Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization of foot-and-mouth disease virus 3C protease: surface mutagenesis and a novel crystal-optimization strategy

James R. Birtley and Stephen Curry*

Biophysics Section, Department of Biological Sciences, Blackett Laboratory, Imperial College, South Kensington Campus, London SW7 2AZ, England

Correspondence e-mail: s.curry@imperial.ac.uk

Received 31 January 2005
Accepted 13 March 2005

Foot-and-mouth disease virus (FMDV) 3C protease ($3C^{pro}$) plays a vital role in virus replication by performing most of the cleavages required to divide the viral polyprotein precursor into its functional component proteins. To date, no structural information has been available for FMDV $3C^{pro}$, which is an attractive target for antiviral drugs. Targeted mutagenesis of surface amino acids identified two Cys residues that were detrimental to solubility and contributed to the time-dependent formation of a proteinaceous skin in samples of purified wild-type protein. Substitution of these amino acids, combined with trimming of the N- and C-termini, yielded a $3C^{pro}$ construct that was amenable to crystallization. High-resolution diffraction (1.9 Å) was only obtained following 'iterative screening' in which commercial crystal screening solutions were used as additives once initial crystallization conditions had been obtained.

1. Introduction

Foot-and-mouth disease virus (FMDV) is a highly contagious pathogen that afflicts cloven-hoofed domestic livestock worldwide (Knowles & Samuel, 2003). Although vaccines are available, the occurrence in recent years of devastating outbreaks in Uruguay, Taiwan and the United Kingdom has stimulated the search for alternative methods of disease control (Grubman & Baxt, 2004). As a member of the picornavirus family, FMDV has a positive-sense single-stranded RNA genome with a single open reading frame that is translated in infected cells as a polyprotein precursor. This polyprotein contains one or more proteases that act to dissect the precursor into the functional components required for viral replication. FMDV 3C protease ($3C^{pro}$) performs ten of the 13 cleavages of the polyprotein and is therefore a key player in the viral replication cycle (Grubman & Baxt, 2004). Like the 3C proteases from other picornaviruses and the homologous NIa protease from tobacco etch virus, TEV (Phan *et al.*, 2002), FMDV $3C^{pro}$ is predicted to be a chymotrypsin-like cysteine protease with a Cys-His-Asp charge-relay system in the active site (Grubman *et al.*, 1995). The dependence of the virus on the activity of $3C^{pro}$ makes the enzyme an attractive target for antiviral drug design.

The precise mechanism of 3C proteases remains to be fully elucidated, largely as a result of the significant deviations from the canonical catalytic triad conformation that have been observed in the 3C proteases from hepatitis A virus, human rhinovirus and poliovirus (Allaire *et al.*, 1994; Matthews *et al.*, 1994; Mosimann *et al.*, 1997). Moreover, although most $3C^{pro}$ enzymes are highly selective for Gln at the P1 position of polypeptide substrates, FMDV $3C^{pro}$ appears not to discriminate between Gln and Glu (Blom *et al.*, 1996). To investigate these issues at the molecular level, we aimed to determine the crystal structure of FMDV $3C^{pro}$. Although this protein is expressed well in *Escherichia coli* (Li *et al.*, 2001), it proved to be relatively insoluble and refractive to crystallization. We report here the efforts made to mutate the protein so as to improve its solubility (without compromising its catalytic activity) and the adoption of a novel but straightforward crystal-optimization strategy that produced a dramatic improvement in crystal quality once initial crystallization conditions had been obtained.

2. Experimental procedures and results

2.1. Expression and purification

The initial FMDV 3C^{pro} expression plasmid (kindly provided by Chris Proud, Wei Li and Graham Belsham; Li *et al.*, 2001) contains 3C^{pro} fused to the three upstream copies of the 3B peptide as a 35 kDa 3B1-3B2-3B3-3C protein corresponding to a fragment from the viral polyprotein (Fig. 1). The first six residues at the N-terminus of 3B1 are missing and the last six residues of 3C^{pro} (^{208}PEPHHE^{213}) have been converted to a 6×His tag to facilitate purification. During expression of the protein in BL21 (DE3) pLysS *E. coli*, the precursor is processed into its constituent peptides, releasing 3C^{pro} (MW = 23 kDa) with an authentic N-terminal sequence.

3C^{pro} was overexpressed for 4 h at 310 K in cultures that had reached an OD at 600 nm of at least 0.6 by addition of 1 m*M* IPTG (isopropyl-β-D-thiogalactopyranoside) and was purified by standard procedures on TALON resin (BD Biosciences). However, the purified protein [in buffer *A*; 50 m*M* HEPES pH 7.0, 200 m*M* NaCl, 1 m*M* 2-mercaptoethanol, 0.01%(*w*/*v*) sodium azide] precipitated within 1–2 h and efforts to manipulate the buffer conditions during purification (*e.g.* by addition of salt or glycerol) gave no significant improvement in solubility. Even after clarification by centrifugation of 3C^{pro} solutions at 16 000*g* for 5 min, the residual soluble protein continued to precipitate over several days when stored in a sealed microfuge tube at 277 K; for example, a solution initially at 7.2 mg ml^{-1} declined to 3.8 mg ml^{-1} over 5 d (data not shown) as a result of precipitation. During this period, an elastic skin was also observed to form on the top surface of the solution. No crystals were ever obtained using this protein preparation.

2.2. Mutagenesis to enhance solubility

We generated a homology model for FMDV 3C^{pro} to facilitate the identification of amino acids that might be substituted by mutagenesis in order to alter the surface properties of the enzyme and thereby enhance its solubility. A structural alignment of HAV and HRV 3C^{pro} (23 and 17% identical to the FMDV enzyme, respectively) was performed in *O* (Jones *et al.*, 1991) and used to optimize the sequence alignment of these proteins with FMDV 3C^{pro}. This alignment was then submitted to the Swissmodel server (Schwede *et al.*, 2003) to generate a homology model based on the crystal structure of HAV 3C^{pro} (PDB code 1hav; Allaire *et al.*, 1994), the closest homologue to FMDV 3C^{pro}.

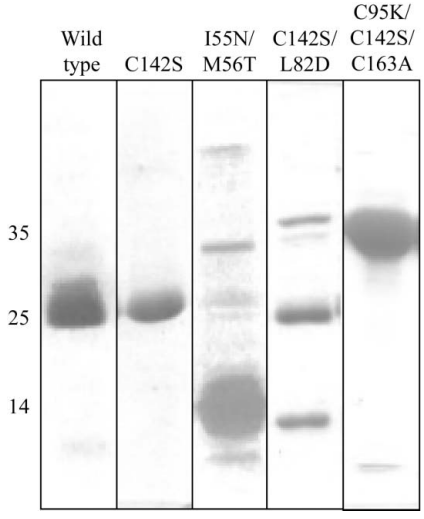

Figure 1
SDS–PAGE analysis of the elution peak of recombinant FMDV 3C^{pro} from TALON resin. The positions of molecular-weight markers are shown on the left-hand side.

Table 1
Surface mutagenesis to improve solubility of FMDV 3C^{pro}.

Mutant	Mutations	Expression/processing†	Polydispersity‡ (%)
Wild type	—	+++	48
M1	I55N/M56T	Incorrect processing	nt
M2	F70S	Incorrect processing	nt
M3	P114S/V115D	Unprocessed	nt
M4	V140S	+++	nt
M5	I55D	—	nt
M6	M81N	+++	33
M7	L82D	+++	55
M8	C95K	+++	37
M9	V124S	+++	47
M10	A133S	+++	35
M11	C142S	+++	25
M11-6	M81N/C142S	+++	19
M11-7	L82D/C142S	+	33
M11-8	C95K/C142S	+++	11
M11-9	V124S/C142S	+++	29
M11-10	A133SC142S	+++	29

† +++, fully processed; +, partially processed; —, insoluble. ‡ nt, not tested.

Selected residues were mutated using the QuikChange Mutagenesis Kit (Stratagene) (Table 1) and expressed and purified essentially as described above. The 3B1-3B2-3B3-3C fusion construct provides a fast read-out of the impact of mutations on the activity of the enzyme. Mutations that had a deleterious effect on the solubility were easily detected since they resulted in most of the protein appearing in the insoluble fraction during purification. Those that had a more subtle effect on the protein could be detected by gel analysis of the product that eluted from the TALON resin following incubation with clarified lysate (Fig. 1). For example, whereas the wild-type 3C^{pro} and the C142S mutant both yielded a predominant band at around 25 kDa corresponding to correctly processed enzyme, the double mutant I55N/M56T gave a much smaller product suggesting incorrect processing, while the C142S/L82D double mutant showed evidence of incomplete processing suggestive of a subtle defect in folding or catalytic activity (or both). Control experiments with an inactive mutant containing a substitution of the active-site nucleophile (C163A) showed that only the 35 kDa precursor was recovered.

Dynamic light-scattering measurements were used to provide a more quantitative assessment of the aggregation behaviour of those mutants that were expressed and processed at least as well as the wild type. Measurements were performed at room temperature using a DynaPro instrument (Protein Solutions, Charlottesville, USA) on 20 µl samples of 3C^{pro} at 2 mg ml^{-1} in buffer *A* with 1 m*M* EDTA. The dynamic light-scattering data were analysed to determine a percentage polydispersity value for each sample. A control sample of monomeric human serum albumin (HSA) had a polydispersity of 11% under our conditions; this compares with 48% for wild-type FMDV 3C^{pro}, which is indicative of a much higher degree of heterogeneous aggregation. Of the single mutants tested, C142S was the least polydisperse (25%), although it exhibited significantly more aggregation than HSA.

In a second round of mutagenesis, mutations from the first round were combined with C142S to produce a series of double mutants. Of these, the C95K/C142S mutant exhibited the least polydispersity (11%), a level comparable with the control protein. However, although it had good solubility and did not precipitate or form a surface skin even after prolonged storage, extensive crystallization trials with this double mutant yielded no crystals.

Table 2
Effect of modification of N- and C-termini on expression and solubility.

Construct†	Terminal sequences (N . . . C)‡	Expression§	Processing enzyme
gam-3C^{pro}(1–213)	gamSGAPPTDL . . . KMKAHVEPEPHHE	+++	TEV^{pro}
gam-3C^{pro}(1–202)	gamSGAPPTDL . . . KM	−	TEV^{pro}
gam-3C^{pro}(1–205)	gamSGAPPTDL . . . KMKAH	+	TEV^{pro}
gam-3C^{pro}(1–207)	gamSGAPPTDL . . . KMKAHVE	+++	TEV^{pro}
gam-3C^{pro}(1–207h)	gamSGAPPTDL . . . KMKAHVEh	+++	TEV^{pro}
gam-3C^{pro}(1–207hh)	gamSGAPPTDL . . . KMKAHVEhh	+++	TEV^{pro}
g-3C^{pro}(1–207)	gSGAPPTDL . . . KMKAHVE	+++	TEV^{pro}/thrombin¶
g-3C^{pro}(1–207h)	gSGAPPTDL . . . KMKAHVEh	+++	TEV^{pro}/thrombin¶
3C^{pro}(2–207)	GAPPTDL . . . KMKAHVE	+++	TEV^{pro}
g-3C^{pro}(3–207)	gPPTDL . . . KMKAHVE	−	TEV^{pro}
g-3C^{pro}(4–207)	gPTDL . . . KMKAHVE	−	TEV^{pro}

† Each of these constructs contains substitutions C95K and C142S to enhance solubility and C163A to inactivate the enzyme (thus preventing 3C from processing the TEV^{pro} cleavage site). Lower case letters in the names indicate amino-acid sequences at the N- or C-termini that are not present in the wild-type 3C^{pro} sequence (see column 2). ‡ The N- and C-terminal sequences of the proteins produced following proteolytic cleavage to remove the N-terminal His tag by TEV^{pro} or thrombin are shown (using single-letter amino-acid codes). § +++, high expression of soluble protein; +, low expression; −, insoluble protein. ¶ Indicates that two versions of the construct were produced with either TEV^{pro}- or thrombin-cleavable histidine tags; note that the processed products have the same N-terminal sequence.

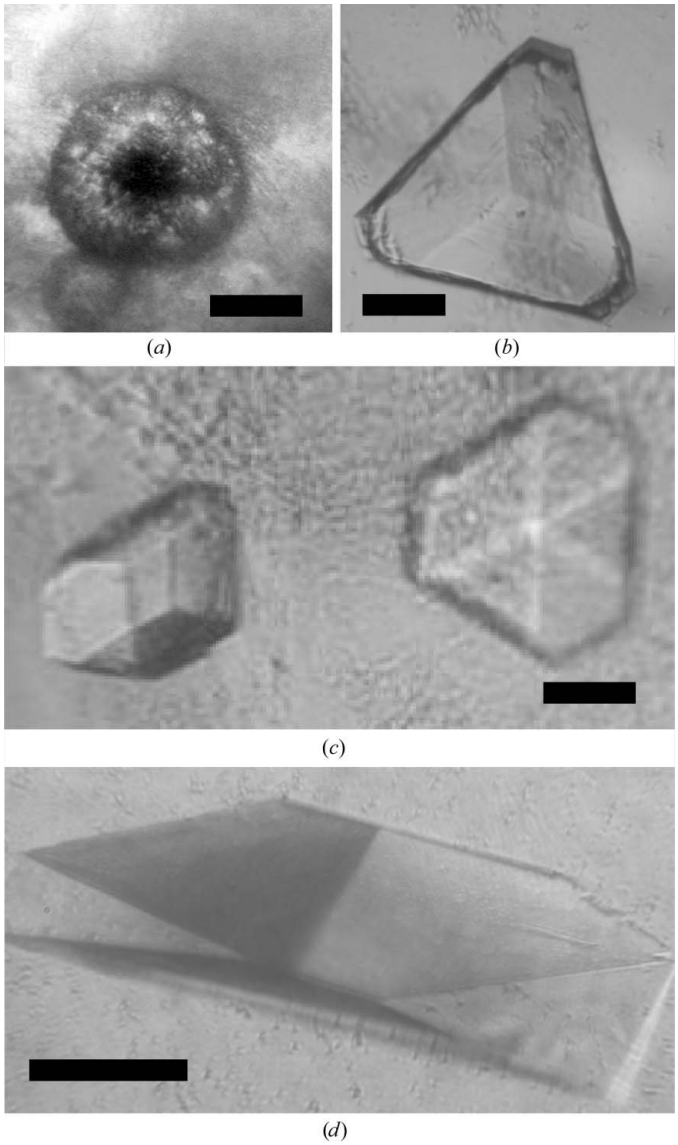

Figure 2
Crystals of recombinant FMDV 3C^{pro}. (*a*) Initial crystals obtained with g-3C^{pro}(1–207) in Crystal Screen 2 condition No. 42 were optimized with (*b*) 1%(*v*/*v*) PEG 1000 or (*c*) 1%(*v*/*v*) Anapoe-X-405. (*d*) Crystals obtained after further additive screening with construct 3C^{pro}(2–207); see text. Scale bar = 75 µm.

At this point, efforts were made to trim the N- and C-termini of the protein. The C-terminus was considered to be particularly amenable to truncation since sequence alignment revealed that FMDV 3C^{pro} was extended by 7–13 residues compared with HAV, HRV and PV 3C proteases; moreover, it had been shown that the last six residues could be substituted by a 6×His tag without loss of activity (Li *et al.*, 2001). Initially, a series of constructs that would produce proteins truncated by four, five, six, eight and 11 residues were subcloned by PCR into a pETM-11 vector (Zou *et al.*, 2003) that adds an N-terminal 6×His tag that can be cleaved off with TEV protease (TEV^{pro}; Parks *et al.*, 1994). These constructs all contained the C95K and C142S substitutions that were required for solubility; in addition, to prevent autolysis of the tag by 3C^{pro}, the enzyme was inactivated by substitution of a third Cys residue, the active-site nucleophile (C163A). The proteins were expressed in BL21(DE3) *E. coli* and purified on TALON resin essentially as described above. The His tag was removed by digestion with 1 mg recombinant His-tagged TEV^{pro} per 10 mg of 3C^{pro}; the TEV^{pro} was removed by binding to TALON resin and the processed 3C^{pro} finally purified by gel filtration on a HR 10/30 Superdex 75 column (Amersham Biosciences) run using buffer *A*. The resulting protein products have three extraneous residues at the N-terminus (Gly-Ala-Met; see Table 2). Expression tests showed that the removal of the last six C-terminal residues was the largest truncation that still resulted in a good yield of soluble protein (typically 10 mg of purified protein per litre of *E. coli* culture). Site-directed mutagenesis, performed using the QuikChange method (Stratagene), was then used to modify the N-terminus of 3C^{pro} while still preserving the histidine tag and the TEV^{pro} cleavage site. Modifications at the N-terminus were more limited in scope: truncations beyond residue 2 resulted in insoluble protein (Table 2) and constructs g-3C^{pro}(1–207) and 3C^{pro}(2–207) were therefore selected for renewed crystallization experiments.

2.3. Crystal screening and optimization

Construct g-3C^{pro}(1–207) (Table 2) yielded several hits within 1–2 d in crystallization trials performed at 277 and 291 K by sitting-drop vapour diffusion with Crystal Screen and Crystal Screen 2 (Hampton Research) and Wizard Screens I and II (Emerald Biosystems). Construct 3C^{pro}(2–207) gave many fewer hits, which only appeared over a timescale of several weeks, and was not used subsequently. Initial crystals of g-3C^{pro}(1–207) were either small (<50 µm) or grew as irregular clusters (Fig. 2*a*). Efforts to optimize

the initial crystallization conditions were most successful for the crystals obtained at 291 K using condition No. 42 of Crystal Screen 2 (1.6 *M* sodium citrate pH 6.5; Fig. 2*a*); addition of 1%(*v*/*v*) PEG 1K or Anapoe-X-405 (which is structurally similar) produced a marked increase in crystal size up to a maximum dimension of around 200 µm (Figs. 2*b* and 2*c*). The improved crystals diffracted anisotropically to a maximum resolution of 3.5 Å and were found to belong to space group *R*3.

In an effort to obtain further improvement of the crystals, a version of g-3C^{pro}(1–207) with a thrombin-cleavable His tag was produced by mutating the ELYFQ/G TEV^{pro} cleavage site to one recognized by thrombin (LVPR/GS) in order to avoid contamination of the protein with impurities in the preparations of TEV^{pro} (data not shown). Although processing of this construct with 50 µg thrombin (Sigma–Aldrich, T4648) per 10 mg of 3C^{pro}, followed by gel filtration (as described above), resulted in a purer final product of 3C^{pro} as judged by SDS–PAGE, this material gave no improvement in crystal quality.

At this point, a new optimization strategy was adopted that permitted a convenient and rapid broadening of the search for additives that might further improve the quality of the crystals. This procedure simply involved the use of the standard crystallization screens as additive screens. To prepare the crystallization solutions, we mixed 70–85 µl of the optimized crystallization solution for thrombin-cleaved g-3C^{pro}(1–207) in a microfuge tube with a smaller volume (15–30 µl) of each of the solutions from the two Hampton Research screens to give a total volume of 100 µl (98 mixtures at each volume ratio). The solutions were vortexed briefly to ensure thorough mixing. In one case, the additive (Crystal Screen 2 solution No. 18) resulted in the immediate appearance of precipitate, most likely owing to the formation of insoluble salts, and this mixture was not used for crystallization. Crystallization trials with the remaining mixtures indicated that the 75:25 volume ratio gave the best yield of crystals. In around 25% of the drops set up at this ratio, crystals were obtained and about half of these exhibited some improvement in terms of size and sharpness (Crystal Screen condition Nos. 3, 4, 10, 28, 32, 35; Crystal Screen 2 condition Nos. 13, 14, 25, 32, 45). Based on visual appearance, addition of around 0.1 *M* ammonium dihydrogen phosphate appeared to produce the greatest improvement. Further refinement of the conditions established that addition of 2 µl 10 mg ml^{-1} protein solution [in 50 m*M* HEPES pH 7.0, 200 m*M* NaCl, 1 m*M* 2-mercaptoethanol, 1 m*M* ethylenediaminetetraacetic acid, 0.01%(*w*/*v*) sodium azide] with 2.5 µl precipitant solution (1.15 *M* sodium citrate pH 6.5, 0.125 *M* ammonium dihydrogen phosphate) and 0.5 µl 10%(*v*/*v*) Anapoe-X-405 over a reservoir containing 100 µl precipitant solution was optimal for crystallization.

This optimized crystallization condition was then used with the g-3C^{pro}(1–207h) construct, which had also proved to be soluble in expression tests (Table 2), and yielded large blade-like crystals that diffracted to 2.25 Å (Fig. 2*d*; Table 3). These crystals were reproduced with selenomethionine-labelled g-3C^{pro}(1–207h) protein, using the same precipitant conditions but with the protein concentration reduced to 4 mg ml^{-1}, and enabled the structure to be determined to 1.9 Å by multi-wavelength anomalous dispersion (Birtley *et al.*, 2005).

Table 3
Data-collection statistics.

Values in parentheses are for the highest resolution shell.

Space group	*R*3
Unit-cell parameters	
$a = b$ (Å)	143.07
c (Å)	44.35
$\alpha = \beta$ (°)	90
γ (°)	120
Resolution range	27.0–2.25 (2.37–2.25)
R_{merge} (%)	5.3 (39.3)
Completeness (%)	99.3 (97.7)
$I/\sigma(I)$	10.3 (2.1)
Observed reflections	93129
Independent reflections	15396

3. Discussion

The solubility problems with recombinant FMDV 3C^{pro} were completely solved by the substitution of two Cys residues (Cys95 and Cys142), which were subsequently found to be the most surface-exposed cysteines of the seven found in the protein sequence (Birtley *et al.*, 2005). Mutation of the remaining Cys residues (Cys31, Cys32, Cys191 and Cys193) to Ala gave no further improvement in protein solubility. This result suggests that for other proteins where low solubility is combined with a time-dependent skin formation (even in the presence of reducing agents) mutagenesis trials on Cys residues may be of benefit. In other cases, even where skin formation has not been reported, substitution of Cys residues has helped to improve protein samples for crystallization (Al-Ayyoubi *et al.*, 2004; Niessing *et al.*, 2004).

Most of the other residues selected for substitution (Ile55, Phe70, Pro114, Val124, Ala133; Table 1) were found to be at least partially buried in hydrophobic contacts with other side chains and this probably accounts for their reduced activity or lack of enhanced solubility (Birtley, 2004). In contrast, Met81, Leu82 and Val140 are all solvent-exposed on the surface of the protein, but their substitution by polar residues could not of course prevent aggregation owing to Cys95 and Cys142.

The iterative crystal-screening strategy described here is clearly applicable to other proteins and has the convenience of using reagents that are likely to be readily available in most crystallography laboratories. It can also be readily implemented in robotic crystallization experiments using existing liquid-handling technology. Variations in relative volumes of the initial condition and the additive component should probably be tested as part of these crystal optimization experiments. It is likely that different patterns of insoluble salts may be obtained upon formulation of the mixtures, depending on the initial conditions that are being modified, but this is straightforward to determine empirically.

We thank Chris Proud, Wei Li and Graham Belsham for 3C^{pro} cDNA, and David Barford and James Parker for help with dynamic light-scattering measurements. We thank Peter Brick for careful reading of the manuscript. We are grateful to staff at the X13 (EMBL/DESY, Hamburg) and BM14 (ESRF, Grenoble) synchrotron beamlines. JRB was supported by an MRC studentship. This work is supported by the Fleming Fund (Imperial College) and the BBSRC. JRB's PhD thesis will be made available upon request as a pdf file.

References

Al-Ayyoubi, M., Gettins, P. G. & Volz, K. (2004). *J. Biol. Chem.* **279**, 55540–55544.
Allaire, M., Chernaia, M. M., Malcolm, B. A. & James, M. N. (1994). *Nature (London)*, **369**, 72–76.
Birtley, J. R. (2004). PhD thesis, University of London, England.
Birtley, J. R., Knox, S. R., Jaulent, A. M., Brick, P., Leatherbarrow, R. J. & Curry, S. (2005). *J. Biol. Chem.* **280**, 11520–11527.
Blom, N., Hansen, J., Blaas, D. & Brunak, S. (1996). *Protein Sci.* **5**, 2203–2216.

Grubman, M. J. & Baxt, B. (2004). *Clin. Microbiol. Rev.* **17**, 465–493.
Grubman, M. J., Zellner, M., Bablanian, G., Mason, P. W. & Piccone, M. E. (1995). *Virology*, **213**, 581–589.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Knowles, N. J. & Samuel, A. R. (2003). *Virus Res.* **91**, 65–80.
Li, W., Ross-Smith, N., Proud, C. G. & Belsham, G. J. (2001). *FEBS Lett.* **507**, 1–5.
Matthews, D. A., Smith, W. W., Ferre, R. A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J. E., Janson, C. A., McElroy, H. E., Gribskov, C. L. & Worland, S. (1994). *Cell*, **77**, 761–771.
Mosimann, S. C., Cherney, M. M., Sia, S., Plotch, S. & James, M. N. (1997). *J. Mol. Biol.* **273**, 1032–1047.
Niessing, D., Huttelmaier, S., Zenklusen, D., Singer, R. H. & Burley, S. K. (2004). *Cell*, **119**, 491–502.
Parks, T. D., Leuther, K. K., Howard, E. D., Johnston, S. A. & Dougherty, W. G. (1994). *Anal. Biochem.* **216**, 413–417.
Phan, J., Zdanov, A., Evdokimov, A. G., Tropea, J. E., Peters, H. K. III, Kapust, R. B., Li, M., Wlodawer, A. & Waugh, D. S. (2002). *J. Biol. Chem.* **27**, 50564–50572.
Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. (2003). *Nucleic Acids Res.* **31**, 3381–3385.
Zou, P., Gautel, M., Geerlof, A., Wilmanns, M., Koch, M. H. & Svergun, D. I. (2003). *J. Biol. Chem.* **278**, 2636–2644.