

Binding of S-Methyl-5'-Thioadenosine and S-Adenosyl-L-Methionine to Protein MJ0100 Triggers an Open-to-Closed Conformational Change in Its CBS Motif Pair

María Lucas¹, José Antonio Encinar², Egoitz Astigarraga Arribas³, Iker Oyenarte¹, Inmaculada Gómez García¹, Danel Kortazar¹, José A. Fernández³, Jose M. Mato⁴, María Luz Martínez-Chantar⁴ and Luis Alfonso Martínez-Cruz^{1*}

¹Unidad de Biología Estructural, CIC bioGUNE, Parque Tecnológico de Bizkaia, Ed. 800, 48160-Derio, Bizkaia, Spain

²Instituto de Biología Molecular y Celular, Universidad Miguel Hernández, Avda. del Ferrocarril s/n, 03202 Elche (Alicante), Spain

³Departamento de Química-Física, Universidad del País Vasco UPV-EHU, Lejona, Bizkaia, Spain

⁴Unidad de Metabolómica, CIC bioGUNE, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (Ciberedh), Parque Tecnológico de Vizcaya, Ed. 801, 48160-Derio, Bizkaia, Spain

Received 16 November 2009;
received in revised form
9 December 2009;
accepted 9 December 2009
Available online
21 December 2009

Edited by I. Wilson

Cystathionine β -synthase (CBS) domains are small motifs that are present in proteins with completely different functions. Several genetic diseases in humans have been associated with mutations in their sequence, which has made them promising targets for rational drug design. The protein MJ0100 from *Methanocaldococcus jannaschii* includes a DUF39 domain of so far unknown function and a CBS domain pair (Bateman domain) at its C-terminus. This work presents the crystallographic analysis of four different states of the CBS motif pair of MJ0100 in complex with different numbers of S-adenosyl-L-methionine (SAM) and S-methyl-5'-thioadenosine (MTA) ligands, providing evidence that ligand-induced conformational reorganization of Bateman domain dimers could be an important regulatory mechanism. These observations are in contrast to what is known from most of the other Bateman domain structures but are supported by recent studies on the magnesium transporter MgtE. Our structures represent the first example of a CBS domain protein complexed with SAM and/or MTA and might provide a structural basis for understanding the molecular mechanisms regulated by SAM upon binding to the C-terminal domain of human CBS, whose structure remains unknown.

© 2009 Elsevier Ltd. All rights reserved.

Keywords: MJ0100; CBS domain; S-adenosylmethionine; methylthioadenosine

*Corresponding author. E-mail address: amartinez@cicbiogune.es.

Abbreviations used: CBS, cystathionine β -synthase; AMPK, 5'-AMP-activated protein kinase; SAM, S-adenosyl-L-methionine; MTA, S-methyl-5'-thioadenosine; GnCl, guanidinium chloride; 3D, three-dimensional; Se-Met, selenomethionine; PDB, Protein Data Bank; ESRF, European Synchrotron Radiation Facility.

Introduction

The cystathionine β -synthase (CBS) domain proteins comprise a large superfamily of evolutionarily conserved proteins that are present in all kingdoms of life.¹ CBS domains are 60-residue-long motifs that were originally discovered in the enzyme CBS and consist of a three-stranded β -sheet and two α -helices packed according to the sequence β - α - β - β - α .^{2,3}

CBS domains regulate enzyme activity based on the concentration of AMP/ATP or other adenosine derivatives, and their importance is underlined by the range of hereditary diseases in humans associated with mutations in their sequence.^{4–11} Thus, they are promising targets for the development of novel drugs.¹² Comparison among CBS domains of iso-functional proteins from different species shows a highly conserved fold despite the low degree of sequence similarity.^{3,13} CBS domains usually occur in tandem pairs, forming a so-called CBS pair or Bateman module¹ in which both CBS subunits are related by a pseudo-2-fold symmetry axis running parallel with the central β -sheets. Some proteins such as 5'-AMP-activated protein kinase (AMPK) and its homologs have revealed tetra-repeat units,¹² but even hepta-repeats have been detected in some genomes.¹⁴ Nowadays, the crystal structures of several tandem CBS domains have been elucidated.^{15–24} In all these structures, both head-to-head and head-to-tail associations of the Bateman modules have been observed,²¹ though the forces directing their final assembly are not known. These assemblies include a total of four CBS domains per structural module and have been baptized as “CBS modules”.²⁵

The cleft between two CBS domains of a Bateman module is the binding site for the adenosyl group.^{18–20,22,23} Interestingly, the crystal structures of the complexes CIC5-ADP/ATP¹⁶ and AMPK-ATP/AMP/ADP/5-aminoimidazole-4-carboxamide ribonucleoside monophosphate^{18–20,22,23} raise the question of whether a CBS pair has one or two nucleotide binding sites, since two symmetry-related cavities are formed at both sides of the central β -sheets.²¹

But what is the effect exerted by a ligand when it binds to a CBS domain? Although there is experimental evidence that the binding of adenosyl groups to the CBS pairs of inosine monophosphate dehydrogenase,¹¹ AMPK,²⁶ and CBS²⁷ has an influence in the catalysis of these proteins, the structural information available does not elucidate how the adenosyl compound binding is transduced to the corresponding catalytic domains. Taking into account the high number of motifs to which CBS domains can be fused (168 different architectures annotated in the Pfam database), it seems reasonable to think that the physiological functions and binding partners of these motifs may vary considerably between different proteins.²⁸ Several studies have revealed a first glimpse of the structural basis of ligand binding by eukaryotic CBS domains, a subject currently attracting considerable interest.^{16,19,20,22,23} In contrast to initial assumptions, recent studies on the transporter MgtE revealed that CBS domain proteins can undergo large conformational changes in the presence of metal ions.^{24,29} The available information seems to indicate that the physicochemical properties of specific ligands are tied to the type of response that these molecules induce in the proteins interacting with them. To gain further insight into the structural features involved in the recognition of new ligands by CBS domain proteins

and the mechanisms by which these proteins modulate the function of other domains, we sought to identify new putative binding sites that might potentially recognize nucleotides different from AMP/ADP/ATP, which have already been studied extensively by other authors.^{14,18,19,22} CBS motifs are unusually abundant in archaea.³⁰ Therefore, organisms such as the hyperthermophile *Methanocaldococcus jannaschii*³¹ offer excellent models for the characterization of novel adenosyl binding sites. The genome of *M. jannaschii* encodes 15 CBS domain proteins†, which differ significantly in their composition and presumably in their abilities to bind to different ligands.

The open reading frame of gene *mj0100* from *M. jannaschii* (UniProtKB/Swiss-Prot entry Q57564) encodes a polypeptide chain of 509 amino acids with a molecular mass of 56,458 Da. Its sequence contains two domains: a DUF39 (Pfam database PF01837) domain (residues 15–320) of so far unknown function and a CBS domain pair (CBS1, residues 392–445; CBS2, residues 458–506). Although no ready hypothesis for function can be assigned, MJ0100 homologs share the same domain composition and are found primarily in Euryarchaeota, and most represented in methanogens. MJ0100 homologs are found in what may be a transcriptional unit with a gene encoding a probable ferredoxin with two Fe₄S₄ signatures including proteins such as MJ0099 from *M. jannaschii*, AF1185 from *Archaeoglobus fulgidus*, MTH854 from *Methanothermobacter thermautotrophicus*, or MmarC7_0603 from *Methanococcus maripaludis* among others. This puts MJ0100 very likely to be an oxidoreductase of some form as methanogenic Archaea, like many bacteria, often group genes with dedicated metabolic functions.

We recently reported the crystallization and the preliminary crystallographic analysis of the CBS domain pair of MJ0100.³² Here, we describe the crystal structure of four different states of the CBS motif pair, in complex with different numbers of *S*-adenosyl-L-methionine (SAM) and *S*-methyl-5'-thioadenosine (MTA) ligands. Our study reveals that each ligand binds at a different site, triggering an “open-to-closed” conformational change that, to our knowledge, represents the first example of a “dynamic” response of a CBS module upon binding of adenosyl groups and reveals two novel SAM- and MTA-binding motifs that should assist in structure-based design of compounds binding at these sites.

Results

Target selection

We set out to compare and contrast the predicted protein sequences of *M. jannaschii* against all known CBS domain protein–ligand complexes from differ-

† www.tigr.org

ent organisms with the aim of identifying novel binding motifs. To address this issue, we first performed multiple sequence alignments³³ and three-dimensional (3D) theoretical models^{34,35} using the available information on public databases. Subsequently, we focused our attention on those amino acid residues that potentially constitute the ligand binding cavities of each protein (Fig. 1) and divided them in three structural blocks: (i) residues from the flexible loop preceding the first strand of the β -sheet of each CBS domain (β 1 and β 5 in Fig. 2), (ii) residues from the second strand of the β -sheet (β 2 and β 6 in Fig. 2), and (iii) residues located at the third strand of each β -sheet (β 7 and β 3 in Fig. 2) and the first two turns of the following helix (H4 and H2 in Fig. 2 for CBS1 and CBS2, respectively). As shown in Fig. 1, the highest diversity was found in the first block. The second group contained a conserved proline residue and maintained the hydrophobic environment that stabilizes the adenine ring of the adenosyl group, but the third block turned out to be the most conserved among the proteins studied. This latter section shelters the recognition motif (G-h-x-S/T-x-S/T-D) for the ribose-phosphate moiety of molecules such as AMP in AMPK,^{18,19} Ta0289 from *Thermoplasma acidophilum*, or PF1953 from *Pyrococcus furiosus*.³⁶ In agreement with previous observations on mammalian AMPK,²³ those binding sites lacking the conserved aspartate that H-bonds the hydroxyls of the ribose ring of the nucleotide (i.e., AMPK_cbs12_site1, PH1780_cbs12_site2, or MJ0868_site2 in Fig. 1) were not included in the list of potential adenosyl-binding sites. Finally, we grouped the putative binding sites according to their plausible chance to bind phosphate- or non-phosphate-containing nucleotides by comparing the residues structurally equivalent to those forming the AMP-binding motif. As shown in Fig. 1, in MJ0100, the Ser/Thr residues preceding the last aspartate of the AMP-binding motif¹⁸ are substituted by a negatively charged (E499) and a bulky hydrophobic residue (W438), respectively, which would interfere with phosphate binding in both sites of the protein, suggesting that they might be more compatible with other adenosine derivatives. Accordingly, MJ0100 was considered as a promising target that potentially might bind nucleotides that do not contain phosphate groups.

Construct design

Previous studies suggested that the full-length MJ0100 has a strong tendency to aggregate, making physical studies of the protein very difficult.³² Alternatively, and based on sequence alignments and on published structures of CBS domain pairs,^{17,32,36,37} a recombinant truncated form of MJ0100 comprising amino acid residues 381–509 (hereafter denoted as C381) was expressed and purified to near homogeneity. This protein, missing ~380 N-terminal residues, forms dimers and does not exhibit the aggregating properties of the full-length species.³² Using this construct, we were able

to obtain the first native and selenomethionine (Se-Met)-labeled protein crystals of C381 belonging to space group $P2_12_12_1$ and diffracting to 3.1 and 3.5 Å, respectively.³² In order to investigate whether we could improve the quality and stability of the crystals, we designed two new constructs, namely, C386 (residues 386–509) and C388 (residues 388–509), using the 3D structure of C381 as a template. We concluded that constructs shorter than C388 would have probably destabilized the resulting species since they would eliminate the interactions existing between the N-terminus and the loop between strand β 5 and helix H3 (see Fig. 2). This was later confirmed by our crystal structures, since these interactions are mainly represented by hydrogen bonds between the main-chain carbonyl oxygen of residues T388, E464, and E466 and the NH peptide of V390, K391, and V468. Additionally, a salt link is formed between K391 (in helix HA) and D465.

Overall structure

The overall structure of the three protein constructs C381, C386, and C388 (hereafter denoted in general as MJ0100c) shows the expected fold for a CBS domain pair (Fig. 2). It comprises two tandemly repeated CBS motifs, which are hereafter denoted as the CBS1 (residues 394–450) and CBS2 (455–509) domains, respectively. CBS1 is made up of three α -helices and a β -sheet composed of four strands of topology 1 \uparrow 2 \uparrow 3 \downarrow 4 \uparrow . Similarly, CBS2 has three α -helices and a β -sheet with three strands 5 \uparrow 6 \uparrow 7 \downarrow . An additional β -strand (β 1*) formed by residues 395–396 is observed in the MJ0100c structure when a ligand molecule occupies the binding cavity (site A) located underneath. The N-terminus of MJ0100c has a short relatively unstructured segment that precedes a single turn of α -helix (denoted as HA) that can be considered to be the beginning of the CBS1 domain structure (Fig. 2). Similarly, a turn related by pseudo-2-fold symmetry (HB) constitutes the beginning of domain CBS2. The two CBS domains are related by approximate 2-fold symmetry and share 34.5% identity over 54 residues where C $^\alpha$ atoms can be superimposed with a 0.321-Å root-mean-square deviation (rmsd). Two clefts measuring approximately 15 Å \times 10 Å \times 8 Å and related by a dyad axis running parallel with the β -sheets constitute the two putative ligand binding regions within each monomer (Fig. 2). These clefts are formed by the β -sheets of both CBS1 and CBS2 domains and are composed of hydrophobic and charged residues. The first cleft (hereafter denoted as site A) is formed by β 2 and β 7, the loop between HA and β 1, the N-terminus of helix H4, and the C-terminus of H1 (Fig. 2). The second cavity (denoted as site B) is formed by β 3 and β 6, the loop between β 4 and β 5, the N-terminus of helix H2, and the C-terminus of H3 (Fig. 2). Within the monomer, ligand binding cavities are not connected, but they are separated by the walls formed by amino acid residues 420–423 and 495–498 in site A and 434–437 and 481–484 in site B. The final

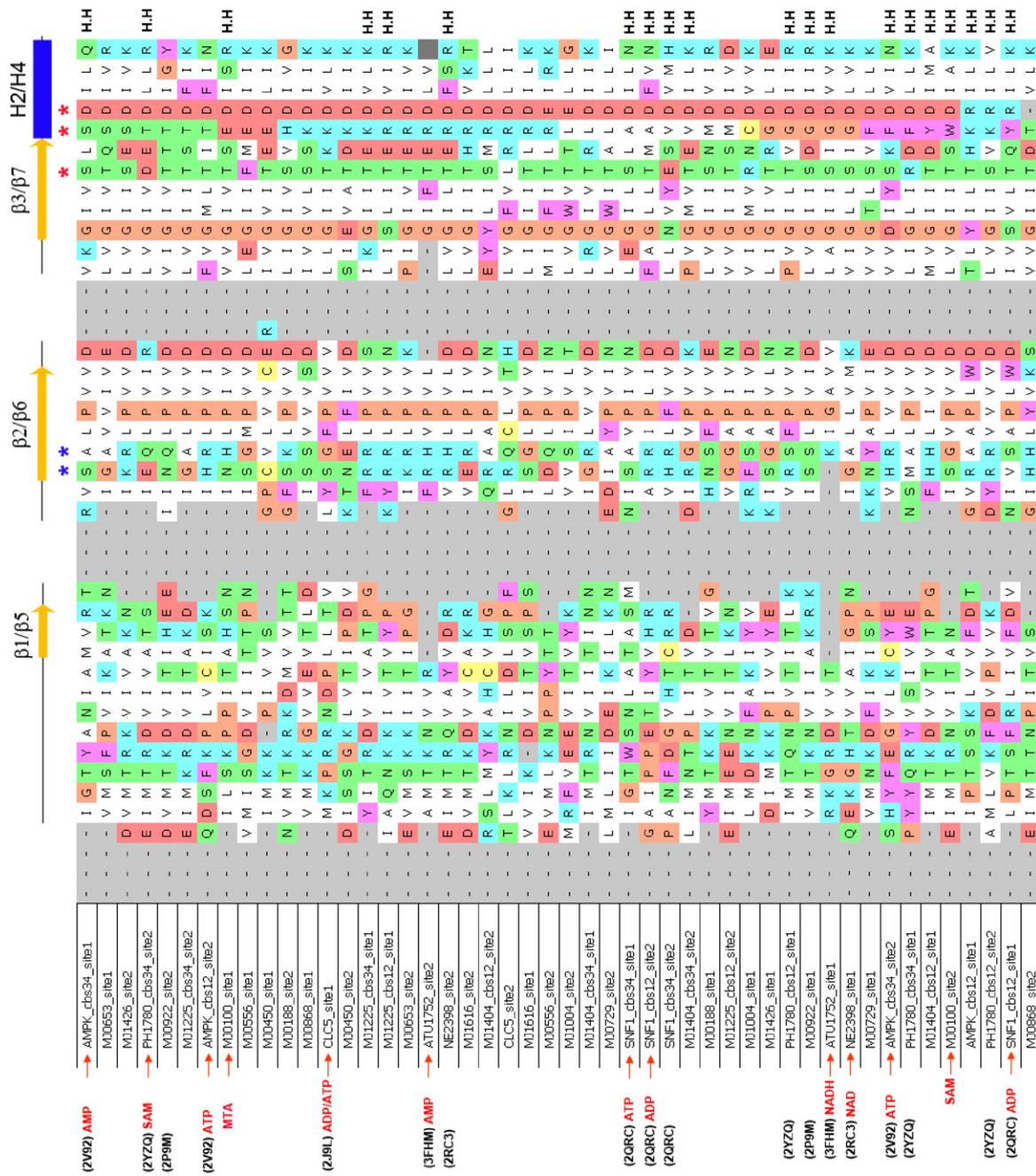


Fig. 1 (legend on next page)

models of the protein chains begin at residues 386–388 and end between residues 504 and 506. The monomers in these crystals have roughly the same conformation, with an rms distance that varies from 0.34 to 0.88 for their equivalent C α atoms, being higher when a monomer with occupied binding sites is compared with a subunit in which the ligand cavities are empty.

Oligomeric state of the protein

The asymmetric units of the six crystal habits studied ($P2_12_12_1$, $P2_12_12$, $P1$, $P6_1$, $C222_1$, and $C222$) contain 10, 4, 4, 2, 2, and 1 molecules, respectively (Table 1). Our crystallographic analyses revealed that in all the cases, two MJ0100c subunits associate to form parallel (“head-to-head”) dimers, which are elliptical disks with an aqueous pore in the center (Fig. 3). These dimers were interpreted as the biological unit, due to the large interfacial surface area between both monomers, which is in agreement with previous gel-filtration and dynamic light-scattering studies on MJ0100c.³² However, in contrast with previous reports on other CBS domain proteins that showed a unique arrangement, four different conformations were found for MJ0100c, which basically differ in the angle adopted between the two associated monomers (A and B) (see Fig. 3). The first type of dimer (hereafter denoted as “Open-1” or OP1) adopts an open disk-like conformation, in which the angle between A and B is approximately 23°. OP1 is represented by the C388 structure determined from $C222_1$ crystals and by one of the two dimers found in the asymmetric unit of the $P1$ species (rmsd = 0.314 Å) (Fig. 3). The dimer interface in OP1 is the smallest (789 Å²) among the four different MJ0100c oligomers and comprises amino acid residues of the helices H1, H2, H1*, and H2* of the CBS1 and CBS1* subdomains. OP1 is mainly stabilized by hydrophobic interactions contributed by aliphatic residues I407, M408, A411, and I415 (from helix H1) and I440, A441, L444, and A445 (from helix H2) of complementary monomers and by hydrogen bonds existing between S437 and S437* and between the main-chain carbonyl oxygen of A445 and the amino group of K412. Difference density could be clearly seen and unambiguously interpreted for the adenosyl moiety of a unique molecule of SAM located in site B of one of the monomers in both crystal forms (Fig. 4; Table 2). An additional unidentified density corresponding to a much smaller molecule was also detected in site B* of the second monomer. The site A and A* cavities

were found empty in OP1. The second type of association (OP2) also adopts an open disk-like conformation but shows a smaller angle (12°) between subunits (Fig. 3). OP2 is represented by the second dimer of C388 found in the asymmetric unit of the $P1$ crystals. As in OP1, the dimer interface (833 Å²) comprises only amino acid residues of helices H1 and H2 of symmetric CBS1 subdomains. Interestingly, two molecules of SAM (instead of one) were found, one in site B of monomer A and the other in site B* of monomer B, respectively, while sites A and A* remained empty (Fig. 3; Table 2). The third conformation (OP3), represented by constructs C388 and C386 (Table 1), crystallizes in space groups $P6_1$ and $C222_1$ and consists of a more compact (still open) disk in which the angle between subunits is reduced to 9° (Fig. 3). The dimer interface (1045 Å²) is bigger than that in OP1 and OP2, and new direct or water-mediated H-bond interactions are formed involving residues N479, S473, S502, R503, and E499. In the $P6_1$ crystals, two molecules of SAM were found in sites B and B*, respectively, with sites A and A* being empty (Fig. 3; Table 2), whereas the $C222_1$ species showed four molecules of MTA in sites A, A*, B, and B*. Finally, the fourth conformation (CL1, “closed-1”), which was independently adopted by the three protein constructs, crystallizes in space groups $P2_12_12_1$, $C222$, $P2_12_12$, and $C222_1$ and can be described as an elliptical closed disk-like parallel dimer that is approximately 55 Å tall, 45 Å wide, and 25 Å thick (Fig. 3; Table 1). The overall structural features of CL1 are highly similar to those previously reported for other “closed” head-to-head dimers such as PAE2072,¹⁴ TA0289,³⁶ or the hypothetical protein Ykul [Protein Data Bank (PDB) code: 1yav]. In CL1, the extensive dimer interface (1518 Å²) involves exposed hydrophobic residues on helices H1 and H2 of subdomain CBS1 packing against their complementary partners of CBS1* and *vice versa*. In contrast with the OP1, OP2, and OP3 conformations, in CL1, the exposed hydrophilic residues D469, S476, and N479 of helix H3 and E499, S502, R503, and K509 of helix H4 of CBS2 interact with their complementary residues of CBS2* and *vice versa*. This indicates that approximately 1450 Å² are buried when the MJ0100c dimer evolves from the fully open to the closed conformation. In CL1, helices H3 and H3* are sandwiched between helices H3*–H4* and H3–H4 of the complementary monomer, respectively. This allows the side chain of residue F505 to be buried in the hydrophobic environment provided

Fig. 1. Sequence alignment of residues constituting the potential binding sites of the *M. jannaschii* CBS proteins against those found in protein–ligand complexes from different organisms. Assuming MJ0100c as reference, the binding cavities can be described as formed by three blocks of residues: (i) the loop following HA + β 1-strand in site A (loop following HB + β 5-strand in site B), (ii) strand β 2 or β 6 in sites A and B, respectively, and (iii) strand β 7 in site A (or β 3 in site B) + helix H4 in site A (H2 in site B). The PDB code for each entry is on the left. For known protein–ligand complexes, the corresponding ligands bound at each site are also indicated. The dimerization orientation of the subunits [head-to-head (H.H) or head-to-tail (H.T)] is indicated on the right. Residues involved in interactions with the phosphate moiety of AMP and included in the AMP-binding motif (Gly-*h-x*-Ser/Thr-*x*-Ser/Thr-Asp)¹⁸ are marked with a red asterisk. Additional residues interacting with the phosphate group and not included in this motif are marked with a blue asterisk.

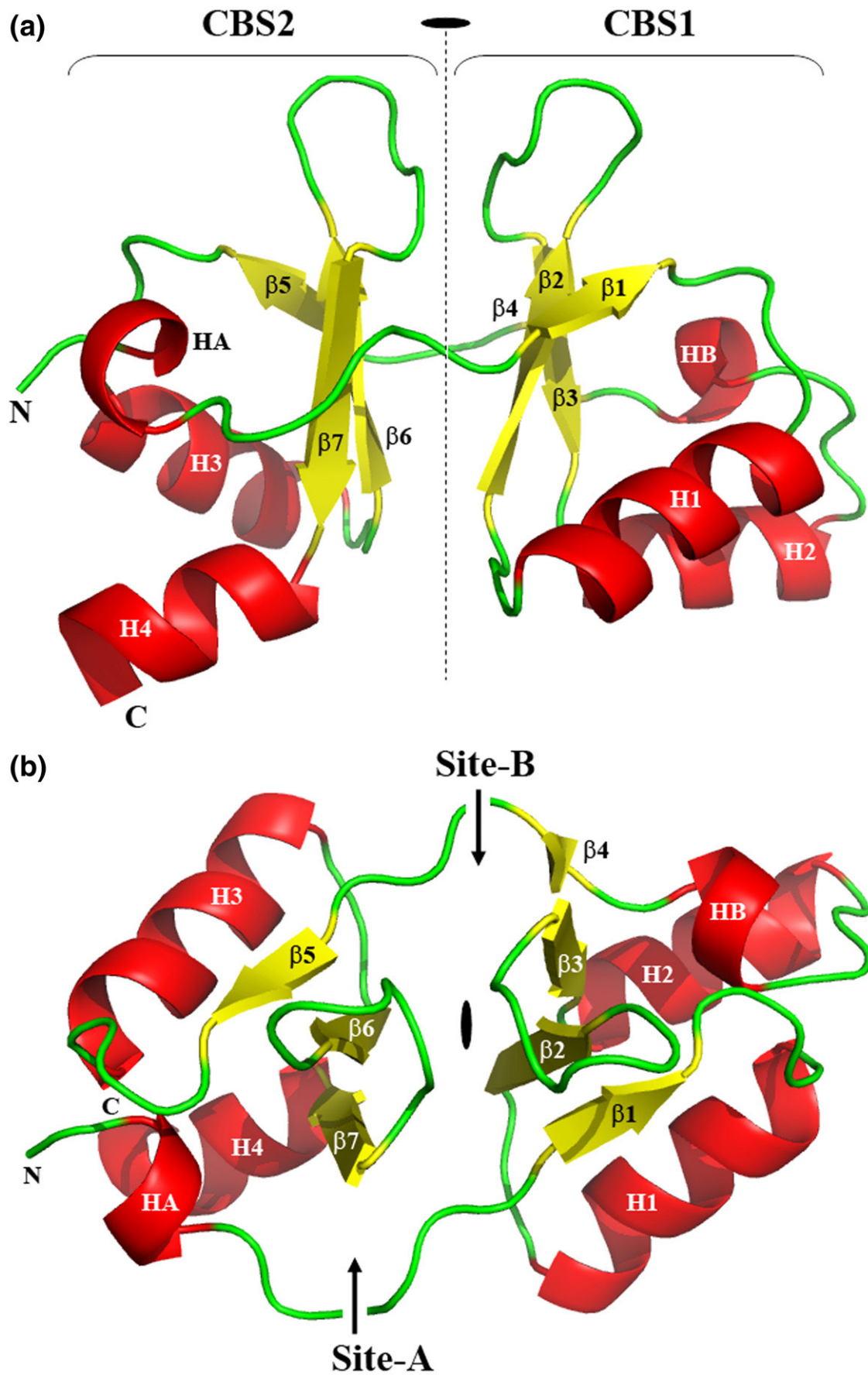


Table 1. Native MJ0100c: statistics for data collection and refinement

Dimer conformer	OP1	OP1, OP2	OP3	OP3, CL1	CL1		
<i>Data collection</i>							
Space group	C222 ₁	P1	P6 ₁	C222 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2	C222
Protein construct	C388	C388	C386	C388	C381	C386	C386
Unit cell parameters							
<i>a</i> (Å)	57.69	46.99	114.65	58.671	80.95	101.66	50.96
<i>b</i> (Å)	83.59	49.70	114.65	165.794	118.93	64.53	71.48
<i>c</i> (Å)	106.40	59.12	45.02	122.997	173.33	86.31	86.37
α (°)	90	80.11	90	90	90	101.661	90
β (°)	90	75.85	90	90	90	64.527	90
γ (°)	90	64.04	120	90	90	86.311	90
Cell volume (Å ³)	513,081.3	121,513.2	512,503.6	1,196,428.7	1,677,463.1	566,191.4	314,637.7
Molecules per asymmetric unit	2	4	2	4	10	4	1
Resolution (Å)	50–2.6	50–1.6	50–2.6	50–2.9	50–3.1	50–1.95	50–1.8
R_{sym}^a (%)	9.2 (22.3)	5.6 (31.2)	12.5 (48.7)	9.6 (47.4)	7.4 (28.5)	11.8 (41.0)	7.1 (27.5)
Mean <i>I</i> / σ <i>I</i>	27.2 (5.1)	21.8 (2.6)	21.4 (3.3)	29.8 (6.2)	18.43 (5.08)	17.1 (3.4)	30.8 (6.1)
Completeness (%)	91.9 (68.9)	95.6 (78.4)	98.2 (87.6)	99.9 (100)	98.2 (99.1)	99.9 (98.8)	98.9 (92.5)
Redundancy	8.2 (5.4)	3.2 (2.3)	11.7 (9.2)	13.8 (13.0)	6.6 (5.7)	7.1 (6.0)	9.4 (6.2)
Radiation wavelength (Å)	0.8726	0.8726	0.8726	0.9790	0.9340	0.8726	0.8726
Mosaicity (°)	0.42	0.38	0.86	0.85	0.65	0.44	0.27
<i>Refinement</i>							
Number of reflections/Unique	229,790/ 8425	361,355/ 61,168	351,592/ 10,654	532,547/ 13,555	201,817/ 31,157	507,094/ 42,417	244,152/ 15,273
Number of working reflections	7364	55,517	9961	12,845	28,783	40,171	14,353
Number of test reflections	355	2796	500	670	1531	2323	752
R_{work}^b (%) / R_{free}^c (%)	22.1/25.9	20.2/22.9	23.1/29.1	23.4/33.4	25.1/34.3	21.5/26.9	21.6/22.9
Number of atoms							
Protein	1871	3732	1890	3733	9629	3803	973
Ligand (SAM/MTA)	54/N/A	135/N/A	54/N/A	54/100	243/40	108/80	27/20
Glycerol	6	30	N/A	N/A	N/A	12	N/A
Water	83	534	43	N/A	N/A	375	172
Average <i>B</i> -factors (Å ²)							
Protein	40.7	21.1	58.6	58.9	94.5	22.5	20.2
Ligand (SAM/MTA)	35.3/N/A	23.3/N/A	62.2/N/A	52.0/63.7	101.8/131.6	21.4/34.4	19.6/28.6
Glycerol	76.3	32.6	N/A	N/A	N/A	34.5	N/A
Water	38.0	35.9	61.7	N/A	N/A	33.8	38.7
rmsds							
Bond lengths (Å)/angles (°)	0.019/1.985	0.008/1.144	0.016/1.720	0.012/1.623	0.016/1.749	0.026/1.905	0.009/1.159
Ramachandran plot statistics (%)							
Residues in most favored regions	94.9	99.4	94.1	97.3	86.6	98.1	100
Residues in additional allowed regions	5.1	0.6	5.9	7.3	9.8	1.9	0
Residues in disallowed regions	0	0	0	0.4	3.7	0	0

One crystal was used per data set. Values in parentheses are for the highest-resolution shell.

N/A, not applicable.

$$^a R_{\text{sym}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$$

$$^b R_{\text{work}} = \frac{\sum |F_o - F_c|}{\sum F_o}$$

$$^c R_{\text{free}} = \frac{\sum |F_o - F_c|}{\sum F_o}, \text{ calculated using a random 5\% of reflections that were not included throughout refinement.}$$

by residues V468, V471, A472, M475, V496, I393, I501, L504, and G506 of the symmetric CBS2* subunit, acting as a “buckle” that contributes to keeping the dimer closed (Fig. S1). Besides a closed conformation, difference density clearly showed the presence of four adenosyl-derivative molecules bound to CL1 (Figs. 3 and 4). These molecules were interpreted as SAM in sites B and B* and as MTA in sites A and A*, respectively. Co-crystallization experiments of construct C388 with MTA showed that the CL1 conformer can also be adopted

when three ligand molecules bind to the protein (one MTA at site A and two SAM at sites B and B*) (Fig. 3; Tables 1 and 2).

A DALI³⁸ search for structural homologs of MJ0100c identified several uncharacterized CBS domain proteins such as NE2398 (2rc3, Z score=18.0, rmsd=1.4 Å), PAE2072 (2rif, Z score=18.0, rmsd=1.8 Å), ATU1752 (3fhm, Z score=17.7, rmsd=1.7 Å), MJ0922 (2p9m, Z score=17.6, rmsd=1.8 Å), RV2626C (1y5h, Z score=17.0, rmsd=1.6 Å), PH1780 (2yzq, Z score=16.5, rmsd=1.9 Å), ST2348

Fig. 2. 3D structure of MJ0100c. (a) Schematic representation of the MJ0100c monomer. The molecule consists of two tandemly repeated CBS domains, namely, CBS1 and CBS2, which are related by a dyad axis (represented with broken lines). The α -helices (red) and β -strands (yellow) in the two domains are numbered. (b) The MJ0100c monomer after 90° rotation around the horizontal axis. Each monomer contains two potential ligand binding cavities located at opposite sides of the central β -sheets (broken lines).

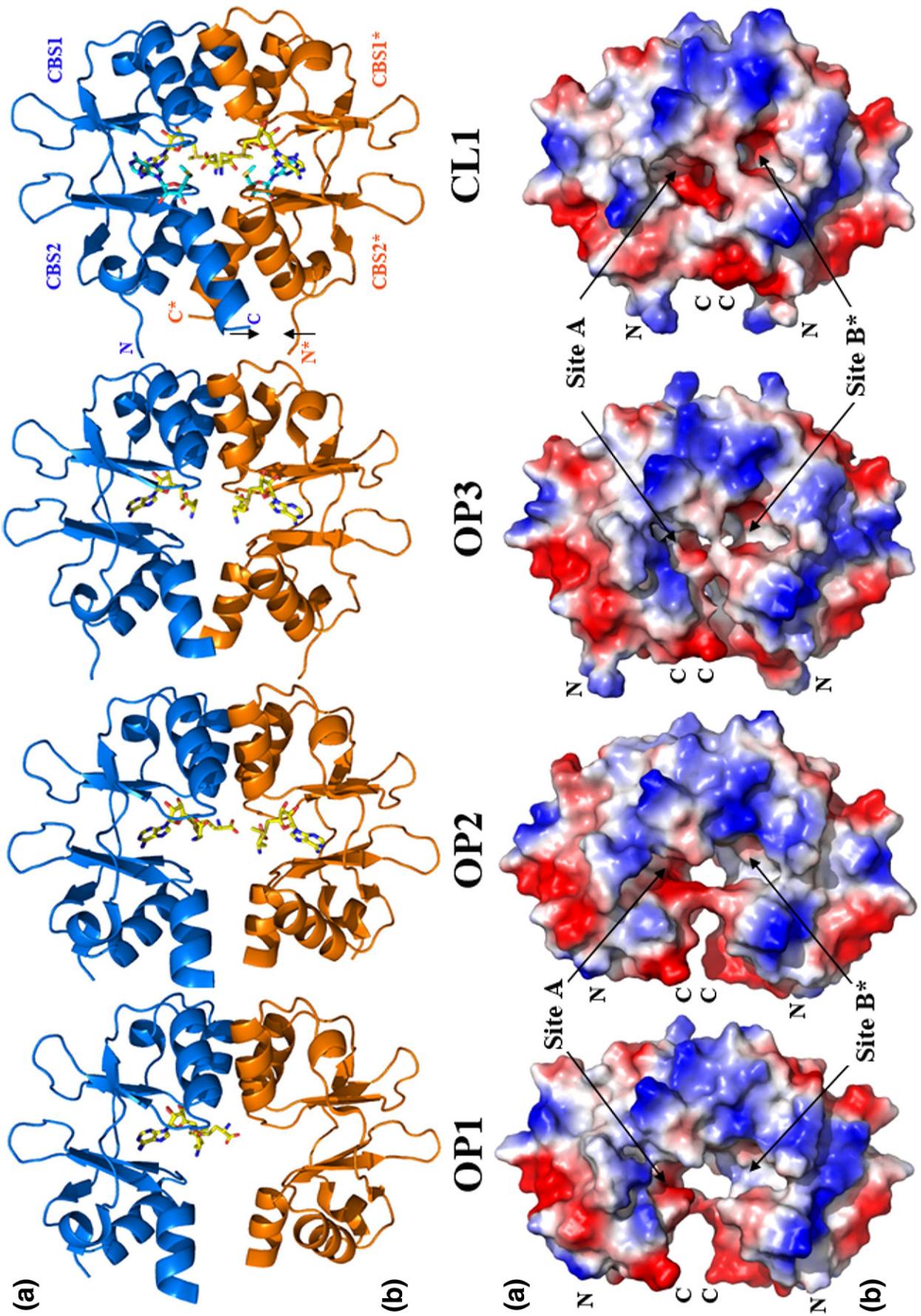


Fig. 3 (legend on next page)

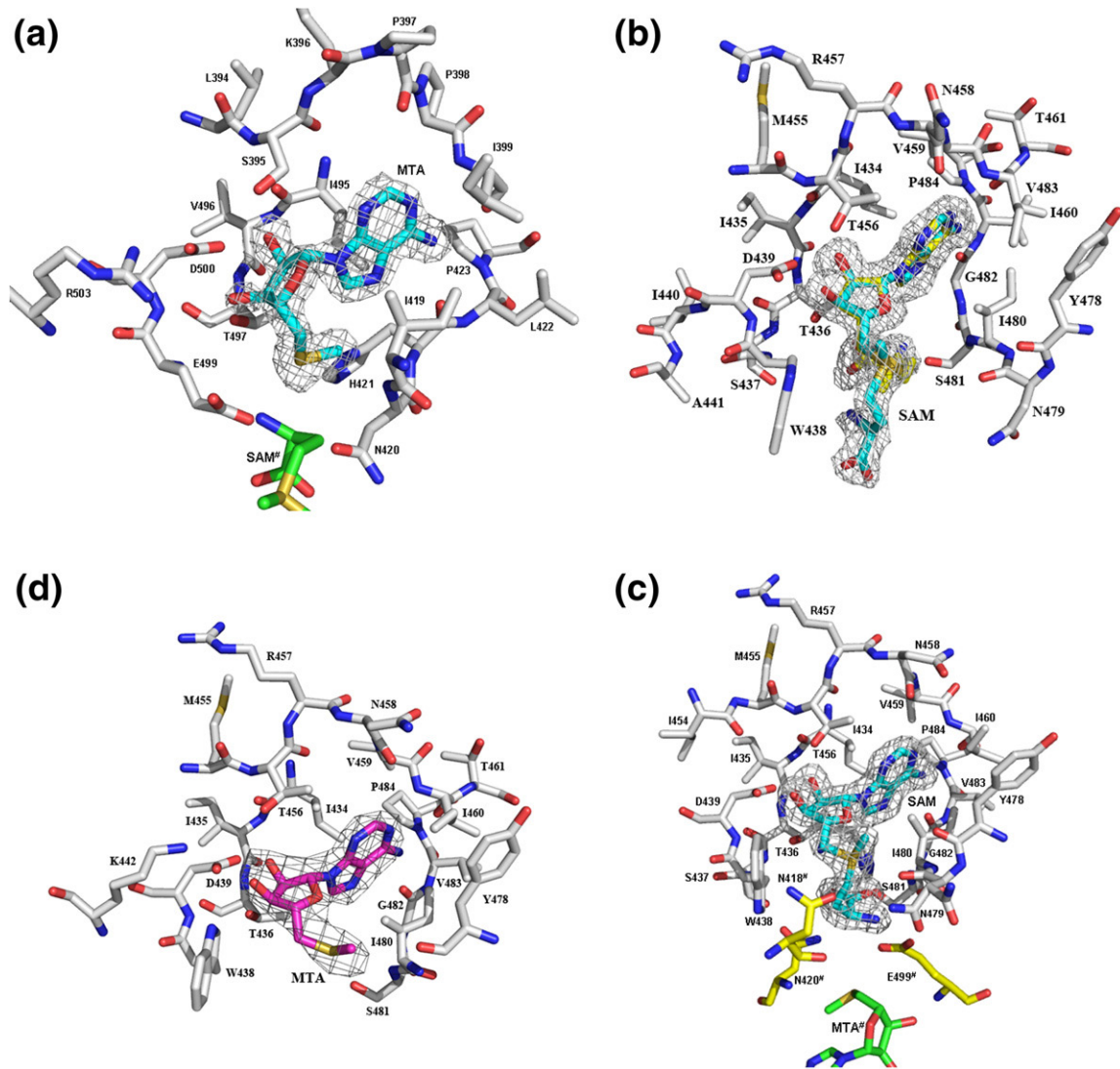


Fig. 4. Ligand binding sites of MJ0100c. (a) MTA-binding site (site A) of the CL1 conformer, as observed in C222 and P2₁2₁2 crystals. MTA is represented in cyan. A molecule of SAM bound at the neighboring site B* is in green. (b) SAM-binding site (site B) of the OP1 conformer, which crystallizes in space groups P1 and C222₁. SAM adopts two different conformations (“extended” (cyan) and “folded” (yellow), which have been superposed. (c) Site B of MJ0100c in the CL1 conformer. The SAM molecule (cyan) is in its folded state. Residues colored *yellow* are donated from the neighboring subunit. A molecule of MTA bound at the neighboring site B* is in *green*. (d) Site B of MJ0100c (CL1 conformer) with bound MTA (cyan). 2F_o - F_c maps are contoured at 1.0 σ.

(2ef7, Z score=15.6, rmsd=2.5 Å), or TA0289 (2qh1, Z score=15.4, rmsd=1.9 Å) as well as proteins of known function such as inosine monophosphate dehydrogenase (2qh1, Z score=15.7, rmsd=2.1 Å), Mg²⁺ transporter MgTE (2yvy, Z score=16.2, rmsd=2.0 Å), or γ-AMPK (AMPK subunit gamma-1) (2uv4, Z score=15.0, rmsd=1.8 Å). All of them form parallel dimers, being ST2348, RV2626C, and PH1780 in an open conformation and the rest forming closed dimers. To our knowledge, MgtE is

the only known case that can adopt an open and a closed conformation.^{24,29} As in MJ0100c, none of the four reported open dimers deposited in the databases have all the potential binding sites occupied.

MJ0100c binds MTA and SAM

During the purification steps of MJ0100c, an absorption peak at 260 nm was observed for those fractions containing the target protein. This peak

Fig. 3. (a) Ribbon representation of the four MJ0100c dimers: OP1, OP2, OP3, and CL1. The displacement of both subunits that takes place during the open-to-closed conformation change is indicated by arrows. (b) Representation of the molecular surface and electrostatic potential of one face of the head-to-head dimers of MJ0100c. There are another two binding sites in the other face of the dimers. The average distances measured from the center of site A to that of site B* (at the same side of the dimer) range from approximately 28 Å in OP1 to 20 Å in CL1. In CL1, the closest approach distances between nucleotides bound at sites A and B* (or A*-B) are 4.7 Å.

Table 2. Summary of the MJ0100c conformations crystallized

MJ0100c conformer	Angle (°)	Site A	Site B	Site A*	Site B*
OP1	23	—	SAM	—	—
OP2	12	—	SAM	—	SAM
OP3	9	—	SAM	—	SAM
OP3	9	MTA	MTA	MTA	MTA
CL1	0	MTA	SAM	MTA	SAM
CL1	0	MTA	SAM	—	SAM

The angle seen between monomers A and B and the nature of the ligands occupying the four sites of each MJ0100c subunit are indicated.

was associated only with MJ0100-derived material since absorption maxima shifted towards 280 nm after denaturation with 6 M guanidinium chloride (GnCl) and subsequent dialysis against 100 mM Hepes, pH 7.5, and 1 mM DTT buffer (Fig. 5f). Taking into account that MJ0100 was annotated as

an unknown function protein, the identification of this species was considered an important issue in characterizing the macromolecule, since it might give clues about its natural function. To identify this material, we subjected it to matrix-assisted laser desorption/ionization time-of-flight/mass spectrometric analysis (see details in Materials and Methods and [Supplementary Material](#)). Accurate mass determination revealed the presence of SAM and its degradation product MTA. It is worth mentioning that SAM and/or MTA was never added during purification or crystallization. Hence, MJ0100c must have scavenged these molecules from its *Escherichia coli* host during overexpression and not released them during purification. In agreement with the mass spectrometry data, the analysis of the $2F_o - F_c$ and $F_o - F_c \exp(i\varphi c)$ electron density maps confirmed the presence of adenosyl-derivative ligands that had been co-crystallized with the macromolecule (Fig. 4). These molecules remained bound to MJ0100c

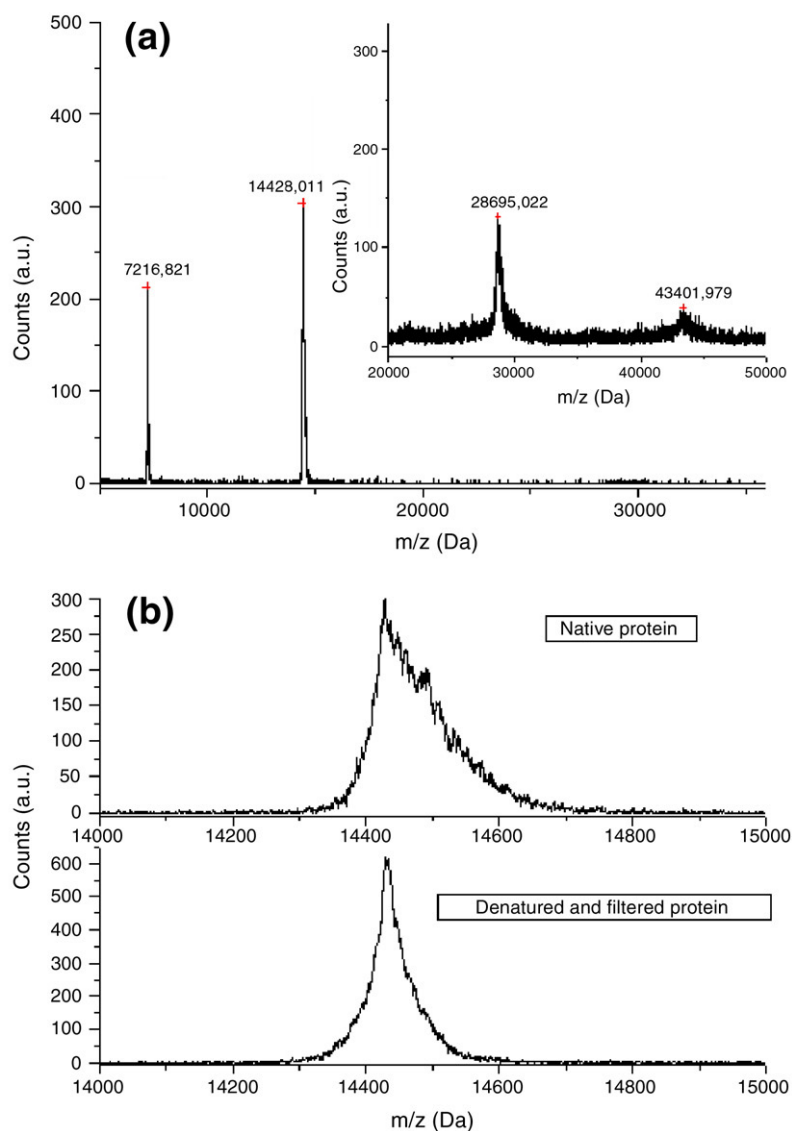


Fig. 5. Mass spectra of (a) native and (b) denatured C381 proteins. The peaks at $m/z = 28,695.022$, $14,428.011$, and 7216.821 Da correspond to $[2M+H]^+$, $[M+H]^+$, and $[M+2H]^{2+}$, respectively. C381 has a theoretical $M_w = 14,426.9$ Da including the first methionine at the N-terminus. (c) Mass spectra of native C381 in the range 0–1000 Da (top) versus a blank (bottom). Peaks at $m/z = 399.085$ and 298.075 correspond to SAM and MTA, respectively. The peak at $m/z = 250.013$ drifts of a fragmentation of SAM involving the loss of the methionyl group and subsequent reorganization of the ribose ring. (d) Mass spectra of the non-proteic filtrate solution obtained after denaturation of C381 with 6 M GnCl. The peaks at $m/z = 399.099$ and 298.067 Da correspond to SAM and MTA, respectively. Peak at $m/z = 250.062$ Da is equivalent to that at 250.013 Da shown in (a). The peak at $m/z = 298.067$ Da observed in blank (which is very weak compared with the peak corresponding to MTA at 298.067 Da) assigned to an interfering mass. (e) Mass spectra of pure SAM (from SIGMA) for comparison. (f) Optical absorption spectra of native and denatured MJ1-381. Blue, native MJ1-381 protein extracted from gel-filtration chromatography. Gray, MJ1-381 denatured with 6 M GnCl. The absorption maxima shift towards 280 nm as a conse-

quence of the loss of the bound ligands. Pink, filtrate containing the non-proteic fraction obtained after removal of denatured MJ1-381.

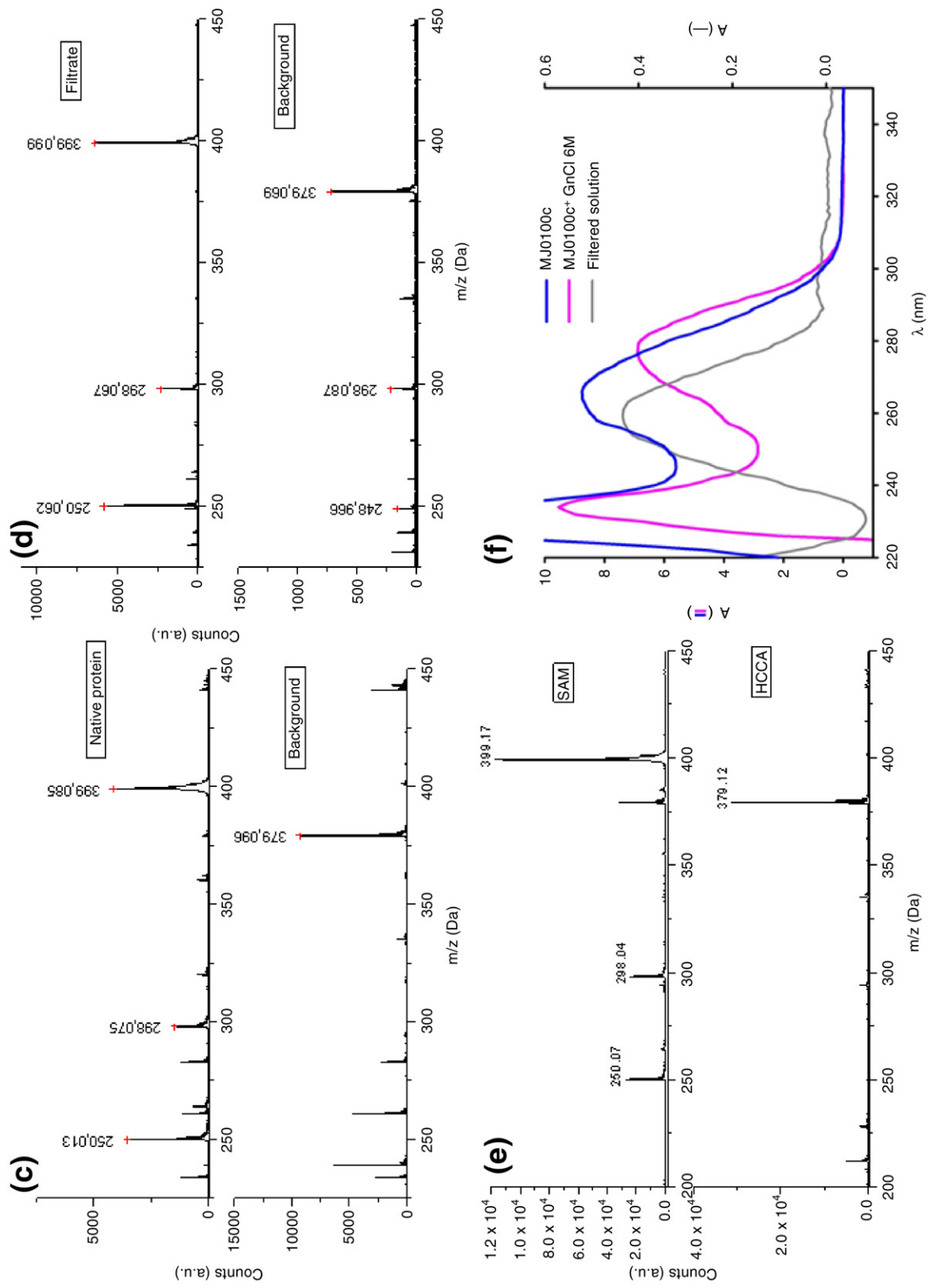


Fig. 5 (legend on previous page)

throughout purification and crystallization. However, the exact chemical nature could not be determined unambiguously with the diffraction data only. The occupancy of the four potential binding cavities varied depending on the crystal habit from one molecule (in *P1* crystals) to two (*P1*, *C222₁*, and *P6₁* crystals), three (*C222₁*), or four (*C222₁*, *P2₁2₁2*, and *C222* crystals), and this coincided with different aperture degrees of the MJ0100c dimers (Fig. 3). Strikingly, not a unique but two different types of adenosyl-derivative molecules, namely, SAM and/or MTA, were observed in different crystal habits, with SAM always being located in site B (and/or B*) and MTA in site A (and/or A*) of the corresponding dimers.

Ligand binding sites

MJ0100c contains two CBS domains that associate to form a compact structure with a cleft between domains. As mentioned above, each subunit shows two different cavities (site A and site B), which are related by a dyad axis and are located at opposite sides of the β -sheets (Fig. 2). These cavities have been proven to be the potential binding site for adenosyl groups in CIC5 and AMPK.^{16,18,20,22,23} In MJ0100c, sites A and B, though clearly distinct, exhibit a high degree of similarity in overall structure regarding the adenosyl binding site (Fig. 4). As in other adenosyl-derivative-bound structures,^{14,18,19,21,36} MTA and SAM are bound by interactions similar to those of AMP/ADP or ATP. In both cases, the adenine base is stabilized in a predominantly hydrophobic binding pocket via stacking interactions with aliphatic residues, namely, I399, I419, L422, and I495 in site A (Fig. 4a) and equivalently I460, I480, and I434 in site B (Fig. 4b and c). The adenine rings are further stabilized by a network of hydrogen bonds between the 6'-amino exocyclic groups and the backbone carbonyl oxygen of residues I399 and H421 in site A and I460 and G482 in site B and also between the 1'-aza groups and the main-chain nitrogen atoms of I399 and I460 in sites A and B, respectively. In addition, there are water-mediated hydrogen-bond interactions between the N1 of the adenine rings and the main-chain nitrogen atoms of I399 (site A) and I460 (site B), respectively, and also between the N3 of the adenine rings and the main-chain carbonyl oxygen atoms of P397, H417, and S395 (in site A) and Y478 and N479 (in site B). As observed in other CBS domain proteins,^{14,20} protein backbone groups (P397, site A; N458, site B) sterically preclude potential accommodation of the two-amino group of guanine derivatives. The ribose moieties are bound in polar pockets in which the free hydroxyl groups are hydrogen bonded to the side-chain carboxyl of D500 and of D439 in sites A and B, respectively, as well as to the hydroxyl of S395 (site A) and T456 (site B). In addition, there are water-mediated hydrogen-bond interactions between the hydroxyl groups of the ribose rings and residues R503 and E499 in site A as well as with K442 in site B. Both K442 and R503, which are located three

positions after the aspartates D439 and D500, are strongly conserved (Fig. 1) and play an important role in stabilizing the position of the aspartates D439 and D500 within the crevice by forming a salt link with them. Though the location of the adenosyl moiety is well maintained in both sites A and B, we found important differences in the position of the methionyl group of SAM in site B, which varied depending on the conformation of the dimer. By this means, SAM adopts two different conformations in our protein complexes. The methionyl group fluctuates between an "extended" and a "folded" form when just one molecule of SAM is bound to the MJ0100c dimer (OP1 conformation) (Fig. 4). On the contrary, a "folded" SAM occupies sites B and B* in the OP3 and CL1 conformations of the dimer (Fig. 4). An intermediate situation is observed in OP2, in which the molecule of SAM is always folded in site B of monomer A, while its alkyl group is fluctuating between an extended and a folded state in site B* of monomer B (Figs. 3 and 4). In its extended state, the methionyl group of SAM runs parallel with the aromatic ring of W438 with its amino and carboxylate groups pointing towards the hydrophilic pocket provided by the side chain of residues N418, N420, and E499 of the second subunit.

In the open dimers OP1 and OP2, SAM adopts an alternative folded conformation (Fig. S3a), which is stabilized through interactions with residues of the CBS2 subdomain and involves direct or water-mediated H-bonds between the carboxylate group of SAM and the side chains of T436, S437, and H421, as well as between the amino group of the ligand and S481, H421, and the main-chain oxygen of S479. Additionally, water-mediated H-bonds are formed between SAM and residues S437 and E499 of the second monomer. In the OP3 and CL1 conformers, SAM is found in a slightly different folded state with respect to that observed in OP1 and OP2 (Fig. S3b). In this case, the carboxylate group of SAM is stabilized via H-bond interactions with the hydroxyl group of S481 and by water-mediated H-bonds with T436, H421, and W438 of the subunit that contains it, as well as with N420 of the complementary subunit. The amino group of SAM H-bonds N479 and additionally interacts with residues E499, N418, and R503 via direct or water-mediated H-bonds.

Positive density clearly indicated the presence of two molecules of MTA in sites A and A* of the closed conformation CL1 of MJ0100c (Fig. 4). A detailed analysis of the amino acid sequence of MJ0100c (Fig. 1) evidenced that sites A and B show some remarkable differences that may be responsible for their specific affinity for MTA and SAM, respectively. These particular features specially affect the β 3- α 2 region, which has been proposed to contain a recognition motif (*GlxS/TxS/TD*, where *x* is any amino acid and *h* is hydrophobic) that determines the affinity for the ribose-phosphate moiety in AMP/ADP/ATP binding proteins.¹⁸ In M0100c, the serine/threonine residues preceding the conserved aspartate groups (D500 in site A and D439 in site B) are E499 and W438 (Fig. 1). Thus,

the presence of a negatively charged (E499) and a bulky hydrophobic residue (W438) nearby the putative location of the ribose rings would be expected to create unfavorable interactions from AMT/ADP/ATP binding and might preclude these sites for binding phosphate-containing ligands. Our structural analysis suggests that MTA could presumably bind to MJ0100 in site B. However, further studies need to be done in order to determine its provenance in our crystals and whether sites A and A* would be suitable for SAM binding. Nevertheless, the 3D structures reported here suggest that accommodation of SAM in site A would be much less stable than that in site B, since it would not be compatible with the H-bond network observed between SAM and MJ0100c in site B. We were also able to obtain crystals of MJ0100c complexed with three adenosyl-derivative molecules per dimer but not with MTA bound only in site A or A*. The network of interactions connecting residues from both MJ0100c subunits suggests that ligand binding may be cooperative. Notably, residues N479 and R503 make interactions with both MTA and SAM molecules in each subunit.

Discussion

Since its discovery in 1997,¹ several 3D structures of CBS domain proteins have been elucidated,^{15–24} including a scarce number of protein–ligand complexes (Fig. 1). Despite this effort, the biological role of these domains is still largely unknown. One of the main reasons is the considerable uncertainty about their potential ligands and the essential features that must be met by the sites interacting with them. Accordingly, we are unaware of the effect of these ligands on their corresponding targets and the molecular mechanisms underlying its regulation. A close examination of the amino acid sequences of CBS domain proteins from different sources (Fig. 1) suggests that each of their potential binding sites might have a specific affinity for different molecules. Several protein–ligand complexes seem to support this idea, including mammalian γ -AMPK (the primary sensor of the intracellular energy status), which contains three AMP binding sites, one of which does not exchange with ATP,²³ or protein ATU1752 from *Agrobacterium tumefaciens*, which complexes NADH and AMP at two different sites (PDB code: 3fhm). The question immediately derived from these data is whether binding of different ligands at different sites might also exert distinct effects in the structure of the corresponding partners. At the moment, two different types of regulatory mechanisms have been proposed. The first one (herein denoted as “static”) claims that the nucleoside portion of the ligand induces essentially no change in the protein structure, the electrostatic potential at the binding site being the most significant property of adenosine nucleotide binding.^{22,23} This “static” response would be involved in processes in which regulation by energy charge would

be advantageous. According to this model, different CBS modules would specifically interact with different effector proteins (or domains) whose activity they regulate, with a variable surface patch contributing to interaction specificity.¹⁴ On the contrary, the second type of mechanism (here denoted as “dynamic”) involves dramatic conformational changes in the protein structure upon ligand binding. This dynamic behavior has just been reported for the cytosolic domain of the magnesium transporter MgtE, which revealed a “closed-to-open” structural transition in the presence/absence of Mg^{2+} ions.^{24,29} The findings of Ishitani *et al.* suggested that the MgtE cytosolic domain senses the Mg^{2+} concentration in the environment, and this enables a hinge motion of the overall CBS module.²⁹ To our knowledge, no dynamic response is known to occur upon binding of adenosine derivatives. The study presented herein provides a first structural characterization of such a case, involving the CBS pair of protein MJ0100 complexed to SAM and MTA. Our crystallographic analysis was conducted on three different constructs containing the CBS pair located at the C-terminus of the protein (named MJ0100c). We found that the topology of the MJ0100c monomer is typical of that seen in other CBS domain pair structures^{21,36,39} (Fig. 2). In agreement with previous studies in solution,³² we established that MJ0100c forms a dimer and that the most likely mode of dimerization is a “head-to-head” association of both subunits. Unexpectedly, four different conformations of the dimer (named OP1, OP2, OP3, and CL1) could be distinguished among the crystals studied (Fig. 3). Three of these conformers (OP1, OP3, and CL1) crystallized in more than one space group (see Fig. 3 and Table 1), suggesting that the observed species might represent stable intermediate states, which could be isolated. The four different conformers basically differed in two main aspects: (i) the angle adopted by the two subunits and (ii) the number of ligands bound to the protein. In a first approach, MJ0100c crystals were grown in the absence of nucleotides; hence, the bound molecules might have been scavenged from its *E. coli* host during over-expression and not released during the purification process. Two different type of ligands were complexed to MJ0100c, namely, MTA (in sites A and A*) and SAM (in sites B and/or B*). In no case did we observe a reversed occupation of these two cofactors or a dimer in which both sites (A and B) of monomer A were occupied, while their complementary A* and B* in the second monomer remained empty. In the same way, we could not isolate a dimer in which only cavities A and A* were occupied.

A detailed analysis of the electron density maps from crystals containing the C381 and C386 constructs indicated that the N-termini might not play a significant role in stabilizing the “closed” conformation of the dimer (this part of the protein was disordered in all the data sets tested). However, since the initial experiments carried out in the lack of ligands yielded conformers OP3 and CL1 only for

the C386 and C381 constructs, we wanted to investigate whether the different length of the polypeptide chain might determine the conformation adopted by the MJ0100c dimer. Accordingly, we crystallized the shortest C388 construct in the presence of SAM and MTA to see whether we were able to displace the MJ0100c dimer towards its closed conformation. Co-crystallization with SAM yielded a shower of needle-shaped crystals not suitable for crystallographic studies. However, a new type of crystals that belong to space group $C222_1$ grew in the presence of MTA. These crystals contained four independent molecules in the asymmetric unit (Table 1) that formed three different dimers upon application of the symmetry operators. As we expected, these species showed a high degree of occupation in their potential binding sites. Two of them were complexed to four molecules of MTA and adopted the OP3 conformation whereas the third dimer contained one MTA molecule in site A and two SAM molecules in site B and adopted the CL1 conformation. As it happened with some other crystals analyzed, the protein must have scavenged these SAM molecules from its *E. coli* host during overexpression and not released them during purification. Thus, we can conclude that MTA by itself is not able to promote the complete closure of the dimer (CL1) even when it occupies all possible sites. On the contrary, this can be achieved by binding two molecules of SAM plus one MTA (Fig. 3).

Unlike other proteins of the family, the structural features of dimeric MJ0100c are significantly different between the unliganded and ligand-bound structures. The implications of the large conformational change observed in MJ0100 are unclear, although they may well play a role in the regulation of the DUF39 domain, which is not present in our constructs but is contained in the full-length MJ0100 protein. Taken together, our data demonstrate that full occupation of the four potential binding cavities seems to stabilize a closed conformer of the homodimer. Conversely, when not all sites are occupied, the CBS module adopts varying degrees of openness based on the number of ligands bound to the protein. The crystal structures presented herein might represent snapshots of the following molecular mechanism:

- (1) First, one molecule of SAM would bind to site B of monomer A through its adenosyl moiety. Initially, the methionyl group of SAM would be ranging between an “extended” conformation, in which practically no interactions with the protein would take place and a “folded” conformation that would establish new inter-

actions with amino acid residues of the monomer to which SAM is bound (OP1 en Fig. 3). These new interactions would help to fix the position of the methionyl group of SAM inside the cavity. As result, a SAM–MJ0100c complex would be formed in which monomer A is slightly contracted, adopting a more compact structure in which CBS1 and CBS2 subdomains rotate with respect to each other by approximately 5° (see Movie 2). This conclusion was obtained after structural comparison of both subunits in the OP1 conformation, one of which shows a bound molecule of SAM at site B, whereas the other lacks the cofactor. Our crystals revealed that this torsion takes place around the loops connecting helix HA with β 1-strand and helix HB with β 5-strand (Fig. 2), without affecting the secondary-structure elements within each block. Moreover, H/D exchange analysis clearly indicated that ligand binding produces a reduction in protein structural dynamics and a more rigid protein core but does not induce secondary structural changes as demonstrated by infrared data (see FTIR Analysis in Supplementary Material). As a consequence of this contraction, monomers A and B get closer to each other, since they are bound through hydrophobic interactions between helices H1 and H2.

- (2) Once the first molecule of SAM is bound and fixed in its folded state at site B of the first subunit, a second molecule of SAM would bind in its extended form to site B* of monomer B through its adenosyl group (OP2 en Fig. 3). Again, the methionyl group of this second molecule of SAM would be oscillating between an “extended” and a “folded” state analogously to what was previously described. Consequently, monomer B would follow the same behavior as that described for the first subunit and the MJ0100c dimer would progress towards conformation OP3 (Fig. 3), in which both molecules of SAM are in its folded state (Fig. 3).
- (3) The last step of the mechanism would take place upon binding of two molecules of MTA at sites A and A* of the CBS module, respectively, and results progressively in the closed conformations OP3 and finally CL1 (Fig. 3).

Fig. 6. Amino acid conservation between human CBS and MJ0100c from *M. jannaschii*. (Top) The figure depicts the degree of conservation between the amino acid residues in both species according to a color coded shown underneath. (Bottom) View at 180° of the same structure. The 3D structure of MJ0100c was used as 3D template. The sequence alignment and color code assignment were done with Clustal W³³ and the ConSeq server [<http://conseq.tau.ac.il>]⁴⁰ respectively, and represented with PyMOL (DeLano Scientific, LLC).

The open-to-closed conformational change observed in MJ0100c is possible due to the intrinsic features of the exposed amino acids of helices H1, H2, H3, and H4, which are directly involved in intersubunit interactions. While hydrophobic residues donated from helices H1 and H2 (Fig. S8) help in maintaining both monomers together, acting as a hinge during the molecular motion, the alternation of hydrophilic and hydrophobic residues in helices H3 and H4 weakens the interaction between the CBS1 domains. Hydrophobic residues such as L504 or F505 located at the C-terminus of helix H4 might contribute to stabilize the dimer in its closed conformation CL1 (Fig. S1). Similar features have been seen previously for protein TM0892 from *Thermotogamaritima* (PDB code: 1VR9) in which the hydrophobic surface of helices α A and α B in the second CBS motif mediates the self-association, whereas the hydrophilic surface of the α A and α B helices in the first CBS motif splays them away from each other in the dimer.²¹

The structure of MJ0100c presented here suggests that residues 494–500 of CBS2 may represent a novel recognition motif (*GlxS/TxhD/E*, where *x* is any amino acid and *h* is a bulky hydrophobic residue) for the SAM molecule (Fig. 4). This motif can also be found in archaeal proteins such as MJ1404, MJ0729, MJ1004, or MJ0188 (Fig. 1). In the same way, residues 433–439 of CBS1 may form a novel recognition motif (*GlxS/TxED*) for MTA. Both motifs are well conserved in the family of proteins whose structure is characterized as consisting of a DUF39 domain plus a tandem CBS pair (Fig. S4), which might be also potentially regulated by SAM and/or MTA. Interestingly, the ligand recognition mechanism shows striking similarities and differences with respect to other SAM-binding proteins. It involves interactions that can be divided into three parts: contacts to the adenine ring, the ribose, and the methionyl moiety. The first two types of contacts are similar to those observed in SAM radical proteins and methyltransferases.⁴⁰ Thus, the adenine interacts via both hydrophobic stacking and hydrogen bonding, whereas the ribose hydroxyl groups typically show the same hydrogen bond with a conserved D or E.⁴¹ This interaction is common to all CBS domain proteins.^{19–25} In contrast, the third type of contacts varies among different species. For example, in *EcBioB*, the amino N and the carboxylate of the methionyl moiety are positioned to hydrogen bond with backbone O atoms and to form a salt bridge with the guanidinium group of an arginine residue. These interactions that modulate the properties of the methionyl moiety of SAM to improve ligation to the four Fe atoms of the Fe₄S₄ cluster are absent in MJ0100, where the methionyl moiety may adopt at least three alternative orientations.

The overall fold of MJ0100c and the fact that it binds SAM invited us to consider as to whether this protein might act as an archaeal homolog of CBS. CBS (EC 4.2.1.22), the first enzyme of the transsulfuration pathway, is a SAM-regulated enzyme that plays a key role in the metabolism of homocysteine

by catalyzing a pyridoxal 5'-phosphate-dependent condensation of serine and homocysteine to give cystathionine, which is subsequently converted to cysteine by the enzyme cystathionase.⁴² In humans, deficiency of CBS activity is the most common cause of homocystinuria, an inherited disease.⁴³ However, although little is known of methionine salvage in methanogens, recent work from Soll's group and others indicates that these organisms synthesize cysteine *de novo* primarily via a phosphoseryl-tRNA route akin to selenocysteine metabolism.^{44,45} With the aim to further investigate whether our structures might provide the structural basis for understanding the molecular mechanisms regulated by SAM in human CBS (for which no crystal structure is available so far), we carefully aligned MJ0100c against human CBS (Fig. 6). Interestingly, a significant number of residues were well conserved between both species, especially at both extremes of the polypeptide chain, with greater similarity between the CBS2 motifs than the CBS1 motifs.

Overall, our structures of the CBS-subunit pair of MJ0100 provide insight into the chemical function of this protein and represent the first reported example of a CBS domain protein complexed with SAM and/or MTA. Further studies are in progress to unravel its biological function.

Materials and Methods

Cloning, mutagenesis, and purification

C386 construct (residues 386–509) was produced by PCR amplification using the plasmid pML1 that carries the CBS pair of MJ0100 (residues 381–509)³² as template and the primers C386f (CACCATGACATTAGTTAAGGATATTTAAGC) and MJ0100r (TCATTTTTTCCCTCCGAA-TAATC). The PCR product was cloned in pET101D plasmid using the Champion pET Directional TOPO Expression Kit (Invitrogen). C388 construct (residues 388–509) was obtained in the same way as described for C386 but the forward primer C388f (CACCATGACATTAGTTAAGGATATTTAAGC) was used. MJ0100c mutant expression constructs were generated in a PCR-based site-directed mutagenesis using the QuikChange Mutagenesis Kit (Stratagene). First, the mutants C381–D439A and C381–D500A were generated using as template the plasmid pML1. The primers D439Af (GGAATAATTA-CATCGTGGGCTATAGCTAAAGCTCTTGCT) and D439Ar (AGCAAGAGCTTTAGCTATAGCCCACGATG-TAATTATTC) were used for introducing the mutation D439A, and the primers D500Af (GGGCATTGTAA-CATCTGAAGCTATCTCAAGATTATTCGGAG) and D500Ar (CTCCGAATAATCTTGAGATAGCTTCA-GATGTTACAATGCCC) were used for the mutation D500A (mutated nucleotides are underlined). The double mutant C381–D439A+D500A was constructed using as template the plasmid carrying the D439A mutation and introducing de D500A mutation with the oligonucleotides D500Af and D500Ar. All plasmid DNAs were transformed and amplified in XL1-Blue bacteria. The entire sequence of the mj0100 cloned fragment was verified by sequencing. The plasmids obtained were transformed into *E. coli* strain BL21-Codon plus (Stratagene) for protein overexpression.

Table 3. Statistics for data collection of Se-Met derivative crystals

Space group	Redundancy (overall/last shell)	Completeness (%) (overall/last shell)	λ (Å)	Mosaicity (°)	R_{sym} (overall/last shell)	$I/\sigma(I)$ (overall/last shell)
$P2_12_12_1$	7.3/6.8	100/100	0.9791	0.4	0.101/0.095	16.1/8.07
Figure of merit (overall/last shell)				0.37/0.26		
Resolution (Å)				50–3.4		
Unit cell parameters: a, b, c (Å)/ α, β, γ (°)				82.407, 120.364, 175.656/90.0, 90.0, 90.0		
Reflections (measured/unique) (Å)				296,565/22,851		
Molecules per asymmetric unit				10		

Protein purification

C381, C386, C388, and the mutant proteins C381–D439A, C381–D500A, and C381–D439A+D500A were purified as described in Ref. 32. All proteins had an identical behavior in the different chromatographic steps and were eluted in the Superdex-75 10/300 column (GE Healthcare) as a dimer. The apo form of C381 was obtained following a denaturation and refolding protocol. One milliliter of C381 at 10 mg/ml was dialyzed against 500 ml GnCl (6 M), 50 mM Hepes, pH 7.0, and 1 mM DTT at room temperature for 6 h using the dialysis membrane CelluSep 6000–8000 Da. This dialysis allows the denaturation of the protein and the diffusion of the ligand out of the membrane in a single step. The refolding was performed in a second dialysis against 1 l buffer (50 mM Hepes, pH 7.0, and 1 mM DTT) for 16 h at 4 °C. The isolation of the ligand bound to purified C381 was performed, denaturing the protein with 6 M GnCl and filtering the solution with a Vivaspin 5000 (Sartorius). The filtered ligand was identified by mass spectrometry.

Mass spectrometry analysis

A saturated solution of α -CHCA in a 1:1 mixture of acetonitrile/water was used as matrix. Samples were mixed with matrix solution in a 1:2 (v/v) ratio. One or two microliters of the mixture was spotted in each well of the stainless steel target plate and co-crystallized by evaporation. Spectra were acquired in positive linear mode with the aid of a Bruker Reflex IV time-of-flight mass spectrometer, employing different sets of voltages for the ion source in order to focus the adequate mass range. Each spectrum is an average of 300 shots. The following samples and their respective blanks were analyzed: (i) native MJ1-381 extracted from gel-filtration chromatography, (ii) MJ1-381 protein denatured with 6 M GnCl, and (iii) filtrate solution containing the non-proteic fraction resulting from denaturation of MJ1-381 with 6 M GnCl (Fig. 5f). In the mass spectrum of the (i) MJ1-381 protein, a series of peaks corresponding to $[M+2H]^{2+}$, $[M+H]^+$, and $[2M+H]^+$ are readily detected (Fig. 5a). This series is less intense in the spectrum of denatured protein with 6 M GnCl (data not shown). Although mass spectrometry matrix-assisted laser desorption/ionization time of flight is not well suited to non-covalent-bonded molecules, obtained spectra were consistent with previous gel-filtration chromatography and dynamic light-scattering experiments, which indicated that MJ0100c forms a dimer in solution.³² As shown in Fig. 5b, the peak $[M+H]^+$ observed by mass spectrometry on sample (i) was considerably broader than the corresponding peak observed on the denatured protein mass spectrum (ii). This is in good agreement with UV spectroscopy experiments, which suggested the presence

of a nucleotide bound to the protein. To confirm the presence of an adenosyl derivative, we acquired mass spectra in the range 0–1000 Da for both the native C381 and the non-protein fraction resulting from denatured protein filtrate (Fig. 5c and d). In both cases, three peaks appeared at $m/z=399.1$, 298.1, and 250.0 Da, which are the characteristic signature of SAM (Fig. 5e). SAM (exact mass, 398.137 Da) degrades into MTA (297.090 Da) and into a fragmentation product with $m/z=250$ Da, probably due to the loss of the methionine chain. To further confirm such assignment, we carried out a recalibration using the procedure described in Ref. 46, using propylenglycol as external calibrant. In this way, an accuracy of ca 10 ppm is achieved. Accurate mass determination (see Supplementary Material) revealed that these peaks could be due to the presence of SAM ($M_w=398,137$) and its degradation products MTA ($M_w=297,090$) and unassigned $m/z=250$ Da. A search in the Biological Magnetic Resonance Data Bank database[‡] using the corrected mass values returned two adenosyl-derivative candidates: (i) 398.113: $C_{15}H_{20}N_5O_6S^+$ ($^{12}C^{14}N$): S-adenosyl-4-methylthio-2-oxobutanoate and (ii) 398.137: $C_{15}H_{22}N_6O_5S$ ($^{12}C^{14}N$): SAM, but only the second candidate can explain all the experimental data. Supporting the presence of SAM, mass spectra of Se-Met-labeled C381 protein revealed three peaks at $m/z=447$, 346, and 250 Da, which correspond to Se-SAM, Se-MTA, and to a fragment resulting out of the loss of the methionyl group of SAM (Fig. S2). The two first peaks differ in 48 mass units from the corresponding trace peaks of 399 Da (SAM) and 298 Da (MTA) detected in the native protein. These differences in nominal mass can be explained from the substitution of sulfur by selenium during the Se-Met C381 overexpression. Consequently, peak at $m/z=250$ Da remains unaltered with respect to the spectra of the native protein since it corresponds to a fragment lacking the selenomethionyl group of Se-SAM.

Crystallization and data collection

Crystals of native C381, C386, and C388 as well as of Se-Met-labeled C381 were obtained by the hanging-drop vapor-diffusion method at 20 °C according to the protocol described in Ref. 32 with slight variations in the protein and precipitant concentrations. Crystals of C381 appeared overnight and belong to space group $P2_12_12_1$. C386 showed three different crystal habits ($P2_12_12_1$, $C222$, and $P6_1$), while C388 crystallized in space groups $P1$ and $C222_1$. The statistics for data collection and crystals parameters are summarized in Tables 1 and 3. All crystals

‡ http://www.bmrb.wisc.edu/metabolomics/mass_query.php

were looped and frozen at 100 K. In order to prevent freezing problems due to physical tensions between inner and outer regions of the C381 samples, we just freeze crystals whose size ranged from 0.1 to 0.5 mm (in the largest dimension). The best result was obtained by soaking the crystals in crystallization solution containing a final glycerol concentration of 25% and a slight increase (5%) of the precipitant.³² Diffraction data of the native proteins were collected at the ID14-2, ID23-1, and ID23-2 beamlines of the European Synchrotron Radiation Facility (ESRF, Grenoble, France). The beam was attenuated with filters to 1.0% of its primary intensity for data sets collected at beamline ID23-2 in order to avoid radiation damage. The single-wavelength anomalous dispersion data set of the C381 Se-Met protein was collected at the BM16 beamline of the ESRF. The data were indexed and integrated with HKL2000 suite⁴⁷ and were further processed with CCP4 programs.⁴⁸

The co-crystallization experiments of C388 with MTA were carried out preincubating the protein (at a concentration of 7 mM) with the ligand (at 14 mM) for 1 h at 4 °C and then set up for crystallization by using the hanging-drop method. The co-crystallization yielded crystals of space group C222₁ (Table 1) containing the OP3 and CL1 conformers.

Structure solution and refinement

For phasing, we used a single-wavelength anomalous dispersion data set of protein C381 at 3.4 Å resolution, collected at the Se peak ($\lambda=0.9791$ Å). C381 contains five methionine sites per molecule, including the methionine in the N-terminal position. Analysis based on the Matthews coefficient⁴⁹ was ambiguous and indicated 8 ($V_M=3.57$) to 10 ($V_M=2.86$) molecules per asymmetric unit, corresponding to a solvent content of 66% and 57%, respectively. Therefore, the search for the selenium positions was carried out with different numbers of anomalous sites with SOLVE.⁵⁰ The solutions were judged by the correlation coefficient and were taken to be a good solution if the correlation coefficient was double than that of wrong solutions of the same run. In spite of the low figure-of-merit value (see Table 3), a correct enantiomer of 10 out of the 50 heavy atom sites could be distinguished by comparing the connectivity and contrast value after a first turn of density modification and refinement of the heavy atom sites using SOLVE.⁵⁰ In the model generated after solvent flattening with DM,⁴⁸ 8 out of the 10 molecules could be identified per asymmetric unit. The initial model of each of the 10 monomers could be further interpreted and completed by hand with the aid of the homologous protein MJ0922 from *M. jannaschii* (PDB code: 2p9m). After several rounds of model building using O⁵¹ and Coot⁵² and simulated annealing and energy minimization using CNS,⁵³ R_{work} dropped to 28% and R_{free} dropped to 0.34. A higher-resolution data set of the native protein to 3.1 Å was collected at beamline ID14-2 (ESRF).³² Since the native crystals were isomorphous to those of the Se-Met-substituted protein, the refined model of the Se-Met C381 protein was taken as starting model for further refinement. Strict NCS restraints were used in the initial stages of the refinement between the 10 copies in the asymmetric unit. In the later stages, these strict restraints were loosened, and restraint individual isotropic B -factors

were refined. The σA -weighted $2F_o - F_c$ maps as well as omit maps were calculated at regular intervals to allow manual rebuilding of insertion/deletion loops and side chains with a different rotamer. The backbone geometry was regularly checked against a structural database with the pep flip option. R_{free} (calculated from 5% of the reflections omitted during refinement) was monitored throughout. Residual weak intensity in a $2F_o - F_c$ map allowed for tracing the backbone of several additional residues of the N- and C-termini and the linker regions between the CBS domains within each monomer. The structure solution of these crystals revealed 10 monomers in the asymmetric unit that were refined to an R/R_F factor of 0.26/0.33 (Table 1). The slightly elevated factors for this resolution and the high B -factors for some residues are probably due to the large fraction of the molecule that is found to be poorly ordered in the crystals. Additionally, some lysine and glutamate residues could not be modeled since poor or inexistent density was observed. The cofactor SAM could be clearly located in the electron density map and was added to the model in the site 2 of 8 out of the 10 molecules in the asymmetric unit. The site 2 of the other two molecules was apparently empty. Although residual electron density could be observed in the site 1 of each molecule, the quality of the data did not allow an unequivocal interpretation of this density. The final model contains residues 386–509 with good geometry and no outliers in the Ramachandran plot.

The crystal structures of C386 and C388 were determined by molecular replacement using the program MOLREP,⁵⁴ with the C381 monomer serving as the search model, and subsequently refined with REFMAC5.⁵⁵ The geometry quality model was assessed with the structure validation tools PROCHECK⁵⁶ and VADAR.⁵⁷ Solvent molecules (all regarded as water) based on higher than 3 σ peaks in the $F_o - F_c$ σA -weighted maps were added gradually and conservatively with regard for their environment including potential interactions with hydrogen-bond partners. This solvent model was further comprehensively checked several times during the refinement by omitting all water molecules that had high B values (>60 Å²) or made either too close contacts with each other or with protein atoms or a too sharp angle with potential hydrogen-bonding partners. The final model contains residues 386–506 with good geometry and no outliers in the Ramachandran plot. Depending on the crystal habit, two cofactors, MTA and SAM, could be clearly located in the electron density maps and were added to the models in sites A/A* (MTA) and B/B* (SAM), respectively (see Table 1). Protein structures were illustrated by using the program PyMOL (DeLano Scientific, LLC, Palo Alto, CA, USA). Full statistics for data collection and refinement are given in Table 1.

Computer analysis

Multiple sequence analysis was performed with Clustal W³³ and represented with UTOPIA v1.4.4.⁵⁸ BLASTP was used to search for protein homologs among the complete sequenced genomes[¶] (TIGR database; Entrez-Genome Database). rmsd differences between the protein structures were calculated with Coot.⁵² 3D predicted models were performed by using the SWISS-MODEL suite

§ <http://www.ccp4.ac.uk>

¶ <http://utopia.cs.manchester.ac.uk/>

¶ <http://archaea.ucsc.edu/>

of programs³⁴ and the Phyre server^{a,35}. Comparison of structural homologs was carried out with DALI.³⁸ Calculation of protein interfaces was done with the PISA server.⁵⁹ Genome analysis was done with STRING 8^{b,60}. Multiple sequence alignment color coded by conservation was performed with the ConSeq server.⁴⁰

Concluding remarks

We have solved the crystal structure of the CBS domain pair of protein MJ0100 from *M. jannaschii* (MJ0100c). Analogously to other proteins of its family, MJ0100c forms stable dimers that arrange in a head-to-head fashion. Our crystallographic analysis revealed two important facts. First, each MJ0100c dimer can manage to occupy its four potential binding sites with two different cofactors, SAM and MTA. MTA was always found to occupy the site A of each subunit, whereas the molecule of SAM was always bound to site B. This fact provides new clues about the possible function of MJ0100 and, up to our knowledge, represents the first reported structure of a CBS domain protein complexed with any of these two ligands. The second relevant observation is that binding of these small molecules triggers a drastic conformational change in the protein, which evolves progressively from an "open" form in the absence of ligands to a "closed" form when the four sites are fully occupied. Seven different crystal habits covering four different scenarios of such a mechanism are reported herein. Therefore, MJ0100c represents the first example in which the role of adenosyl binding to a CBS protein seems not to be directed to change the electrostatic potential of the corresponding cavities but, instead, to trigger a conformational change of the CBS module. This effect is similar to that recently reported for the ion transporter MgtE in the presence of Mg²⁺ ions.

Accession numbers

Coordinates and structure factors of the four conformational states have been deposited in the PDB with accession numbers 3KPB, 3KPC, and 3KPD.

Acknowledgements

We deeply thank Dr. Joseph A. Krzycki from the Department of Microbiology at the Ohio State University and Dr. Rowena G. Matthews from University of Michigan for fruitful discussions. We also thank Adriana Rojas for maintenance of the in-house equipment and the staff of ESRF beamlines BM16, ID14-2, ID23-1, and ID23-2 for support during synchrotron data collection. This research was supported by program grants from 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 Ministry of Education (SAF2005-00855), and SICI CONSOLIDER Project (CSD2008-00005) as well as from a postdoctoral fellowship from CIC bioGUNE and from Spanish Ministry of Science and Innovation (BFU2008-00602/BMC) to J.A.E.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jmb.2009.12.012](https://doi.org/10.1016/j.jmb.2009.12.012)

References

- Bateman, A. (1997). The structure of a domain common to archaeobacteria and the homocystinuria disease protein. *Trends Biochem. Sci.* **22**, 12–13.
- Sintchak, M. D., Fleming, M. A., Futer, O., Raybuck, S. A., Chambers, S. P., Caron, P. R. *et al.* (1996). Structure and mechanism of inosine monophosphate dehydrogenase in complex with the immunosuppressant mycophenolic acid. *Cell*, **85**, 921–930.
- Zhang, R., Evans, G., Rotella, F. J., Westbrook, E. M., Beno, D., Huberman, E. *et al.* (1999). Characteristics and crystal structure of bacterial inosine-5'-monophosphate dehydrogenase. *Biochemistry*, **38**, 4691–4700.
- Shan, X., Dunbrack, R. L., Jr., Christopher, S. A. & Kruger, W. D. (2001). Mutations in the regulatory domain of cystathionine beta synthase can functionally suppress patient-derived mutations in cis. *Hum. Mol. Genet.* **10**, 635–643.
- Bowne, S. J., Sullivan, L. S., Blanton, S. H., Cepko, C. L., Blackshaw, S., Birch, D. G. *et al.* (2002). Mutations in the inosine monophosphate dehydrogenase 1 gene (IMPDH1) cause the RP10 form of autosomal dominant retinitis pigmentosa. *Hum. Mol. Genet.* **11**, 559–568.
- Pusch, M. (2002). Myotonia caused by mutations in the muscle chloride channel gene CLCN1. *Hum. Mutat.* **19**, 423–434.
- Haug, K., Warnstedt, M., Alekov, A. K., Sander, T., Ramirez, A., Poser, B. *et al.* (2003). Mutations in CLCN2 encoding a voltage-gated chloride channel are associated with idiopathic generalized epilepsies. *Nat. Genet.* **33**, 527–532.
- Konrad, M., Vollmer, M., Lemmink, H. H., van den Heuvel, L. P., Jeck, N., Vargas-Poussou, R. *et al.* (2000). Mutations in the chloride channel gene CLCNKB as a cause of classic Bartter syndrome. *J. Am. Soc. Nephrol.* **11**, 1449–1459.
- Cleiren, E., Bénichou, O., Van Hul, E., Gram, J., Bollerslev, J., Singer, F. R. *et al.* (2001). Albers-Schönberg disease (autosomal dominant osteopetrosis, type II) results from mutations in the CLCN7 chloride channel gene. *Hum. Mol. Genet.* **10**, 2861–2867.
- Blair, E., Redwood, C., Ashrafian, H., Oliveira, M., Broxholme, J., Kerr, B. *et al.* (2001). Mutations in the gamma(2) subunit of AMP-activated protein kinase cause familial hypertrophic cardiomyopathy: evidence for the central role of energy compromise in disease pathogenesis. *Hum. Mol. Genet.* **10**, 1215–1220.
- Ignoul, S. & Eggermont, J. (2005). CBS domains: structure, function, and pathology in human proteins. *Am. J. Physiol.: Cell Physiol.* **289**, C1369–1378.

^a <http://www.sbg.bio.ic.ac.uk/phyre/>

^b <http://string-db.org/>

12. Scott, J. W., Hawley, S. A., Green, K. A., Anis, M., Stewart, G., Scullion, G. A. *et al.* (2004). CBS domains form energy-sensing modules whose binding of adenosine ligands is disrupted by disease mutations. *J. Clin. Invest.* **113**, 274–284.
13. Estévez, R., Schroeder, B. C., Accardi, A., Jentsch, T. J. & Pusch, M. (2003). Conservation of chloride channel structure revealed by an inhibitor binding site in CIC-1. *Neuron*, **38**, 47–59.
14. King, N. P., Lee, T. M., Sawaya, M. R., Cascio, D. & Yeates, T. O. (2008). Structures and functional implications of an AMP-binding cystathionine beta-synthase domain protein from a hyperthermophilic archaeon. *J. Mol. Biol.* **380**, 181–192.
15. Meyer, S. & Dutzler, R. (2006). Crystal structure of the cytoplasmic domain of the chloride channel CIC-0. *Structure*, **14**, 299–307.
16. Meyer, S., Savaresi, S., Forster, I. C. & Dutzler, R. (2007). Nucleotide recognition by the cytoplasmic domain of the human chloride transporter CIC-5. *Nat. Struct. Mol. Biol.* **14**, 60–67.
17. Markovic, S. & Dutzler, R. (2007). The structure of the cytoplasmic domain of the chloride channel CIC-Ka reveals a conserved interaction interface. *Structure*, **15**, 715–725.
18. Day, P., Sharff, A., Parra, L., Cleasby, A., Williams, M., Hörer, S. *et al.* (2007). Structure of a CBS-domain pair from the regulatory gamma1 subunit of human AMPK in complex with AMP and ZMP. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **63**, 587–596.
19. Amodeo, G. A., Rudolph, M. J. & Tong, L. (2007). Crystal structure of the heterotrimer core of *Saccharomyces cerevisiae* AMPK homologue SNF1. *Nature*, **449**, 492–495.
20. Jin, X., Townley, R. & Shapiro, L. (2007). Structural insight into AMPK regulation: ADP comes into play. *Structure*, **15**, 1285–1295.
21. Rudolph, M. J., Amodeo, G. A., Iram, S. H., Hong, S. P., Pirino, G., Carlson, M. & Tong, L. (2007). Structure of the Bateman2 domain of yeast Snf4: dimeric association and relevance for AMP binding. *Structure*, **15**, 65–74.
22. Townley, R. & Shapiro, L. (2007). Crystal structures of the adenylate sensor from fission yeast AMP-activated protein kinase. *Science*, **315**, 1726–1729.
23. Xiao, B., Heath, R., Saiu, P., Leiper, F. C., Leone, P., Jing, C. *et al.* (2007). Structural basis for AMP binding to mammalian AMP-activated protein kinase. *Nature*, **449**, 496–500.
24. Hattori, M., Tanaka, Y., Fukai, S., Ishitani, R. & Nureki, O. (2007). Crystal structure of the MgtE Mg²⁺ transporter. *Nature*, **448**, 1072–1075.
25. Mahmood, N. A., Biemans-Oldehinkel, E. & Poolman, B. (2009). Engineering of ion sensing by the CBS module of the ABC transporter OpuA. *J. Biol. Chem.* **284**, 14368–14376.
26. Cheung, P. C., Salt, I. P., Davies, S. P., Hardie, D. G. & Carling, D. (2000). Characterization of AMP-activated protein kinase gamma-subunit isoforms and their role in AMP binding. *Biochem. J.* **346**, 659–669.
27. Finkelstein, J. D., Kyle, W. E., Martin, J. L. & Pick, A. M. (1975). Activation of cystathionine synthase by adenosylmethionine and adenosylethionine. *Biochem. Biophys. Res. Commun.* **66**, 81–87.
28. Pimkin, M. & Markham, G. D. (2008). The CBS subdomain of inosine 5'-monophosphate dehydrogenase regulates purine nucleotide turnover. *Mol. Microbiol.* **68**, 342–359.
29. Ishitani, R., Sugita, Y., Dohmae, N., Furuya, N., Hattori, M. & Nureki, O. (2008). Mg²⁺-sensing mechanism of Mg²⁺ transporter MgtE probed by molecular dynamics study. *Proc. Natl Acad. Sci. USA*, **105**, 15393–15398.
30. Martínez-Cruz, L. A., Encinar, J. A., Kortazar, D., Prieto, J., Gómez, J., Fernández-Millán, P. *et al.* (2009). The CBS domain protein MJ0729 of *Methanocaldococcus jannaschii* is a thermostable protein with a pH-dependent self-oligomerization. *Biochemistry*, **48**, 2760–2776.
31. Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G. *et al.* (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science*, **273**, 1058–1073.
32. Lucas, M., Kortazar, D., Astigarraga, E., Fernández, J. A., Mato, J. M., Martínez-Chantar, M. L. & Martínez-Cruz, L. A. (2008). Purification, crystallization and preliminary X-ray diffraction analysis of the CBS-domain pair from the *Methanococcus jannaschii* protein MJ0100. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **64**, 936–941.
33. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H. *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.
34. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). The SWISS-MODEL Workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, **22**, 195–201.
35. Kelley, L. A. & Sternberg, M. J. E. (2009). Protein structure prediction on the web: a case study using the Phyre server. *Nat. Protoc.* **4**, 363–371.
36. Proudfoot, M., Sanders, S. A., Singer, A., Zhang, R., Brown, G., Binkowski, A. *et al.* (2008). Biochemical and structural characterization of a novel family of cystathionine beta-synthase domain proteins fused to a Zn ribbon-like domain. *J. Mol. Biol.* **375**, 301–315.
37. Ragunathan, P., Kumarevel, T., Agari, Y., Shinkai, A., Kuramitsu, S., Yokoyama, S. & Ponnuraj, K. (2008). Crystal structure of ST2348, a CBS domain protein, from hyperthermophilic archaeon *Sulfolobus tokodaii*. *Biochem. Biophys. Res. Commun.* **375**, 124–128.
38. Holm, L. & Sander, C. (1995). DALI: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480.
39. Sharpe, M. L., Gao, C., Kendall, S. L., Baker, E. N. & Lott, J. S. (2008). The structure and unusual protein chemistry of hypoxic response protein 1, a latency antigen and highly expressed member of the DosR regulon in *Mycobacterium tuberculosis*. *J. Mol. Biol.* **383**, 822–836.
40. Berezin, C., Glaser, F., Rosenberg, J., Paz, I., Pupko, T., Fariselli, P. *et al.* (2004). ConSeq: the identification of functionally and structurally important residues in protein sequences. *Bioinformatics*, **20**, 1322–1324.
41. Schubert, H. L., Blumenthal, R. M. & Cheng, X. (2003). Many paths to methyltransferase: a chronicle of convergence. *Trends Biochem. Sci.* **28**, 329–335.
42. Meier, M., Janosik, M., Kery, V., Kraus, J. P. & Burkhard, P. (2001). Structure of human cystathionine beta-synthase: a unique pyridoxal 5'-phosphate-dependent heme protein. *EMBO J.* **20**, 3910–3916.
43. Mudd, S. H., Levy, H. L. & Kraus, J. P. (2001). Disorders of transsulfuration. In *The Metabolic and Molecular Bases of Inherited Disease* (Scriver, C. R., Beaudet, A. L., Sly, W. S., Valle, D., Childs, B., Kinzler,

- K. W. & Vogelstein, B., eds), vol. 1, pp. 2007–2056. McGraw-Hill, New York, NY.
44. Kitabatake, M., So, M. W., Tumbula, D. L. & Soll, D. (2000). Cysteine biosynthesis pathway in the archaeon *Methanosarcina barkeri* encoded by acquired bacterial genes? *J. Bacteriol.* **182**, 143–145.
 45. Sauerwald, A., Zhu, W., Major, T. A., Roy, H., Palioura, S., Jahn, D. *et al.* (2005). RNA-dependent cysteine biosynthesis in archaea. *Science*, **307**, 1969–1972.
 46. Gobom, J., Mueller, M., Egelhofer, V., Theiss, D., Lehrach, H. & Nordhoff, E. (2002). A calibration method that simplifies and improves accurate determination of peptide molecular masses by MALDI-TOF MS. *Anal. Chem.* **74**, 3915–3923.
 47. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology, Vol. 276: Macromolecular Crystallography, part A* (Carter, C. W. & Sweet, R. M., eds), pp. 307–326, Academic Press, New York, NY.
 48. Cowtan, K. (1994). *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, **31**, 34–38.
 49. Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.
 50. Terwilliger, T. C. (2003). SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol.* **374**, 22–36.
 51. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for the building of protein models in electron density maps and the location of errors in these models. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **47**, 110–119.
 52. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132.
 53. Brunger, A. T., Adams, P. D., Clore, G. M., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S. *et al.* (1998). Crystallography & NMR System (CNS), a new software suite for macromolecular structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **54**, 905–921.
 54. Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025.
 55. Vagin, A. A., Steiner, R. S., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F. & Murshudov, G. N. (2004). REFMAC5 dictionary: organisation of prior chemical knowledge and guidelines for its use. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2284–2295.
 56. Laskowski, R. A., Moss, D. S. & Thornton, J. M. (1993). Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.* **231**, 1049–1067.
 57. Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R. F., Sykes, B. D. & Wishart, D. S. (2003). VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res.* **31**, 3316–3319.
 58. Pettifer, S., Wolstencroft, K., Alper, P., Attwood, T., Coletta, A., Goble, C. *et al.* (2007). myGrid and UTOPIA: an integrated approach to enacting and visualising in silico experiments in the life sciences. *Lect. Notes Bioinformatics*.
 59. Krissinel, E. & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797.
 60. Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J. *et al.* (2009). STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* **37**, D412–416.