The crystal structure of archaeal serine hydroxymethyltransferase reveals idiosyncratic features likely required to withstand high temperatures

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · DECEMBER 2014

Impact Factor: 2.63 · DOI: 10.1002/prot.24697

6 AUTHORS, INCLUDING:

Veronica Morea
Italian National Research Council
57 PUBLICATIONS **1,379** CITATIONS
SEE PROFILE

Sebastiana Angelaccio
Sapienza University of Rome
39 PUBLICATIONS **485** CITATIONS
SEE PROFILE

Fulvio Saccoccia
Uppsala University
16 PUBLICATIONS **149** CITATIONS
SEE PROFILE

Roberto Contestabile
Sapienza University of Rome
57 PUBLICATIONS **821** CITATIONS
SEE PROFILE

PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

The crystal structure of archaeal serine hydroxymethyltransferase reveals idiosyncratic features likely required to withstand high temperatures

Francesco Angelucci,[1]* Veronica Morea,[2]* Sebastiana Angelaccio,[3] Fulvio Saccoccia,[3,4] Roberto Contestabile,[3] and Andrea Ilari[2]

[1] Department of Life, Health and Environmental Sciences, University of L'Aquila, P.le Salvatore Tommasi 1, L'Aquila, Italy

[2] National Research Council of Italy, Institute of Molecular Biology and Pathology, c/o Department of Biochemical Sciences "A. Rossi Fanelli", Sapienza University of Rome, P.le Aldo Moro 5, Rome, Italy

[3] Department of Biochemical Sciences "A. Rossi Fanelli", Sapienza University of Rome, P.le Aldo Moro 5, Rome, Italy

[4] Institute Pasteur-Fondazione Cenci Bolognetti, c/o Sapienza University of Rome, P.le Aldo Moro 5, Rome, Italy

ABSTRACT

Serine hydroxymethyltransferases (SHMTs) play an essential role in one-carbon unit metabolism and are used in biomimetic reactions. We determined the crystal structure of free (apo) and pyridoxal-5′-phosphate-bound (holo) SHMT from *Methanocaldococcus jannaschii*, the first from a hyperthermophile, from the archaea domain of life and that uses H_4MPT as a cofactor, at 2.83 and 3.0 Å resolution, respectively. Idiosyncratic features were observed that are likely to contribute to structure stabilization. At the dimer interface, the C-terminal region folds in a unique fashion with respect to SHMTs from eubacteria and eukarya. At the active site, the conserved tyrosine does not make a cation-π interaction with an arginine like that observed in all other SHMT structures, but establishes an amide-aromatic interaction with Asn257, at a different sequence position. This asparagine residue is conserved and occurs almost exclusively in (hyper)thermophile SHMTs. This led us to formulate the hypothesis that removal of frustrated interactions (such as the Arg-Tyr cation-π interaction occurring in mesophile SHMTs) is an additional strategy of adaptation to high temperature. Both peculiar features may be tested by designing enzyme variants potentially endowed with improved stability for applications in biomimetic processes.

Proteins 2014; 82:3437–3449.

Key words: serine hydroxymethyltransferase; archaea; biocatalysis; cation-pi interaction; local frustration; amide-aromatic contact.

INTRODUCTION

Serine hydroxymethyltransferases (SHMTs, EC 2.1.2.1) are ubiquitous and extensively studied enzymes, playing an important role in cellular one-carbon metabolism and attracting major interest due to their chemical versatility. Because of their essential role in cell replication, they represent a well-known target for cancer chemotherapy.[1–3] Using pyridoxal-5′-phosphate (PLP) as cofactor, SHMTs catalyze the reversible transfer of CB of L-serine to tetrahydropteroylglutamate (H_4PteGlu) to form glycine and 5,10-methylene-H_4PteGlu by a modified retroaldol cleavage reaction.[4,5] An imine exchange between the side-chain NH_2 of an invariant lysine residue (internal aldimine) and the NH_2 of L-serine substrate anchors the latter to the C4' of PLP (external aldimine). The nucleophilic attack of the N_5 atom of H_4PteGlu to the CB of the L-serine substrate yields glycine and 5,10-methylene-H_4PteGlu. The aldimine exchange takes place by a series

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: European Community's Seventh Framework Programme; Grant number: FP7/2007–2013 (BioStruct-X [N°283570]); Grant sponsor: "FIRB/Proteomica 2007"; Grant number: protRBRN07BMCT; Grant sponsor: International FIRB; Grant number: RBIN06E9Z8.

F. Angelucci and V. Morea contributed equally to this work

*Correspondence to: Francesco Angelucci; Department of Health, Life and Environmental Sciences, University of L'Aquila, P.le Salvatore Tommasi 1, 67100 L'Aquila, Italy. E-mail: francesco.angelucci@univaq.it and Veronica Morea; National Research Council of Italy, Institute of Molecular Biology and Pathology, c/o Department of Biochemical Sciences "A. Rossi Fanelli", Sapienza University of Rome, P.le Aldo Moro 5, 00185 Rome, Italy. E-mail: veronica.morea@uniroma1.it

Received 5 March 2014; Revised 9 September 2014; Accepted 10 September 2014
Published online 26 September 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24697

of acid/base reactions and is likely accomplished by the interaction between the phosphate group of PLP and the phenolic oxygen of a conserved tyrosine residue (Y55 in *Escherichia coli* SHMT, *ec*SHMT).[6] The pK_a of this tyrosine is particularly low, due to the fact that the phenolate ion is stabilized by a cation-π interaction with a conserved arginine and a polar interaction with a conserved histidine (R235 and H228 in *ec*SHMT). In addition to the physiological reaction, SHMT is able to catalyze other H_4PteGlu-independent reactions, such as aldol cleavage, transamination, decarboxylation, and racemization, using different amino acids or amino acid derivatives as substrates. As C—C bond formation and cleavage reactions are at the heart of organic synthesis, SHMT enzymes are used in several biotechnological applications such as the synthesis of optically active β-hydroxy-α,ω-diamino acid derivatives.[7,8] The wealth of available information concerning the biochemical and structural properties of SHMTs from different sources, combined with enzyme modification by rational protein engineering, can be exploited to improve catalytic and physical properties and develop novel catalytic functions.[9]

To date, three-dimensional (3D) structures have been experimentally determined only for SHMTs from eubacterial and eukaryotic organisms (Supporting Information Table S1). These structures display a highly conserved overall fold, with monomers organized into obligate homodimers (in eukarya the enzyme occurs as dimer of dimers). The active-site is located at the dimeric interface and is delimited by amino acid residues contributed by both dimer subunits.[10,11]

Here, we report the first 3D structure of SHMT from an archaeon, the hyperthermophilic methanogen *M. jannaschii* (*mj*SHMT). It has been previously shown that this protein conserves residues involved in catalysis and possesses catalytic properties matching those of the bacterial and eukaryotic counterparts, suggesting that both its active site structure and catalytic mechanism are conserved.[12] However, *mj*SHMT presents some peculiarities. SHMTs from eubacteria and eukarya use H_4PteGlu as the pteridine cofactor in the hydroxymethyltransferase reaction, whereas *mj*SHMT uses tetrahydromethanopterin (H_4MPT), a coenzyme involved in methanogenesis and a more efficient carrier of C1 groups in methanogens and other archaea. Additionally, PLP-bound (holo) *mj*SHMT has an optimal reaction temperature higher than that of the mesophilic *ec*SHMT and is extremely resistant to denaturing agents, retaining residual structure even in 10*M* urea at neutral pH[12]. In the same conditions, apo-*mj*SHMT (without bound PLP) also maintained residual structure, although to a lesser extent than the holo form (Roberta Chiaraluce and Valerio Consalvi, personal communication). These features make *mj*SHMT, like other enzymes from hyperthermophilic bacteria, which are stable under extreme chemical-physical conditions, an ideal candidate for biotechnological applications.[12,13]

The 3D structure of *mj*SHMT was solved both in the PLP-holo and apo forms, at 3.0 and 2.83 Å resolution, respectively. The crystal structure of holo-*mj*SHMT was used to build a model of the complex with H_4MPT. Extensive structure analyses and comparisons with SHMT structures from bacteria and eukarya allowed us to identify conserved and variable features in SHMTs from the three domains of life, and identify structural features putatively underlying the idiosyncratic functional and stability properties of *mj*SHMT. The active site of *mj*SHMT does have, as predicted, a similar architecture to mesophilic bacterial and eukaryotic SHMTs, and most residues required for catalysis are structurally conserved. However, significant differences were detected in: (i) the structural arrangement of the ~20 residues C-terminal region, which is unique among known SHMTs; (ii) the main-chain conformation and side-chain identity of regions involved in H_4PteGlu binding by SHMTs from other species, based on which we identified putative structural determinants of *mj*SHMT specificity for the H_4MPT cofactor; and (iii) the identity of some active site residues and interactions that might affect protein stability. In particular, the conserved active site cation-π interaction observed in all other SHMT structures[6] is replaced by a putative amide-aromatic interaction between Y50 and N257. Since N257 is only present and completely conserved in SHMTs from (hyper)thermophilic archaea, we hypothesized that decreased local frustration may represent yet another mechanism exploited by proteins to enhance their chemical-physical stability.

MATERIALS AND METHODS

Protein expression and purification

*mj*SHMT has been expressed and purified as previously reported.[12]

Crystallization

Two *mj*SHMT samples (~10mg/mL each) in phosphate buffered saline with and without 2mM PLP, were screened for crystallization conditions. *Mj*SHMT was purified in the holo form, but the cofactor was not visible in the structure solved without addition of PLP, either because of intrinsic enzyme tendency to lose the cofactor and/or dissociation induced by crystallization conditions (e.g., high salt concentration and slightly acidic pH). Indeed, loss of PLP during the crystallization process has been previously reported for PLP-dependent enzymes such as *Chromobacterium violaceum* ω-transaminase[14] and human DOPA decarboxylase.[15]

Protein solutions were mixed with equal amounts of reservoir solutions. In the absence of PLP, *mj*SHMT crystals grew in 1 month in 1.8*M* NaH_2PO_4/K_2HPO_4 pH 5.0 and were cryoprotected with the same reservoir solution

Table I
Crystallographic Data Reduction, Processing, and Refinement

	apo-*mj*SHMT	holo-*mj*SHMT
Crystallization conditions	NaH_2PO_4/K_2HPO_4 1.8 M pH 5.0	$(NH_4)_2SO_4$ 1.8M + NH_4F 0.1M
PDB code	4BHE	4UQV
Space group	$P2_12_12_1$	$P2_1$
Unit cell dimensions (Å)	$a = 86.59$, $b = 110.16$, $c = 110.69$	$a = 123.11$, $b = 47.15$, $c = 344.0$, $\beta = 90.02$
Resolution range (Å)	50.0–2.83 (2.95–2.83)[a]	50.0–3.0 (3.18–3.0)
mean I/sigma(I)	13.93 (2.5)	6.05 (1.2)
Completeness (%)	98.5 (98.9)	99.4 (97.2)
Multiplicity	5.9 (5.5)	9.1 (8.7)
Rmerge (raw)	0.07 (0.50)	0.28(1.45)
Rmerge (detwinned)	—	0.11
CC1/2	99.5 (80.8)	92.2 (35.4)
Twin fraction	—	0.5
No. mol/ASU	2	12
Refinement statistics		
Resolution range (Å)	50.0–2.83	49.15–3.00
No. reflections	24195	80956
R/Rfree	0.24/0.29	0.20/0.24
No. atoms	6412	40388
RMSD bond length	0.006	0.006
RMSD bond angle	0.99	1.23

[a]Values given in parentheses refer to the highest resolution shell.

plus 30% glycerol. In the presence of 2m*M* PLP, crystals grew within 1 week in $(NH_4)_2SO_4$ 1.8*M* + NH_4F 0.1*M* and were flash-cooled in Na-malonate 40%.

X-ray data collection

Data for *mj*SHMT in the free state (apo-*mj*SHMT) and in complex with PLP (holo-*mj*SHMT) were collected at 100 K at the ESRF (Grenoble, France) and BESSY (Berlin, GE) synchrotron sources, respectively. The best apo-*mj*SHMT dataset was collected at ESRF as 1.0° oscillation frames using the ADSC detector on ID23-1 beamline at a wavelength of 0.972 Å. Data analysis performed with DENZO[16] indicated that the crystals belongs to the P212121 space group with unit cell dimensions: $a = 86.59$ Å; $b = 110.16$ Å; $c = 110.69$ Å. Data were scaled using SCALEPACK[16] and are 98.5% complete at 2.83 Å resolution, with R_{merge}= 0.07. The crystal contains two monomers per asymmetric unit, with $V_M = 2.7$ Å^3 Da^{-1} and 55% solvent content.

Crystals of holo-*mj*SHMT were collected at BESSY as 0.25° oscillation frames using MARCCD detector on ID14-1 beamline at a wavelength of 0.918. Data reduction and scaling were performed by XDS.[17] The data set gave the possibility to index the diffraction pattern as monoclinic or orthorhombic group, because of the very small difference of the distortion index. The highest symmetry group was initially used to integrate reflections. However, Phenix/XTRIAGE[18] gave indication of possible merohedral twinning by means of L-test and Stanley factor.[19] Therefore, the data set was indexed with both space groups and both reflection files were used to solve the structure. However, only the data set indexed in the monoclinic space group provided acceptable R-factor values during refinement procedures (see below and Table I). The unit cell dimensions were: $a = 123.11$ Å; $b = 47.15$ Å; $c = 344.0$ Å; $\beta = 90.02°$. As reported by Hamdane *et al.*[20] and Angelucci *et al.*,[21] the beta angle is very close to 90 degrees and the crystal is twinned with a twin law -h, -k, l. The CC1/2 value (see Table I) was chosen as the criterion to determine the high-resolution limit.[22] The crystal contains 12 monomers per asymmetric unit, with $V_M = 1.8$ Å^3 Da^{-1} and 30.2% solvent content.

Structure refinement and analysis

The structure of apo-*mj*SHMT was solved by molecular replacement using the program PHASER.[23] The 3D structure of SHMT from *(Geo)Bacillus stearothermophilus* (*bs*SHMT) in Protein Data Bank (PDB)[24] coordinate file 1KL2 was used as search model because: (i) *bs*SHMT has the highest sequence identity (39%) with *mj*SHMT as measured by BLAST;[25] and (ii) coordinate file 1KL2 contains the only structure of wild-type *bs*SHMT determined in complex with both PLP and pteridine cofactors (see Supporting Information Table S1).

Refinement of atomic coordinates and displacement parameters was carried out using Refmac5.[26] Model building was performed using the program COOT.[27] Repeated cycles of model building and refinement were necessary to build several protein regions, including the C-termini. The structure was refined at 2.83 Å resolution, to an R_{factor} (working set) of 24.4% and an R_{free} of 29.7%. The geometry was checked using the program PROCHECK.[28] The allowed or generously allowed regions of the Ramachandran plot comprise 100% of the residues. Atomic coordinates and structural factors have been deposited in the PDB with accession code 4BHD (see Table I).

According to Chandra *et al.*[29] and Saccoccia *et al.*,[30] twinned intensities of holo-*mj*SHMT were directly used for the molecular replacement procedure. The crystal structure of apo-*mj*SHMT was used as search model. PHASER found 12 monomers in the asymmetric unit, according to the solvent content analysis (see above). Refinements and data detwinning were performed with Phenix/REFINE,[18] providing the twin operator. TLS refinement was performed using three groups for each polypeptide chain.[31] The program calculated the twin fraction to be close to 0.5. As a further control, Refmac5[26] finds similar twinning fractions and estimates the detwinned R_{merge} reported in Table I. Model building was performed using COOT.[27] A summary of refinement parameters is given in Table I. The geometry was checked using the program PROCHECK.[28] Atomic coordinates

and structure factors have been deposited in the PDB with accession code 4UQV (see Table I).

Modeling of holo-mjSHMT in complex with H_4MPT

As H_4MPT is not commercially available, we built a model of *mj*SHMT in complex with H_4MPT. We used the G/H dimer of holo-*mj*SHMT since, together with dimer I/J, it is the most complete (see Supporting Information Table S2), and the structure of *bs*SHMT dimer in PDB coordinate file 1KL2 as a template for the pteridine cofactor. This structure was chosen for both the aforementioned and the following reasons: (i) *bs*SHMT is the most extensively characterized SHMT representative; 39 different structures (containing 41 nonequivalent monomers) are available, determined for wild-type or mutant forms, either in the free state or in complex with different combinations of ligands[32] (Supporting Information Table S1); and (ii) based on the structure analyses described below, this is the SHMT structure most similar to *mj*SHMT among those determined in complex with a H_4PteGlu derivative (see Supporting Information Table S3).

The A/B chains of *bs*SHMT were optimally superimposed on the G/H dimer of holo-*mj*SHMT by least square fit superposition of CA atoms using the Protein structure comparison service Fold at the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/ssm)[33] (root-mean square deviation (RMSD) value = 1.9 Å over 749 atom pairs). H_4MPT coordinates were extracted from the crystal structure of $N5$,N_{10}-Methenyl-tetrahydromethanopterin cyclohydrolase from *Archaeoglobus fulgidus* (PDB code: 4GVQ)[34] and manually superimposed on H_4PteGlu in complex with *bs*SHMT exploiting the common structural features of the two molecules. First, the pteridine rings of H_4MPT and H_4PteGlu were optimally superimposed to achieve a correct orientation of H_4MPT pteridine ring with respect to PLP, which has been demonstrated to be a requirement for the progress of the catalytic reaction.[5] Second, the superposition between the 1,4-substituted phenyl rings common to both cofactors was optimized to avoid clashes with protein residues. The conformation of the hydrophilic anionic portion of H_4MPT was not modified with respect to the H_4PteGlu template because no positively charged cluster was found at the rim of the cofactor binding site. Finally, structure idealization of the protein-cofactor complex was carried out.[26]

Structure and sequence comparisons

The coordinate files of all the available, experimentally determined SHMT structures (Supporting Information Table S1) were downloaded from the PDB.

Structurally conserved regions between apo-*mj*SHMT A-B chains or holo-*mj*SHMT A-L chains (SCRs, Supporting Information Table S2) and between all apo- and holo-*mj*SHMT monomers (SCRs-*mj*, Supporting Information Table S2), were identified as follows. The CA atoms of structurally equivalent residues were optimally superimposed by least-square fit superposition using the lsqkab program[35] of the CCP4 package.[36] The pair of equivalent residues with the highest CA–CA distance was removed and structural superposition of the remaining regions performed. The procedure was repeated until all CA–CA distances of equivalent residue pairs were below 3.0 Å. In the case of holo-*mj*SHMT, comprising six dimers, all other chains (i.e., A—F and H—L) were compared to chain G, for which all residues could be fitted in the electron density map.

The Protein structure comparison service Fold[33] was used to generate: (i) a multiple structure alignment of 102 nonidentical SHMT monomers, that is, the two apo- and twelve holo-*mj*SHMT, listed in Supporting Information Table S2, plus the 88 nonidentical monomers in the 58 SHMT structures from eukarya and bacteria available from the PDB and listed in Supporting Information Table S1; and (ii) pair-wise structural alignments between holo-*mj*SHMT G chain and 13 nonredundant SHMT monomers, which were combined in the alignment reported in Supporting Information Table S4. The 13 nonredundant monomers aligned to *mj*SHMT G chain are representative of the SHMTs from different sources available from the PDB. To obtain this nonredundant set, when more than one nonidentical structure from the same species and/or cellular compartment was available from the PDB, monomers were selected according to the following criteria: (i) wild-type over mutant structures; (ii) presence of PLP, amino acid and pteridine cofactor; (iii) chain completeness.

The Protein Interfaces, Surfaces and Assemblies (PISA) server[37] was used to calculate interface areas and estimate the solvation free energy gain on interface formation ($\Delta^i G$).

The volume of the folate cofactor binding cavity of *mj*SHMT and *bs*SHMT was calculated using the Computed Atlas of Surface Topography of proteins (CASTp) server.[38]

Protein substructures similar to the 19-residue C-terminal tail of *mj*SHMT ("P-motif") were searched for using the following servers: VAST (Vector Alignment Search Tool);[39] RASMOT 3D PRO;[40] ProBis 2012;[41] SALAMI;[42] DALI;[43] FATCAT (Flexible structure AlignmenT by Chaining Aligned fragment pairs allowing Twists).[44] Depending on the server, searches are performed either in the whole PDB or in a nonredundant PDB-subset comprising proteins with less than 100, 95, or 90% sequence identity.

The Naccess V2.1.1—Atomic Solvent Accessible Area Calculations program (http://www.bioinf.manchester.ac.uk/naccess/)[45] was used to calculate surface accessible surface area (SASA) of SHMT residues in subunits listed in Supporting Information Table S4 and interacting

Table II
Variance of rmsd values Calculated After Pair-Wise Structure Superimpositions of 102 Nonredundant SHMT Monomers (see Table S3)

DoL	Total variance $(RMSD)^2$	Number of comparisons (N_c)	Normalized Variance $(RMSD^2/N_c)$
A2A	101.8	91	1.1
B2B	1362.9	2278	0.6
E2E	41.2	190	0.2
A2B	2288.5	952	2.4
A2E	727.1	280	2.6
B2E	1737.6	1360	1.3
A2A+B2B	1464.7	2369	0.6
A2A+E2E	143	281	0.5
B2B+E2E	1404.1	2468	0.6

The Cα atoms of 306 residues that are structurally conserved in all the monomers (see Table S4) were used for all pair-wise structure superimpositions and RMSD value calculations. DoL: domain of life to which the superimposed monomer structures belong. A, archaea (*mj*SHMT only), 14 subunits; B, bacteria, 68 subunits; E, eukarya, 20 subunits; A2A, B2B, and E2E, monomers belonging to the same domain of life were compared to one another. A2B, A2E, and B2E, monomers belonging to each domain of life were compared to monomers belonging to a different domain of life.

monomers. Based on side-chain SASA values, residues were assigned to different protein regions: (i) SASA ≤ 20 Å^2: buried; (ii) SASA > 20 Å^2 in the dimer: solvent accessible; (iii) difference in SASA between monomer and dimer > 20 Å^2: dimeric interface. According to this definition, while buried residues are unique, a small number of residues (whose side-chains had SASA > 20 Å^2 in the dimer in spite of losing > 20 Å^2 SASA on dimer formation) resulted to be both solvent accessible and located at the interface. Eukaryotic SHMT residues involved in the interface between the two dimers were excluded from the solvent accessible set.

Structure visualizations and analyses were performed using InsightII (Accelrys). Residues were defined to be in contact if they have at least two atoms at a distance ≤ 4.0 Å from each other.

SHMT sequences assigned to different taxonomic groups (see Table III) were downloaded from the NCBI web site (http://www.ncbi.nlm.nih.gov/) and locally aligned using ClustalX.[46]

RESULTS AND DISCUSSION

Structure determination

We determined the crystal structure of *mj*SHMT in the absence of PLP (apo-*mj*SHMT) at 2.83 Å resolution and in complex with PLP (holo-*mj*SHMT) at 3 Å resolution (Fig. 1). Refinement statistical parameters are reported in Table I.

Crystals of apo-*mj*SHMT and holo-*mj*SHMT belong to the orthorhombic and monoclinic space groups, respectively. Crystals of the monoclinic form were revealed to be twinned by pseudomerohedry, with a percentage of twinning close to 50%. The small distortion from the orthorhombic space group (the distortion of the beta angle is only 0.02 degrees) is not unusual, and similar cases are reported in the literature.[20,21] However, the structure in the monoclinic form was refined to an R_{free} of 0.24. The refinement statistical parameters reported in Table I demonstrate that the space group is correct.

Structural comparison of apo- and holo-*mj*SHMT subunits

The asymmetric units of apo- and holo-*mj*SHMT crystals contain one and six functional dimers, respectively. In the model built for the holo-protein, chain G and J are complete while the other chains lack just a few (from two to nine) residues (Supporting Information Table S2). These belong almost entirely to surface loops or, in just two chains, are in contact with the phosphate moiety of PLP at the dimer interface. In both chains of apo-*mj*SHMT, a large region (residues 49–60) corresponding to a PLP binding loop is missing; additionally, a loop belonging to the binding site for the pteridine cofactor (residues 126–138) and a surface loop interacting with it (residues 106–111) are not visible in chain B.

Comparisons between all pairs of apo- and holo-*mj*SHMT monomers show that a large portion of the structure has the same conformation in all monomers (see Materials and Methods). The CA atoms of the structurally conserved regions [SCRs-*mj*, Fig. 1(C) and Supporting Information Table S2], comprising 354 residues, can be optimally superimposed to equivalent residues of a different monomer yielding RMSD values of 0.5 Å for apo-*mj*SHMT A versus B chain; in the range 0.7–1.0 Å for holo-*mj*SHMT G chain versus all other chains; and in the range 0.7–1.6 Å for all apo-*mj*SHMT versus all holo-*mj*SHMT chains (Supporting Information Table S3). The regions outside the SCRs-*mj*, whose conformation varies between different monomers, comprise residues 1–2, 45, 47–61, 106–111, 120, 124–138, 160–161, 209–210, 241–244, 251–258, 308–310, 321–327, 352–357, 392–393, 428 [Fig. 1(C)]. These residues are located within loops either on the protein surface or belonging to the binding sites for the PLP cofactor (missing in apo-*mj*SHMT) and/or H_4MPT (missing in both structures).

Comparison with known SHMT structures

To investigate the extent of structural similarity between the archaeal *mj*SHMT and SHMTs from different domains of life all holo- and apo-*mj*SHMT monomers were compared with all the nonidentical monomers present in the SHMT structures from eukarya and bacteria available from the PDB (see Supporting Information Table S1).

The automatically generated multiple structural alignment of the 102 nonidentical SHMT monomers (data

Table III
Frequency of Residues Occurring at Positions equivalent to *mj*SHMT Q232 and N257 in the Multiple Sequence Alignment of 1665 SHMT Sequences from Different Sources

DoL	Source	Nb. Seqs.	*mj*−232	*mj*−257
E		505	R	G (99)[a]
B		988	R[b]	G
A		156	R (37)	G (37)
			Q (62)	N (58)
			N (1)	S (5)
	Halophilacee	1	R	G
	Euryarchaeotes			
	Halobacteriaceae	32	R	G
	Methanosarcinales	16	R	G
	Methanomicrobiales	8	R	G
	Methanococcales	6	Q	N
	Methanobacteriales	9	Q	N
	Methanocellales	3	Q	N
	Methanopyrales	1	Q	N
	Aciduliprofundum	3	Q	N
	Archaeoglobales	5	Q	N
	Thermococcales	14	Q	N
	Thermoplasmatales	3	Q	N
		1[c]	N	S
	Methanomethylophylus	1[d]	N	S
	Korarchaeacee	1	Q	N
	Crenarchaeotes	46	Q	N
		1[e]		S
	Cenarchaeales	1	Q	S
	Nitrosopumilales	3	Q	S
	Nitrososphaerales	1	Q	S

DoL, domain of life; A, archaea; B, bacteria; E, eukarya. Nb. Seqs., number of sequences. *Mj*−232 and *mj*−257, residues at positions equivalent to *mj*SHMT Q232 and N257, respectively. When residues are not 100% conserved, percentages of residue occurrence are indicated in parenthesis.
[a]The remaining 5 sequences have Q (3), V (1), or R (1).
[b]K in *M. haemofelis*.
[c]*T. archaeon*.
[d]*Candidatus M. alvus*.
[e]*A. saccharovorans*.

not shown) highlighted the occurrence of relatively large conformational variations. These changes are likely due to the fact that the aligned structures have been determined in different states of ligation (i.e., in the presence or absence of PLP, pteridine cofactor and amino acid substrates), and include mutant, as well as wild-type, structures (see Supporting Information Table S1). Nevertheless, a large portion of the SHMT monomer, comprising 306 residues, was found to assume a conserved main-chain conformation in all the analyzed structures (SCRs-all, see Supporting Information Table S4). The RMSD values calculated after all-against-all SHMT monomer pair-wise superpositions of CA atoms within the SCRs-all were all below 2.0 Å (Supporting Information Table S3). Analysis of these RMSD values showed that, even in the SCRs-all, structures from organisms belonging to each domain of life are more similar to one another than to structures belonging to a different life domain (Table II). Additionally, SHMT structures from the bacterial and eukaryotic domains of life appear to be more similar to one another than to the archaeal *mj*SHMT (Table II).

Overall fold

In agreement with the relatively high sequence similarity with the other SHMTs of known 3D structure (~31–38% sequence identity in the SCRs-all, see Supporting Information Table S4) the overall fold (Fig. 1) resembles that of eukaryotic and bacterial SHMTs. In particular, the two large domains, that is, the catalytic (residues 25–282, made of three layers α/β/α) and C-terminal (residues 283–410, with an α/β fold) domain, and the N-terminal α-helix (residues 1–24), which packs with the C-terminal domain and with the opposite monomer, are highly conserved with respect to *bs*SHMT-1KL2.

In addition to these conserved regions, *mj*SHMT contains a 19-residue C-terminal tail (411–429), which is paired with the equivalent region of the twofold symmetry related subunit and with the conserved N-terminal α-helix [Fig. 1(B)]. The conformation of this distinguishing region resembles the shape of an upper case "P" letter and shows neither structural nor sequence similarity with other SHMTs (Supporting Information Table S4). Additionally, database searches performed using several structure comparison programs (see Materials and Methods) did not highlight clear similarities of the "P-motif" with other known protein structures (see below).

Catalytic site

Residues interacting with PLP in holo-*mj*SHMT are highlighted in Supporting Information Table S4. As previously reported, most of the residues known to assist PLP binding and catalysis conserve their identity in *mj*SHMT.[12] Comparison of holo-*mj*SHMT structure with SHMTs from other species allowed us to identify three sets of residues in *mj*SHMT: (i) involved in identical or very similar interactions with the PLP cofactor to those occurring in other structures (Fig. 2); (ii) expected to be involved in interactions with amino acid substrates or pteridine cofactor (that are not present in the structure of *mj*SHMT), based on the conservation of residue main-chain conformation and identity of whole side-chains or side-chain functional groups; (iii) potentially involved in H_4MPT binding, based on their structural location with respect to residues in contact with H_4PteGlu in other SHMTs (Supporting Information Table S4). Although the large majority of functional interactions occurring in other SHMT structures either are, or are expected to be conserved in *mj*SHMT, a few intriguing differences were observed.

We previously demonstrated that the proton exchanging ability of *ec*SHMT Y55 is modulated by a cation-π interaction[47–50] with R235, and that this interaction is conserved in all SHMT structures available at the time.[6] Indeed, the R235Q mutation in *ec*SHMT resulted in a reduction of the catalytic efficiency by about five orders of magnitude and an increase in stability with respect to

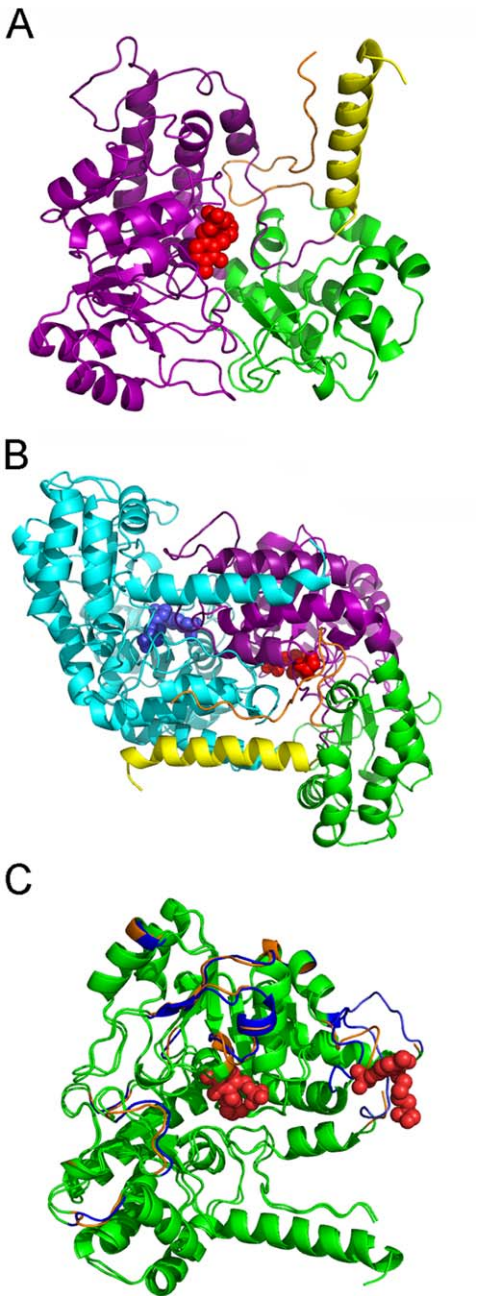

Figure 1

3D structure of *mj*SHMT. **A)** The G monomer of holo-*mj*SHMT is colored by domain type: yellow, small N-terminal α-helix (residues 1–24); purple, catalytic domain (residues 25–282); green, C-terminal domain (residues 283–410); orange, 24 C-terminal residues (411–429). PLP is shown as spheres and colored red. **B)** G/H dimer. The G monomer is colored as in panel A, the H monomer is cyan and the PLP blue. **C)** Ribbon representation of *mj*SHMT holo G and apo A monomers after least square structure superposition of the structurally conserved regions (SCRs-*mj*, see Methods and Supporting Information Table S3). The SCRs-*mj* of both proteins are colored green, the remaining regions of the holo and apo dimers are blue and orange, respectively. Both PLP molecules bound to holo-*mj*SHMT G/H dimer are shown as sticks and colored red. The picture was generated with PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC).

the wild type enzyme.[6] Based on these results, we hypothesized that R235 may be a "frustrated residue," providing a functional advantage at the expense of a detrimental effect on protein stability, due to the net positive charge of the arginine.[51] Interestingly, while *ec*SHMT Y55 is conserved in *mj*SHMT (Y50), *ec*SHMT R235 is replaced by *mj*SHMT Q232 (Fig. 2 and Supporting Information Table S4), whose side-chain is too short to interact with Y50. In eight of the twelve holo-*mj*SHMT subunits of the asymmetric unit the π-cloud of the invariant Y50 is in contact with the side-chain amide moiety of N257 (see Figs. 2 and S1), the distance between the centroid of the aromatic ring and the CG atom of the amide moiety of the two residues being in the range 3.8–4.2 Å and the Y50-N257 CA-CA distance being highly conserved (5.7 ± 0.6 Å). Although in some subunits both orientations of N257 side-chain amide moiety appear to be compatible with other groups present in the chemical environment, the positioning of the ND2 atom toward the aromatic ring of Y50 would allow an attractive amide-aromatic interaction between these side-chains to be established, and the OD1 of N257 to be at a distance >4 Å from the centroid of Y50. Conversely, the alternative orientation would place the OD1 of N257 and the Y50 centroid at a distance <4 Å, resulting in a repulsive interaction.[52] Amide-aromatic contacts are characterized by variable stereochemistry and can achieve binding energies similar to or higher than those involved with cation-π interactions over a wide configurational space.[52,53] This high interaction energies have been ascribed to the contribution of both van der Waals (due to π-π stacking) and electrostatic (due to the electron charge distributions in both groups) components.[52,53] In the remaining four subunits an amide-aromatic interaction was not detected, either because N257 (in chains K and L) or both N257 and Y50 (in chain C) are not visible in the electron density map and/or because Y50 is shifted from the canonical position that it assumes in all other SHMT structures (in chains J, K and L).[6]

Inspection of a multiple sequence alignment of SHMTs from a few eukaryotic and prokaryotic species had previously suggested that the substitution of *ec*SHMT R235 with glutamine might be a general feature of methanogenic bacteria.[12] To exhaustively investigate the extent of conservation and phylogenetic distribution of the residues involved in the cation-π and amide-aromatic interactions, respectively, we generated a comprehensive multiple sequence alignment of SHMT sequences from ~1650 species (data not shown) and calculated the frequency of occurrence of residues at positions equivalent to *mj*SHMT Y50, Q232, and N257 (Table III). The tyrosine residue equivalent to *mj*SHMT Y50 is completely conserved across all domains of life and the position equivalent to *mj*SHMT Q232 (and *ec*SHMT R235) is arginine in all bacterial and eukaryotic species.

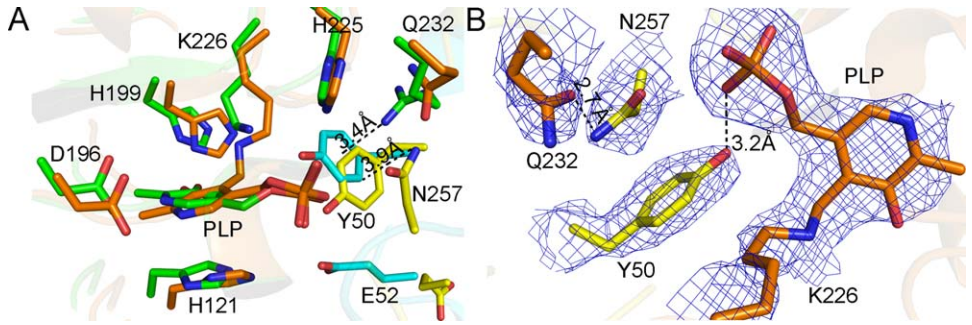

Figure 2

Catalytic site of holo-*mj*SHMT. **A)** Comparison with the catalytic site of *bs*SHMT. Residue numbering refers to *mj*SHMT. The A/B subunits of holo-*mj*SHMT and *bs*SHMT are shown as yellow/orange and cyan/green ribbons, respectively. Residues belonging to the catalytic sites and the bound PLP molecules are shown as sticks and colored by atom type: N, blue; O, red; C, orange/yellow and cyan/green for the A and B subunits of *mj*SHMT and *bs*SHMT, respectively. Dashed lines indicate: (i) in *mj*SHMT, the distance between the centroid of Y50 aromatic ring and the ND2 atom of N257, involved in the putative amide-aromatic interaction; (ii) in *bs*SHMT, the cation-π interaction between Y51 and R232 (structurally equivalent to *mj*SHMT Y50 and Q232, respectively). **B)** 2Fo-Fc electron density map contoured at 1σ (in blue) of holo-*mj*SHMT Y50 and N257 residues in the A monomer, involved in the putative amide-aromatic interaction, and Q232 and PLP-bound K226 in the B monomer. Dashed lines indicate the distances between ND2 of N257 and OE1 of Q232, and between Y50 side-chain oxygen and the phosphate moiety of PLP.

Conversely, arginine and glutamine occur in less than 40% and more than 60% of the SHMT sequences from archaeal organisms, respectively. Strikingly, in nearly all SHMT sequences where the residue at position equivalent to *mj*SHMT Q232 is glutamine the residue equivalent to N257 is asparagine, whereas in SHMTs where the residue corresponding to *mj*SHMT Q232 is arginine the residue equivalent to N257 is glycine, as observed in the known structures (Table III). This strong residue covariation supports the hypothesis that in SHMTs presenting a tyrosine-asparagine pair at positions corresponding to *mj*SHMT Y50-N257, amide-aromatic contacts occur in place of the classic cation-π interaction present in the large majority of known SHMTs. It is interesting to notice that almost all organisms where the Y50-N257 pair occurs are either hyperthermophiles (87%) or thermophiles (7%), whereas the tyrosine-arginine pair, involved in the cation-π interaction, is mostly present in SHMTs from mesophiles and halomesophiles (90%), very few psychrophiles (7%) and thermophiles (3%) and no hyperthermophile. Both structure and sequence analysis results are in agreement with our previous hypothesis that R235 of *ec*SHMT is a frustrated residue,[6] and extend this hypothesis to other SHMTs with arginine at the equivalent position. Additionally, these data led us to formulate the hypothesis that the amide-aromatic interaction between Y50 and N257 in *mj*SHMT and possibly other SHMTs from organisms that live in extreme environments, might have been selected because of its ability to increase stability due to the elimination of the net charge of the interacting arginine-tyrosine pair (as reported by Elcock).[54] The possible loss of catalytic efficiency consequent to the loss of the arginine might be compensated for, in (hyper)thermophiles, by a number of factors, including the effect of high temperatures on the rate of the catalytic reaction and polarization of the tyrosine OH group due to the amide-aromatic interaction with the asparagine.

In apo-*mj*SHMT, the loops surrounding the active site are highly flexible and only partially visible in the electron density map. In the active site, Y50, E52, and Y60 are not visible, as well as several residues putatively involved in pteridine cofactor binding. Residues N257 and, in the B subunit, H121, which in holo-*mj*SHMT are involved in the putative amide-aromatic contact with Y50 and stacking with the PLP cofactor, respectively, are oriented toward the empty PLP binding site cavity.

H_4MPT binding site

*Mj*SHMT is the first reported structure of SHMT using H_4MPT as the preferred pteridine cofactor. H_4PteGlu can also bind to *mj*SHMT and take part in the

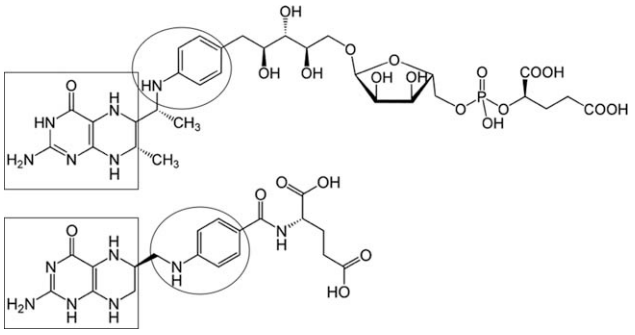

Figure 3

Structural formulas of H_4MPT (top) and H_4PteGlu (bottom). Boxes and circles highlight similar structural features, namely the pteridine and 1,4-substituted phenyl rings, respectively.

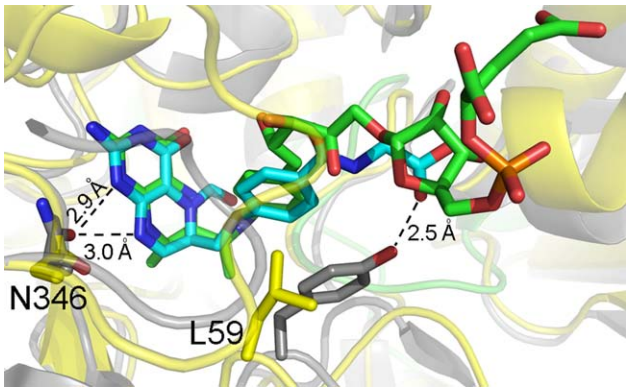

Figure 4

Comparison between the active sites of *mj*SHMT (yellow cartoons), where H_4MPT has been modelled, and *bs*SHMT (grey cartoons), whose structure has been determined in complex with [6R]-5-formyl-5,6,7,8-tetrahydrofolate (PDB code: 1KL2). Residue numbering refers to *mj*SHMT. The cofactors are shown as sticks and colored by atom type: N, O and P atoms are blue, red and orange, respectively; C atoms are green in H_4MPT and cyan in H_4PteGlu. Dashed lines indicate hydrogen bonds between *bs*SHMT N341 (structurally equivalent to *mj*SHMT N346) and nitrogen atoms of H_4PteGlu pteridine ring, and between Y60 (structurally equivalent to *mj*SHMT L59) and the terminal carboxylate moiety of the cofactor.

hydroxymethyltransferase reaction, but this is 450-fold less efficient than in the presence of H_4MPT.[12]

H_4MPT shows both structural similarities and significant differences with respect to H_4PteGlu (Fig. 3). Both cofactors comprise one pteridine and one phenyl ring. However, H_4MPT is characterized by: (i) larger bulk, in that it contains several branched methyl and hydroxyl groups and one sugar-phosphate moiety between the phenyl ring and the terminal carboxylate group; and (ii) increased flexibility, due to the higher degrees of freedom of its backbone; for example, the phenyl ring, which is linked to a methylene group, is more mobile than in H_4PteGlu, where the electron withdrawing C(O)NH moiety is able to reduce ring rotation due to resonance effects.

A model of the complex between H_4MPT and *mj*SHMT was built and compared with SHMT structures determined in complex with H_4PteGlu (Fig. 4). The aim was to identify *mj*SHMT residues likely to be involved in cofactor binding and potentially responsible for the specificity toward H_4MPT over H_4PteGlu. The folate binding site is localized in a deep pocket between the two subunits of the dimer. Comparative analysis of the *mj*SHMT and *bs*SHMT folate binding pockets indicates that the former is substantially larger (the volumes being 724 and 357 $Å^3$, respectively) due to differences in backbone architecture. Although the crystal structure of *bs*SHMT bound to 5-formyl-H_4PteGlu (PDB code: 1KL2) has been used for the comparison, the cavity volume of the enzyme in the absence of the cofactor (PDB code: 1KKJ) is very similar. Specific clusters of positive amino acid residues that may interact with terminal polar groups of the cofactor were not detected at the protein rim. However, the lack of specific contacts and consequent disorder of the terminal portion of the cofactor is a common feature shared by 3D structures of unrelated enzymes that use H_4MPT.[34,55]

In agreement with the structural similarity of the pteridine and phenyl rings of the two cofactors, most of the residues that contact these chemical groups in other SHMT structures are either conserved in *mj*SHMT or replaced by residues with similar chemical-physical properties, enabling them to establish analogous interactions with the cofactors (Supporting Information Table S4). As an example, the crucial asparagine (N341 in *bs*SHMT) that is hydrogen bonded to nitrogen atoms of the heterocycle, and the phenylalanine (F251 in *bs*SHMT) that packs with the phenyl ring of the cofactor, are conserved in *mj*SHMT (N346 and F251) (Fig. 4) and expected to interact with the folate cofactor in the same way. Non conserved *mj*SHMT residues that might establish conserved interactions include: the hydrophobic side-chain groups of L172, which share the ability of alanine and cysteine residues of other SHMTs to interact with the pteridine ring of the folate cofactor; N354, N355, S356, and D357, whose structural location is compatible with the establishment of polar interactions with the hydroxyl groups or terminal carboxylate moieties of H_4MPT; and V116 and I122 main-chain atoms (O and N, respectively) that may bind the primary amino substituent and carboxylic oxygen of H_4MPT pteridine ring (Fig. 3) like the equivalent atoms of leucine residues conserved in other SHMTs (e.g., *bs*SHMT residues L117 and L123) bind the corresponding chemical groups of H_4PteGlu. Interestingly, the tyrosine residue corresponding to *bs*SHMT Y60, which binds H_4PteGlu in SHMT structures from bacterial and eukaryotic species, is replaced by L59 in *mj*SHMT. This tyrosine residue, conserved in 93% of the analyzed SHMT sequences (see Materials and Methods), forms a stacking interaction with the p-aminobenzoic acid ring of the cofactor and a hydrogen-bond with the terminal carboxylate group. The side-chain of leucine at position 59 of *mj*SHMT, which is present in ~1% of SHMT sequences, may still establish a hydrophobic interaction with the phenyl ring of H_4MPT. Consistent with the lack of hydrogen bonding ability of this residue, the terminal carboxylate group of H_4MPT is located at a greater distance with respect to H_4PteGlu. Accordingly, all SHMTs where *mj*SHMT L59 is conserved are from methanogenic archaea, many of which also use H_4MPT as cofactor. Furthermore, in 5% of the analyzed sequences, belonging to both bacterial and archaeal (methanogenic and nonmethanogenic) species, L59 is replaced by phenylalanine. The side-chain of this residue can establish a stacking interaction with the phenyl ring of the cofactor similar to that of Y60 in *bs*SHMT but,

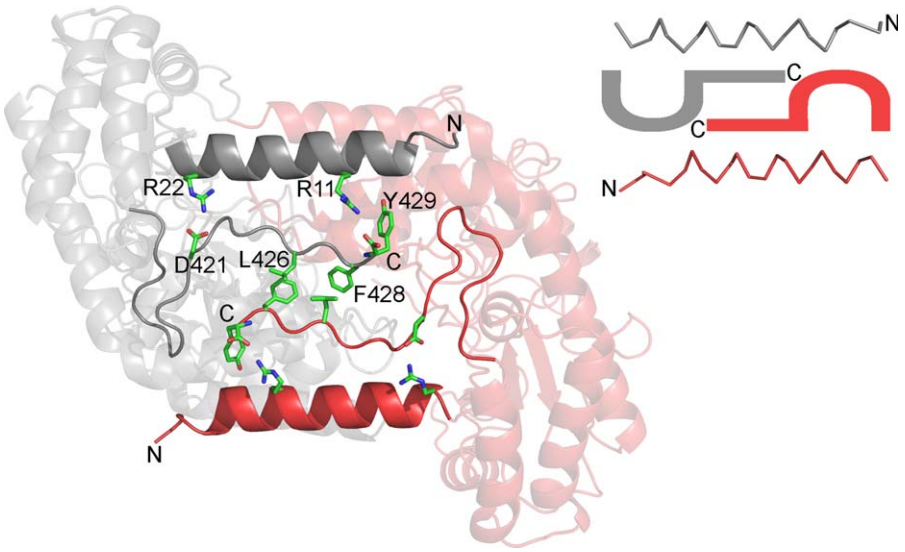

Figure 5

Cartoon representation of *mj*SHMT physiological dimer. The two monomers are colored grey and red. The N-terminal α-helix and the unique *mj*SHMT C-terminal region at the dimer interface are shown in bright colors; the conserved catalytic and C-terminal domains of both monomers are shaded. Residues involved in salt-bridges (between the N- and C-terminus within each monomer) or hydrophobic contacts (between the C-termini of the two monomers) are shown as sticks and colored by atom type: N, O, and C atoms are blue, red, and green, respectively. The conformation of the unique *mj*SHMT C-terminal regions is schematically depicted at the top-right corner.

like *mj*SHMT L59, is deprived of hydrogen bonding ability. The remaining 1% of SHMT sequences contains residues whose side-chain properties are similar to either leucine or tyrosine (i.e., able to establish either hydrophobic interactions, like valine and cysteine, or both stacking and hydrogen bond interactions, like histidine and tryptophan). These data strongly suggest that the presence of a residue that either does or does not possess side-chain hydrogen-bond properties at this position correlates with preferential usage of H_4PteGlu and H_4MPT, respectively, as a cofactor in the hydroxymethyltransferase reaction.

C-terminal region

The last 19 amino acids at the C-terminus of *mj*SHMT do not show any structural or sequence similarity to other SHMT proteins (see Supporting Information Table S4). Intriguingly, structure comparison programs did not detect regions with significantly similar conformation in protein structures with different folds either (see Materials and methods).

This apparently unique C-terminal region resembles the shape of an uppercase "P" letter and it will be indicated as P-motif hereafter (Fig. 5). The P-motif contributes 12% of the total *mj*SHMT dimer interface (1044 vs. 8840 Å^2, respectively). As shown in Figure 5, the P-motifs of the twofold symmetry related subunits interact at the dimer interface in a head-to-tail fashion, with the straight stretch of one motif contacting the corresponding portion of the other. The two P-motifs are embedded between the two N-terminal alpha helices, generating a continuous array of interactions between the N-termini and the C-termini of the two monomers. The N-terminus and C-terminus of the same monomer are connected by a series of salt-bridges, whereas the interactions between the two C-termini are mainly hydrophobic (Fig. 5). Theoretical calculations support the importance of this network of interactions. The PISA server estimates a difference of 9 kcal/mol between the free energy of assembly dissociation (Δ^iG) of the holo-*mj*SHMT G/H dimer (−62 kcal/mol) and that of a truncated form computationally generated by deleting the P-motifs of both monomers (−53 kcal/mol).

It has been demonstrated that the quaternary assemblies of SHMTs from bacteria are stabilized mainly by interactions of their C-terminal domains.[56] Additionally, enzymes that must resist to extreme physical-chemical conditions need to maintain the stability of their quaternary structure. This is especially true in case of enzymes whose active site is formed by residues contributed to by adjacent subunits, as is the case for PLP enzymes. Indeed, hydrophobic interactions at intersubunit interfaces have been proposed, both based on experimental[57,58] and computational[59,60] studies to contribute to the stabilization of oligomeric proteins from (hyper)thermophilic organisms. In this light, the peculiar C-terminal fold of *mj*SHMT may be exploited to engineer new

chimeric SHMTs or, possibly, other enzymes sharing the same overall fold (e.g., those belonging to the GABA-aminotransferase-like family, PLP-dependent transferases superfamily or PLP-dependent transferase-like fold, according to the classification in the SCOP database)[61] with improved chemical-physical stability.

Additional factors putatively contributing to *mj*SHMT thermostability

A number of physical-chemical features have been previously reported to be involved in protein adaptation to extreme temperatures. In particular, three strategies have been proposed to be mainly responsible for thermal stability in SHMTs[60]: (i) increased number of charged residues on the protein surface; (ii) increased hydrophobicity of the core; and (iii) decrease of thermolabile residues exposed to the solvent. This hypothesis was based on the results of the analysis of 53 sequences and three structures (from *E. coli*, *H. sapiens* and *O. cuniculus*) of mesophilic SHMTs, and of the sequences and molecular models built for 10 thermophilic/hyperthermophilic SHMTs, including *mj*SHMT.

Taking advantage of the now available 3D structure of *mj*SHMT and of SHMT structures from the nine bacterial and four eukaryotic species, we performed a comparative analysis of residues present on the protein surface, interior and dimeric interface (Supporting Information Table S5). The most relevant difference between *mj*SHMT and the other structures appeared to be the higher number of Glu residues exposed on the protein surface (37 in *mj*SHMT vs. a maximum of 30 and 28 in bacterial and eukaryotic structures, respectively). The higher frequency of Glu residues in thermophilic versus mesophilic organisms has been previously reported and explained in terms of formation of additional salt-bridges.[60] However, visual inspection indicated that *mj*SHMT contains only one and two ion-pairs more than *ec*SHMT and *bs*SHMT, respectively, suggesting that the contribution of this interaction to *mj*SHMT thermostability is limited.

We detected a number of additional differences between *mj*SHMT and the mesophilic counterparts, but all of them were rather small. These include some previously highlighted features,[60] such as: (i) number of charged side-chains other than Glu on the protein surface (in *mj*SHMT Lys is only slightly more represented, while Asp and Arg are both within the range observed in other structures); (ii) number of hydrophobic side chains in the core (with respect to the ranges of these residues in bacterial and eukaryotic structures, *mj*SHMT had just one more Met); (iii) number of exposed thermolabile residues (although the number of Asn and Cys in *mj*SHMT is low, it falls within the range observed in other structures, and there is one more Gln); (iv) number of His residues at the interface (two and one more than the maximum number in bacterial and eukaryotic homologues, respectively). Additionally, slight variations were observed in the distribution of other residues (such as decreased number of Gly and Thr on the surface and of Val in the core, and increased number of Ser in the core). Determination and analysis of additional thermophile/hyperthermophile structures would be required to determine whether these differences may be significant.

CONCLUSIONS

SHMT is a key enzyme in cellular one-carbon pathway and representatives from many living organisms, from bacteria to higher plants and mammals, have been extensively studied. However, structural characterization of SHMTs from either archaea or hyperthermophilic organisms had not been reported to date.

The structural study at the atomic level of SHMT from *M. jannaschii* increases our knowledge about the mechanism of adaptation of this bacterium that lives in extreme conditions, at temperatures close to 100°C and high pressure. Moreover, due to its stability under extreme chemical-physical conditions, *mj*SHMT may be used in biomimetic reactions and is ideally suited for industrial processes. Therefore, knowledge of structure-function relationships in this protein assumes a biotechnological relevance.

*Mj*SHMT is the first SHMT structure to be solved both in the free (apo) and PLP-bound (holo) state. PLP binding induced only local conformational changes in *mj*SHMT, while both the overall architecture and quaternary structure were unaffected. This is at odds with what is observed for fold-type I PLP-dependent DOPA decarboxylase, where dramatic structural changes associated with cofactor binding were observed.[15] Importantly, comparison of *mj*SHMT with available SHMT structures and sequences revealed several idiosyncratic features of this protein, based on which it is possible to rationalize its peculiar functional and stability properties.

At the active site, the 'frustrated' arginine residue, which in mesophiles is involved in a cation-π interaction with an invariant tyrosine residue, is substituted by a noncharged glutamine residue. We have previously shown that the arginine to glutamine substitution in *ec*SHMT partially preserves the acidic properties of Y55 (Y50 in *mj*SHMT) necessary for catalysis, while relieving the destabilizing effect due to the presence of a positive charge. Examination of the *mj*SHMT structure revealed that Y50 is in contact with N257, which is not present in previously determined SHMT structures, and establishes an amide-aromatic interaction with it. This observation led us to hypothesize that removal of the frustrated Tyr-Arg interacting pair present in mesophiles with a Tyr-Asn pair in *mj*SHMT and, possibly, other (hyper)thermophiles may be a novel strategy used to achieve high thermal stability.

The X-ray structure of *mj*SHMT also sheds light on the reasons why tetrahydromethanopterin (H_4MPT) is preferred by methanogenic archaea over H_4PteGlu, which is used by eubacterial and eukaryotic organisms. The main differences with respect to homologous proteins consist in a wider cofactor binding pocket and a few single point mutations. Among them, a discriminating feature appears to be the presence of L59 in *mj*SHMT and other methanogens in place of the tyrosine (Y60 in *bs*SHMT), which is conserved in almost all other species. While tyrosine forms both a stacking interaction with the p-aminobenzoic acid ring and a hydrogen bond with the terminal carboxylate of H_4PteGlu, leucine may allow a better accommodation of the more flexible and bulky H_4MPT in the catalytic pocket. In agreement with the greater distance of the terminal carboxylate moiety of H_4MPT compared to H_4PteGlu, leucine has no hydrogen bond ability.

Further, the structure of *mj*SHMT allowed us to identify a new C-terminal P-motif, which is likely to play an important role in the stabilization of the dimeric interface of *mj*SHMT. Analysis of amino acid residue composition in surface, buried and interface protein regions highlighted the higher number of solvent accessible Glu residues as the only feature, among those previously proposed to be responsible for thermal stability in SHMTs, that was clearly different between *mj*SHMT and the mesophilic homologues.

Future studies will be aimed at evaluating the roles of the peculiar structural features of *mj*SHMT proposed in the present work by determining the structural, functional and thermodynamic properties of *ad hoc* designed protein variants. Following experimental validation, *mj*SHMT structure-function-stability features highlighted by this study will be exploited to design new proteins with improved chemical-physical properties for biotechnological and biomedical applications.

ACKNOWLEDGMENTS

The authors thank Prof. Andrea Bellelli and Dr. Alessandro Paiardini for useful discussions. The authors acknowledge the Helmholtz-Zentrum Berlin—Electron storage ring BESSY II and the European Synchrotron Radiation Facility (ESRF, Grenoble, FR) for provision of synchrotron radiation; and the Biocrystal facility of the National Research Council of Italy for screening of crystallization conditions. The authors declare no conflict of interest.

REFERENCES

1. Amadasi A, Bertoldi M, Contestabile R, Bettati S, Cellini B, di Salvo ML, Borri-Voltattorni C, Bossa F, Mozzarelli A. Pyridoxal 5′-phosphate enzymes as target for therapeutic agents. Curr Med Chem 2007;14:1291–324.
2. Daidone F, Florio R, Rinaldo S, Contestabile R, di Salvo ML, Cutruzzolà F, Bossa F, Paiardini A. In silico and in vitro validation of serine hydroxymethyltransferase as a chemotherapeutic target of the antifolate drug pemetrexed. Eur J Chem 2011;46:1616–1621.
3. di Salvo ML, Contestabile R, Paiardini A, Maras B. Glycine consumption and mitochondrial serine hydroxymethyltransferase in cancer cells: the heme connection. Med Hypotheses 2013;80:633–636.
4. Appaji RN, Ambili M, Jala VR, Subramanya HS, Savithri HS. Structure-function relationship in serine hydroxymethyltransferase. Biochim Biophys Acta 2003;1647:24–29.
5. Schirch V, Szebenyi DM. Serine hydroxymethyltransferase revisited. Curr Opin Chem Biol 2005;9:482–487.
6. Vivoli M, Angelucci F, Ilari A, Morea V, Angelaccio S, di Salvo ML, Contestabile R. Role of a conserved active site cation-pi interaction in Escherichia coli serine hydroxymethyltransferase. Biochemistry 2009;48:12034–12046.
7. Gutierrez ML, Garrabou X, Agosta E, Servi S, Parella T, Joglar J, Clapés P. Serine hydroxymethyltransferase from Streptococcus thermophilus and L-threonine aldolase from Escherichia coli as stereocomplementary biocatalysts for the synthesis of beta-hydroxy-alpha,omega-diamino acid derivatives. Chemistry 2008;14:4647–4656.
8. Zhao GH, Li H, Liu W, Zhang WG, Zhang F, Liu Q, Jiao QC. Preparation of optically active β-hydroxy-α-amino acid by immobilized Escherichia coli cells with serine hydroxymethyltransferase activity. Amino Acids 2011;40:215–220.
9. Angelaccio S. Extremophilic SHMTs: from structure to biotechnology. BioMed Res Int 2013;2013:851428.
10. Paiardini A, Bossa F, Pascarella S. Evolutionarily conserved regions and hydrophobic contacts at the superfamily level: the case of the fold-type I, pyridoxal-5′-phosphate-dependent enzymes. Protein Sci 2004;13:2992–3005.
11. Contestabile R, Paiardini A, Pascarella S, di Salvo ML, D'Aguanno S, Bossa F. L-Threonine aldolase, serine hydroxymethyltransferase and fungal alanine racemase. A subgroup of strictly related enzymes specialized for different functions. Eur J Biochem 2001;268:6508–6525.
12. Angelaccio S, Chiaraluce R, Consalvi V, Buchenau B, Giangiacomo L, Bossa F, Contestabile R. Catalytic and thermodynamic properties of tetrahydromethanopterin-dependent serine hydroxymethyltransferase from Methanococcus jannaschii. J Biol Chem 2003;278:41789–41797.
13. Atomi H, Sato T, Kanai T. Application of hyperthermophiles and their enzymes. Curr Opin Biotechnol 2011;22:618–626.
14. Humble MS, Cassimjee KE, Håkansson M, Kimbung YR, Walse B, Abedi V, Federsel HJ, Berglund P, Logan DT. Crystal structures of the Chromobacterium violaceum ω-transaminase reveal major structural rearrangements upon binding of coenzyme PLP. FEBS J 2012;279:779–792.
15. Giardina G, Montioli R, Gianni S, Cellini B, Paiardini A, Voltattorni CB, Cutruzzolà F. Open conformation of human DOPA decarboxylase reveals the mechanism of PLP addition to Group II decarboxylases. Proc Natl Acad Sci USA 2011;108:20514–20519.
16. Otwinowski Z, Minor Z. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 1997;276:307–326.
17. Kabsch W. XDS. Acta Crystallogr D Biol Crystallogr 2010;66:125–32.
18. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 2010;66:213–221.
19. Stanley E. The identification of twins from intensity statistics. J Appl Crystallogr 1972;5:191–194.
20. Hamdane D, Lechauve C, Marden MC, Golinelli-Pimpaneau, B. Pseudo-merohedral twinning in monoclinic crystals of wild-type human brain neuroglobin. Acta Crystallogr D Biol Crystallogr 2009; 65:388–389.
21. Angelucci F, Saccoccia F, Ardini M, Boumis G, Brunori M, Di Leandro L, Ippoliti R, Miele AE, Natoli G, Scotti S, Bellelli, A. Switching between the alternative structures and functions of a 2-

Cys peroxiredoxin, by site-directed mutagenesis. J Mol Biol 2013; 425:4556–4568.
22. Karplus PA, Diederichs K. Linking crystallographic model and data quality. Science 2012;336:1030–1033.
23. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr 2007;40:658–674.
24. Rose PW, Bi C, Bluhm WF, Christie CH, Dimitropoulos D, Dutta S, Green RK, Goodsell, DS, Prlic A, Quesada M, Quinn, GB, et al. The RCSB Protein Data Bank: new resources for research and education. Nucleic Acids Res 2013;41:D475–D482.
25. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
26. Murshudov GN, Skubák, P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr 2011;67:355–367.
27. Emsley P, Lohkamp B, Scott W, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr 2010;66:486–501.
28. Laskowski RA, Macarthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.
29. Chandra N, Acharya KR, Moody PCE. Analysis and characterization of data from twinned crystals. Acta Crystallogr 1999;D55:1750–1758.
30. Saccoccia F, Di Micco P, Boumis G, Brunori M, Koutris I, Miele AE, Morea V, Sriratana P, Williams DL, Bellelli A, Angelucci F. Moonlighting by different stressors: crystal structure of the chaperone species of a 2-Cys peroxiredoxin. Structure 2012;20:429–439.
31. Painter J, Merritt EA. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D Biol Crystallogr 2006;62:439–450.
32. Trivedi V, Gupta A, Jala VR, Saravanan P, Rao GS, Rao NA, Savithri HS, Subramanya HS. Crystal structure of binary and ternary complexes of serine hydroxymethyltransferase from Bacillus stearothermophilus: insights into the catalytic mechanism. J Biol Chem 2002; 277:17161–17169.
33. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions Acta Crystallogr 2004;D60:2256–2268.
34. Upadhyay V, Demmer U, Warkentin E, Moll J, Shima S, Ermler U. Structure and Catalytic Mechanism of N(5),N(10)-Methenyl-tetrahydromethanopterin Cyclohydrolase. Biochemistry 2012;51:8435–8443.
35. Kabsch, W. A solution for the best rotation to relate two sets of vectors. Acta Crystallogr 1976;A32:922–923.
36. Winn MD, Ballard CC, Cowtan, KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. Overview of the CCP4 suite and current developments. Acta Crystallogr 2011;D67:235–42.
37. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol 2007;372:774–797.
38. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res 2006;34:W116–W118.
39. Madej T, Lanczycki CJ, Zhang D, Thiessen PA, Geer RC, Marchler-Bauer A, Bryant SH. MMDB and VAST+: tracking structural similarities between macromolecular complexes. Nucleic Acids Res 2014; 42:D297–D303.
40. Debret G, Martel A, Cuniasse P. RASMOT-3D PRO: a 3D motif search webserver. Nucleic Acids Res 2009;37:W459–W464.
41. Konc J, Janezic, D. ProBiS-2012: web server and web services for detection of structurally similar binding sites in proteins. Nucleic Acids Res 2012;40:W214–W221.
42. Schenk G, Margraf T, Torda AE. Protein sequence and structure alignments within one framework. Algorithms Mol Biol 2008;3:4.
43. Holm L, Rosenström P. Dali server: conservation mapping in 3D. Nucleic Acids Res 2010;38:W545–W549.
44. Ye Y, Godzik A. Flexible structure alignment by chaining aligned fragment pairs allowing twists. Bioinformatics 2003;19:ii246–ii255.
45. Hubbard SJ, Thornton JM. 'NACCESS', Computer Program, Department of Biochemistry and Molecular Biology, University College London, 1993.
46. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. Bioinformatics 2007;23:2947–2948.
47. Meyer EA, Castellano RK, Diederich F. Interactions with aromatic rings in chemical and biological recognition. Angew Chem Int Ed Engl 2003;42:1210–1250.
48. Johnson KA, Angelucci F, Bellelli A, Hervé M, Fontaine J, Tsernoglou D, Capron A, Trottein F, Brunori M. Crystal structure of the 28 kDa glutathione S-transferase from Schistosoma haematobium. Biochemistry 2003;42:10084–10094.
49. Angelucci F, Baiocco P, Brunori M, Gourlay L, Morea V, Bellelli A. Insights into the catalytic mechanism of glutathione S-transferase: the lesson from Schistosoma haematobium. Structure 2005;13:1241–1246.
50. Baiocco P, Gourlay LJ, Angelucci F, Fontaine J, Hervé M, Miele, AE, Trottein F, Brunori M, Bellelli, A. Probing the mechanism of GSH activation in Schistosoma haematobium glutathione-S-transferase by site-directed mutagenesis and X-ray crystallography. J Mol Biol 2006;360:678–689.
51. Ferreiro DU, Hegler JA, Komives EA, Wolynes, PG. Localizing frustration in native proteins and protein assemblies. Proc Natl Acad Sci USA 2007;104:19819–19824.
52. Duan G, Smith Jr VH, Weaver DF. Characterization of aromatic-amide(side-chain) interactions in proteins through systematic ab initio calculations and data mining analysis. J Phys Chem A 2000; 104:4521–4532.
53. Hughes RM, Waters ML. Effects of lysine acylation in a b-hairpin peptide: Comparison of an amide-π and a cation-π interaction. J Am Chem Soc 2006;128:13586–13591.
54. Elcock AH. Prediction of functionally important residues based solely on the computed energetics of protein structure. J. Mol. Biol 2001;312:885–896.
55. Acharya P, Warkentin E, Ermler U, Thauer RK, Shima, S. The structure of formylmethanofuran: tetrahydromethanopterinformyltransferase in complex with its coenzymes. J Mol Biol 2006;357:870–879.
56. Bhatt AN, Khan MY, Bhakuni V. The C-terminal domain of dimeric serine hydroxymethyltransferase plays a key role in stabilization of the quaternary structure and cooperative unfolding of protein: domain swapping studies with enzymes having high sequence identity. Protein Sci 2004;13:2184–2195.
57. Kirino H, Aoki M, Aoshima M, Hayashi Y, Ohba M, Yamagishi A, Wakagi T, Oshima T. Hydrophobic interaction at the subunit interface contributes to the thermostability of 3-isopropylmalate dehydrogenase from an extreme thermophile, Thermus thermophilus. Eur J Biochem 1994;220:275–281.
58. Villeret V, Clantin B, Tricot C, Legrain C, Roovers M, Stalon V, Glansdorff N, Van Beeumen JF. The crystal structure of Pyrococcus furiosus ornithine carbamoyltransferase reveals a key role for oligomerization in enzyme stability at extremely high temperatures. Proc Natl Acad Sci USA 1998;17:2801–2806.
59. Maugini E, Tronelli D, Bossa F, Pascarella S. Structural adaptation of the subunit interface of oligomericthermophilic and hyperthermophilic enzymes. Comput Biol Chem 1999;33:137–148.
60. Paiardini A, Gianese G, Bossa F, Pascarella S. Structural plasticity of thermophilic serine hydroxymethyltransferases. Proteins 2003;50,122–134.
61. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540.