Protein Engineering, Design & Selection, 2015, 1–12
doi: 10.1093/protein/gzv026
Original Article

OXFORD

Original Article

Conserved water molecules in bacterial serine hydroxymethyltransferases

Teresa Milano, Martino Luigi Di Salvo, Sebastiana Angelaccio, and Stefano Pascarella*

Dipartimento di Scienze Biochimiche 'A. Rossi Fanelli', Università La Sapienza, Roma 00185, Italy

*To whom correspondence should be addressed. E-mail: stefano.pascarella@uniroma1.it
Edited by Haruki Nakamura

Received 25 September 2014; Revised 16 April 2015; Accepted 17 April 2015

Abstract

Water molecules occurring in the interior of protein structures often are endowed with key structural and functional roles. We report the results of a systematic analysis of conserved water molecules in bacterial serine hydroxymethyltransferases (SHMTs). SHMTs are an important group of pyridoxal-5′-phosphate-dependent enzymes that catalyze the reversible conversion of L-serine and tetrahydropteroylglutamate to glycine and 5,10-methylenetetrahydropteroylglutamate. The approach utilized in this study relies on two programs, ProACT2 and WatCH. The first software is able to categorize water molecules in a protein crystallographic structure as buried, positioned in clefts or at the surface. The other program finds, in a set of superposed homologous proteins, water molecules that occur approximately in equivalent position in each of the considered structures. These groups of molecules are referred to as 'clusters' and represent structurally conserved water molecules. Several conserved clusters of buried or cleft water molecules were found in the set of 11 bacterial SHMTs we took into account for this work. The majority of these clusters were not described previously. Possible structural and functional roles for the conserved water molecules are envisaged. This work provides a map of the conserved water molecules helpful for deciphering SHMT mechanism and for rational design of molecular engineering experiments.

Key words: buried and cleft water, conserved water molecules, molecular dynamics, pyridoxal 5′-phosphate, serine hydroxymethyltransferase

Introduction

Serine hydroxymethyltransferase (SHMT, EC 2.1.2.1) is a pyridoxal-5′-phosphate (PLP)-dependent enzyme that catalyzes the reversible conversion of L-serine and tetrahydropteroylglutamate (or tetrahydrofolate, H_4PteGlu) to glycine and 5,10-methylenetetrahydropteroylglutamate (Schirch *et al.*, 1985). This enzyme plays a central role in one-carbon unit metabolism and for that reason it represents a potential target for chemotherapeutics (Daidone *et al.*, 2011; di Salvo *et al.*, 2013). SHMT also catalyzes the H_4PteGlu-independent cleavage of many 3-hydroxyamino acids and the decarboxylation of aminomalonate, at rates similar to that of H_4PteGlu-dependent serine cleavage. These features confer to SHMT a significant biotechnological potential (Angelaccio *et al.*, 2012). SHMT is a ubiquitous enzyme and its sequence and structure were conserved during divergent evolution. Many primary structures of SHMTs from *Eucarya*, *Eubacteria* and *Archaea* are available and several crystallographic structures are deposited in data banks. Also, SHMTs from extremophilic sources have been purified and their functional and structural properties have been characterized (Angelaccio, 2013; Angelaccio *et al.*, 2014). SHMT belongs to the fold type-I superfamily of PLP-dependent enzymes, a very complex group of proteins arising from an intricate evolutionary process that produced the impressive functional variety we observe today (Paiardini *et al.*, 2004). These enzymes are known to exist as stand-alone homo-oligomers (mostly dimers or tetramers) although several other forms, e.g. domains in a multidomain molecular

architectures, have been recently described (Bramucci *et al.*, 2011; Edayathumangalam *et al.*, 2013; Milano *et al.*, 2013). Because of its peculiar structural and functional characteristics, its complex and versatile enzymatic mechanism, broad substrate and reaction specificity and its pivotal role between amino acid and folate metabolism, SHMT has been under the focus of scientific community for decades. Despite the attention paid to this protein, analysis of the presence of conserved water molecules within SHMT structures, as well as in the other fold type-I enzymes, has been often neglected. Water molecules play an important role in protein structure, function and stability (Ball, 2008). For example, as reported by a vast scientific literature, water is the driving force in the hydrophobic effect, mediates protein–protein or protein–ligand interactions and can be involved in the catalytic mechanism. Moreover, the presence of water molecules in key points of protein structure may influence enzyme response to inhibitors and modulators (de Beer *et al.*, 2010; Ahmed *et al.*, 2011). Conservation of the position of water molecules in homologous structures is a strong indication that these molecules possess a structural and/or functional role. Therefore, analysis of conservation of water molecules can provide useful hints. There is active interest in this aspect of protein structure, as witnessed by recent literature reports about software for analysis of water molecules in proteins (Hu and Lill, 2014; Patel *et al.*, 2014). Investigations on the presence of conserved water molecules are often hampered by the lack of a significant number of homologous structures solved at a resolution possibly better than 2.0 Å. In the case of bacterial SHMTs, there are several three-dimensional structures with suitable properties deposited in Protein Data Bank (PDB). In this report, we examine the occurrence of conserved water molecules inside bacterial SHMT protein scaffold and analyze the evolutionary conservation of side chains interacting with these water molecules. We provide a map of the location of the conserved water molecules and of their interactions while attempting to delineate their possible functions.

Materials and methods

Bacterial SHMT structures were collected from the PDB (Rose *et al.*, 2013). After PDB scrutiny, 11 structures were selected on the basis of better resolution, the absence of point mutations and non-physiological ligands (Table I). For ease of programing, water molecules were renumbered consecutively and each of the 11 structures was analyzed by the ProACT2 program (Williams *et al.*, 1994). This program is conceived for structural analysis of proteins and provides a set of functions covering several protein structural features. Among the available functions, we utilized those related to the calculation of solvent accessibilities of water molecules and their classification into buried, cleft or surface classes. Buried water molecules were defined as those molecules contained in cavities not connected to protein surface. Water molecules found in spaces connected to the surface of proteins were considered cleft waters. The bacterial structures in their dimeric form, including the entire set of crystallographic water molecules, were superposed with PyMOL (The PyMOL Molecular Graphics System, Version 1.6 Schrödinger, LLC) and analyzed with the program WatCH (Sanschagrin and Kuhn, 1998), a tool for the identification of clusters of conserved water sites in a set of superposed protein structures. In this context, a 'cluster' of conserved water molecules consists of water molecules from each of the considered protein structures that are found in an approximately equivalent position, i.e. the clustering is between structures and a distance cutoff is applied during clustering such that there is only one member of the cluster from each structure. In our case, a cluster was considered interesting for further analysis if it contained at least 7 water molecules from different structures (from a possible maximum of 11). Each water cluster found by WatCH was assigned to one of the three possible categories buried, cleft or surface using the following rules: (a) if at least eight water molecules of the cluster were individually assigned to the buried category by ProACT2 then the cluster was labeled as buried; (b) else if at least seven water molecules were assigned to the cleft category then the cluster was labeled as cleft. If none of the two conditions were met, the cluster was labeled as surface. Secondary structures were assigned with the program STRIDE (Heinig and Frishman, 2004). Potential H-bonds involving the water molecules belonging to each cluster were predicted with the program HBplus (McDonald and Thornton, 1994). Interfacial water molecules were considered those interacting with atoms from both subunits. An approximate but fast estimate of interaction energy of buried water molecules was obtained with the program Dowser-3 (Zhang and Hermans, 1996). Data analyses were carried out by means of Perl and Bash scripts under Mac OS X operating system.

Multiple sequence alignments were calculated with the program MUSCLE (Edgar, 2004) and were analyzed with Jalview (Waterhouse *et al.*, 2009). Sequences were retrieved from the UniProt (Magrane and Consortium, 2011) resource or from UniRef50 archive (Suzek *et al.*, 2007).

Molecular dynamics (MD) and trajectory analysis were carried out with the software NAMD (Phillips *et al.*, 2005) and VMD (Humphrey

Table I. List of bacterial SHMT structures utilized in the work

PDB code	Source	Form[a]	Resolution (Å)	Water[b]
1DFO	*Escherichia coli*	PLP (external aldimine) + FFO	2.4	388 (4)
2DKJ	*Thermus thermophilus*	PLP	1.15	896 (2)
1KKJ	*Geobacillus stearothermophilus*	PLP	1.93	226 (1)
3ECD	*Burkholderia pseudomallei*	Apo	1.6	1607 (4)
3GBX	*Salmonella typhimurium*	Apo	1.8	450 (2)
3H7F	*Mycobacterium tuberculosis*	PLP	1.5	962 (2)
3N0L	*Campylobacter jejuni*	Apo	1.8	504 (2)
3PGY	*Staphylococcus aureus*	Apo	1.92	1057 (4)
4J5U	*Rickettsia rickettsii*	PLP	1.7	771 (2)
4N0W	*Burkholderia cenocepacia*	PLP	1.65	1494 (4)
4P3M	*Psychromonas ingrahamii*	Apo	1.85	548 (2)

[a]PLP means internal aldimine; Apo denotes the apo form; FFO is the 5-formyltetrahydrofolate.
[b]Number of water molecules included in the PDB file. In parentheses, the number of subunits in the original PDB file is indicated.

et al., 1996), respectively, run under the Mac OS X operating system. The approach employed was modeled onto the work by Mustata *et al.* (2003) and Mustata and Briggs (2004) who analyzed the structural role of water molecules in alanine racemase, a PLP-dependent enzyme. The MD simulations were carried out on the dimer in the apo form of the structure 4N0W corresponding to *Burkholderia cepacia* SHMT (Table I). The dimer along with its crystallographic water molecules was immersed in the smallest possible water sphere able to contain the entire assembly. The CHARMM36 force field (Best *et al.*, 2012) and the TIP3P (Jorgensen *et al.*, 1983) water model were employed. Simulation was carried out without periodic boundary conditions, at a constant temperature (310 K), a cutoff at 12 Å and a time step of 2 fs. The system was initially energy minimized for 1000 steps of conjugate gradient minimizer, followed by relaxation with 100 ps of MD and again minimized for 1000 steps. Finally, the entire system has been simulated for 2 ns of MD. The functional role of selected water clusters was tested by removing the corresponding crystallographic water molecule from 4N0W. MD has been run again under identical conditions and trajectories compared with those obtained for the structure possessing the complete set of crystallographic water molecules. Equilibration of the system was indicated by convergence of total energy and temperature while conformational stability was assessed by monitoring the backbone root mean square deviation (RMSD) variations versus time of simulation.

Results

All 11 selected SHMT structures, except 1DFO, possess resolution higher than 2.0 Å (Table I). Sequence identity among these bacterial SHMTs ranges from 44 to 84% (Supplementary Table SI). Water clusters were assigned to one of three classes using the categorization operated by the program ProACT2 on individual molecules, following a majority rule as described in 'Materials and Methods' section. A stricter criterion was applied to the 'buried' category, in order to select only the *bona-fide* buried cluster subset. The clusters found are reported in Table II labeled with an arbitrary index. Lower indices indicated clusters in the inner region of the SHMT dimers. Symmetrical clusters (i.e. clusters of water molecules located in symmetrical positions in each subunit of the dimer) were assigned a unique index. There is a slight variation in the subunit orientation within the dimers in each structure, which may influence the water cluster definition depending on the method used for the rigid body superposition. For this reason, the 11 SHMTs were superposed using the alternative method provided by the MultiProt server (Shatsky *et al.*, 2004). The cluster identification was repeated using the new superposition. Results indicated that, in this case, only 138 clusters compared with the 143 found with the PyMOL superposition were collected. For this reason, PyMOL superposition was taken for subsequent analyses. Ideally, superposition of the two subunits should be done independently to remove any possible effect connected to the slight variations of their reciprocal orientation. However, comparison of the results of the structural superposition obtained with the two mentioned methods strongly suggests that, in any case, the expected differences are small.

Buried water clusters

Table II reports the list of clusters categorized as described earlier. Overall, 14 symmetrical clusters of conserved buried water molecules were observed (Fig. 1) in SHMT dimers corresponding to 28 individual clusters (Table II). Only one asymmetrical buried cluster was detected in a position located in the subunit A of the *Escherichia coli* SHMT. Ten buried clusters contained water molecules conserved in at least 10 of the 11 considered structures. Residue numbers reported in the text and in the tables refer, if not otherwise specified, to the numbering system of *E.coli* SHMT sequence (1DFO).

Cluster 5 interconnects the side chains of the highly conserved residues Asn345 and Arg363 (Fig. 2). Arg363 is responsible for the formation of a salt bridge to the α-carboxylic group of the substrate when it is bound to the cofactor as an external aldimine (Fig. 2). The water molecules of this cluster interact with the ε-nitrogen atom of the Arg363 side chain. Noteworthy, the same interaction takes place also in those structures solved with an internal aldimine or in the apo form (Table I). This fact strongly suggests that this water molecule is essential for properly orienting and stabilizing the Arg363 side chain also in the absence of interaction to the substrate α-carboxylic group.

Clusters 27 and 28 are interconnected by H-bonds and are predicted to interact only with main chain peptide bonds. They are located between the N-terminus of the short α-helix encompassed by sequence positions 366–372 and the turn at position 33–36. These clusters lie at the interface between the large and small domains that constitute the SHMT monomer (Fig. 2). Cluster 26 is ~4.5 Å from cluster 28. They interact with the same turn at sequence positions 33–36.

Clusters 11 and 12 are interconnected by H-bond and interact with the side chains of conserved Gln261 and Gln94 residues, and to the peptide bonds of Gly262 and Gly58, respectively. The former loop is part to the phosphate binding cup (Denesyuk *et al.*, 2003), i.e. the cavity that accommodates and stabilizes the cofactor phosphate group through polar interactions to the phosphate oxygens, while the latter forms part of the wall of the active site of the adjacent subunit (Fig. 3). Clusters 13 and 14 connect, through H-bond to the main chains, the loop 262–264 interacting with the cluster 12, to the phosphate binding cup loop (Fig. 3).

It is well established that water molecules occurring at protein subunit interface are important for the stability and function of the oligomer (Li *et al.*, 2013). In SHMT, four conserved buried water clusters, denoted by the indices 6, 7, 14 and 37 in Table II, were observed at the dimer interface. Cluster 37 is H-bonded to the side chain of a conserved Glu residue (corresponding to Glu22 in the protein 1DFO) and bridges the first N-terminal α-helix of a subunit to the main chain of the other subunit in correspondence of the peptide bond connecting residue Gly67 at the beginning of the α-helix located between positions 68 and 86 in 1DFO (Fig. 4). Cluster 6 is H-bonded to the side chain of residue Gln47 (non-conserved in the most distant SHMTs even if often replaced by polar residues Ser, Thr or Glu as shown in Supplementary Fig. S1) and to the peptide bonds of the other subunit connecting residues Gln47 and Ser49 (Fig. 4). Cluster 7 is connected through H-bonds to the side chains of Gln47 and His267 belonging to two different subunits. Gln47 is part of α-helix 41–48, while His267 belongs to the interfacial α-helix encompassed by the positions 266–279 (Fig. 4). His267 is conserved in 62% of the considered SHMT sequences. However, it is often replaced by polar residues Asn, Glu or Gln (Supplementary Fig. S1). Buried cluster 14 is in contact with the residue Arg235 from the other subunit and it is H-bonded to the backbone donors and acceptors in sequence positions 51, 54 and 266.

Three water clusters, numbered 8, 9 and 10, form an interconnected system in the large domain, in the proximity of the binary symmetry axis at the subunit interface even though they are not predicted to interact to any residue of the other subunit. Clusters 9 and 10 bind to the side chains of residues Lys272 and His96, respectively (Fig. 5), which are not conserved across bacterial SHMTs (Table II and Supplementary Fig. S1).

Table II. List of conserved water clusters in the selected SHMT structures

Clusters indices[a]	Cat[b]	Cons[c]	Missing structures[d]	Water molecule index[e]	Interacting side chains[f]	Average number of H-bond to main chain[g]	Dowser energy[h] (kcal/mol)
1	C	10	4P3M	A545; B628	–	2.7	−16.2
2	C	11	–	A704; B613	Asn102 (91%) Asp200 (100%)	0.6	−1.7
3	CI	11	–	B500; A506	SerA97 (85%) HisB96 (58%) PLP-OP1	1.8	−15.2
4	CI	9	3ECD 3N0L	A608; A590	Arg235 (83%) PLP-OP3	1.0	−3.6
5	B	11	–	A510; B501	Asn345 (85%) Arg363 (98%)	2.3	−19.8
6	BI	9	4J5U 3N0L	A560; B520	Gln47 (35%)	2.0	−13.0
7	BI	10	3ECD	B568; A630 (4N0W)	Gln47 (35%) His267 (62%)	1.2	−12.0
8	B	11	–	A535; B507	A567;B575	2.0	−16.4
9	B	11	–	A567; B575	Lys272 (67%) B507;A535;B533; A557	0.9	−18.0
10	B	11	–	A557; B533	His96 (58%) A567;B575	2.0	−12.5
11	B	9	3GBX 3ECD	A572; B725	Gln94 (87%) Gln261 (80%) A528	1.0	−18.7
12	B	10	3ECD	A528; B683	Gln94 (87%) A572	2.7	−12.0
13	B	9	3ECD 3GBX	A573; B628	–	3.0	−13.5
14	BI	11	–	A512; B502	–	3.4	−15.0
15	CI	10	3N0L	A623; B893	Thr52 (54%) A643;B660	2.0	−10.0
16	CI	10	3N0L	A643; B660	Asn53 (83%) A623;B893 B553	1.0	−8.4
17	C	10	3N0L	A542; B553	B660;A643	1.5	−9.4
18	C	10	1DFO	C620 (4N0W) A651 (4N0W)	Glu37 (84%) (4N0W)	2.7	−16.7
19	CI	9	3N0L 4J5U	B877; C742 (4N0W)	Glu36* (84%)	1.6	−8.8
20	C	8	4N0W 3ECD 4J5U	A587; B591	Tyr140 (55%)	1.0	0.0
21	C	11	–	A644; B598	–	3.6	−17.7
22	C	10	4P3M	A637; B564	His129 (83%)	2.3	−10.1
23	C	8	4P3M 2DKJ 3ECD	A837; B559	Tyr145 (89%)	1.5	−11.0
24	C	7	4J5U 3N0L 1DFO 3PGY	A513; C770 (4N0W)	Tyr38 (34%)	1.4	−11.8
25	CI	9	3N0L 1DFO	C641 (4N0W) A614 (4N0W)	Asn24 (4N0W) (<10%)	1.0	0.0
26	C	9	3N0L 3PGY	A540; B558	–	2.2	−9.0
27	B	11	–	A555; B509	A583;B514	3.2	−16.7
28	B	11	–	A583; B514	A555;B509	3.1	−15.0
29	C	10	3N0L	A580; B588	Thr39 (10%) B819	2.0	−13.5
30	C	10	4P3M	A644; B598	–	3.6	−17.7
31	C	9	3N0L 1DFO	B819; C636 (4N0W)	Glu278 (70%) Arg371 (<10%) B588;B597	0	−24.0

Continued

Table II. Continued

Clusters indices[a]	Cat[b]	Cons[c]	Missing structures[d]	Water molecule index[e]	Interacting side chains[f]	Average number of H-bond to main chain[g]	Dowser energy[h] (kcal/mol)
32	C	7	1DFO 3ECD 4P3M 3H7F	A662 (4N0W) C764 (4N0W)	Arg233 (68%) (4N0W)	0.1	0.0
33	C	9	3N0L 4P3M	A570; B597	B819	2.1	−11.0
34	C	8	3ECD 1DFO 3PGY	A741 (4N0W) B721	–	1.0	0.0
35	C	11	–	A522; B532	Glu75 (100%) Asn92 (84%) A754;B666	1.0	−4.8
36	S	8	4P3M 3GBX 3DKJ	B842; A953 (4N0W)	Arg81* (80%)	0.0	0.0
37	BI	10	3GBX	A544; B529	Glu22 (90%)	3.0	−26.5
38	C	9	3ECD 3N0L	A556; B593	–	2.4	−5.0
39	C	11	–	A687; B641	–	2.0	0.0
40	C	9	1KKJ 4J5U	A663; B677	–	2.0	−3.2
41	C	10	3H7F	A645; B757	–	2.0	−2.1
42	C	8	3ECD 1DFO 1KKJ	A672; C736 (4N0W)	–	1.0	0.0
43	C	8	3PGY 3GBX 3ECD	A985; B782	Asp72 (95%) A726; B708;B733	0.0	0.0
44	C	8	3ECD 3GBX 3PGY	A726; B733	Glu69 (59%) A985;B782	2.0	0.0
45	C	9	3GBX 4P3M	A754; B666	Asn253 (40%) A522;B532	1.0	−5.3
46	C	10	3PGY	B818; A643 (4N0W)	–	2.0	0.0
47	S	8	1DFO 3ECD 3GBX	A650 (4N0W) C703 (4N0W)	–	1.0	0.0
48	C	8	1DFO 3ECD 3GBX	A656 (4N0W) B708	B782	1.0	0.0
49	C	8	1DFO 3ECD 3GBX	C645 (4N0W) A712 (4N0W)	Arg25 (83%)	0.0	0.0
50	C	9	1DFO 3H7F	C695 (4N0W) A657 (4N0W)	–	1.4	0.0
51	C	9	1DFO 4P3M	A667 (4N0W) C746 (4N0W)	–	1.9	0.0
52	B	7	4N0W 4P3M 1DFO 3GBX	A622 (1KKJ) B622 (1KKJ)	–	3.1	−16.7
53	C	9	1DFO 4P3M	A621 (4N0W) C628 (4N0W)		3.2	−4.0
54	C	11	–	A635; C603 (4N0W)	Asp12 (84%)	1.0	0.0
55	C	9	3ECD 3GBX	B810; A867 (4N0W)	B766	1.1	0.0
56	S	10	1DFO	A665 (4N0W) C781 (4N0W)	–	1.0	0.0
57	S	7	3H7F 1DFO 1KKJ 4P3M	A735 (4N0W) C662 (4N0W)	–	1.0	0.0

Continued

Table II. Continued

Clusters indices[a]	Cat[b]	Cons[c]	Missing structures[d]	Water molecule index[e]	Interacting side chains[f]	Average number of H-bond to main chain[g]	Dowser energy[h] (kcal/mol)
58	C	8	3GBX 4P3M 1DFO	A754 (4N0W) C709 (4N0W)	–	2.0	0.0
59	S	8	3PGY 3N0L 1DFO	A894 (4N0W) C827 (4N0W)		0.9	0.0
60	C	8	3ECD 1KKJ 1DFO	B979; A946 (4N0W)	B766	1.0	0.0
61	C	9	3ECD 1DFO	A723 (4N0W) C785 (4N0W)	–	1.0	0.0
62	C	7	3ECD 1KKJ 3N0L 1DFO	C751 (4N0W) A963 (4N0W)	–	1.7	0.0
63	S	9	1DFO 1KKJ	A654 (4N0W) C657 (4N0W)		1.1	0.0
64	C	9	1KKJ 4J5U	A550	–	2.0	0.0
65	C	7	3PGY 1KKJ 2DKJ 4J5U	A724	–	0.3	−2.3
66	C	8	4N0W 3N0L 3H7F	A889	–	0.9	0.0
67	B	7	3ECD 4J5U 1KKJ 4P3M	A701	Tyr287 (84%) Gln288 (20%)	0.4	−6.8
68	S	7	4P3M 2DKJ 1KKJ 4J5U	B611	Glu281 (<5%)	1.0	0.0
69	S	8	4P3M 3ECD 3PGY	B871	–	1.0	0.0
70	C	7	3PGY 1KKJ 3GBX 3ECD	B766	B810 B979	0.0	0.0
71	C	7	4J5U 1DFO 3H7F 3N0L	A698 (4N0W)	–	0.9	0.0
72	C	7	4P3M 3ECD 3PGY 1DFO	A1028 (4N0W)	–	1.0	0.0
73	C	7	3GBX 4P3M 1DFO 3N0L	C696 (4N0W)	–	1.0	0.0
74	C	8	1KKJ 3ECD 4J5U	B690	–	1.2	0.0
75	C	8	3ECD 3N0L 3DKJ	A885	–	1.0	0.0

Continued

Table II. Continued

Clusters indices[a]	Cat[b]	Cons[c]	Missing structures[d]	Water molecule index[e]	Interacting side chains[f]	Average number of H-bond to main chain[g]	Dowser energy[h] (kcal/mol)
76	C	7	3H7F 3PGY 3ECD 4N0W	B638	Asp325 (<15%)	2.0	−11.5
77	S	7	4P3M 1DFO 3ECD 3GBX	A736 (4N0W)	–	1.3	0.0
78	CI	8	3H7F 1DFO 3GBX	A814 (4N0W)	Glu332 (50%) (4N0W)	1.7	0.0
79	C	7	4P3M 1DFO 3ECD 3GBX	C679 (4N0W)	Asp153 (<5%) (4N0W) Tyr156 (60%) (4N0W)	1.0	0.0
80	C	8	3N0L 1DFO 4P3M	C809 (4N0W)	Glu298 (34%) (4N0W) C746 (4N0W)	0.7	0.0

[a]Cluster index; clusters from index 64 to 80 are non-symmetrical.
[b]Cluster category: 'B' indicates buried, 'C' cleft, 'S' surface. 'I' means interface.
[c]Number of structures contributing a water molecule to the cluster; in case of symmetrical clusters, the highest number is reported.
[d]PDB codes of the SHMTs missing the water molecule corresponding to the cluster.
[e]PDB index of the water molecule of the *E.coli* SHMT (1DFO); if missing, the representative water molecule has been taken from another structure (PDB code in parentheses). For symmetrical clusters, both molecules are reported.
[f]Side chains interacting with the cluster water molecules; conservation degree is indicated in parentheses; asterisk indicates a side chain from the other subunit; interacting water molecules are indicated only if belonging to another conserved cluster.
[g]Total number of H-bonds to main chain observed in all the water molecules contained in the cluster divided by the number of water molecules.
[h]Average Dowser energy of the water molecules in the cluster. Energy is defined only for buried or deep cleft water molecules.

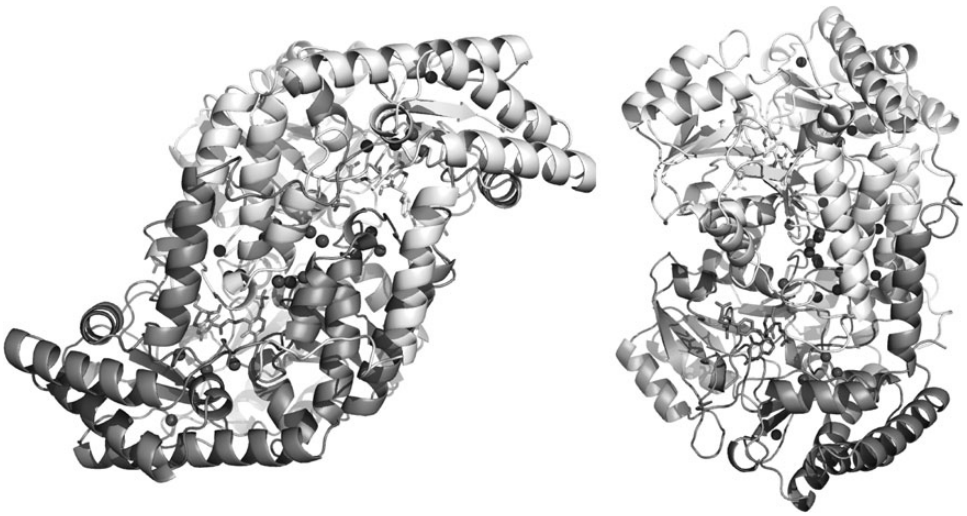

Fig. 1 Position of buried conserved water clusters. SHMT from *E.coli* (PDB code 1DFO) is displayed as cartoon model with the two subunits denoted by different gray colors. PLP and 5-formyltetrahydrofolate cofactors are displayed as sticks. Water clusters are reported as dark spheres. Left panel displays the dimer oriented with the binary symmetry axis approximately perpendicular to the sheet plane, while the right panel reports the dimer oriented with the axis parallel to the plane.

Cleft and surface water clusters

The overall distribution of conserved cleft water molecule clusters in the SHMT dimers is shown in Fig. 6. Forty-three symmetrical and 13 single conserved cleft clusters were found in the 11 SHMT set. Among all the clusters, we examined in more details only those related by dyad symmetry, interacting with conserved side chain residues (at least 50% conserved in the sequence alignment of Supplementary Fig. S1) and possibly containing at least 9 molecules of the possible 11 (Table II).

A few cleft clusters occur at the dimer interface in proximity of the buried clusters described in the above section and contribute to the

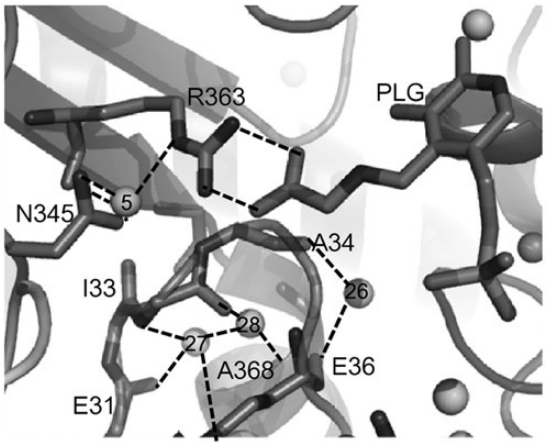

Fig. 2 Buried water clusters in proximity of *E.coli* SHMT active site. Protein backbone is shown as cartoon models colored in dark and light gray to distinguish the two subunits. Main or side chains interacting with the water clusters are represented as stick models. Positions of the buried water clusters, calculated as the mean of the coordinates of each of the cluster's member water molecule, are indicated by gray spheres labeled according to Table II. Black dashed lines display the interactions predicted by HBplus for the clusters discussed in the text. Residue numbering refers to the *E.coli* SHMT structure (1DFO). Label PLG indicates the external aldimine formed by the substrate Gly and the cofactor PLP. One-letter code is used to indicate amino acid residues. The figure displays the conserved clusters in proximity of the cofactor binding site. Cluster 5 interacts with the side chain of the Arg363 that binds the α-carboxylate group of the external aldimine. Clusters 26–28 interact with the loop lining the cofactor phosphate group.

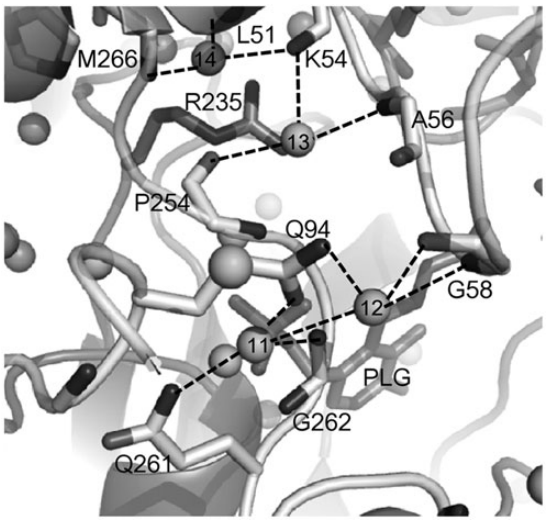

Fig. 3 Buried water clusters in proximity of the phosphate binding cup of *E.coli* SHMT active site. Display and drawing styles are as described in the legend of Fig. 2. Clusters 11–14 are reported. These water clusters interact with the loop taking part in the formation of the phosphate-binding cup.

stabilization of the SHMT assembly (Fig. 4). In particular, clusters 15–19 interact with groups from both subunits. Cluster 15 H-bonds to the side chain of Thr52 (54% conserved) and to the backbone in correspondence of Gly233 from the other subunit. Cluster 16 interacts with cluster 15 and is H-bonded to the side chain of Asn53 (83% conserved), to the cluster 17, and to the backbone at Asn37 of the other subunit. Cluster 17 is, in turn, connected to the backbone groups of

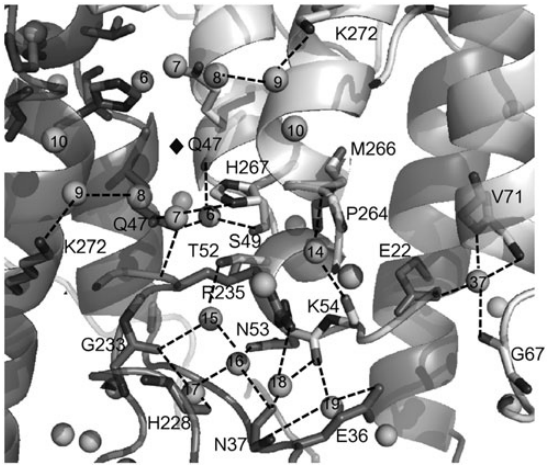

Fig. 4 Dimer interface of *E.coli* SHMT displaying the positions of the conserved buried water clusters. Display and drawing styles are as described in the legend of Fig. 2. Diamond indicates the approximate positions of the 2-fold axis of symmetry, perpendicular to the sheet plane. Conserved buried clusters 6, 7, 14 and 37 at the subunit interface are reported. Cluster 6 is H-bonded to the side chain of residue Gln47 and to peptide bonds of the other subunit. Cluster 7 interacts to the side chains of Gln47 and His267 from the two different subunits. Buried cluster 14 is in contact with the residue Arg235 from the other subunit and it is H-bonded to backbone donors and acceptors in sequence positions 51, 54 and 266. Cluster 37 H-bonds to the side chain of a conserved Glu residue bridging the first N-terminal α-helix of a subunit to the main chain of the other subunit. The buried clusters numbered 8–10, form an interconnected system in the large domain, in the proximity of the binary symmetry axis at the subunit interface. Cleft interface clusters 15–19 interact with groups from both subunits. Cluster 15 H-bonds to the side chain of Thr52 and to the backbone of the other subunit. Cluster 16 interacts with cluster 15 and is H-bonded to the side chain of Asn53, to the cluster 17, and to the backbone of the other subunit. Cluster 17 is connected to the backbone groups of adjacent subunit. Cluster 18 is bonded to the residue Arg236 and to backbone donor and acceptor groups of the other subunit. Cleft cluster 19 is H-bonded to the side chain of Glu36.

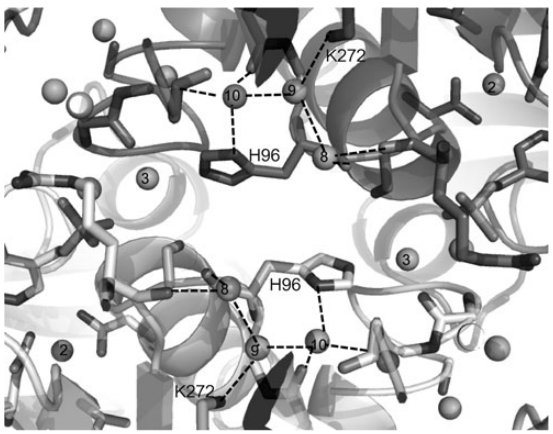

Fig. 5 Buried water clusters in proximity of the binary axis of symmetry at the dimer interface of *E.coli* SHMT active site. Display and drawing styles are as described in the legend of Fig. 2. The buried water clusters 8–10, form an interconnected system in the large domain, in the proximity of the binary symmetry axis at the subunit interface. They are not predicted to interact to any residue of the other subunit. Clusters 9 and 10 bind to the side chains of residues Lys272 and His96, respectively.

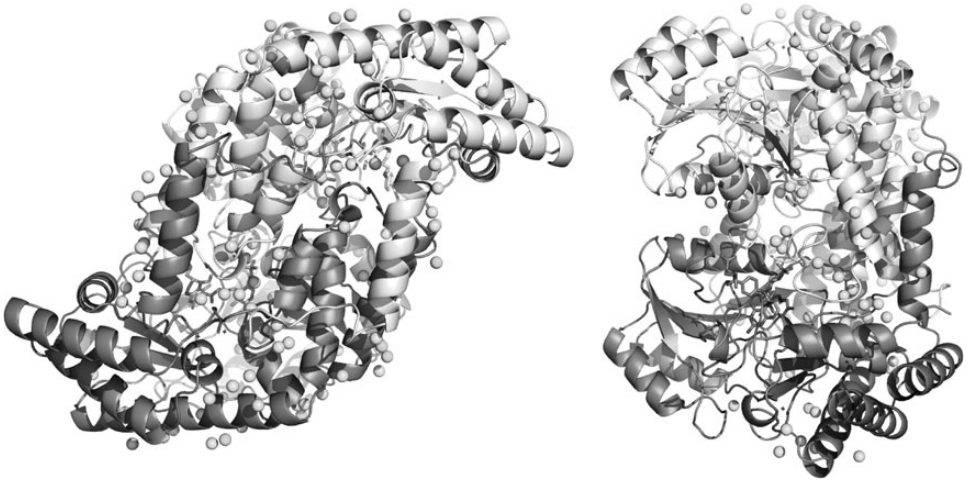

Fig. 6 Position of cleft conserved water clusters. SHMT from *E.coli* (PDB code 1DFO) is displayed as cartoon model with the two subunits denoted by different colors. PLP and 5-formyltetrahydrofolate cofactors are displayed as sticks. Water clusters are reported as gray spheres. Left panel displays the dimer oriented with the binary symmetry axis approximately perpendicular to the sheet plane, while the right panel reports the dimer oriented with the axis parallel to the plane.

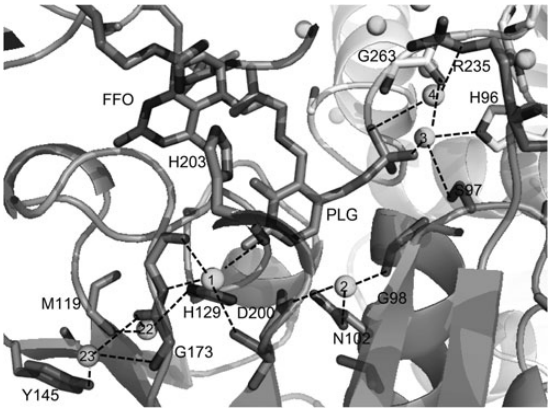

Fig. 7 Cleft water clusters localized at the cofactor binding site of *E.coli* SHMT. Display and drawing styles are as described in the legend of Fig. 2. FFO stands for 5-formyltetrahydrofolate. Cleft clusters 3 and 4 interact with the cofactor phosphate group. Moreover cluster 3 is H-bonded to Ser97 and His96 of the other subunit, while cluster 4 binds the side chain of Arg235 of the same subunit and the main chain of the adjacent subunit. Cluster 2 is close to the cofactor molecule and interacts with residues Asn102, Asp200 and Thr224. Cluster 1 is close to the methyl substituent of the PLP ring and interacts with the main chain peptide bonds. Cluster 22 interacts with His129 and with cluster 23.

the residue His288 and Gly233 of the adjacent subunit. Cluster 18 is bonded to the side chain of the residue Arg236 (90% conserved) and to the backbone donor and acceptor groups of residues Lys54 and Asn37 of the other subunit. Finally, cluster 19 is H-bonded to the side chain of Glu36 (84% conserved) and to the backbone of Asn37 of one subunit and to carbonyl oxygen of Lys54.

Several of the cleft clusters are also located in the cofactor binding site. Cleft clusters 3 and 4 interact with the cofactor phosphate group (Fig. 7). In particular, cluster 3 is H-bonded to the side chain of Ser97 and His96 of the other subunit, while cluster 4 binds the side chain of residue Arg235 of the same subunit and the main chain, in correspondence of Gly263 of the adjacent subunit. Ser97 is highly conserved in bacterial SHMTs while His96 is more variable although it is often replaced by a Tyr residue in the most distant bacterial SHMTs (Supplementary Fig. S1). These side chains and their interactions are components of the phosphate binding cup (Denesyuk *et al.*, 2003). Cluster 2 is positioned in vicinity of the cofactor molecule (Fig. 7) and corresponds to a water molecule already reported in other type-I PLP-dependent enzymes, such as aspartate aminotransferase (Okamoto *et al.*, 1994). The water molecules belonging to this cluster (occurring in all 11 SHMT structures) interact through H-bonds with the side chains of conserved residues Asn102, Asp200 and Thr224. In particular, Asp200 interacts with the nitrogen atom of the cofactor pyridine ring, as it does in several other PLP-dependent enzymes of fold type-I, such as aspartate aminotransferases (Jager *et al.*, 1994). Close to the methyl substituent of the PLP ring, lies cluster 1. The water molecules contained in this cluster interact with the main chain peptide bonds of residues Phe174, Asp200 and His203. Its position within the SHMT structures suggests this cluster be important for stabilizing the active site wall and the peptide backbone of residues directly involved in the catalytic process, namely Asp200 and His203 (Talwar *et al.*, 2000; Griswold and Toney, 2011). Cluster 22 interacts with His126 (83% conservation), which is H-bonded to the side chain of Asp200 and with cluster 23. This cluster interacts with the conserved Tyr145, conserved in 89% of SHMT sequences (Fig. 7). Apparently, cluster 22 belongs to a network of interactions able to stabilize the active site conformation and the orientation of the catalytically relevant side chain of Asp200. Clusters 35 and 45 (Fig. 8) contain water molecules H-bonded to each other. Moreover, cluster 35 (present in all SHMT structures) establishes an H-bond to the side chain of Glu75 (100% conserved) belonging to the α-helix in sequence positions 68–86, to Asn92 (84% conserved) and to the main chain peptide bond of Val93. Cluster 45 is connected to cluster 35 via H-bond and interacts with the side chain of Asn253, included in the α-helix 246–256, and to the main chain at residue Ala91 of the β-strand encompassed by the positions 90–92. The residue Asn253 is not identically conserved in the other SHMTs, although when substituted it is often replaced by an Asp, which can likely preserve the water coordination. Cluster 43 should be considered along with clusters 35 and 45 because indirectly connected to Glu75. Cluster 43 interacts with the conserved residue Asp72 and with water molecules of the cleft clusters 44 and 48. Cluster 44 interacts with the nitrogen atom of the peptide bond connecting Arg63 to Lys62

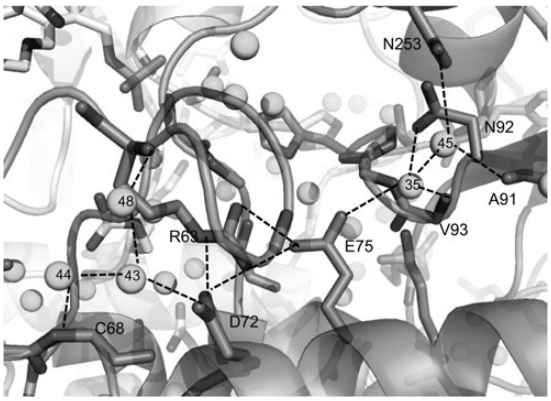

Fig. 8 Cleft water clusters in the interior of *E.coli* SHMT. Display and drawing styles are as described in the legend of Fig. 2. Cluster 35 establishes an H-bond to Glu75, Asn92 and to the main chain peptide bond of Val93. Cluster 45 is connected to cluster 35 and interacts with Asn253. Cluster 43 is connected to clusters 44, 48 and to Asp72. Cluster 43 interacts with Asp72 and with cluster 44.

(Fig. 8). Arg63 is a conserved residue and its side chain interacts with Glu75, which is involved in the bond to cleft cluster 35.

Six conserved symmetrical clusters were identified as completely exposed, i.e. belonging to the exposed category (Table II). Of these, one is predicted to interact with the side chain of Arg81, a conserved residue (Supplementary Fig. S2).

Molecular dynamics

MD was employed to assess the putative role of water molecules belonging to selected conserved clusters. Exhaustive test of all the conserved water clusters found during this work was beyond the aims of our research. We instead choose for subsequent analysis two clusters taken among those most populated by water molecules, buried, interacting with conserved side chains and displaying a low Dowser energy. Among the clusters fulfilling the above criteria were (Table II) clusters 5 (Fig. 2) and 11 (Fig. 3). Noteworthy, Cluster 5 is in contact, in 1DFO, with Arg343 that interacts with the α-carboxylate of the substrate when this is bound at the active site as internal aldimine. Cluster 11 is interesting because interacts with the side chain of Gln261 (1DFO) that belongs to the loop constituting part of the phosphate binding cup (Denesyuk *et al.*, 2003), i.e. the cavity that accommodates and stabilizes the cofactor phosphate group of the other subunit. The water molecules of the structure 4N0W corresponding to each cluster were alternatively removed and the MD simulation run in conditions identical to those used for the original structure. In the case of cluster 11, also the interacting cluster 12 was removed. Conformational variations were analyzed with the Timeline plugin of VMD software and were reported in terms of RMSD variations versus simulation time, of the residues interacting with the cluster water molecule. As for cluster 5 (Supplementary Fig. S3), removal of the corresponding water molecule in 4N0W (id A944) induced a marked increase in the conformational fluctuation of the residues Asn343 (corresponding to 1DFO Asn345) and Arg361 (Arg363) which may suggest destabilization of the active site and possibly perturbation of the binding of substrate carboxylate (Fig. 9). The possible effect of the removal of water molecules belonging to the clusters 11 (water A619) and 12 (A606) (Supplementary Fig. S4) is the increased fluctuation of residue Gln261 (corresponding to Gln259 in 4N0W), while

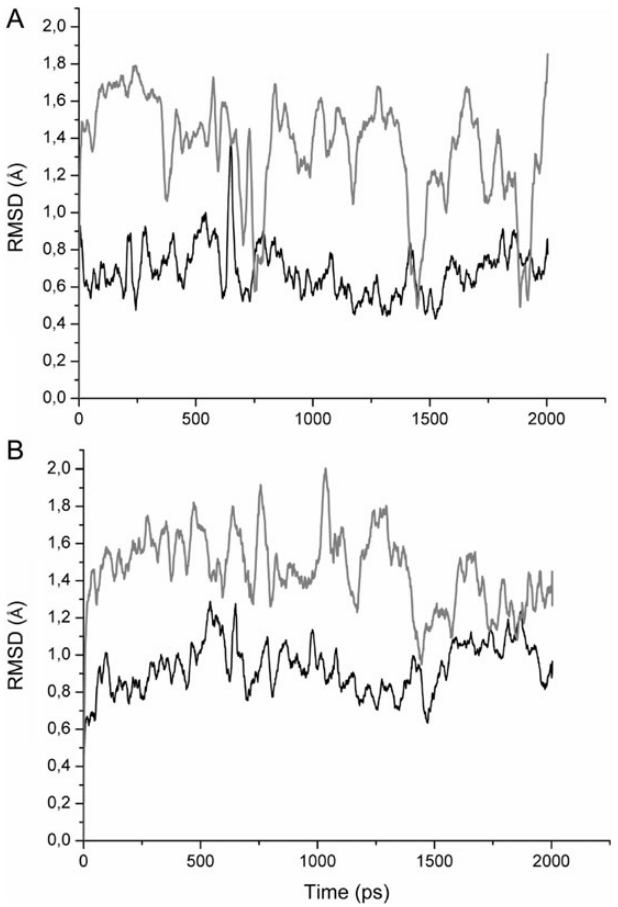

Fig. 9 RMSD variation. RMSD fluctuation with respect to the starting conformation versus time of simulation. Gray and black lines refer to the MD simulation without and with the cluster 5 water molecule in 4N0W, respectively. Data have been smoothed using adjacent averaging over a 21-point window. The results for (A) 4N0W Asn343 (corresponding to 1DFO Asn345) and (B) Arg361 (1DFO Arg363) are reported.

Gln94 (Gln95) seems to be largely unaffected (Supplementary Fig. S5). It should be mentioned that the side chain of this residue is predicted to interact with the side chain of Glu76. This interaction may stabilize the conformation of Gln95 after removal of the conserved water molecule. On the contrary, Gln259 displays only an interaction of its side chain to the carbonyl oxygen of the peptide bond of Gln95.

Discussion

The work reported herein presents a systematic analysis of the occurrence of conserved water molecules in bacterial SHMTs using static structures solved by X-ray crystallography. Conservation analysis of water molecules strongly hints at a structural and/or functional role (Debler *et al.*, 2009) and for this reason there is an active interest in the study of such aspects of protein structure, as witnessed by new softwares recently described in the literature (Hu and Lill, 2014; Patel *et al.*, 2014). For example, the presence of conserved water molecules at enzyme active sites can significantly influence ligand binding modes and affinities (Poornima and Dean, 1995; Davies *et al.*, 2012). Moreover, awareness of the presence of conserved water molecules

is helpful for accurate homology modeling of members of the same protein family (Henriques *et al.*, 1997). However, proper analysis of conserved water molecules in a set of proteins is often hampered by the lack of a sufficient number of high-resolution coordinate sets. In the case of bacterial SHMT, PDB contains several structures suitable for systematic comparative studies. However, it should be emphasized that sufficient structural information is available now for several other protein families such as, for example, α-amylase for which 28 homologous structures from Archaea, bacteria and eukaryotes are known, glyceraldehyde-3-phosphate dehydrogenase (30 structures), phosphopyruvate hydratase (19 structures) and ribulose-bisphosphate carboxylase (20 structures). Moreover, for other families, structures solved by neutron diffraction are also available: for example, human carbonic anhydrase, inorganic pyrophosphatase from *Thermococcus thioreducens* and dihydrofolate reductase from *E.coli*. The analysis here described can be applied to these families as well.

In our analysis we used two programs, ProACT2 (Williams *et al.*, 1994) and WatCH (Sanschagrin and Kuhn, 1998), which are able to categorize water molecules within a protein structure on the basis of their solvent accessibility and are capable to find clusters of conserved water molecules in a set of superposed homologous protein structures. In our hands, these programs proved to be efficient and easy to use. The combination of these two programs, along with PyMOL and Perl parsing scripts, allowed us to build up a pipeline able to detect conserved water clusters employing minimal computational resources. With this approach, we were able to find a number of water clusters conserved in a set of bacterial SHMTs, the majority of which were not explicitly described before. Overall, we found 143 single conserved clusters. A half of these clusters contain water molecules from at least 9 SHMT structures among the 11 available in the data set. The most conserved are the buried and the deep cleft clusters, also those at the subunit interface. About half of the buried water clusters interact with conserved polar side chains and four of them are part of the subunit interface involved in dimer formation (Table II). The presence of interactions between conserved water molecules (either in buried cavities or clefts) and one or more side chains of invariant protein residues suggests a possible explanation for the conservation of these residues, which in many cases is not obvious. As an example, the buried clusters at the subunit interface (Fig. 4) interact with the highly conserved side chain of residue Glu22, and with the less conserved residues Gln47 and His267. This suggests that these interface water clusters and their interacting side chains play an essential role in dimer formation and stability. This might be particularly true for the cluster interacting with residue Glu22 (cluster 37), which is predicted to bind very tightly (Dowser energy was −26.0 kcal/mol, very low if compared with the value of −12.0 kcal/mol, considered to be the minimum energy for a cavity to be occupied by a stable water molecule). Other clusters have functions that might be putatively related to stabilization of active site residues. An example is represented by the buried cluster 5 (Dowser energy −20.0 kcal/mol) (Fig. 2) that stabilizes the side chain of residue Arg363, responsible for binding the substrate α-carboxylate and for properly orienting the bond to be cleaved, according to the Dunathan hypothesis of PLP-dependent catalytic mechanism (Dunathan, 1966). Results of the MD experiments concerning cluster 5 strongly support its role in the stabilization of the interacting side chains.

Cleft water clusters (Table II) are found in higher number and the majority of them are located near the protein surface (Fig. 6). Here, only those pairs interacting with SHMT side chain residues and related by dyadic symmetry or occurring at the active site were considered more closely. Cluster 3 (Fig. 7) anchors the cofactor phosphate group to the loop encompassed by positions 261–264 of the symmetrically related subunit, through an H-bond to His96. This cluster corresponds approximately to the so-called position *h* of the second layer of the PLP phosphate binding cup (Denessiouk *et al.*, 1999; Denesyuk *et al.*, 2003). Likewise, cleft cluster 4 anchors the cofactor phosphate oxygen in position 3 to the residue Arg235 and to the peptide oxygen of Gly263 of the other subunit. The presence of an H-bond between cluster 4 and Arg235 may enhance the cation–π interaction to Tyr55 (Vivoli *et al.*, 2009) and attenuate the frustration of this residue (Florio *et al.*, 2011). This cluster may be approximately associated with the position *i* of the second layer of the phosphate binding cup (Denessiouk *et al.*, 1999; Denesyuk *et al.*, 2003). It should be noted that positions *h* and *i* of the phosphate binding cup are occupied by a water molecule also in the enzyme L-*allo*-threonine aldolase, closely related to SHMT (Kielkopf and Burley, 2002; di Salvo *et al.*, 2014).

The two cleft clusters 1 and 2 (Fig. 7), positioned in proximity of the PLP binding site, may play a significant and critical role for catalysis since they interact with the residue Asp200, supposedly responsible for the stabilization of the protonation state of the pyridine nitrogen of the cofactor and the modulation of its electrophilicity (Griswold and Toney, 2011).

Conserved cleft clusters 35, 45 and 43 (Fig. 8) form a network of interactions which, along with the salt bridges involving conserved residues Arg63, Asp72 and Glu75, stabilizes the Gly-rich loop encompassed by sequence positions 58–67. This loop hosts two important residues: Tyr64 and Tyr65. The first tyrosine makes a staking interaction to the benzoic ring of the tetrahydrofolate cofactor, while the other residue is deemed to have a role in the conversion of a closed active site to an open structure (Contestabile *et al.*, 2000). It may be speculated that the stabilizing interactions among conserved cleft water clusters and Tyr65 side chain are essential for the stabilization of the conformation of this important loop, also in the absence of substrates or cofactors.

MD has been applied to hint at the possible role of at least two conserved water clusters predicted to be particularly stable in the interior of the SHMT structure. In both cases, results of the simulations pointed at a stabilization of the side chains of functionally important residues such as Arg363. Since not all clusters interact with conserved side chains, this role cannot obviously be generalized.

In conclusion, the often neglected examination of conserved water molecules in a protein family can help unraveling several aspects of the stability, and the function of an enzyme that may be otherwise difficult to explain. However, it should be pointed out that our approach is meant to detect only highly conserved water molecules; this implies that water molecules only transiently bound to the enzyme, e.g. during catalysis, are not captured by our analysis although they may be relevant for SHMT function. An example is represented by the water molecule localized in the active site near a conserved Glu residue (Glu57 in *e*SHMT) (Szebenyi *et al.*, 2004). This water molecule hydrogen bonds to the formyl oxygen atom (distance 2.91 Å) of 5-formyltetrahydrofolate, to one of the carboxylate oxygen atoms of Glu57 (from the adjacent subunit) (2.60 Å), and to Nε2 atom of the imidazole ring of His126, which is stacked on the PLP ring (Scarsdale *et al.*, 2000). Glu57, which is conserved in both bacterial and eukaryotic SHMTs, was proved to be involved in the mechanism of folate-dependent reactions catalyzed by SHMT. This water molecule, which is absent in the apo SHMT structures and could not be singled out in our analysis, may help stabilize the functional conformation of the ligands in the active site.

Supplementary data

Supplementary data are available at *PEDS* online.

Acknowledgments

This work has been partially funded by the Italian MIUR (Ministero dell'Istruzione, Università, Ricerca). We are grateful to the anonymous referees for their useful and constructive criticisms that helped improving our work.

References

Ahmed,M.H., Spyrakis,F., Cozzini,P., Tripathi,P.K., Mozzarelli,A., Scarsdale,J. N., Safo,M.A. and Kellogg,G.E. (2011) *PLoS ONE*, **6**, e24712.

Angelaccio,S. (2013) *Biomed. Res. Int.*, **2013**, 851428.

Angelaccio,S., Florio,R., Consalvi,V., Festa,G. and Pascarella,S. (2012) *Int. J. Mol. Sci.*, **13**, 1314–1326.

Angelaccio,S., Dworkowski,F., Di Bello,A., Milano,T., Capitani,G. and Pascarella,S. (2014) *Proteins*, **82**, 2831–2841.

Ball,P. (2008) *Chem. Rev.*, **108**, 74–108.

Best,R.B., Zhu,X., Shim,J., Lopes,P.E., Mittal,J., Feig,M. and Mackerell,A.D., Jr. (2012) *J. Chem. Theory Comput.*, **8**, 3257–3273.

Bramucci,E., Milano,T. and Pascarella,S. (2011) *Biochem. Biophys. Res. Commun.*, **415**, 88–93.

Contestabile,R., Angelaccio,S., Bossa,F., Wright,H.T., Scarsdale,N., Kazanina, G. and Schirch,V. (2000) *Biochemistry*, **39**, 7492–7500.

Daidone,F., Florio,R., Rinaldo,S., Contestabile,R., di Salvo,M.L., Cutruzzola, F., Bossa,F. and Paiardini,A. (2011) *Eur. J. Med. Chem.*, **46**, 1616–1621.

Davies,N.G., Browne,H., Davis,B., *et al.* (2012) *Bioorg. Med. Chem.*, **20**, 6770–6789.

de Beer,S.B., Vermeulen,N.P. and Oostenbrink,C. (2010) *Curr. Top. Med. Chem.*, **10**, 55–66.

Debler,E.W., Muller,R., Hilvert,D. and Wilson,I.A. (2009) *Proc. Natl Acad. Sci. U.S.A.*, **106**, 18539–18544.

Denessiouk,K.A., Denesyuk,A.I., Lehtonen,J.V., Korpela,T. and Johnson,M.S. (1999) *Proteins*, **35**, 250–261.

Denesyuk,A.I., Denessiouk,K.A., Korpela,T. and Johnson,M.S. (2003) *Biochim. Biophys. Acta*, **1647**, 234–238.

di Salvo,M.L., Contestabile,R., Paiardini,A. and Maras,B. (2013) *Med. Hypotheses*, **80**, 633–636.

di Salvo,M.L., Remesh,S.G., Vivoli,M., Ghatge,M.S., Paiardini,A., D'Aguanno, S., Safo,M.K. and Contestabile,R. (2014) *FEBS J.*, **281**, 129–145.

Dunathan,H.C. (1966) *Proc. Natl Acad. Sci. U.S.A.*, **55**, 712–716.

Edayathumangalam,R., Wu,R., Garcia,R., *et al.* (2013) *Proc. Natl Acad. Sci. U. S.A.*, **110**, 17820–17825.

Edgar,R.C. (2004) *BMC Bioinformatics*, **5**, 113.

Florio,R., di Salvo,M.L., Vivoli,M. and Contestabile,R. (2011) *Biochim. Biophys. Acta*, **1814**, 1489–1496.

Griswold,W.R. and Toney,M.D. (2011) *J. Am. Chem. Soc.*, **133**, 14823–14830.

Heinig,M. and Frishman,D. (2004) *Nucleic Acids Res.*, **32**, W500–W502.

Henriques,E.F., Ramos,M.J. and Reynolds,C.A. (1997) *J. Comput. Aided Mol. Des.*, **11**, 547–556.

Hu,B. and Lill,M.A. (2014) *J. Comput. Chem.*, **35**, 1255–1260.

Humphrey,W., Dalke,A. and Schulten,K. (1996) *J. Mol. Graph.*, **14**, 33–38, 27–38.

Jager,J., Moser,M., Sauder,U. and Jansonius,J.N. (1994) *J. Mol. Biol.*, **239**, 285–305.

Jorgensen,W.L., Chandrasekhar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) *J. Chem. Phys.*, **79**, 926–935.

Kielkopf,C.L. and Burley,S.K. (2002) *Biochemistry*, **41**, 11711–11720.

Li,Z., He,Y., Liu,Q., Zhao,L., Wong,L., Kwoh,C.K., Nguyen,H. and Li,J. (2013) *BMC Bioinformatics*, **14**(Suppl 16), S11.

Magrane,M. and Consortium,U. (2011) *Database*, **2011**: bar009.

McDonald,I.K. and Thornton,J.M. (1994) *J. Mol. Biol.*, **238**, 777–793.

Milano,T., Paiardini,A., Grgurina,I. and Pascarella,S. (2013) *BMC Struct. Biol.*, **13**, 26.

Mustata,G. and Briggs,J.M. (2004) *Protein Eng. Des. Sel.*, **17**, 223–234.

Mustata,G.I., Soares,T.A. and Briggs,J.M. (2003) *Biopolymers*, **70**, 186–200.

Okamoto,A., Higuchi,T., Hirotsu,K., Kuramitsu,S. and Kagamiyama,H. (1994) *J. Biochem.*, **116**, 95–107.

Paiardini,A., Bossa,F. and Pascarella,S. (2004) *Protein Sci.*, **13**, 2992–3005.

Patel,H., Gruning,B.A., Gunther,S. and Merfort,I. (2014) *Bioinformatics*, **30**, 2978–2980.

Phillips,J.C., Braun,R., Wang,W., *et al.* (2005) *J. Comput. Chem.*, **26**, 1781–1802.

Poornima,C.S. and Dean,P.M. (1995) *J. Comput. Aided Mol. Des.*, **9**, 521–531.

Rose,P.W., Bi,C., Bluhm,W.F., *et al.* (2013) *Nucleic Acids Res.*, **41**, D475–D482.

Sanschagrin,P.C. and Kuhn,L.A. (1998) *Protein Sci.*, **7**, 2054–2064.

Scarsdale,J.N., Radaev,S., Kazanina,G., Schirch,V. and Wright,H.T. (2000) *J. Mol. Biol.*, **296**, 155–168.

Schirch,V., Hopkins,S., Villar,E. and Angelaccio,S. (1985) *J. Bacteriol.*, **163**, 1–7.

Shatsky,M., Nussinov,R. and Wolfson,H.J. (2004) *Proteins*, **56**, 143–156.

Suzek,B.E., Huang,H., McGarvey,P., Mazumder,R. and Wu,C.H. (2007) *Bioinformatics*, **23**, 1282–1288.

Szebenyi,D.M., Musayev,F.N., di Salvo,M.L., Safo,M.K. and Schirch,V. (2004) *Biochemistry*, **43**, 6865–6876.

Talwar,R., Jagath,J.R., Rao,N.A. and Savithri,H.S. (2000) *Eur. J. Biochem.*, **267**, 1441–1446.

Vivoli,M., Angelucci,F., Ilari,A., Morea,V., Angelaccio,S., di Salvo,M.L. and Contestabile,R. (2009) *Biochemistry*, **48**, 12034–12046.

Waterhouse,A.M., Procter,J.B., Martin,D.M., Clamp,M. and Barton,G.J. (2009) *Bioinformatics*, **25**, 1189–1191.

Williams,M.A., Goodfellow,J.M. and Thornton,J.M. (1994) *Protein Sci.*, **3**, 1224–1235.

Zhang,L. and Hermans,J. (1996) *Proteins*, **24**, 433–438.