Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Purification, crystallization and preliminary crystallographic analysis of the CBS-domain protein MJ1004 from *Methanocaldococcus jannaschii*

Iker Oyenarte,[a]‡ María Lucas,[b]‡ Inmaculada Gómez García[a] and Luis Alfonso Martínez-Cruz[a]*

[a]Structural Biology Unit, CIC bioGUNE, Parque Tecnológico de Bizkaia, Edificio 800, 48160 Derio, Bizkaia, Spain, and [b]Center for Integrated Protein Sciences and Munich Center for Advanced Photonics at the Gene Center, Department of Biochemistry, Ludwig-Maximilians-University Munich, Feodor-Lynen-Strasse 25, 81377 Munich, Germany

‡ These authors contributed equally to this work.

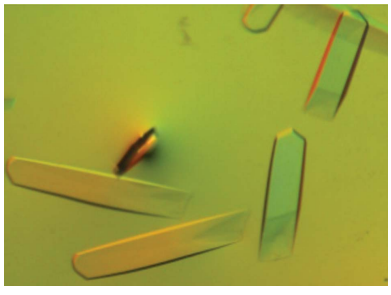
Correspondence e-mail: amartinez@cicbiogune.es

Received 23 October 2010
Accepted 20 December 2010

The purification and preliminary crystallographic analysis of the archaeal CBS-domain protein MJ1004 from *Methanocaldococcus jannaschii* are described. The native protein was overexpressed, purified and crystallized in the monoclinic space group $P2_1$, with unit-cell parameters a = 54.4, b = 53.8, c = 82.6 Å, β = 106.1°. The crystals diffracted X-rays to 2.7 Å resolution using synchrotron radiation. Matthews-volume calculations suggested the presence of two molecules in the asymmetric unit that are likely to correspond to a dimeric species, which is also observed in solution.

1. Introduction

CBS domains are small motifs that as such have no defined function(s) but regulate the activity of a wide variety of proteins in organisms ranging from bacteria to humans (Bateman, 1997; Scott *et al.*, 2004). CBS domains are known to bind adenosine ligands (Scott *et al.*, 2004), metal ions (Hattori *et al.*, 2007; Ishitani *et al.*, 2008) and DNA (Aguado-Llera *et al.*, 2010) with different affinities. Their role as energy-sensing modules has been widely suggested and they are also involved in the gating of osmoregulatory proteins (Biemans-Oldehinkel *et al.*, 2006), the transport and binding of Mg^{2+} (Hattori *et al.*, 2007; Ishitani *et al.*, 2008), the modulation of intracellular trafficking of chloride channels (Carr *et al.*, 2003), nitrate transport (De Angeli *et al.*, 2009) and as 'internal inhibitors' of pyrophosphatase activity (Jämsen *et al.*, 2010; Tuominen *et al.*, 2010). Despite their low degree of sequence similarity, CBS motifs share common structural properties, either isolated or as domains of larger proteins (Zhang *et al.*, 1999; Estévez *et al.*, 2003). Interestingly, mutations in their amino-acid sequence have been associated with human diseases (OMIM 146690, MIM 602643, MIM 180105 and MIM 600858, where OMIM refers to the online version of *Mendelian Inheritance in Man* at http://www.ncbi.nlm.nih.gov/omim and MIM refers to the 1998 edition of the printed version; McKusick, 1998; Bowne *et al.*, 2002; Scott *et al.*, 2004; Kemp, 2004), which enhances them as potential targets for drug development. To date, more than 30 structures of CBS-domain proteins have been deposited in the databases (see Supplementary Table 1[1]). In all these structures two CBS domains associate maintaining pseudo-twofold symmetry (Lucas *et al.*, 2010) to form a compact fold known as a 'Bateman domain' (Fig. 1; Kemp, 2004). The cleft between the two CBS subunits is a binding site for adenosyl groups (Fig. 1). In turn, two Bateman domains usually dimerize in a head-to-head (HH) or a head-to-tail (HT) manner, forming a disc-shaped structure that contains four CBS motifs and has been named a 'CBS module' (Mahmood *et al.*, 2009; Fig. 1). The forces that direct the final assembly as well as the relationship between conformation and function currently remain unknown (Rudolph *et al.*, 2007), but the exposed amino-acid residues of the interficial α-helices are expected to play a relevant role (Fig. 1). The properties of these residues have been suggested to determine the 'aperture' or 'closure'



[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: FW5287).

that is observed in different CBS modules (Rudolph *et al.*, 2007). Obviously, the type of association is important because it affects not only the size and shape of the binding sites but also their potential affinity for specific ligands. However, despite an enormous experimental effort, only 5% of all CBS-domain entries in the PDB (3lv9, 1pbj and 1o50; Midwest Center for Structural Genomics, unpublished work; Miller *et al.*, 2004) show HT-oriented modules. Strikingly, a truncated form of human γ-AMPK containing only its second CBS pair (PDB code 2uv4) self-associates to form an HT dimer (Day *et al.*, 2007). This lack of information limits our knowledge of the molecular mechanisms underlying the regulatory role exerted by CBS domains. In the course of our studies (Fernández-Millán *et al.*, 2008; Lucas *et al.*, 2008, 2010; Martínez-Cruz *et al.*, 2009, 2011; Gómez-García *et al.*, 2009, 2010; Aguado-Llera *et al.*, 2010) and with the aim of devising new approaches that will overcome this deficiency, we carefully examined all available entries in the PDB (Supplementary Table 1) and compared the structural features of HH-oriented species with those observed in HT modules. We focused on proteins containing two tandemly repeated CBS motifs, since four tandemly repeated assemblies (*e.g.* the proteins AMPK or MJ1225) necessarily adopt an HH orientation (Xiao *et al.*, 2007; Amodeo *et al.*, 2007; Gómez-García *et al.*, 2010). The study immediately revealed several aspects that attracted our attention.

(i) The available structures belong to a wide variety of organisms for which no systematic structural analysis has been carried out (with the exception of humans).

(ii) HH orientations are preferable when CBS pairs are fused to other protein domains.

(iii) The nature of the exposed residues from interficial α-helices participating in dimerization of Bateman domains is similar (mainly

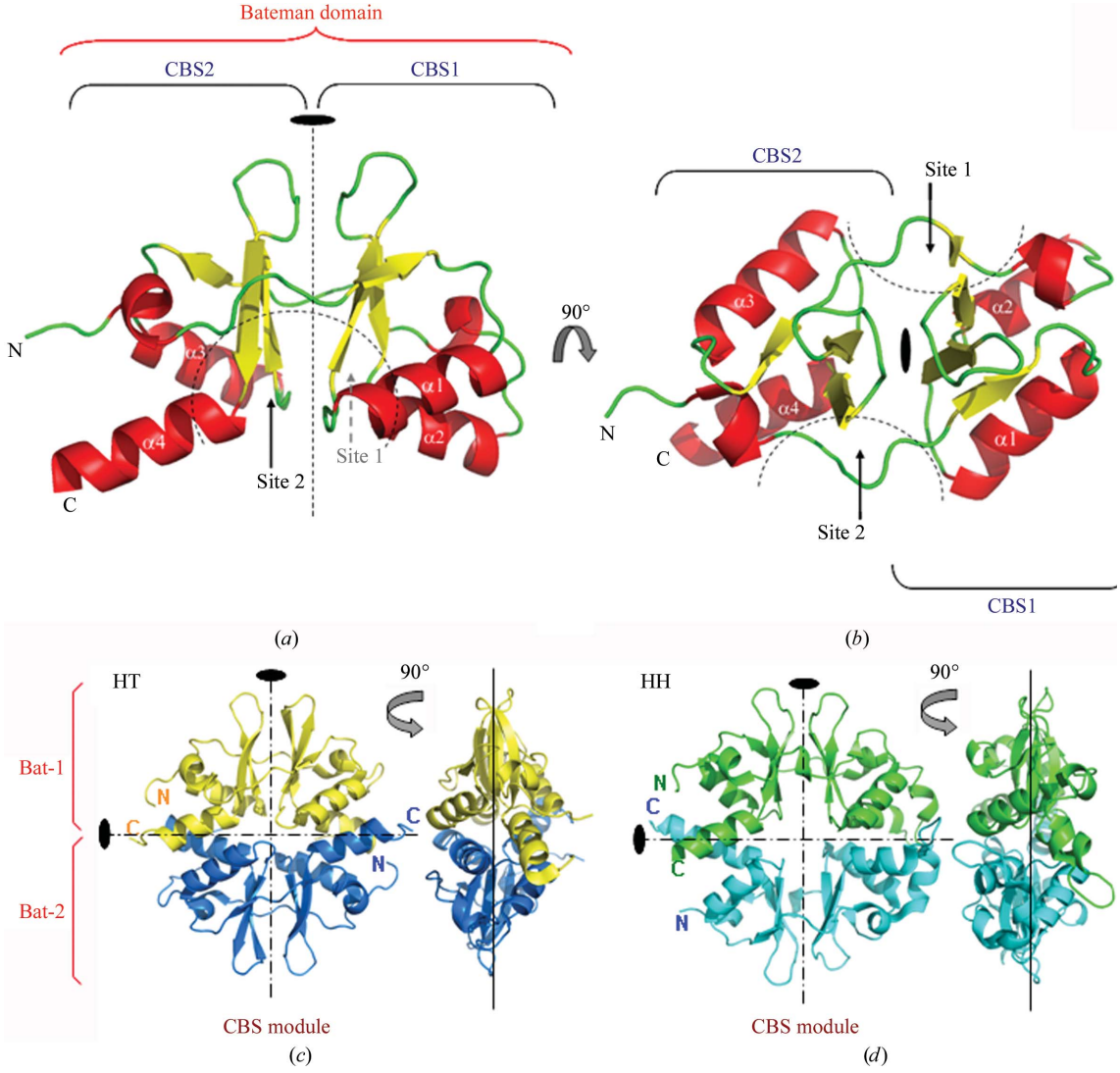


Figure 1
(*a*, *b*) Three-dimensional structure of a Bateman domain. The protein MJ0100 from *M. jannaschii* was used as a template (Lucas *et al.*, 2010). (*a*) A Bateman domain consists of two tandemly repeated CBS domains, namely CBS1 and CBS2, which are related by a dyad axis (represented by broken lines). The α-helices (α1–α4) that usually intervene in dimerization of the Bateman domains are numbered. (*b*) The same structure after a 90° rotation around the horizontal axis. Each Bateman domain contains two potential ligand-binding cavities located on opposite sides of the central β-sheets (broken lines). (*c*, *d*) Ribbon representation of the most common orientations adopted by Bateman domains in CBS-domain proteins. Upon dimerization, Bateman domains form a disc-shaped structure that contains four CBS motifs and is known as a 'CBS module', in which the two Bateman domains (Bat-1 and Bat-2) lie approximately within the same plane (represented by a straight continuous vertical line). Dotted–dashed lines indicate the dyad axis between the CBS1 and CBS2 subunits within a Bateman domain (vertical axis) or between two Bateman domains (horizontal axis). (*c*) HT orientation corresponding to PDB entry 1o50. (*d*) HH orientation corresponding to PDB entry 1yav (D. Kumaran & S. Swaminathan, unpublished work). The same structures are also represented after a 90° rotation around the vertical axis. As depicted, the main interactions between two Bateman domains occur between α-helices (α1–α4).

Table 1
Data-processing statistics for $MJ1004_{full}$.

A native data set was collected on synchrotron beamline ID23.2 at the European Synchrotron Research Facility (ESRF), Grenoble, France. Values in parentheses are for the outer resolution shell. $R_{p.i.m.}$ and $R_{meas}/R_{r.i.m.}$ were calculated with *SCALA* from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994).

Beamline	ID23.2
Wavelength (Å)	0.873
Total no. of reflections	168619
No. of unique reflections	12996
Resolution range (Å)	79–2.7
Space group	$P2_1$
Unit-cell parameters (Å)	
a (Å)	54.4
b (Å)	53.8
c (Å)	82.6
β (°)	106.1
Mosaicity (°)	0.3
Completeness (%)	99.1 (93.0)
Multiplicity	7.2 (5.3)
R_{merge}† (%)	8.6 (29.4)
$R_{p.i.m.}$‡ (%)	5.1 (15.3)
$R_{meas}/R_{r.i.m.}$§ (%)	13.8 (42.3)
Mean $I/\sigma(I)$	23.5 (4.9)
Wilson B value (Å^2)	77.5

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is the weighted average intensity for all observations i of reflection hkl. ‡ $R_{p.i.m.} = \sum_{hkl}[1/(N-1)]^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where N is the number of observations of reflection hkl (Weiss, 2001; Evans, 2006). § $R_{meas} = R_{r.i.m.} = \sum_{hkl}[N/(N-1)]^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$.

hydrophobic amino acids) in both the HH and the HT arrangements, making it difficult to infer a preferred orientation.

(iv) HT-oriented CBS modules always adopt a 'closed' conformation (see Lucas *et al.*, 2010) which is strongly stabilized by hydrophobic interactions between α-helices (Fig. 1).

(v) The length of the peptide chain in HT modules is approximately the same in both CBS domains within each Bateman domain. In contrast to proteins such a γ-AMPK (Xiao *et al.*, 2007) or MJ1225 (Gómez-García *et al.*, 2010), there are no chain insertions that break the twofold symmetry that exists between the CBS subunits of each Bateman domain.

(vi) Strikingly, the amino-acid sequence identity between the two CBS motifs of the same Bateman domain is not homogeneous for proteins from a specific organism (see Fig. 2). More interestingly, the sequence identity is significantly lower in HT-oriented molecules (less than 20%) than in HH assemblies (Fig. 2), at least in prokaryotic molecules.

We then wondered whether exploiting these observations might help us to identify new potential HT candidates. Accordingly, we initiated a systematic structural study using a unique organism as a model. At this point, we should mention that three new entries (PDB codes 3pc2, 3pc3 and 3pc4) corresponding to the crystal structure of the enzyme cystathionine β-synthase from *Drosophila melanogaster* (dCBS) were deposited in the Protein Data Bank during the review process of the present manuscript. In the crystal, dCBS forms dimeric species in which Bateman domains associate in an HT manner (Koutmos *et al.*, 2010). Based on this structure, the sequence identity between the two CBS motifs of this protein (residues 385–445 and 452–511) is 11.6%.

CBS motifs are unusually abundant in archaea (Bult *et al.*, 1996). Therefore, organisms such as the hyperthermophile *Methanocaldococcus jannaschii* offer excellent models for the characterization of novel adenosyl-binding sites and ligands (Lucas *et al.*, 2010; Gómez-García *et al.*, 2010; Aguado-Llera *et al.*, 2010; Martínez-Cruz *et al.*, 2011) as well as for the identification of the molecular mechanisms involved (Lucas *et al.*, 2010). We believe that they might be also excellent candidates on which to perform systematic studies aimed at identifying novel HT-oriented species and characterizing their structural/functional properties. The *M. jannaschii* genome encodes 15 CBS-domain proteins (http://www.tigr.org; Fig. 2), which differ significantly in their composition and presumably in their ligand-binding abilities (Martínez-Cruz *et al.*, 2009; Lucas *et al.*, 2010). A close examination of their amino-acid sequences reveals the presence of two different groups: (i) one group with very short sequences that are not fused to other domains (so-called 'stand-alone' CBS-domain proteins, *e.g.* MJ0729; Martínez-Cruz *et al.*, 2009) and (ii) a second group with longer amino-acid sequences that are fused to other protein motifs (*e.g.* MJ0100; Lucas *et al.*, 2010). Only two members of this latter group (MJ1404 and MJ1225) contain four CBS domains in tandem (Gómez-García *et al.*, 2010), as in the γ-subunit of γ'-AMP-activated protein kinase (AMPK; Amodeo *et al.*, 2007). Among the *M. jannaschii* proteins, three molecules show significantly lower values for the identity between their two CBS subunits: MJ0450, MJ1004 and MJ0868 (Fig. 2). The open reading frame of gene *mj1004* (UniProtKB/Swiss-Prot entry Q58410) encodes a polypeptide chain of 214 amino acids with a molecular mass of 24 585 Da. Its sequence is formed by a CBS-domain pair (CBS1, residues 7–65; CBS2, residues 69–129), an unknown region (residues 123–181) and a transmembrane region (residues 182–204) as predicted using *TMHMM*2 (http://smart.embl-heidelberg.de/help/smart_glossary.shtml#TMHMM). MJ1004 is currently annotated as an uncharacterized hypothetical protein. Here, we describe the successful purification, crystallization and preliminary analysis of MJ1004 as a first step towards elucidating its three-dimensional structure and validating whether the low sequence identity between its CBS subunits leads to the adoption of an HT assembly of Bateman domains as observed in the other three reported cases.

2. Materials and methods

2.1. Cloning, expression and purification of MJ1004

Two different constructs, $MJ1004_{full}$ encoding the full-length native protein (residues 1–214; MW = 24 585 Da; pI = 9.01) and $MJ1004_{126}$ encoding a truncated form (residues 1–126; MW = 14 638 Da; pI = 8.59), of MJ1004 from *M. jannaschii* were used in this study. Selection of the truncated construct $MJ1004_{126}$ with potential crystallizability was made after the careful analysis of several candidates using the *XtalPred* crystallizability prediction server (http://ffas.burnham.org/XtalPred-cgi/xtal.pl; Slabinski *et al.*, 2007) and on the basis of multiple alignments against CBS-domain-containing proteins of known three-dimensional structure. The MJ1004 protein was amplified by PCR using genomic cDNA as a template, which was kindly provided by Sung-Hou Kim, University of Berkeley, California, USA. The primers used for $MJ1004_{full}$ were 5′-CACCACATATGAAAGTTAGGGATTTAATGGATAAGAAT-3′ (forward) and 5′-ATCGGATCCTCAACCCCCTTTCCTTCTTTT-3′ (reverse) and those for $MJ1004_{126}$ were 5′-CACCATGAAAGTTAGGGATTTAATGGATAAGAATTTTGCC-3′ (forward) and 5′-CTATTCATAGAGCTTAGCTAATGTTTTGACAACATC-3′ (reverse). The amplified DNAs were cloned into pET101/D-TOPO vector (Invitrogen). The DNA sequences were verified by sequencing and were then transformed into chemically competent BL21 (DE3) Star One Shot *Escherichia coli* cells (Invitrogen; Studier & Moffatt, 1986; Grunberg-Manago, 1999). Starter cultures were grown overnight at 310 K in Luria–Bertani (LB) medium containing 100 mg ml^{-1} ampicillin and 25 mg ml^{-1} chloramphenicol. The starter cultures were diluted into 2 l LB medium containing 100 mg ml^{-1} ampicillin and 25 mg ml^{-1} chloramphenicol. Overexpression of the

target proteins was induced by the addition of IPTG to a final concentration of 0.5 m*M* when the optical density of the culture at 600 nm reached 0.5. Expression was allowed to proceed for 3 h. The cells were harvested by centrifugation at 4000*g* for 15 min at 277 K. The cell pellet was resuspended in 20 ml lysis buffer (50 m*M* HEPES pH 7.0, 1 m*M* EDTA, 1 m*M* β-mercaptoethanol, 1 m*M* benzamidine, 0.1 m*M* PMSF with added DNAse) and lysed by sonication in a Labsonic P sonicator (Sartorius) for 10–12 s at 90% amplitude, keeping the cells on ice in order to prevent overheating. The cell debris was separated by ultracentrifugation at 120 000*g* in a 70 Ti rotor (Beckman) for 25 min at 277 K. A similar protocol was followed for the production of $MJ1004_{126}$. Since the source organism of MJ1004 is a hyperthermophile, the first purification step consisted of a heat shock in which the clarified lysate from the previous centrifugation step was heated at 348 K for 30 min. The *E. coli* proteins precipitated by the heat shock were removed by centrifugation at 4000*g* in an SX4250 rotor (Beckman) for 15 min at 277 K. The supernatant was filtered through a 0.22 µm filter and then injected onto a 5 ml HiTrap SP column (GE Healthcare) at a flow rate of 1 ml min^{-1}, washed with several column volumes of buffer *A* (50 m*M* HEPES pH 7.0, 1 m*M* EDTA, 1 m*M* β-mercaptoethanol) and eluted with a stepped gradient of buffer *B* (50 m*M* HEPES pH 7.0, 1 *M* NaCl, 1 m*M* EDTA, 1 m*M* β-mercaptoethanol) over 30 min. The fractions of interest were combined and concentrated using Vivaspin centrifugal concentrators (5000 molecular-weight cutoff) to a volume of approximately 0.5 ml and injected onto a HiLoad Superdex 75 16/60 Prep Grade column (GE Healthcare). The protein was eluted at a flow rate of 0.5 ml min^{-1} with an isocratic gradient of buffer *C*

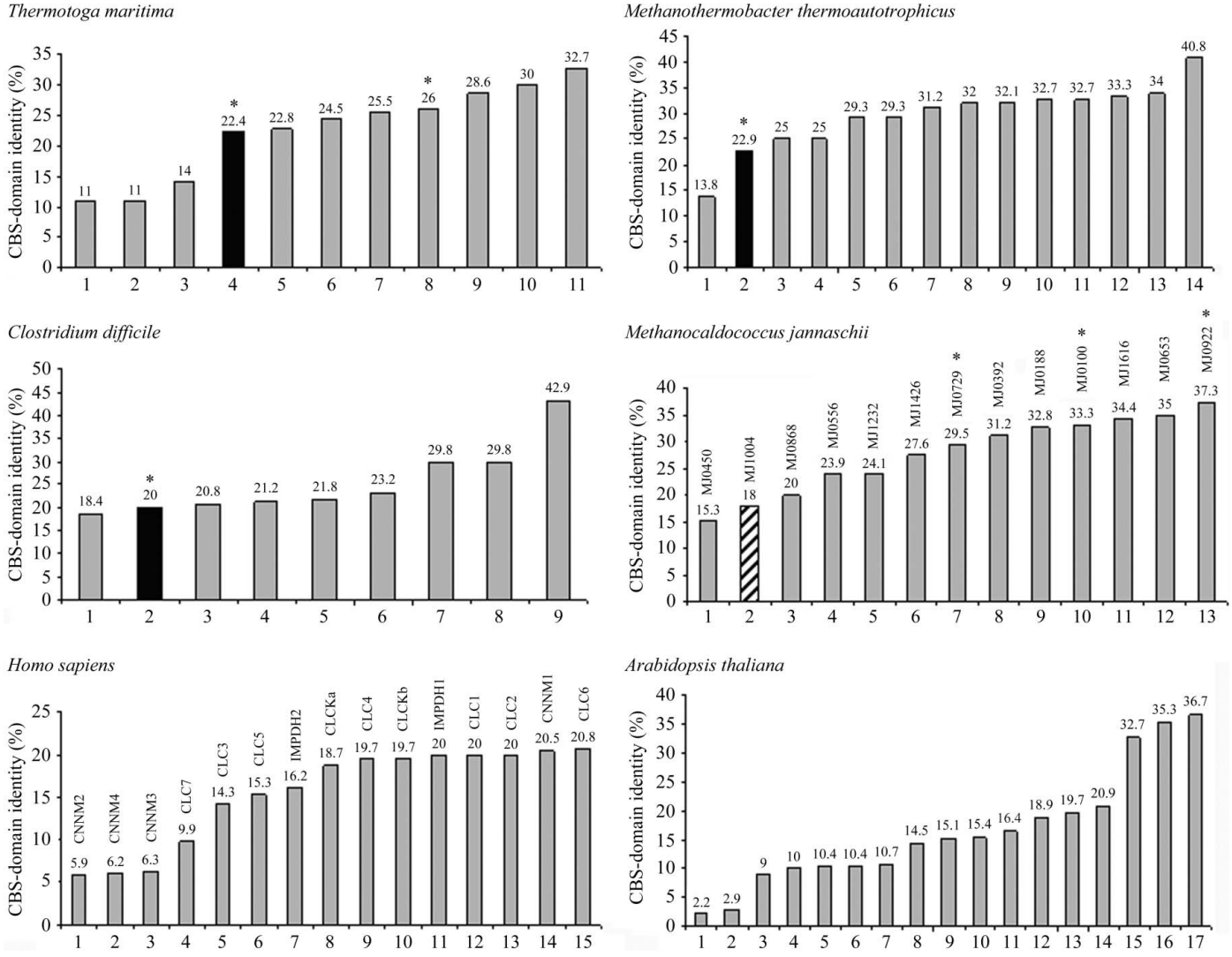


Figure 2
Sequence identity (%) observed between the two CBS subunits of each Bateman domain in proteins from different organisms. The sequences of the CBS motifs (for which the limits are those defined in the Pfam database; Finn *et al.*, 2010) were aligned with *EMBOSS* (Rice *et al.*, 2000). Proteins for which crystal structures are known are indicated by asterisks. Proteins adopting an HT orientation are indicated by black bars. The names of all *M. jannaschii* proteins have been added to help the reader. The numbering shown on the *x* axis corresponds to the following proteins (UniProt database codes). *Thermotoga maritima*: 1, UPI000016610F; 2, PPAC; 3, Q9WZU9; 4, Q9X033 (TM0935); 5, Q9X0P4; 6, Q9X0M4; 7, Q9WZH4; 8, Q9WZZ4 (TM0892); 9, Q9WZT6; 10, Q9X168; 11, Q9X175. *Methanothermobacter thermautotrophicus*: 1, O26717; 2, O27659; 3, O27294; 4, O27343; 5, Q9P9J1; 6, O26506; 7, O26907; 8, O27616; 9, O27912; 10, O26943; 11, O26835; 12, O26245; 13, O26229; 14, O26740. *Clostridium difficile*: 1, Q18A90; 2, Q183U2; 3, Q18D73; 4, Q185K9; 5, Q181Z7; 6, Q17ZY8; 7, Q182C2; 8, UPI0000D72419; 9, Q180W2. *Methanocaldococcus jannaschii*: 1, Q57892; 2, Q58410; 3, Q58278; 4, Q57976; 5, Q58629; 6, Q58821; 7, Q58139; 8, Q57837; 9, Q57647; 10, Q57564; 11, Q59011; 12, Q58069; 13, Q58332. *Homo sapiens*: 1, Q9H8M5; 2, Q6P4Q7; 3, Q8NE01; 4, P51798; 5, P51790; 6, P51795; 7, P12268; 8, P51800; 9, P51793; 10, P51801; 11, P20839; 12, P35523; 13, P51788; 14, Q9NRU3; 15, P51797. *Arabidopsis thaliana*: 1, Q8GZA4; 2, Q9SSD0; 3, CLCE; 4, CLCC; 5, CLCA; 6, Q9M1T1; 7, Q9LG18; 8, Q9LK65; 9, CLCG; 10, CLCB; 11, UMP3; 12, CLCD; 13, CLCF; 14, Q9XI37; 15, O23193; 16, Q9C5D0; 17, O49493.

(50 m*M* HEPES pH 7.0, 200 m*M* NaCl, 1 m*M* EDTA, 1 m*M* β-mercaptoethanol). Fractions that contained pure $MJ1004_{full}$ (or $MJ1004_{126}$) were pooled and concentrated using Vivaspin concentrators (5000 molecular-weight cutoff) to a final concentration of 80 mg ml^{-1} for crystallization trials. The concentrations of the proteins were estimated using the Bradford assay (Bradford, 1976). During the purification procedure, we observed that $MJ1004_{full}$ tended to aggregate in the absence of β-mercaptoethanol and/or NaCl. The oligomerization state and the homogeneity of the purified $MJ1004_{full}$ and $MJ1004_{126}$ proteins were analyzed using dynamic light scattering (DLS). All DLS experiments were carried out at 293 K at a laser wavelength of 620 nm using a DynaPro Titan (Wyatt Technology) and were analysed with the *DYNAMICS* v.6.7.7.9 software package. Protein samples were dialysed against 50 m*M* HEPES pH 7.0, 1 m*M* β-mercaptoethanol, 1 m*M* EDTA and filtered using a 0.1 µm filter (Whatman, Maidstone, England). Experiments were carried out using protein concentrations of 33 and 11 µ*M* for $MJ1004_{full}$ and $MJ1004_{126}$, respectively. Hydrodynamic radii of 2.9 and 2.4 nm were obtained for $MJ1004_{full}$ and $MJ1004_{126}$, respectively. The estimated molecular weights of these samples were 41 and 27 kDa, respectively, suggesting that a dimeric species is formed in both cases. The polydispersity was observed to be 13.9 and 13.7%, respectively.

SDS–PAGE gel bands containing the corresponding $MJ1004_{full}$ and $MJ1004_{126}$ constructs were subjected to mass-spectrometric analysis (as detailed in the supplementary material; Shevchenko *et al.*, 1996). Ultimately, $MJ1004_{full}$ and $MJ1004_{126}$ expressed in *E. coli* were wild type and did not contain any mutations in the amino-acid sequence according to the *M. jannaschii* genome-sequence database (http://cmr.jcvi.org/tigr-scripts/CMR/GenomePage.cgi?database=arg).

2.2. Crystallization of MJ1004

Dynamic light scattering indicated the presence of essentially monodisperse (<20% polydispersity) solutions of dimeric $MJ1004_{full}$ and $MJ1004_{126}$ species. Initial screening for crystals was performed with a variety of commercial screens (Crystal Screen, Crystal Screen 2, JCSG Core Suites I–IV, JCSG+ Suite, Index and PACT from Hampton Research, Molecular Dimensions and Qiagen) using the sitting-drop vapour-diffusion technique in 96-well plates at the High-Throughput Crystallization facility of CIC bioGUNE. The drops consisted of 0.1 µl protein solution at 80 mg ml^{-1} (in 100 m*M* HEPES pH 7.5, 1 m*M* β-mercaptoethanol, 1 m*M* EDTA) and 0.1 µl reservoir solution and were equilibrated over a reservoir volume of 70 µl at a constant temperature of 291 K. Initial experiments yielded needles of $MJ1004_{full}$ that grew in 15–20% ethanol, 100 m*M* HEPES pH 7.5 in 1 d. After optimization, crystals of $MJ1004_{full}$ with three different habits (bars, elongated prisms and flat square prisms) were obtained (Fig. 3). The $MJ004_{full}$ protein crystallized over a wide range of concentrations (from 65 to 90 mg ml^{-1}), with higher concentrations giving greater nucleation and numerous but smaller crystals. Very often several habits grew within the same crystallization drop, but all of them had the same unit-cell parameters. The best diffraction-quality crystals (elongated prisms) diffracted X-rays to 2.7 Å resolution and were grown at 295 K by mixing equal amounts (1 µl) of protein solution (65 mg ml^{-1} in 100 m*M* HEPES pH 7.5, 1 m*M* β-mercaptoethanol, 1 m*M* EDTA) and reservoir solution (100 m*M* HEPES pH 7.5 and 25% ethanol) in a 1:1 ratio; the crystals appeared in 1–2 h and grew to maximum dimensions of about 0.5 × 0.5 × 0.7 mm within 2 d. The addition of detergents and/or additives (using screens from Hampton Research) did not improve the quality of the crystals. Alternatively, we tested dehydration of the crystals by slowly increasing the ethanol concentration (in steps of 5%) in the reservoir solution over varying periods of time (from 1 h to 3 d) prior to flash-cooling in liquid nitrogen and/or the addition of cryoprotectant, with no further success. Finding optimal cryogenic conditions for flash-cooling $MJ004_{full}$ crystals was straightforward and did not require significant experimental effort. The best results were obtained using two different approaches. The first approach involved soaking the crystals in crystallization solution containing a final glycerol concentration of 25% and a slight increase in ethanol concentration to 30% prior to flash-cooling by direct immersion into liquid nitrogen at 93 K. The second procedure involved covering the crystallization drop with 5 µl paraffin oil, Paratone-N or DC200 silicone oil (Fluka) just after opening the cover slip of the crystallization well. This technique yielded optimal results and avoided crystal dehydration and fracture (Pflugrath, 2004). Once the crystallization drop had been covered with oil, the crystals were carefully moved from the inner aqueous solution to the surrounding silicone oil, in which they were stable. Water surrounding the crystal surface was then carefully removed by gently displacing the crystal within the silicone oil with the help of a loop. The crystals could then be flash-cooled by

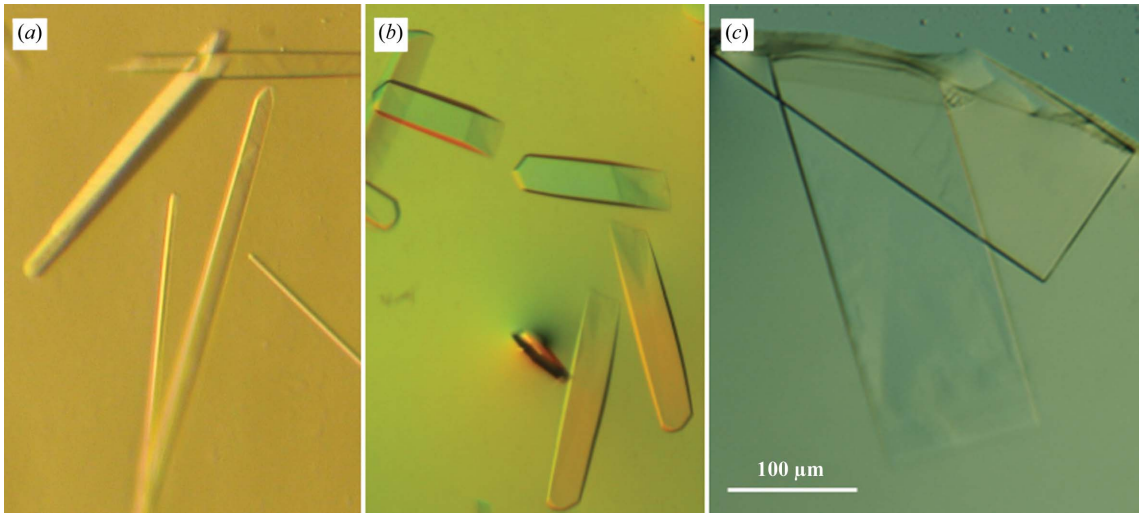


Figure 3
Crystals of $MJ1004_{full}$ showing different habits: (*a*) bars, (*b*) elongated prisms and (*c*) square plates. The highest resolution (2.7 Å) was obtained with the elongated prisms.

immersion in liquid nitrogen prior to data collection. To check whether the cryoprotectant (glycerol or silicone oil) was responsible for the limited resolution of the data sets, we analyzed the diffraction images both at room temperature and after freezing the crystals. Similar results were obtained, suggesting that the resolution limit was intrinsic to the crystals. Despite significant experimental effort, we could not obtain crystals of the $MJ1004_{126}$ construct.

2.3. Preliminary crystallographic analysis

Prior to data collection, the crystals were transferred to crystallization buffer containing 25% glycerol as a cryoprotectant for a few seconds before being flash-cooled by direct immersion in liquid nitrogen at 93 K. Crystals were mounted for X-ray data collection using either CryoLoops (Hampton Research) or MicroMounts (MiTeGen). Initial data sets at 4 Å resolution were collected in-house using a CCD detector mounted on a Microstar-H rotating-anode X-ray generator (Bruker) operated at 60 kV and 100 mA with Helios optics and a copper target (Cu $K\alpha$; λ = 1.542 Å). Further data collection was then carried out on beamline ID23.2 at the ESRF synchrotron (Grenoble, France). Diffraction data were processed using *HKL*-2000 (Otwinowski & Minor, 1997). Preliminary analysis of the data sets was performed using the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). Native crystals of $MJ1004_{full}$ diffracted to 2.7 Å resolution (Table 1; Fig. 4) and belonged to space group $P2_1$, with unit-cell parameters a = 54.4, b = 53.8, c = 82.6 Å, β = 106.1° The presence of one or two molecules within the asymmetric unit gives Matthews coefficients (Matthews, 1968) of 4.7 and 2.4 Å^3 Da^{-1}, respectively, and solvent contents of 74 and 48%, respectively. We estimated that two molecules per asymmetric unit was the most probable value for the $MJ1004_{full}$ crystals. It is likely that these two molecules correspond to a dimeric species of MJ1004 and form a CBS module (Lucas *et al.*, 2010; Gómez-García *et al.*, 2010). Data-collection statistics are summarized in Table 1. Structural determination of the MJ1004 model is currently in progress.

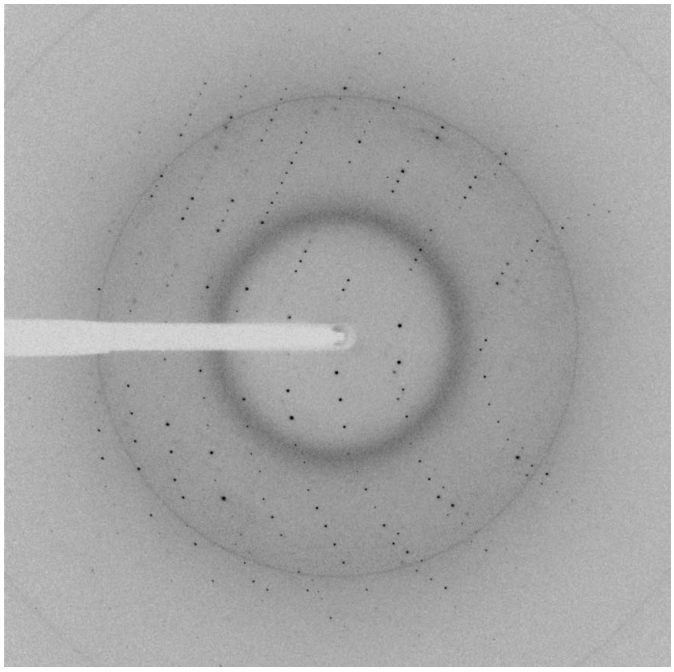


Figure 4
Representative X-ray diffraction image from a full-length native $MJ1004_{full}$ crystal. The dimensions of the native crystals were 0.4 × 0.5 × 0.2 mm and it was exposed for 1 s over a 1° oscillation range. The edge of the detector corresponds to a resolution of 1.7 Å. The diffraction ring observed at around 7–8 Å is caused by the presence of silicone oil, which was used as a cryoprotectant in freezing the native crystals.

We thank Professor Sung-Hou Kim, University of California at Berkeley, USA for providing us with genomic DNA from *M. jannaschii* and the staff of ESRF beamline ID23.2 for support during synchrotron data collection. We also thank Dr Felix Elortza of the Proteomics Service at CIC bioGUNE for mass-spectrometric analysis and Dr Adriana Rojas for maintenance of the in-house X-ray equipment. This work was supported by grants from Departamento de Educación, Universidades e Investigación del Gobierno Vasco (PI2010-17), the Basque Government (ETORTEK IE05-147, IE07-202), Diputación Foral de Bizkaia (Exp. 7/13/08/2006/11 and 7/13/08/2005/14), the Spanish Ministerio de Ciencia e Innovación (MICINN; BFU2010-17857) and the MICINN CONSOLIDER-INGENIO 2010 Program (CSD2008-00005).

References

Aguado-Llera, D., Oyenarte, I., Martínez-Cruz, L. A. & Neira, J. L. (2010). *FEBS Lett.* **584**, 4485–4489.
Amodeo, G. A., Rudolph, M. J. & Tong, L. (2007). *Nature (London)*, **449**, 492–495.
Bateman, A. (1997). *Trends Biochem. Sci.* **22**, 12–13.
Biemans-Oldehinkel, E., Mahmood, N. A. & Poolman, B. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 10624–10629.
Bowne, S. J., Sullivan, L. S., Blanton, S. H., Cepko, C. L., Blackshaw, S., Birch, D. G., Hughbanks-Wheaton, D., Heckenlively, J. R. & Daiger, S. P. (2002). *Hum. Mol. Genet.* **11**, 559–568.
Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
Bult, C. J. *et al.* (1996). *Science*, **273**, 1058–1073.
Carr, G., Simmons, N. & Sayer, J. (2003). *Biochem. Biophys. Res. Commun.* **310**, 600–605.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Day, P., Sharff, A., Parra, L., Cleasby, A., Williams, M., Hörer, S., Nar, H., Redemann, N., Tickle, I. & Yon, J. (2007). *Acta Cryst.* D**63**, 587–596.
De Angeli, A., Moran, O., Wege, S., Filleur, S., Ephritikhine, G., Thomine, S., Barbier-Brygoo, H. & Gambale, F. (2009). *J. Biol. Chem.* **284**, 26526–26532.
Estévez, R., Schroeder, B. C., Accardi, A., Jentsch, T. J. & Pusch, M. (2003). *Neuron*, **38**, 47–59.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Fernández-Millán, P., Kortazar, D., Lucas, M., Martínez-Chantar, M. L., Astigarraga, E., Fernández, J. A., Sabas, O., Albert, A., Mato, J. M. & Martínez-Cruz, L. A. (2008). *Acta Cryst.* F**64**, 605–609.
Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunesekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L., Eddy, S. R. & Bateman, A. (2010). *Nucleic Acids Res.* **38**, D211–D222.
Gómez García, I., Kortázar, D., Oyenarte, I., Mato, J. M., Martínez-Chantar, M. L. & Martínez-Cruz, L. A. (2009). *Acta Cryst.* F**65**, 813–817.
Gómez-García, I., Oyenarte, I. & Martínez-Cruz, L. A. (2010). *J. Mol. Biol.* **399**, 53–70.
Grunberg-Manago, M. (1999). *Annu. Rev. Genet.* **33**, 193–227.
Hattori, M., Tanaka, Y., Fukai, S., Ishitani, R. & Nureki, O. (2007). *Nature (London)*, **448**, 1072–1075.
Ishitani, R., Sugita, Y., Dohmae, N., Furuya, N., Hattori, M. & Nureki, O. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 15393–15398.
Jämsen, J., Baykov, A. A. & Lahti, R. (2010). *Biochemistry*, **49**, 1005–1013.
Kemp, B. E. (2004). *J. Clin. Invest.* **113**, 182–184.
Koutmos, M., Kabil, O., Smith, J. L. & Banerjee, R. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 20958–20963.
Lucas, M., Encinar, J. A., Arribas, E. A., Oyenarte, I., García, I. G., Kortazar, D., Fernández, J. A., Mato, J. M., Martínez-Chantar, M. L. & Martínez-Cruz, L. A. (2010). *J. Mol. Biol.* **396**, 800–820.
Lucas, M., Kortazar, D., Astigarraga, E., Fernández, J. A., Mato, J. M., Martínez-Chantar, M. L. & Martínez-Cruz, L. A. (2008). *Acta Cryst.* F**64**, 936–941.
Mahmood, N. A., Biemans-Oldehinkel, E. & Poolman, B. (2009). *J. Biol. Chem.* **284**, 14368–14376.

Martínez-Cruz, L. A., Encinar, J. A., Kortazar, D., Prieto, J., Gómez, J., Fernández-Millán, P., Lucas, M., Arribas, E. A., Fernández, J. A., Martínez-Chantar, M. L., Mato, J. M. & Neira, J. L. (2009). *Biochemistry*, **48**, 2760–2776.
Martínez-Cruz, L. A., Encinar, J. A., Sevilla, P., Oyenarte, I., Gómez-García, I., Aguado-Llera, D., García-Blanco, F., Gómez, J. & Neira, J. L. (2011). *Protein Eng. Des. Sel.* **24**, 161–169.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
McKusick, V. A. (1998). *Mendelian Inheritance in Man. A Catalog of Human Genes and Genetic Disorders*, 12th ed. Baltimore: Johns Hopkins University Press.
Miller, M. D. *et al.* (2004). *Proteins*, **57**, 213–217.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pflugrath, J. W. (2004). *Methods*, **34**, 415–423.
Rice, P., Longden, I. & Bleasby, A. (2000). *Trends Genet.* **16**, 276–277.
Rudolph, M. J., Amodeo, G. A., Iram, S. H., Hong, S. P., Pirino, G., Carlson, M. & Tong, L. (2007). *Structure*, **15**, 65–74.
Scott, J. W., Hawley, S. A., Green, K. A., Anis, M., Stewart, G., Scullion, G. A., Norman, D. G. & Hardie, D. G. (2004). *J. Clin. Invest.* **113**, 274–284.
Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. (1996). *Anal. Chem.* **68**, 850–858.
Slabinski, L., Jaroszewski, L., Rychlewski, L., Wilson, I. A., Lesley, S. A. & Godzik, A. (2007). *Bioinformatics*, **23**, 3403–3405.
Studier, F. W. & Moffatt, B. A. (1986). *J. Mol. Biol.* **189**, 113–130.
Tuominen, H., Salminen, A., Oksanen, E., Jämsen, J., Heikkilä, O., Lehtiö, L., Magretova, N. N., Goldman, A., Baykov, A. A. & Lahti, R. (2010). *J. Mol. Biol.* **398**, 400–413.
Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.
Xiao, B., Heath, R., Saiu, P., Leiper, F. C., Leone, P., Jing, C., Walker, P. A., Haire, L., Eccleston, J. F., Davis, C. T., Martin, S. R., Carling, D. & Gamblin, S. J. (2007). *Nature (London)*, **449**, 496–500.
Zhang, R., Evans, G., Rotella, F. J., Westbrook, E. M., Beno, D., Huberman, E., Joachimiak, A. & Collart, F. R. (1999). *Biochemistry*, **38**, 4691–4700.