Structure determination of human semaphorin 4D as an example of the use of MAD in non-optimal cases

Robert M. Esnouf,*
Christopher A. Love, Karl Harlos,
David I. Stuart and E. Yvonne
Jones

Division of Structural Biology, University of Oxford, Henry Wellcome Building for Genomic Medicine, Roosevelt Drive, Oxford OX3 7BN, England

Correspondence e-mail: robert@strubi.ox.ac.uk

Semaphorins are an important class of signalling molecules involved in axon guidance, immune function and angiogenesis. They are characterized by having an extracellular sema domain of about 500 residues. The steps involved in the determination of the structure of human semaphorin 4D are described here as a case study of selenium MAD phasing in a difficult case with low symmetry, moderate diffraction and low selenium content. A particular feature of this study was the large number of diffraction images required to give data of sufficient quality for structure determination and these data are re-analyzed here to investigate the effects of radiation damage on eventual data quality and to suggest strategies for successful MAD phasing in similar difficult cases.

1. Introduction

The semaphorins form a large and widespread class of signalling molecules characterized by an extracellular sema domain (Kolodkin *et al.*, 1993). The class is further subdivided into eight subclasses according to organism, domain content and membrane linkage (Semaphorin Nomenclature Committee, 1999). Other signalling molecules, notably plexins and the receptor tyrosine kinases MET and RON, also contain a sema domain (Winberg *et al.*, 1998) and taken together these molecules form the semaphorin superfamily (reviewed in Gherardi *et al.*, 2004). Semaphorins are involved in the regulation of several processes such as axon guidance, immune function and angiogenesis. Signalling is achieved by the formation of sema domain-mediated complexes which typically involve the other members of the semaphorin superfamily, plexins. The semaphorins most extensively characterized functionally are those of class 3, which are secreted vertebrate proteins and include an Ig-like domain and a basic motif along with the sema domain. They function by sema domain-mediated binding to members of the plexin family (reviewed in Raper, 2000). The class 4 semaphorins are also vertebrate proteins and bind to plexins, but in addition to the sema domain they contain an Ig-like domain, a transmembrane region and a cytoplasmic region (Tamagnone *et al.*, 1999).

The sema domain comprises approximately 500 residues and was initially predicted to form a single structural domain (Kolodkin *et al.*, 1993), although it was observed that the C-terminal 54 residues of the domain share some sequence similarity with a similar region in integrins. This C-terminal region was subsequently suggested to form the separate PSI (plexin, semaphorin and integrin) domain (Bork *et al.*, 2001), also referred to as a cysteine-rich domain (CRD; reviewed in

Gherardi *et al.*, 2004). Although the structure of the integrin $\alpha V \beta 3$ had previously been determined, disorder in the crystal structure obscured the PSI-domain fold (Xiong *et al.*, 2001). Thus, no structural information was available to provide a framework for understanding the functioning of any sema-domain-containing protein.

Semaphorin 4D (SEMA4D; also known as CD100) is a 150 kDa glycoprotein that is expressed in lymphocytes, brain, kidney and heart (Hall *et al.*, 1996) and has been shown to function in B-cell activation, T-cell priming (Shi *et al.*, 2000) and in axon guidance with plexin-B1 as its high-affinity receptor (Tamagnone *et al.*, 1999). It forms a homodimer on the cell surface, partly stabilized by a disulfide linkage. Several constructs of SEMA4D were studied in order to find a soluble SEMA4D (sSEMA4D) construct that was homodimeric, would bind to plexin-B1 and was suitable for structure determination. The selected construct comprised residues 1–657 and formed a tight homodimer despite the truncation having removed Cys687, the residue responsible for the disulfide linkage immediately prior to the start of the trans-membrane region. This construct eventually led to a structure determination at 2.0 Å resolution revealing a basic seven-bladed β-propeller topology. The structure is decorated with several significant insertions and the CRD nestles tightly against the side of the propeller. Details of this structure and its implications for the understanding of signalling by sema-domain-containing proteins have been discussed elsewhere (Love *et al.*, 2003; Gherardi *et al.*, 2004). In this paper, attention is focused on the steps involved in the structure deter-mination itself and particularly in the initial experimental phasing using the technique of multiwavelength anomalous dispersion (MAD). Difficulties were associated with the need to express the selenomethionated protein eukaryotically, the glycosylation of the protein, the relatively small number of methionine residues (seven in 657 residues), the low symmetry and pleomorphism of the crystals and the relatively poor quality of the diffraction obtained from the vast majority of crystals.

2. Experimental section

2.1. Cloning, expression and purification

The production of the sSEMA4D construct has previously been described in detail (Love *et al.*, 2003). Briefly, a fragment of the gene encoding the selected construct of SEMA4D was amplified, the product (including a C-terminal KHHHHHH purification tag) subcloned into the glutamine synthase-encoding pEE14 expression vector and transfected into Lec3.2.8.1 Chinese hamster ovary cells. Native protein expression levels of 3–4 mg l^{-1} were obtained and the protein was purified using Ni–NTA agarose affinity chromatography followed by gel filtration. Selenomethionine (SeMet) labelling was carried out by growing cell lines expressing sSEMA4D to confluence in roller bottles before removing the medium, rinsing the monolayer and then adding methionine-free Dulbecco's modification of Eagle's medium (DMEM) containing 30 mg l^{-1} selenomethionine, 2 mM sodium buty-
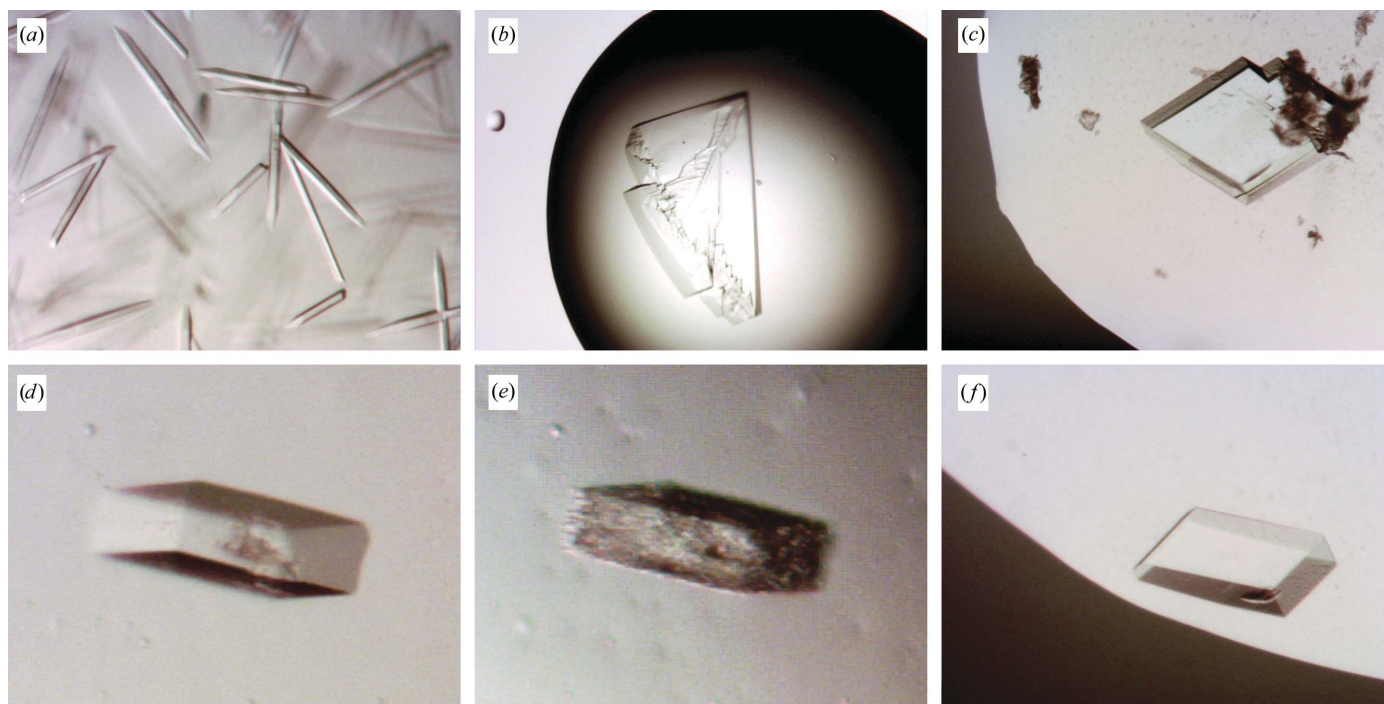

Figure 1
Crystals of sSEMA4D. (*a*) Hexagonal needles obtained with 0.1 M MES pH 6.0, 1 M LiCl and 20%(*w/v*) PEG 6000. (*b*) Thin rectangular plates obtained with 0.1 M HEPES pH 7.0, 1 M LiCl and 20%(*w/v*) PEG 6000. (*c*) Thicker rectangular tiles obtained with 0.1 M Tris–HCl pH 8.0, 1 M LiCl and 20%(*w/v*) PEG 6000. (*d*) A crystal obtained from 0.2 M Na$_2$SO$_4$ and 20%(*w/v*) PEG 3350 after 3 d. (*e*) The same crystal after 5 d. (*f*) Crystal of SeMet sSEMA4D obtained in 0.2 M NH$_4$F and 20%(*w/v*) PEG 3350 by seeding with native sSEMA4D microcrystals obtained under the same conditions.

rate and 5%(*v/v*) foetal calf serum. After purification in the same manner as before, a yield of ~1.6 mg l^{-1} was obtained.

2.2. Crystallization

Purified native sSEMA4D was concentrated to 10 mg ml^{-1} in 0.1 *M* Tris–HCl pH 8.0 and 0.1 *M* NaCl. Initial crystallization screening using Hampton sparse-matrix crystallization kits was by the vapour-diffusion method with sitting drops (1 + 1 µl) on microbridges and yielded crystals of several morphologies (Fig. 1). The hexagonal needles (Fig. 1*a*) did not diffract X-rays beyond 20 Å resolution at synchrotron sources. The plates and tiles (Figs. 1*b* and 1*c*) showed much better diffraction with a high-resolution limit of 3–3.5 Å, but had to be produced fresh for each data-collection trip owing to their instability (Figs. 1*d* and 1*e*). Crystals were only obtained with the fully glycosylated protein containing the C-terminal His tag. Attempts to grow crystals after having removed either the sugars (using EndoH) or the tag (using carboxypeptidase A), or indeed both, failed to produce any crystals despite numerous trials. Crystals of SeMet sSEMA4D did not grow as readily as those of native sSEMA4D, but could be induced by microseeding with native crystals (Fig. 1*f*).

2.3. Data collection

All data collection was carried out at the European Synchrotron Radiation Facility (ESRF), Grenoble, France either on the public JSBG beamlines or on BM14, the UK CRG beamline dedicated primarily to MAD structure determination. Crystals were harvested from the crystallization trials at the beamline and quickly passed through a well containing the cryoprotectant perfluoropolyether XR-75 (Interchim) prior to flash-cooling to 100 K to reduce the effects of radiation damage. About one in 15 crystals yielded useful, if anisotropic, diffraction (Fig. 2) and the unit-cell parameters were found to be very variable. Several data sets were used for the eventual structure determination (Table 1). Data set NAT was collected from a native sSEMA4D crystal at ID14-EH1 using an ADSC Quantum 210 CCD detector. The data that were merged to form PK-COMB as well as the PK1, REM and PK2 data sets were collected from SeMet sSEMA4D
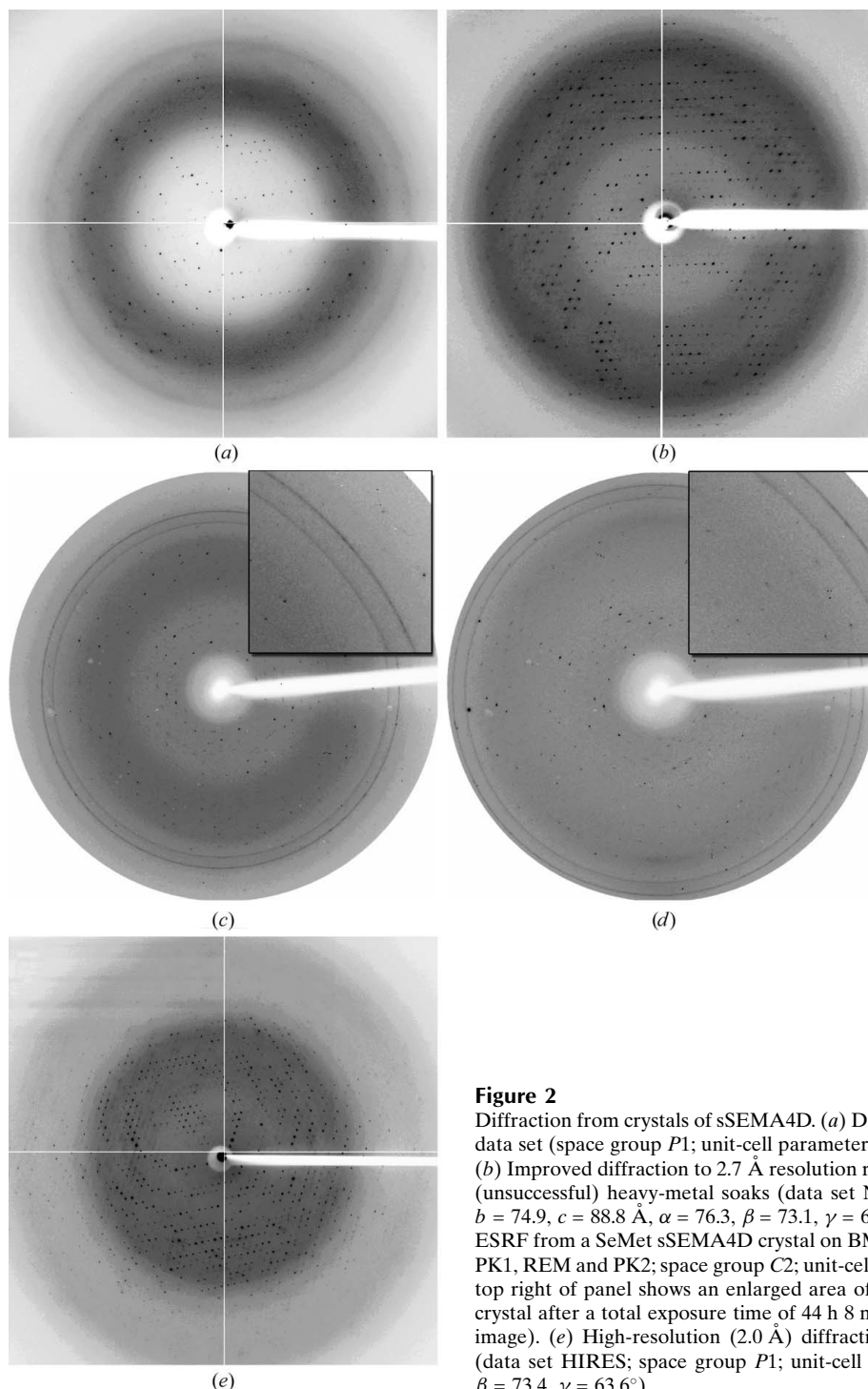

Figure 2
Diffraction from crystals of sSEMA4D. (*a*) Diffraction to 3.5 Å resolution yielding an initial native data set (space group *P*1; unit-cell parameters $a \simeq 79$, $b \simeq 79$, $c \simeq 91$ Å, $\alpha \simeq 85$, $\beta \simeq 76$, $\gamma \simeq 61°$). (*b*) Improved diffraction to 2.7 Å resolution resulting from unexpected dehydration resulting from (unsuccessful) heavy-metal soaks (data set NAT; space group *P*1; unit-cell parameters $a = 73.3$, $b = 74.9$, $c = 88.8$ Å, $\alpha = 76.3$, $\beta = 73.1$, $\gamma = 63.9°$). (*c*) Image number 1 obtained on BM14 at the ESRF from a SeMet sSEMA4D crystal on BM14 showing diffraction to 3.0 Å resolution (data sets PK1, REM and PK2; space group *C*2; unit-cell parameters $a = 71.5$, $b = 127.1$, $c = 87.2$ Å, $\beta = 103.7°$; top right of panel shows an enlarged area of the image). (*d*) Image number 2648 from the same crystal after a total exposure time of 44 h 8 min (top right of panel shows an enlarged area of the image). (*e*) High-resolution (2.0 Å) diffraction from a dehydrated crystal of native sSEMA4D (data set HIRES; space group *P*1; unit-cell parameters $a = 73.3$, $b = 76.8$, $c = 89.4$ Å, $\alpha = 77.4$, $\beta = 73.4$, $\gamma = 63.6°$).

Table 1
Crystallographic data sets obtained from native sSEMA4D (NAT) and SeMet sSEMA4D (all other data sets).

PK-COMB is a data set obtained by combining data from three different crystals having similar unit-cell parameters. The data sets PK1, REM and PK2 were all collected from a single crystal. Values in parentheses are for the outer (highest resolution) data shells *i.e.* 2.80–2.70 Å for NAT, 3.21–3.10 Å for PK-COMB, 3.15–3.04 Å for PK1, 3.31–3.20 Å for REM, 3.52–3.40 Å for PK2 and 2.07–2.00 Å for HIRES.

	NAT	PK-COMB	PK1	REM	PK2	HIRES
Beamline	ID14-EH1	BM14	BM14	BM14	BM14	ID29
Resolution limit (Å)	2.7	3.1	3.04	3.2	3.4	2.0
Space group	*P*1	*P*1	*C*2	*C*2	*C*2	*P*1
Wavelength (Å)	0.934	0.979	0.979	0.886	0.979	0.975
Observations	69114	271279	527379	360652	443125	375205
Unique reflections	41743	28435	28390	24342	19469	108424
Completeness (%)	89.4 (63.7)	100 (100)	99.5 (94.7)	100 (99.9)	96.1 (60.8)	97.2 (90.9)
$I/\sigma(I)$	9.5 (1.3)	13.6 (3.6)	21.4 (5.1)	20.7 (7.6)	26.5 (6.2)	17.6 (1.8)
R_{merge}†	6.7	20.4	10.6	11.1	10.5	9.9

† $R_{merge} = \sum |I - \langle I \rangle| / \sum \langle I \rangle$.

crystals at BM14 using a 133 mm MAR CCD detector. Finally, the HIRES data set was collected at ID29 using an ADSC Quantum 210 CCD detector using a SeMet sSEMA4D crystal that had been dehydrated by increasing the PEG 3350 concentration in the reservoir from 20 to 50%(*w/v*) over a period of 3 d prior to harvesting. All data reduction was carried out using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997).

2.4. Structure determination

Data from three crystals were merged to form a combined anomalous data set for data collected at the peak of the selenium f'' curve, PK-COMB. *Shake-and-Bake* (Weeks & Miller, 1999) was used to search for selenium sites, with *CNS* (Brünger *et al.*, 1998) used to identify further sites (Fig. 3*a*). Attempts at phase improvement in *GAP* (J. Grimes and DIS, unpublished program) used PK-COMB both alone and in a cross-averaging protocol with the NAT data set. These procedures yielded improved maps that showed a dimer envelope and suggested a largely β-structure, but they remained uninterpretable (Fig. 3*b*).

Data for data sets PK1, REM and PK2 were scaled using the 'no merge original index' option in *SCALEPACK* to allow final scaling and merging with *XPREP* (Bruker AXS). *SHELXD* (Schneider & Sheldrick, 2002) was used to identify selenium sites from the PK1 data and the initial maps were phased using *SHELXE* (Schneider & Sheldrick, 2002) followed by phase improvement in *RESOLVE* (Terwilliger, 2003*a,b*). At this point, the basic β-propeller structure of the sema domain was apparent (Fig. 3*c*). Cycles of manual rebuilding using *O* (Jones *et al.*, 1991), refinement using *X-PLOR* (Brünger, 1992) and density modification using *DMMULTI* (Cowtan & Zhang, 1999) were leading toward a refined structure. The collection of the HIRES data set allowed a more direct strategy: an initial molecular-replacement solution was obtained with *AMoRe* (Navaza, 2001) using the model in its current state of refinement and was followed by automatic tracing of the structure using *ARP/wARP* (Morris *et al.*, 2002). Final refinement used *X-PLOR* and *CNS* to model glycosylated residues. Atomic coordinates

and structure factors for the high-resolution data have been deposited in the Protein Data Bank with accession code 1olz (Fig. 3*d*).

3. Results and discussion

3.1. Initial attempts at structure solution

Although crystals of sSEMA4D could be grown readily (Fig. 1), collection of diffraction data from them proved consistently difficult. Not only did the crystals need to be fresh (Figs. 1*d* and 1*e*), but even then only about one in 15 crystals yielded usable diffraction. This diffraction tended to be rather anisotropic and only extended to a limit of ~3.5 Å resolution (Fig. 2*a*). The crystals belonged to space group *P*1 with a dimer in the asymmetric unit, implying a solvent (plus sugar) content of about 63%. The non-crystallographic twofold conferred pseudo-*C*2 symmetry on the crystals. A search for heavy-metal derivatives showed that the crystals had potential for substantial pleomorphism induced by the soaking experiments, but none of these experiments led to a useful derivative. However, one soak resulted in a crystal with a reduced solvent (plus sugar) content of 56% that showed significantly improved diffraction (Fig. 2*b*), which was treated as the reference native data set (NAT; Table 1).

In the absence of either suitable molecular-replacement search models or heavy-atom phasing information, it was decided to express SeMet-labelled sSEMA4D eukaryotically and to attempt to use MAD phasing. The construct contained seven methionine residues (including the N-terminal one, which is most likely disordered) in 657 residues and eukaryotic expression was likely to lead to somewhat incomplete SeMet incorporation (technical difficulties prevented the degree of incorporation from being assessed by mass spectrometry). Furthermore, the non-isomorphism of the crystals meant that it would be difficult to merge data between crystals and the low symmetry of the crystals would make it hard to obtain the required accuracy of measurements from a single crystal given the poor diffraction quality and the effects of radiation damage. On the positive side, we expected that non-crystal-

lographic symmetry would allow density-modification proce-dures to produce significant phase improvements.

Crystals of SeMet sSEMA4D were obtained by seeding with native microcrystals (Fig. 1*f*) and tested on both ID14-EH4 and BM14 at the ESRF. On ID14-EH4, short exposures were used with little attenuation of the beam so that the crystals did not last long enough to give data at a second wavelength. Unfortunately, even at the first wavelength, collected at the peak of the selenium edge, the anomalous signal was barely detectable. On BM14, with a much less intense but very stable beam, several peak data sets were collected and for three of these data sets the unit-cell para-meters were sufficiently similar to allow them to be combined to give the data set PK-COMB (Table 1) with a detectable anomalous signal to at least 4 Å resolution; attempts were made to solve the structure by SAD using this data set.

Computational searches for 14 selenium sites (seven in each monomer) eventually produced a solution for the PK-COMB data set in which ten of the sites obeyed the expected twofold non-crystallographic symmetry (a very useful check for likely correctness in difficult cases as well as for filtering potential weak sites). Refinement of these sites and analysis of residual maps added one further pair of weak sites (Fig. 3*a*). Increas-ingly complex phase-improvement procedures were then attempted. Solvent flattening was initially performed alone, followed by combination with real-space electron-density averaging to exploit the twofold non-crystallographic symmetry. As the maps remained uninterpretable, cross-
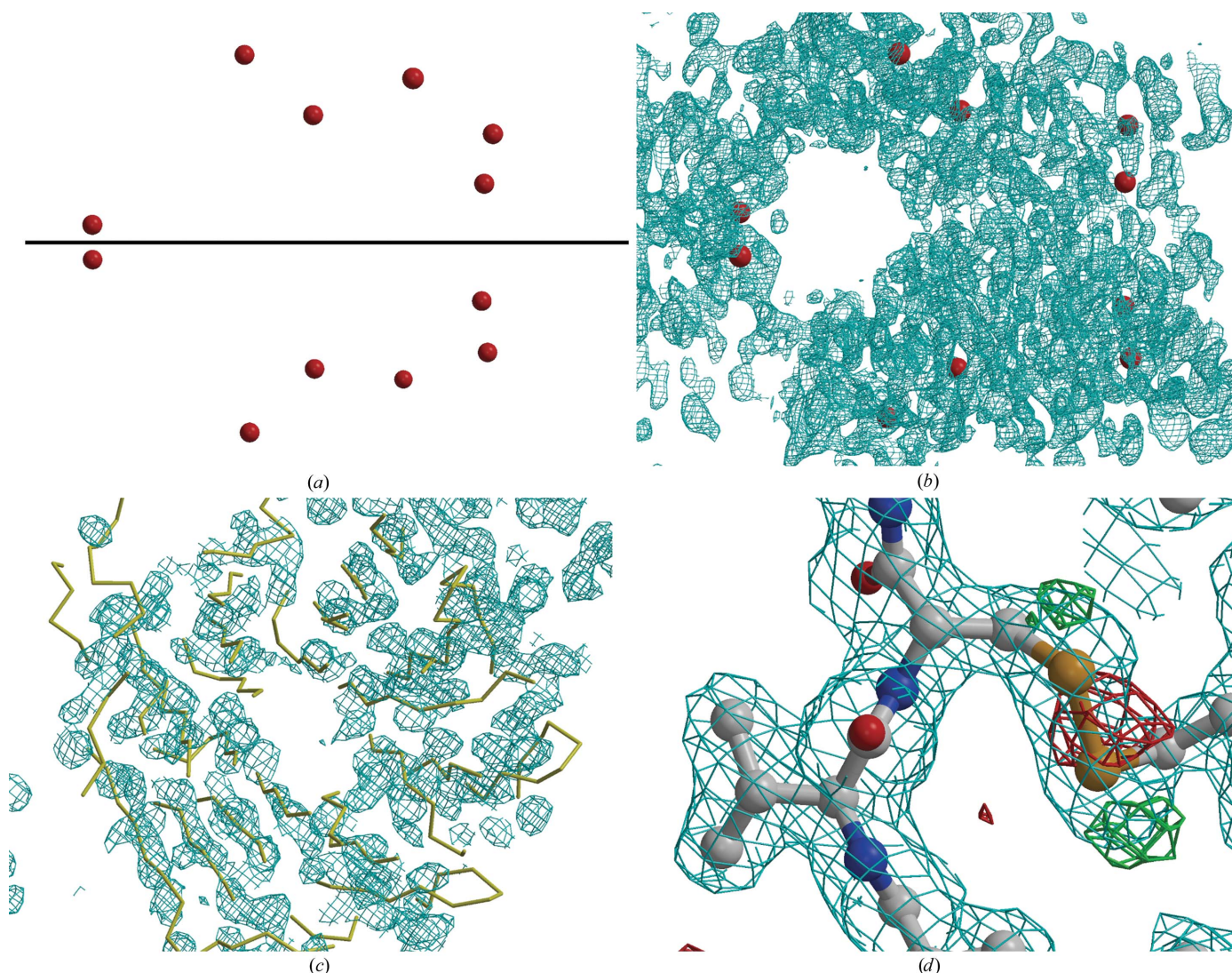

Figure 3
Phasing of sSEMA4D data using MAD. (*a*) Selenium sites obtained from the PK-COMB combined peak data set demonstrating the expected twofold non-crystallographic symmetry (shown by the horizontal line; the pair of sites toward the left-hand edge are the low-occupancy sites). (*b*) Electron-density map (contoured at 1σ) for the phased PK-COMB data set after density modification and cross-crystal form averaging against NAT (spheres mark Se-atom sites). (*c*) Electron-density map (contoured at 1.2σ) obtained from the MAD-phased PK1 data set after phase improvement with *RESOLVE* showing the basic β-propeller structure (yellow lines show a roughly superposed C^α trace for the $\alpha V\beta3$ integrin β-propeller). (*d*) Sample of $2F_{obs} - F_{calc}$ electron density (blue lines, contour level 1σ) and $F_{obs} - F_{calc}$ difference density (red lines, contour level -3σ; green lines, contour level $+3\sigma$) obtained using the HIRES data set and the final refined model showing radiation damage to the disulfide linkage between residues B260 and B305. Figure prepared using *BOBSCRIPT* (Esnouf, 1997, 1999) and rendered with *RASTER3D* (Merritt & Murphy, 1994).

crystal form averaging against the unphased NAT data set (and eventually even including the original 3.5 Å data set) was also attempted to try to exploit the observed non-isomorphism. Defining the procedure for cross-crystal form averaging was made easier by the $P1$ symmetry and very clear twofold non-crystallographic symmetry: between crystal forms there was no need to find a translational operator and the rotational operator was defined simply by the change in orientation between the twofold axes. Having a proper non-crystallographic twofold operator also meant that the electron density could simply be averaged in a dimer envelope without having to define a volume for a monomer. However, beyond revealing a molecular envelope containing a large hole and a structure consisting primarily of β-strands, the maps remained uninterpretable (Fig. 3b) even after these phase-improvement procedures.

3.2. MAD data collection at BM14

Next, freshly prepared SeMet sSEMA4D was crystallized by seeding (see §2), yielding crystals that, although small, were rather thicker (Fig. 1f). Several of these new crystals were exposed on BM14 at the ESRF. Although the fifth crystal showed diffraction similar to that which had led to the PK-COMB data set, it was decided not to pursue data collection. The next crystal to show diffraction of similar quality (Fig. 2c) was the 31st one tested! Remarkably, it also showed a change in space group to $C2$ arising from the non-crystallographic twofold 'clicking' into alignment with the unit-cell edges and a concomitant reduction in solvent (plus sugar) content to about 53%. This was the first crystal to show this effect out of more than 150 examined. Not only did the diffraction extend to slightly higher resolution ($\sim$3.0 Å) than before, but the higher symmetry meant that the redundancy in observations built up more rapidly, thereby improving the chances of performing a MAD experiment on a single crystal.

The anomalous scattering ratio, $\langle|\Delta F|\rangle/\langle F_P\rangle$, is approximately $(f''2^{1/2}n^{1/2})/(Z_{eff}N^{1/2})$, where N is the total number of non-H atoms, Z_{eff} is the effective atomic number of these atoms ($\sim$6.7 for proteins), n is the number of anomalous scatters and f'' is the imaginary part of the anomalous scattering (Hendrickson & Teeter, 1981). For sSEMA4D ($N \simeq 5000$, $n = 5$ ignoring the N-terminal methionine and the weak site observed previously and $f'' = -6.5$), the anomalous
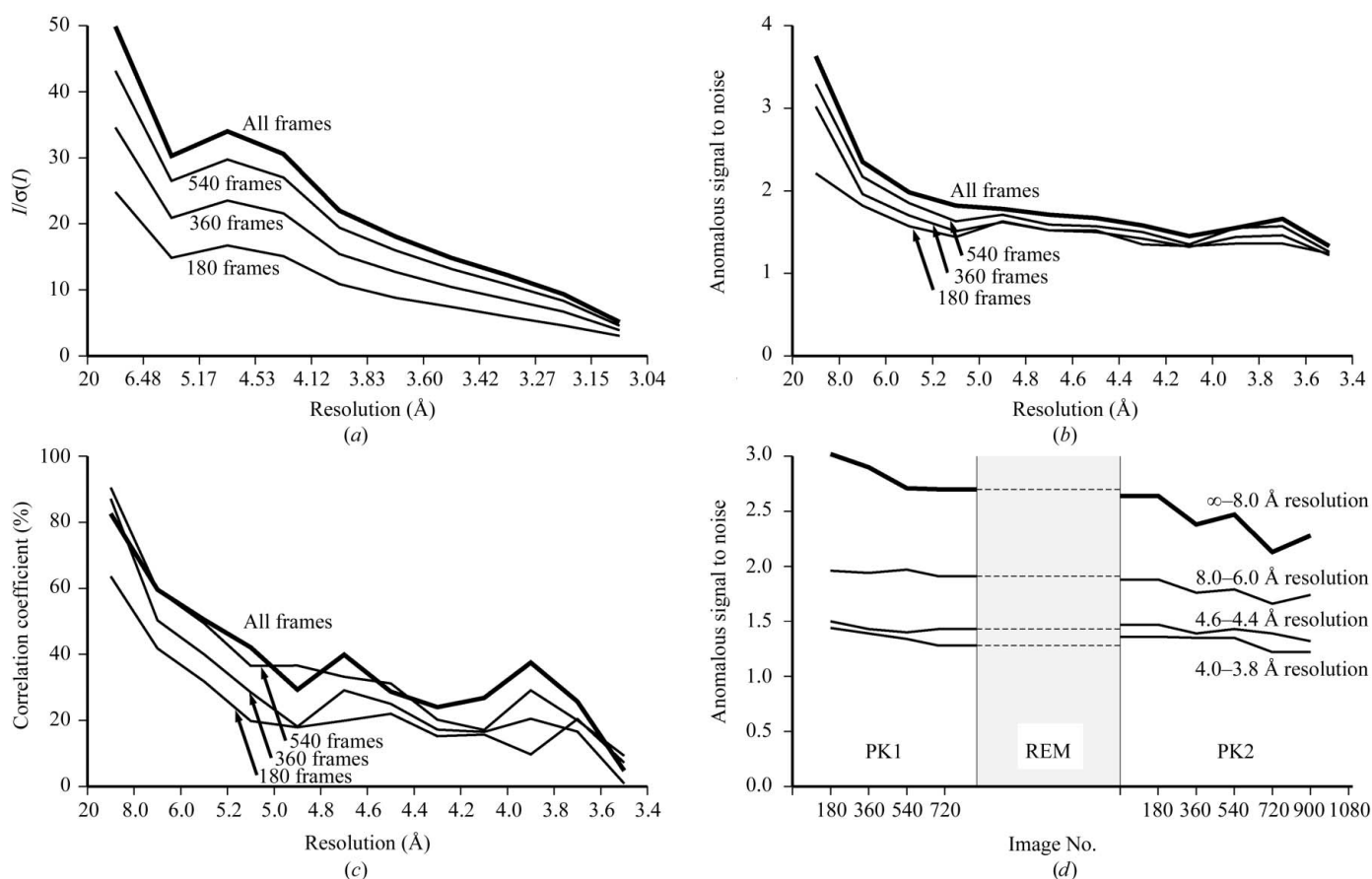

Figure 4
Analysis of the anomalous signal in the data obtained from the $C2$ crystal. (a) $I/\sigma(I)$ in resolution shells for data set PK1 as judged by *SCALEPACK* for increasing numbers of images. (b) The anomalous signal for increasing numbers of images from data set PK1 as judged by the variances of F^+ and F^- reported by *XPREP*. (c) The correlation coefficient in the anomalous signal between data sets PK1 and PK2 for increasing numbers of images in resolution shells as judged by *XPREP*. (d) The anomalous signal for different resolution shells obtained from batches of 360 images (centred on each data point) at different times during data collection as judged by the variances of F^+ and F^- reported by *XPREP* (the gap corresponds to the images collected for REM which are excluded from this analysis; between PK1 and PK2 the crystal-to-detector distance was increased from 200 to 230 mm, giving an overall improvement in signal to noise).

Table 2
MAD data-collection strategy for the *C*2 crystal of SeMet sSEMA4D on BM14.

The ESRF was operating in 16-bunch mode and no attenuation was used. All images were 1° oscillations (the mosaicity of the crystal varied between 0.5° and 0.75°) with 60 s exposures and were collected in batches covering 60° followed by a skip forward of 150°. Thus, the first 30 images of a batch provided the 'inverse-beam' images of the last 30 images of the previous batch, meaning that Friedel opposites were collected close in time (and hence in radiation dose). Data set name refers to the data-processing statistics summarized in Table 1.

	First peak	High-energy remote	Second peak	Totals
Data-set name	PK1	REM	PK2	
No. of images	870	679	1099	2648
Total exposure time	14 h 30 min	11 h 19 min	18 h 19 min	44 h 8 min
Wavelength (Å)	0.979	0.886	0.979	
Crystal-to-detector distance (mm)	200	230	230	
Resolution at edge (Å)	3.0	3.1	3.4	
Observations	527379	360652	443125	1331156

scattering ratio was estimated to be 4.3%, implying that a signal-to-noise ratio of about 23 would be required to solve the structure. With the ESRF running in 16-bunch mode giving reduced intensity and thus requiring longer exposures, a back-of-the-envelope calculation based on processing the first few images showed that in the remaining time allocation it would not be possible to perform a proper experiment, so after discussion with beamline staff the experiment was kindly allowed to run into time allocated for beamline-development studies. The pseudo-inverse-beam data-collection strategy is summarized in Table 2 and led to the data sets PK1, REM and PK2 (Table 1). Data were processed during data collection and the build-up in the anomalous signal was monitored along with the correlation between the anomalous signals in the PK1 and PK2 data sets (Fig. 4). The target was to achieve an anomalous signal-to-noise ratio of about 1.3 at 3.5 Å resolution, as judged by *XPREP* (Bruker AXS), for the peak data sets. Despite the large number of images and an actual exposure time of over 44 h, the crystal was still diffracting to about 3.5 Å resolution even for the last image (Fig. 2*d*).

3.3. Structure solution

To allow for the effects of radiation damage, the PK1 and PK2 data sets were treated as separate wavelengths throughout. Since the *C*2 crystal contained a monomer in the asymmetric unit, only seven selenium sites were searched for (§2.4) and a set of five sites were easily identified which were shown to correspond to the original ten sites found for PK-COMB. Initial phasing of the data and phase improvement (§2.4) produced maps showing a clear solvent boundary and a tight dimer. When viewed from the correct orientation, these maps revealed the basic topology of the sema domain, a seven-bladed *β*-propeller similar to that of integrins (Fig. 3*c*). Based on the integrin structure (Xiong *et al.*, 2001), an initial model for sSEMA4D was built and, owing to the limited resolution of the data, laborious rounds of manual rebuilding and refinement (§2.4) were required. Refinement was progressing stea-

dily with about 500 residues traced and a crystallographic residual (*R* factor) of 31.4% when experiments with crystal dehydration provided a shortcut to the refined structure.

3.4. High-resolution data

To try to exploit the pleomorphism experimentally, crystals grown from freshly prepared sSEMA4D were subjected to a dehydration protocol of increasing the PEG 3350 reservoir concentration (§2.3). We used a very similar protocol in studies on HIV-1 reverse transcriptase to improve the high-resolution diffraction limit from 3.4 to 2.1 Å (Stammers *et al.*, 1994; Esnouf *et al.*, 1998). Remarkably, a similar improvement was obtained again (Fig. 2*e*), allowing a 2.0 Å resolution data set to be collected (HIRES in Table 1) from a crystal of space group *P*1 having very similar unit-cell parameters to the crystal which yielded data set NAT. Surprisingly, this crystal was not as dehydrated as the *C*2 crystal studied at BM14 with an estimated 56% solvent (plus sugar) content. Molecular replacement was used to fit the partially refined model into the new data and automated procedures were used to build a new model (§2.4). Refinement still needed considerable manual intervention, but a final model (without sugars) was rapidly produced for 1242 residues (3–200 and 205–627 in each monomer) with 841 water molecules. This model had a crystallographic *R* factor of 20.6% ($R_{\mathrm{free}} = 27.0\%$) for all data from 20–2.0 Å resolution and a root-mean-square deviation in bond lengths of 0.007 Å and in bond angles of 1.4° (Love *et al.*, 2003; Fig. 3*d*).

4. Conclusions

The difficulties in the structure determination of human semaphorin 4D are unfortunately not uncommon in the analysis of large multidomain proteins and complexes: crystals may be difficult to produce and unstable, domain flexibility may lead to problems of non-isomorphism but also open up possibilities for dehydration protocols and phase improvement, diffraction may be weak and anisotropic and when working with low-symmetry space groups it may be difficult to obtain sufficient data redundancy (for MAD in particular) before the effects of radiation damage become too severe. Whilst there is no single best strategy for coping with these various problems and even exploiting them to advantage, some lessons can be drawn from this case study.

Firstly, it is important to have an idea of what quality of data is required, how it is going to be achieved from the crystals available and how long it will take. Although synchrotron beam time is an expensive resource, erring on the side of minimal data collection can easily be a false economy which delays structure solution and unnecessarily drains the equally valuable time of expert crystallographers. Doing half an experiment will probably yield zero results! In the extreme case reported here, the *C*2 crystal spent more than 2.5 d on the beamline and success was largely thanks to the indulgence of the BM14 staff. Secondly, radiation damage is re-emerging as one of the biggest obstacles to structure solution. If attenua-

tion of bright beamlines is necessary to allow the beamline motors enough time to function accurately then it should be done. Crystals should receive the minimum radiation dose consistent with collecting the quality and resolution of data required. If, say, 3.0 Å resolution data are required for MAD phasing (and a high-resolution native data set is available), then pushing the detector back and decreasing the exposure time will in general produce better data even if the crystal is capable of diffraction to higher resolution (Table 2; Fig. 4d). Our experience is that more lower dose images can eventually provide better anomalous data than fewer higher dose images (data not shown). Thirdly, radiation damage is not necessarily all bad since it can effectively provide a new derivative (Ravelli et al., 2003). Simply treating early and late data sets separately (such as PK1 and PK2 in this case) can allow phase-improvement programs to modify site occupancies separately and give better starting maps.

Finally, dehydration is probably feasible for many crystal forms and particularly where some innate flexibility of the protein structure in the crystal allows it. Where the diffraction is limited to moderate resolution (and this cannot be attributed to damage by cryoprotection solutions or crystal handling), controlled dehydration either using dehydrating solutions (as above) or humidity-controlling devices (such as the Free Mounting System, Proteros Biostructures, Martinsried, Germany) can sometimes produce a dramatic improvement (Kiefersauer et al., 2000).

References

Bork, P., Doerks, T., Springer, T. A. & Snel, B. (2001). *Trends Biochem. Sci.* **24**, 261–263.

Brünger, A. T. (1992). *X-PLOR Version 3.1. A System for X-ray Crystallography and NMR.* Yale University, Connecticut, USA.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Cowtan, K. D. & Zhang, K. Y. (1999). *Prog. Biophys. Mol. Biol.* **72**, 245–270.

Esnouf, R. M. (1997). *J. Mol. Graph. Model.* **15**, 132–134.

Esnouf, R. M. (1999). *Acta Cryst.* D**55**, 938–940.

Esnouf, R. M., Ren, J., Garman, E. F., Somers, D. O'N., Ross, C. K., Jones, E. Y., Stammers, D. K. & Stuart, D. I. (1998). *Acta Cryst.* D**54**, 938–953.

Gherardi, E., Love, C. A., Esnouf, R. M. & Jones, E. Y. (2004). *Curr. Opin. Struct. Biol.* **14**, 669–678.

Hall, K. T., Boumsell, L., Schultze, J. L., Boussiotis, V. A., Dorfman, D. M., Cardoso, A. A., Bensussan, A., Nadler, L. M. & Freeman, G. J. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 11780–11785.

Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kiefersauer, R., Than, M. E., Dobbek, H., Gremer, L., Melero, M., Strobl, S., Dias, J. M., Soulimane, T. & Huber, R. (2000). *J. Appl. Cryst.* **33**, 1223–1230.

Kolodkin, A. L., Matthes, D. J. & Goodman, C. S. (1993). *Cell*, **75**, 1389–1399.

Love, C. A., Harlos, K., Mavaddat, N., Davis, S. J., Stuart, D. I., Jones, E. Y. & Esnouf, R. M. (2003). *Nature Struct. Biol.* **10**, 843–848.

Merritt, E. A. & Murphy, M. E. P. (1994). *Acta Cryst.* D**50**, 869–873.

Morris, R. J., Perrakis, A. & Lamzin, V. S. (2002). *Acta Cryst.* D**58**, 968–975.

Navaza, J. (2001). *Acta Cryst.* D**57**, 1367–1372.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Raper, J. A. (2000). *Curr Opin. Neurobiol.* **10**, 88–94.

Ravelli, R. B., Leiros, H. K., Pan, B., Caffrey, M. & McSweeney, S. (2003). *Structure*, **11**, 217–224.

Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.

Semaphorin Nomenclature Committee (1999). *Cell*, **97**, 551–552.

Shi, W., Kumanogoh, A., Watanabe, C., Uchida, J., Wang, X., Yasui, T., Yukawa, K., Ikawa, M., Okabe, M., Parnes, J. R., Yoshida, K. & Kikutani, H. (2000). *Immunity*, **13**, 633–642.

Stammers, D. K., Somers, D. O'N., Ross, C. K., Kirby, I., Ray, P. H., Wilson, J. E., Norman, M., Ren, J. S., Esnouf, R. M., Garman, E. F., Jones, E. Y. & Stuart, D. I. (1994). *J. Mol. Biol.* **242**, 586–588.

Tamagnone, L., Artigiani, S., Chen, H., He, Z., Ming, G. I., Song, H., Chedotal, A., Winberg, M. L., Goodman, C. S., Poo, M., Tessier-Lavigne, M. & Comoglio, P. M. (1999). *Cell*, **99**, 71–80.

Terwilliger, T. C. (2003a). *Acta Cryst.* D**59**, 38–44.

Terwilliger, T. C. (2003b). *Acta Cryst.* D**59**, 45–49.

Weeks, C. M. & Miller, R. (1999). *J. Appl. Cryst.* **32**, 120–124.

Winberg, M. L., Noordermeer, J. N., Tamagnone, L., Cornoglio, P. M., Spriggs, M. K., Tessier-Lavigne, M. & Goodman, C. S. (1998). *Cell*, **95**, 903–916.

Xiong, J. P., Stehle, T., Diefenbach, B., Zhang, R., Dunker, R., Scott, D. L., Joachimiak, A. Goodman, S. L. & Arnaout, M. A. (2001). *Science*, **294**, 339–345.