is despite the fact that the two proteins differ by only three amino acids within the three EGF domains. The precise role of the CD97–CD55 interaction is still unknown. It may possibly be involved in adhesion and signalling within the inflammatory and immune responses (Gray *et al.*, 1996); for example, both CD97 and CD55 are present in arthritic joints (Hamann *et al.*, 1999). Other studies have suggested that CD97–CD55 complex may play a role in cell migration and co-stimulation during multiple sclerosis (MS; Visser *et al.*, 2002).

In the present study, the crystallization and preliminary crystallographic analysis of three extracellular EGF domains (EGF-1,2,5) of EMR2 are reported. The function of EGF-containing molecules is dependent on the relative orientations of the multiple domains and, to date, most structural information is limited to pairs. This will be the first structure of three consecutive archetypal EGF domains (unlike thrombomodulin) and therefore will prove particularly useful in increasing understanding of the structure–function relationship. With the structure of CD55 (Lukacik *et al.*, 2004), it could also be used to model the binding interaction between EMR2 and CD55, thus furthering understanding of why EGF and SCR modules are so well suited for protein interactions, which is currently hindered by the complete lack of structures of EGF–SCR complexes.

## 2. Protein expression and purification

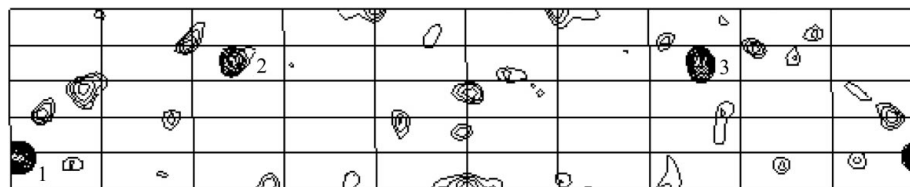
The three extracellular EGF domains (EGF-1,2,5; isoform E; residues 25–167) of the

CD55-binding splice variant of EMR2 (15 579 Da) were expressed in *Escherichia coli* using a His<sub>6</sub>-tag-based inducible expression system as previously described (Lin *et al.*, 2001). Following Ni<sup>2+</sup> affinity purification under denaturing conditions, peptides were reduced, purified and refolded according to a well established *in vitro* refolding protocol (Handford *et al.*, 1991). After purification, EMR2 was lyophilized, reconstituted into water and concentrated to 68.6 mg ml<sup>−1</sup>. The concentration was confirmed using the calculated extinction coefficient at 280 nm (21 000 M<sup>−1</sup> cm<sup>−1</sup> computed from the amino-acid composition using the EXPASY server; Appel *et al.*, 1994).

## 3. Crystallization

Crystallization was performed using the sitting-drop vapour-diffusion method at 298 K in 16-well plates. Drops were prepared by mixing 1 µl of protein solution with an equal volume of crystallization solution and were equilibrated against 300 µl crystallization solution. Conditions were screened using Crystal Screen reagent kits (Hampton Research). Initial leads were followed and rod-shaped native EMR2 crystals were obtained using 0.1–0.175 M calcium acetate, 12–16% (w/v) PEG 8000, 0.1 M sodium cacodylate buffer pH 6.5.

The inclusion of Ca<sup>2+</sup> in the crystallization conditions for EMR2 suggested co-crystallization as a possibility for heavy-atom derivatization. Lanthanide metals are recommended Ca<sup>2+</sup> replacements and a variety were tested, along with several other divalent cations. EMR2 co-crystals were successfully grown in the presence of both 0.1–0.125 M barium chloride (Fig. 1) and 0.1 M strontium chloride.



**Figure 2**  
 $\nu = 0.5$  Harker section of the anomalous Patterson computed between 10 and 2.6 Å using the Ba<sup>2+</sup> crystal data. The map is contoured from 1.5 $\sigma$  in 0.5 $\sigma$  intervals. Main peaks, labelled 1 (6.6 $\sigma$ ), 2 (4.7 $\sigma$ ) and 3 (5.3 $\sigma$ ), are the self-peaks corresponding to the three Ba atoms.

## 4. X-ray diffraction experiments and discussion

X-ray diffraction data have been collected from EMR2 crystals at the European Synchrotron Radiation Facility (ESRF) in Grenoble. Data sets were collected from native Ca<sup>2+</sup> crystals at beamline ID14-2, Ba<sup>2+</sup> crystals at beamline BM14 and Sr<sup>2+</sup> crystals at beamline ID14-2. The crystals were mounted in a fibre loop and transferred briefly to cyroprotectant mother liquor containing 30% ethylene glycol. The crystals were then flash-cooled in a stream of nitrogen at 100 K (Oxford Cryosystems Cryostream, Eynsham, England). All data collected were autoindexed and integrated using the program *MOSFLM* (Collaborative Computational Project, Number 4, 1994) and scaled and merged using the program *SCALA* (Collaborative Computational Project, Number 4, 1994). The Ca<sup>2+</sup> crystals are monoclinic, belong to space group  $P2_1$  and diffract to 2.6 Å. The unit-cell parameters are  $a = 35.6$ ,  $b = 61.5$ ,  $c = 81.0$  Å,  $\beta = 84.2^\circ$ . The calculated packing parameter,  $V_M$  (Matthews, 1968), is 2.8 Å<sup>3</sup> Da<sup>-1</sup>, assuming two molecules in the asymmetric unit. This value corresponds to a solvent content of 56%. The Ba<sup>2+</sup> and Sr<sup>2+</sup> crystals also belong to space group  $P2_1$  and have very similar parameters to the Ca<sup>2+</sup> crystals. Full details of data-collection and processing statistics for all crystals are presented in Table 1.

An anomalous difference Patterson calculated at Harker section  $\nu = 0.5$  between 10 and 2.6 Å from the Ba<sup>2+</sup> crystal data (Fig. 2) displays three strong peaks,

suggesting the presence of three Ba atoms per asymmetric unit. This is one less than might be expected, given the fact that there are two Ca<sup>2+</sup>-binding sites and two copies of the molecular per asymmetric unit. Attempts to phase using this Ba<sup>2+</sup> solution are now under way. We then hope to solve the Ca<sup>2+</sup> and Sr<sup>2+</sup> structures *via* molecular replacement.

RJMA and PL are funded by MRC studentships, PR by BBSRC grant No. 43/B16601 (to SML) and VK by MRC grant No. G000164 (to PAH).

## References

- Appel, R. D., Bairoch, A. & Hochstrasser, D. F. (1994). *Trends Biochem. Sci.* **19**, 258–260.
- Barlow, P. N., Norman, D. G., Steinkasserer, A., Horne, T. J., Pearce, J., Driscoll, P. C., Sim, R. B. & Campbell, I. D. (1992). *Biochemistry*, **31**, 3626–3634.
- Barlow, P. N., Steinkasserer, A., Norman, D. G., Kieffer, B., Wiles, A. P., Sim, R. B. & Campbell, I. D. (1993). *J. Mol. Biol.* **232**, 268–284.
- Brodbeck, W. G., Mold, C., Atkinson, J. P. & Medof, M. E. (2000). *J. Immunol.* **165**, 3999–4006.
- Campbell, I. D. & Bork, P. (1993). *Curr. Opin. Struct. Biol.* **3**, 385–392.
- Casasnovas, J. M., Larvie, M. & Stehle, T. (1999). *EMBO J.* **18**, 2911–2922.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst. D* **50**, 760–763.
- Downing, A. K., Knott, V., Werner, J. M., Cardy, C. M., Campbell, I. D. & Handford, P. A. (1996). *Cell*, **85**, 597–605.
- Gray, J. X., Haino, M., Roth, M. J., Maguire, J. E., Jensen, P. N., Yarme, A., Stetler-Stevenson, M. A., Siebenlist, U. & Kelly, K. (1996). *J. Immunol.* **157**, 5438–5447.

- Hamann, J., Eichler, W., Hamann, D., Kerstens, H. M., Poddighe, P. J., Hoovers, J. M., Hartmann, E., Strauss, M. & van Lier, R. A. (1995). *J. Immunol.* **155**, 1942–1950.
- Hamann, J., Vogel, B., van Schijndel, G. M. & van Lier, R. A. (1996). *J. Exp. Med.* **184**, 1185–1189.
- Hamann, J., Wishaupt, J. O., van Lier, R. A., Smeets, T. J., Breedveld, F. C. & Tak, P. P. (1999). *Arthritis Rheum.* **42**, 65065–65068.
- Handford, P. A., Mayhew, M., Baron, M., Winship, P. R., Campbell, I. D. & Brownlee, G. G. (1991). *Nature (London)*, **351**, 164–167.
- Lin, H. H., Stacey, M., Hamann, J., Gordon, S. & McKnight, A. J. (2000). *Genomics*, **67**, 188–200.
- Lin, H. H., Stacey, M., Saxby, C., Knott, V., Chaudhry, Y., Evans, D., Gordon, S., McKnight, A. J., Handford, P. & Lea, S. (2001). *J. Biol. Chem.* **276**, 24160–24169.
- Lublin, D. M. & Atkinson, J. P. (1989). *Annu. Rev. Immunol.* **7**, 35–58.
- Lukacik, P., Roversi, P., White, J., Esser, D., Smith, G. P., Billington, J., Williams, P. A., Rudd, P. M., Wormald, M. R., Harvey, D. J., Crispin, M. D., Radcliffe, C. M., Dwek, R. A., Evans, D. J., Morgan, B. P., Smith, R. A. & Lea, S. M. (2004). *Proc. Natl. Acad. Sci. USA*, **101**, 1279–1284.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Mayhew, M., Handford, P., Baron, M., Tse, A. G., Campbell, I. D. & Brownlee, G. G. (1992). *Protein Eng.* **5**, 489–494.
- Murthy, K. H., Smith, S. A., Ganesh, V. K., Judge, K. W., Mullin, N., Barlow, P. N., Ogata, C. M. & Kotwal, G. J. (2001). *Cell*, **104**, 301–311.
- Norman, D. G., Barlow, P. N., Baron, M., Day, A. J., Sim, R. B. & Campbell, I. D. (1991). *J. Mol. Biol.* **219**, 717–725.
- Rao, Z., Handford, P., Mayhew, M., Knott, V., Brownlee, G. G. & Stuart, D. (1995). *Cell*, **82**, 131–141.
- Rees, D. J., Jones, I. M., Handford, P. A., Walter, S. J., Esnouf, M. P., Smith, K. J. & Brownlee, G. G. (1988). *EMBO J.* **7**, 2053–2061.
- Schwarzenbacher, R., Zeth, K., Diederichs, K., Gries, A., Kostner, G. M., Laggner, P. & Prassl, R. (1999). *EMBO J.* **18**, 6228–6239.
- Smallridge, R. S., Whiteman, P., Werner, J. M., Campbell, I. D., Handford, P. A. & Downing, A. K. (2003). *J. Biol. Chem.* **278**, 12199–12206.
- Visser, L., de Vos, A. F., Hamann, J., Melief, M. J., van Meurs, M., van Lier, R. A., Laman, J. D. & Hintzen, R. Q. (2002). *J. Neuroimmunol.* **132**, 156–163.
- Werner, J. M., Knott, V., Handford, P. A., Campbell, I. D. & Downing, A. K. (2000). *J. Mol. Biol.* **296**, 1065–1078.
- Wiles, A. P., Shaw, G., Bright, J., Perczel, A., Campbell, I. D. & Barlow, P. N. (1997). *J. Mol. Biol.* **272**, 253–265.