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-Cruz¹*

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

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

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.

Introduction

The cystathionine β-synthase (CBS) domain proteins comprise a large superfamily of evolutionarily conserved proteins that are present in all kingdoms of life.³ 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.
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.\(^{1-11}\) 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.\(^{5,13}\) CBS domains usually occur in tandem pairs, forming a so-called CBS pair or Bateman module\(^{1}\) in which both CBS subunits are related by a pseudo-2-fold symmetry axis running parallel with the central \(\beta\)-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/ATP\(^{16}\) and AMPK-ATP/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 \(\beta\)-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 CBS\(^{27}\) 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.\(^{14,16,19,22}\) CBS motifs are unusually abundant in archaea.\(^{30}\) Therefore, organisms such as the hyperthermophile Methanocaldococcus jannaschii\(^{11}\) offer excellent models for the characterization of novel adenosyl binding sites. The genome of \(M.\) jannaschii encodes 15 CBS domain proteins,\(^{\dagger}\) 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 15–202) and a CBS module (residues 203–309) that each ligand binds at a different site, triggering a “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.

\(\dagger\) www.tigr.org

**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-
ent organisms with the aim of identifying novel binding motifs. To address this issue, we first performed multiple sequence alignments and three-dimensional (3D) theoretical models 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 (β1 and β5 in Fig. 2), (ii) residues from the second strand of the β-sheet (β2 and β6 in Fig. 2), and (iii) residues located at the third strand of each β-sheet (β3 and β7 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 two turns of the following helix (H4 and H2 in Fig. 2). 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 block. The second group contained a conserved residue (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 sequence conservation and potential functional relevance to the AMP-binding motif. As shown in Fig. 1, in MJ0100, the Ser/Thr residues preceding the last aspartate of the AMP-binding motif are substituted by a negatively charged residue (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. Alternatively, and based on sequence alignments and on published structures of CBS domain pairs, 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. 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\(^1\)2\(^1\)2\(^1\), and diffracting to 3.1 and 3.5 Å, respectively. 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 β5 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\(2\alpha\)3\(\beta\)1\(4\). Similarly, CBS2 has three α-helices and a β-sheet with three strands 5\(\beta\)1\(6\). An additional β-strand (β1*) 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 × 10 Å × 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 β2 and β7, the loop between HA and β1, the N-terminus of helix H4, and the C-terminus of H1 (Fig. 2). The second cavity (denoted as site B) is formed by β3 and β6, the loop between β4 and β5, 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...
Structure of MJ0100 in Complex with SAM and MTA

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 Cα 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 (P2₁2₁2₁, P2₁2₁2, P₁, P₆₁, C₂₂₂₁, 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. 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 C222₁ 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 1407, 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. 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 P6₁ and C222₁ 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 P6₁ 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 C222₁ 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 P2₁2₁2₁, C222₁, and P2₁2₁2₂, 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, TA0289, and the hypothetical protein Yk1 [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 is 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+ (I-strand in site A (loop following HB+ (J-strand in site B), (ii) strand 1/2 or 16 in sites A and B, respectively, and (iii) strand 17 in site A (or 13 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-x-Ser/Thr-x-Ser/Thr-Asp) 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.
Fig. 2 (legend on next page)
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 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 DALI 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=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 (2yqz, 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).](image-url)
Structure of MJ0100 in Complex with SAM and MTA
(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), Mg\textsuperscript{2+} transporter MgTE (2yvy, Z score = 16.2, rmsd = 2.0 Å), or \( \gamma \)-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.\textsuperscript{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...
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 $F_o - F_c \exp(i\phi_c)$ 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

Table 2. Summary of the MJ0100c conformations crystallized

<table>
<thead>
<tr>
<th>MJ0100c conformer</th>
<th>Angle (°)</th>
<th>Site A</th>
<th>Site B</th>
<th>Site A*</th>
<th>Site B*</th>
</tr>
</thead>
<tbody>
<tr>
<td>OP1</td>
<td>23</td>
<td>SAM</td>
<td>—</td>
<td>SAM</td>
<td>—</td>
</tr>
<tr>
<td>OP2</td>
<td>12</td>
<td>SAM</td>
<td>—</td>
<td>SAM</td>
<td>—</td>
</tr>
<tr>
<td>OP3</td>
<td>9</td>
<td>MTA</td>
<td>MTA</td>
<td>MTA</td>
<td>MTA</td>
</tr>
<tr>
<td>CL1</td>
<td>0</td>
<td>MTA</td>
<td>MTA</td>
<td>SAM</td>
<td>—</td>
</tr>
<tr>
<td>CL1</td>
<td>0</td>
<td>MTA</td>
<td>SAM</td>
<td>—</td>
<td>SAM</td>
</tr>
</tbody>
</table>

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

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 $[M+2H]^+$, 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 consequence of the loss of the bound ligands. Pink, filtrate containing the non-proteic fraction obtained after removal of denatured MJ1-381.
Structure of MJ0100 in Complex with SAM and MTA

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, P2₁/2, 2, 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,¹⁴,¹⁶,¹⁹,²¹,²⁶ 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 G₄₆₂ 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 AMP 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 β₃–α₂ region, which has been proposed to contain a recognition motif (GlxS/TyS/TD, where x is any amino acid and r is hydrophobic) that determines the affinity for the ribose-phosphate moiety in AMP/ADP/ATP binding proteins.¹⁸ In MJ0100c, 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 Mg2+ ions.24,29 The findings of Ishitani et al. suggested that the MgtE cytosolic domain senses the Mg2+ 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 structures21,36,39 (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 C2221 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 MJ0100c. 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 places around the loops connecting helix HA with β1-strand and helix HB with β5-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 αA and αB in the second CBS motif mediates the self-association, whereas the hydrophilic surface of the αA and αB 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/TxEd/E, where x is any amino acid and l 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.23 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.21 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 FeS4 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 transulfuration 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 Soil’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) as template and the primers C386f (CACCATGACATTAGTTAGGAATTATTAAAGCTCTTGCT) and MJ0100r (TCATTITTTTCCTCCGGAAATATTCC). The PCR product was cloned in pETI01D 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 (CACCATGACATTAGTTAGGAATTATTCAAGCTCTTGCT) and D439Ar (AGCCAAGAGCTTTAGCTATAGCCTGAGG) were used for the mutation D439Ar (CTCCCAATATATGCTAGGGGCCACGATGATATTTCC). C388 construct (residues 388–509) was obtained in the same way as described for C386 but the forward primer C388f (CACCATGACATTAGTTAGGAATTATTCAAGCTCTTGCT) and D439Ar (CTCCCAATATATGCTAGGGGCCACGATGATATTTCC) were used for the mutation D439Ar (CTCCCAATATATGCTAGGGGCCACGATGATATTTCC). C388 construct (residues 388–509) was obtained in the same way as described for C386 but the forward primer C388f (CACCATGACATTAGTTAGGAATTATTCAAGCTCTTGCT) and D439Ar (CTCCCAATATATGCTAGGGGCCACGATGATATTTCC) were used for the mutation D439Ar (CTCCCAATATATGCTAGGGGCCACGATGATATTTCC).
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 propienglycol 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: \( \text{C}_{15}\text{H}_{22}\text{N}_{6}\text{O}_{5}\text{S} (^{12}\text{C}_{14}\text{N}) \); S-adenosyl-4-methylthio-2-oxobutanoate and (ii) 398.137: \( \text{C}_{15}\text{H}_{22}\text{N}_{5}\text{O}_{5}\text{S} (^{12}\text{C}_{14}\text{N}) \); 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. 5f). 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 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 \( P2_12_12_1, C22_2 \), and \( P6_3 \), while C388 crystallized in space groups \( P1 \) and \( C22_2 \). The statistics for data collection and crystals parameters are summarized in Tables 1 and 3. All crystals

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

<table>
<thead>
<tr>
<th>Space group</th>
<th>Redundancy (overall/last shell)</th>
<th>Completeness (%) (overall/last shell)</th>
<th>( \lambda ) (Å)</th>
<th>Mosaicity (°)</th>
<th>( R_{	ext{free}} ) (overall/last shell)</th>
<th>( I/\sigma(I) ) (overall/last shell)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P2_12_12_1 )</td>
<td>7.3/6.8</td>
<td>100/100</td>
<td>0.9791</td>
<td>0.4</td>
<td>0.37/0.26</td>
<td>16.1/8.07</td>
</tr>
</tbody>
</table>

Figure of merit (overall/last shell)
Resolution (Å)

Unit cell parameters:
- \( a, b, c \) (Å)/\( \alpha, \beta, \gamma \) (°)
- Reflections (measured/unique) (Å)
- Molecules per asymmetric unit

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, denaturating 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 an ion source in order to focus the adequate mass range. Each spectrometer, employing different sets of voltages for the mass, 398.137 Da) degrades into MTA (297.090 Da) and unassigned peak at \( m/z = 250 \) Da. A search in the Biological Magnetic Resonance Data Bank database using the corrected mass values returned two adenosyl-derivative candidates: (i) 398.113: \( \text{C}_{15}\text{H}_{22}\text{N}_{6}\text{O}_{5}\text{S} (^{12}\text{C}_{14}\text{N}) \); S-adenosyl-4-methylthio-2-oxobutanoate and (ii) 398.137: \( \text{C}_{15}\text{H}_{22}\text{N}_{5}\text{O}_{5}\text{S} (^{12}\text{C}_{14}\text{N}) \); 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. 5f). 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 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 \( P2_12_12_1, C22_2 \), and \( P6_3 \), while C388 crystallized in space groups \( P1 \) and \( C22_2 \). 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 P212121 (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 (λ=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 =2.86) to 10 (V_M =3.57) 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 correlation coefficient and were taken to be a good and refinement of the heavy atom sites using SOLVE.50 In the later stages, these strict restraints were comprehensively checked several times during the refinement between the 10 copies in the asymmetric unit. In the asymmetric unit that were refined to an R_Rfree 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 nonexistent 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 MOLREP54 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_calcul – F_obs 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 W58 and represented with UTOPIA v1.4.4.59 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.60 3D predicted models were performed by using the SWISS-MODEL suite

¶ http://archaea.usc.edu/
§ http://www.ccp4.ac.uk

http://utopia.cs.manchester.ac.uk/
of programs\textsuperscript{34} and the Phyre server\textsuperscript{a,35} Comparison of structural homologs was carried out with DALLI\textsuperscript{38}. Calculation of protein interfaces was done with the PISA server.\textsuperscript{39} Genome analysis was done with STRING \textsuperscript{8} and \textsuperscript{60} Multiple sequence alignment color coded by conservation was performed with the ConsSeq server.\textsuperscript{60}

**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 Mg\textsuperscript{2+} 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 Proyect (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

**References**


