

Research article

Open Access

Characterization of the cofactor-binding site in the SPOUT-fold methyltransferases by computational docking of S-adenosylmethionine to three crystal structures

Michal A Kurowski, Joanna M Sasin, Marcin Feder, Janusz Debski and Janusz M Bujnicki*

Address: Bioinformatics Laboratory, International Institute of Molecular and Cell Biology, Trojdena 4, 02-109 Warsaw, Poland

Email: Michal A Kurowski - michal@genesilico.pl; Joanna M Sasin - asia@genesilico.pl; Marcin Feder - marcin@genesilico.pl; Janusz Debski - jasiek@genesilico.pl; Janusz M Bujnicki* - iamb@genesilico.pl

* Corresponding author

Published: 14 March 2003

Received: 19 December 2002

BMC Bioinformatics 2003, 4:9

Accepted: 14 March 2003

This article is available from: <http://www.biomedcentral.com/1471-2105/4/9>

© 2003 Kurowski et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: There are several evolutionarily unrelated and structurally dissimilar superfamilies of S-adenosylmethionine (AdoMet)-dependent methyltransferases (MTases). A new superfamily (SPOUT) has been recently characterized on a sequence level and three structures of its members (Igz0, Iipa, and Ik3r) have been solved. However, none of these structures include the cofactor or the substrate. Due to the strong evolutionary divergence and the paucity of experimental information, no confident predictions of protein-ligand and protein-substrate interactions could be made, which hampered the study of sequence-structure-function relationships in the SPOUT superfamily.

Results: We used the computational docking program AutoDock to identify the AdoMet-binding site on the surface of three MTase structures. We analyzed the sequence divergence in two distinct lineages of the SPOUT superfamily in the context of surface features and preferred cofactor binding mode to propose specific function for the conserved residues.

Conclusion: Our docking analysis has confidently predicted the common AdoMet-binding site in three remotely related proteins structures. In the vicinity of the cofactor-binding site, subfamily-conserved grooves were identified on the protein surface, suggesting location of the target-binding/catalytic site. Functionally important residues were inferred and a general reaction mechanism, involving conformational change of a glycine-rich loop, was proposed.

Background

S-adenosyl-L-methionine (AdoMet or SAM) is the most commonly used donor of methyl groups in cellular alkylation reactions and is second only to ATP in the variety of reactions it serves as a cofactor (review: [1]). The AdoMet methyl group is bound to a charged sulfur atom (Figure 1), which thermodynamically destabilizes the molecule and makes it very reactive. The ΔG° in the reac-

tion of hydrolysis: AdoMet + homocysteine (Hcy) to S-adenosylhomocysteine (AdoHcy) + methionine is -17 kcal/mol [2]. RNA methylation is particularly diverse, with over 20 different methylated nucleosides identified in virtually all types of RNA molecules (review: [3]). The most abundant is methylation of 2'-hydroxyl groups of ribose. Among all, nearly 100 different posttranscriptional RNA

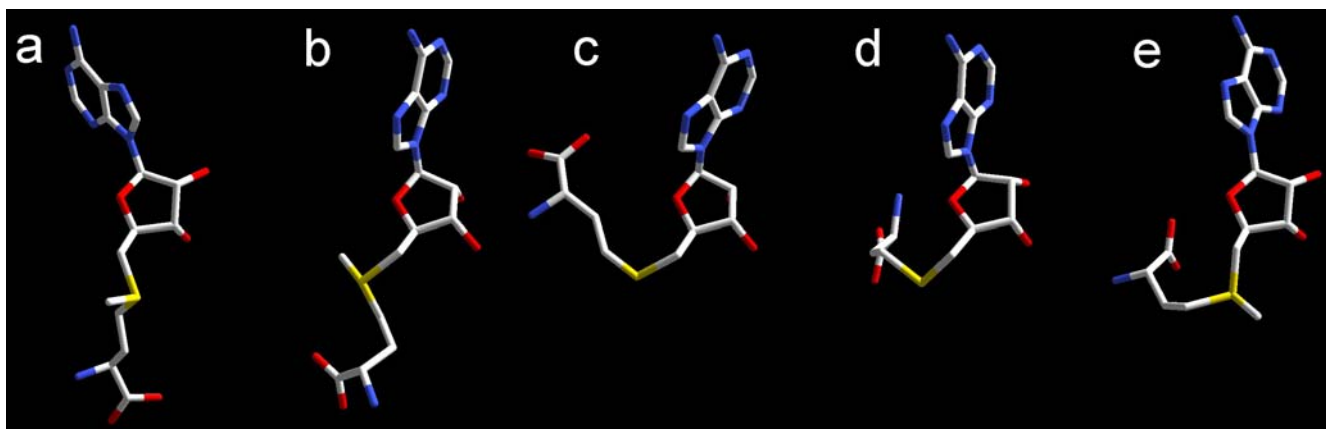


Figure 1

Comparison of the AdoMet/AdoHcy conformations in different MTase structures. a) "classical" Rossmann-fold MTase DpnM (2dpm); b) the MethH reactivation domain (1msk); c) CbiF MTase (1cbf); d) SET-superfamily MTase (1mt6), e) the best docked solution obtained in this study for the SPOUT-supefamily member 1gz0.

modifications, 2'-O-methylation is second only to pseudouridine formation.

The reactions of methyl transfer are catalyzed by AdoMet-dependent methyltransferases (MTases), which act on substrates as varied as nucleic acids, proteins, lipids, and small molecules (comprehensive review: [4]). Most of the known MTases, whose structures were solved by X-ray crystallography or NMR (currently over 30 structures in the Protein Data Bank) belong to a large superfamily related to Rossmann-fold proteins [5,6]. The "classical" Rossmann-fold proteins (RFP), which bind NAD(P), and the Rossmann-fold MTases (RFM), which bind AdoMet, use structurally equivalent and evolutionarily conserved cofactor-binding site and they interact with the adenosine and ribose moieties of their ligands in a very similar manner. In RFM, AdoMet assumes an extended conformation (Figure 1a). Nearly all RFM and RFP exhibit analogous hydrophobic packing against the adenine rings and RFM and NAD-binding RFP coordinate one or both of the adenosine ribose hydroxyls by Asp/Glu (in NADP one of the ribose hydroxyls is phosphorylated and no such bonding can occur in NADP-binding RFP). The methionine moiety of AdoMet has no counterpart in NAD(P) and is bound in a unique way by RFM: in motif I, another conserved Asp/Glu residue coordinates the amino group of methionine by a water-mediated contact, while the glycine-rich region forms a loop (G-loop) with some residues in "disallowed" region of the Ramachandran plot, which accommodates the "sidechain" of AdoMet [6,7]

There are several groups of AdoMet-dependent MTases, which neither share the RFM/RFP fold nor are structurally

or evolutionarily related to one another. Because of their independent evolutionary origin, they should be classified as "superfamilies", regardless of the relatively scarce number of well-characterized representatives. The activation domain of methionine synthase (MetH) [8] and the B12 biosynthetic enzyme CbiF [9] are single examples of structurally characterized representatives of superfamilies with alternative folds that can support AdoMet-dependent methyl transfer reactions (review: [10]). In MetH, AdoMet assumes an extended conformation (Figure 1b), which is distinct from that observed in RFM. The adenine ring is stacked between two tyrosines, but the polar protein-ligand interactions include interactions with conserved Arg residues [8], which is distinct from RFM. In the CbiF structure, AdoHcy (a product of hydrolysis of AdoMet) assumes a folded conformation (Figure 1c), its adenine moiety is not enclosed by hydrophobic amino acids, while the ribose hydroxyls and amino and carboxylate groups of homocysteine interact with main chain NH and CO groups rather than with Asp/Glu [9]. In the recently solved structures of SET-superfamily members histone:lysine N-MTase Set7/9 [11] and Rubisco:lysine N-MTase [12] AdoHcy also assumes a folded conformation (Figure 1d). Its ribose hydroxyls do not make hydrogen bonds with the protein and the amino and carboxyl groups of homocysteine make several contacts with the protein backbone and only one with the side chain (of a conserved Asn residue). The N6 and N7 atoms of adenine are hydrogen-bonded to the main chain amide and carbonyl groups, however the hydrophobic adenine-binding pocket is not present in all members of the SET superfamily. It is obvious that while the AdoMet-binding site is partially conserved within the individual MTase

superfamilies, there is little resemblance between both the cofactor conformation and protein-ligand interactions in unrelated enzymes.

Recently, another superfamily of AdoMet-dependent MTases has been defined based on bioinformatics analyses and dubbed SPOUT for the two major lineages: SpoU and TrmD [13]. All experimentally characterized members of this superfamily act on RNA: the SpoU relatives are 2'-O-ribose MTases [14] and orthologs of TrmD are tRNA:m¹G MTases [15]. Three structures of SPOUT-superfamily members have been reported: 23S rRNA:G2251 2'-O-MTase RlmB [16](1gz0 in PDB), hypothetical RNA 2'-O-MTase with unknown specificity [17](1ipa in PDB) and hypothetical RNA-binding protein MT1 [18]. All these proteins comprise two domains, which exhibit different spatial arrangements. The smaller domain is not conserved and exhibits structural similarity to various unrelated RNA-binding proteins. The large, conserved domain (the actual SPOUT domain) exhibits a novel and unusual fold with a deep knot.

Anatharaman et al [13] studied the sequence conservation in the SPOUT superfamily and hypothesized that the AdoMet-binding site corresponds to a glycine-rich loop (G-loop), localized in the C-terminal part of the SPOUT domain. Subsequent determination of the aforementioned crystal structures revealed that the moderately conserved region localized C-terminally to the G-loop corresponds to a topological knot [16–18]. Unfortunately, none of the published structures of SPOUT MTases include the cofactor. In the MAD electron density map of RlmB, a segment of unassigned density was observed in the knotted region, which was interpreted as a noncovalently bound small molecule [16]. Nevertheless, the authors were not able to unequivocally identify this molecule in the native structure, nor were they convinced to model it as AdoMet or AdoHcy using the data obtained from cocrystallization experiments. Hence, the precise localization of the cofactor-binding site and the mode of protein-ligand interactions in the SPOUT superfamily remain obscure.

Results and Discussion

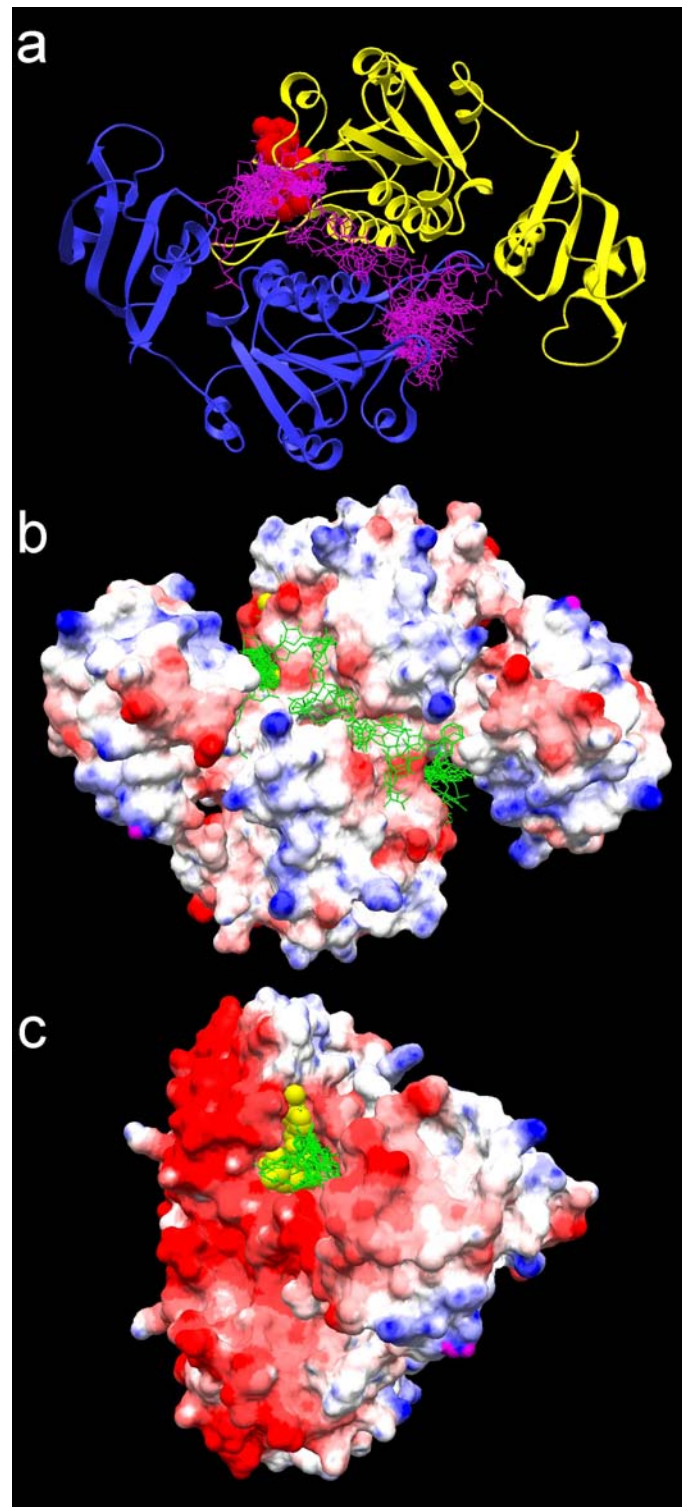
Identification of the AdoMet-binding site

To aid the experimental analysis of the SPOUT MTases in the absence of appropriate co-crystal structures, we decided to investigate the binding of AdoMet to the three available crystal structures (1ipa, 1gz0, and 1k3r) using the computational docking program AutoDock 3.05 [19]. In this approach, the surface of the protein is explored to identify fields that are energetically most favorable for interaction with the flexible ligand. According to the crystallographic analyses, the biologically relevant form is a homodimer, hence we used MTase dimers as targets in all

docking simulations. Since the asymmetric unit of 1gz0 contained 8 monomers, of which two were partially disordered, we selected a representative dimer, comprising chains A and F.

Initially, we disregarded all previous predictions of the AdoMet-binding site and carried out the search for all potential AdoMet-binding sites for a whole-molecule grid. The preliminary docking analysis was carried out using the Genetic Algorithm (GA) and Lamarckian Genetic Algorithm (LGA) methods (see Methods for details). Each docking experiment consisted of a series of 100 simulations, each producing a docking solution. The solutions were first sorted in terms of the similarity of the structures. Each set of solutions having a root-mean-square (rms) deviation of all atoms of less than 0.5 Å in pairwise comparisons was designated as a cluster of solutions. Only the lowest-energy solution of each cluster was retained. The experiment was repeated for all three structures (1ipa, 1gz0, and 1k3r).

Figure 2 shows the results of the preliminary docking analysis using the LGA algorithm, obtained for the 1gz0 structure with the grid encompassing the whole dimer. The distributions of solutions obtained with the GA and LGA methods for all three structures were qualitatively similar with respect to distribution of ligand conformations (data not shown). Nevertheless, the energy values of GA solutions were significantly higher than those of LGA (best solutions for 1gz0, 1ipa, and 1k3r [kcal/mol]: -6.64, -7.58, and -9.61 for GA vs -9.09, -9.73, and -11.38 for LGA), suggesting that the pseudo-Solis-Wets local search procedure was very efficient in finding local energetic minima. These results revealed that SPOUT-superfamily members exhibit strong preference to bind AdoMet in the same location, in a deep, negatively charged groove on the protein surface formed by three loops in the knotted C-terminal region. The lowest-energy preliminary docking solutions were always found in this region, regardless of the structure and the algorithm used, although the conformational details and orientation of the ligand with respect to the groove did vary; the most common lowest-energy conformation is indicated in Figure 2. Other docking solutions mapped to the interface between the catalytic domains of the MTase dimer (Figure 2), but they all exhibited significantly less favorable energies of protein-ligand interactions (-7 kcal/mol and higher), suggesting that they are less likely to correspond to physiologically relevant ligand-binding sites. These results confirmed the earlier sequence-based, low-resolution predictions of the cofactor-binding site [13,16] and allowed us to carry out a refined docking analysis focused on a defined structure fragment.

**Figure 2**

Preliminary (global) docking results mapped onto the structure of the Igz0AF dimer. a) Protein shown in the "cartoon" representation. Blue and yellow indicate different monomers. Docking solutions are shown in purple. The lowest-energy solution is shown in red. b) and c) docking solution mapped onto the protein surface (viewed from two different angles), colored by the electrostatic potential (blue = -5 kT, red = +5 kT). Docking solutions are shown in green, the best solution is shown in yellow.

To refine the prediction of the AdoMet binding mode in 1ipa, 1k3r and 1gz0, we calculated new affinity grids for each of these structures. Positions of the best-scoring ligands obtained in the initial calculations were chosen as the center of grids of dimensions $20 \text{ \AA} \times 20 \text{ \AA} \times 20 \text{ \AA}$, with grid points separated by 0.375 \AA and the docking simulations were repeated as described in Methods. Since we aimed at identification of the true energetic minimum, only the LGA method was used. Briefly, in the refined protocol, the solutions obtained from subsequent runs were first sorted in terms of the similarity of the structures to identify clusters and the analysis was carried out until the total number of clusters for a given docking experiment reached 50.

Conformation and interactions of AdoMet in the cofactor-binding pocket of SPOUT MTases

Figure 3 shows the populations of 20 lowest-energy docking solutions for all three structures, as obtained in the course of the LGA search. The values of the estimated energy of protein-ligand interactions for the best docking solutions are summarized in Table 2; the ligand-binding pockets of 1gz0, 1ipa, and 1k3r are shown in Figure 4. According to the results of our docking simulations, AdoMet binds to all SPOUT MTases in the same folded conformation, similar to that observed for AdoHcy in unrelated CbiF and SET MTases, and different from the extended conformation typical for the RFM superfamily (Figure 1). Nevertheless, in all "top 20" solutions for each SPOUT MTase, the ribose moiety of AdoMet is in the C2'-endo conformation, similarly to the RFM structures, while in CbiF it is C3'-endo (1cbf) and in the SET superfamily C1'-exo (Set79; 1mt6) or C2'-exo (Rubisco LSMT; 1mlv). In nearly all low-energy docking solutions, the cofactor adopts an *anti* conformation about the glycosidic bond (typical for all MTase structures solved to date (Figure 1; review: [6]), which maximizes the burial of the adenine moiety (Figure 3).

Despite similar conformations, the protein-ligand interactions in the CbiF, SET and SPOUT superfamilies are different. The key interactions between the three SPOUT MTases and AdoMet, present in the lowest-energy docking solution and in the majority of sub-optimal solutions, are listed in Table 3a. It is striking that despite the structural variability and extensive sequence divergence, many protein-ligand interactions were found to be conserved in all three cases. The adenine ring of AdoMet lies between the backbone of invariant Gly (G201 in 1gz0) and a variable (usually small) residue (A175 in 1gz0). The docking models explain why G201 is invariant – the adenine ring stacks against the C- α atom of Gly and any side-chain at this position would block the narrow groove. On the other hand, the other face of the adenine ring stacks against the side-chain of A175 (or its counterpart in the two other struc-

tures), which allows for different substitutions, depending on the space available to accommodate the side-chain in a way that it does not clash with the cofactor. The N6 and N7 groups of adenine are exposed to the solvent and not make any conserved interactions with the protein. N1 and N3 are usually surrounded by hydrophobic sidechains (N1 is solvent-exposed only in 1k3r) and do not form recurring hydrogen bonds. It seems that adenine moiety of the cofactor is recognized by SPOUT MTases mainly based on the shape complementarity, rather than any specific interactions.

Unlike the adenine ring, the ribose and methionine moiety of AdoMet seem to make conserved interactions with the protein, although nearly all of them are to the backbone carbonyl and amide groups rather than to side-chains. This agrees with the observation that there are no truly invariant residues in the cofactor-binding site of all SPOUT MTases [[13]; our unpublished data]. In all top-scoring docked solutions, the ribose 2'-hydroxyl invariably hydrogen-bonds to the backbone amide group of M202 in 1gz0 or a homologous (variable) residue in the other structures. The 3'-hydroxyl hydrogen-bonds to the backbone carbonyl of a semi-conserved Thr residue in 1ipa and 1k3r (but not in 1gz0) and to the backbone carbonyl of a nearly invariant G196 in 1gz0 (G218 in 1k3r; in 1ipa, the H-bond with G215 is present in many solutions, but not in the top-scoring one). In all three structures, the α -carboxyl group of methionine hydrogen-bonds to the backbone amide of an aliphatic residue, whose side-chain forms a floor of the cofactor-binding pocket (I216 in 1gz0). In 1gz0 and 1ipa, the backbone carbonyl of the same residue binds also the α -amino group of methionine. Instead, the α -amino group of the methionyl moiety in the 1k3r docked complex is hydrogen-bonded to the side-chains of N234 and Q239.

Spatial relationship of the AdoMet binding site and the predicted catalytic site

In the course of the analysis of sequence conservation in the entire SPOUT superfamily, including representatives with unknown structure, we noticed that the C-terminal motifs implicated in cofactor-binding, are generically conserved among all superfamily members, while the N-terminal motifs show subfamily-specific patterns of conservation (JMB and JD, unpublished data). The same is true for close homologs of the three structures analyzed in this work, suggesting possible correlation with binding of different substrates or/and different mechanisms of catalysis in 1gz0 and 1ipa on the one hand, and 1k3r on the other. Similar differential conservation of cofactor-binding and substrate-binding/catalytic regions is known to be correlated with functional differences in the RFM superfamily of MTases [6]. The availability of the crystal

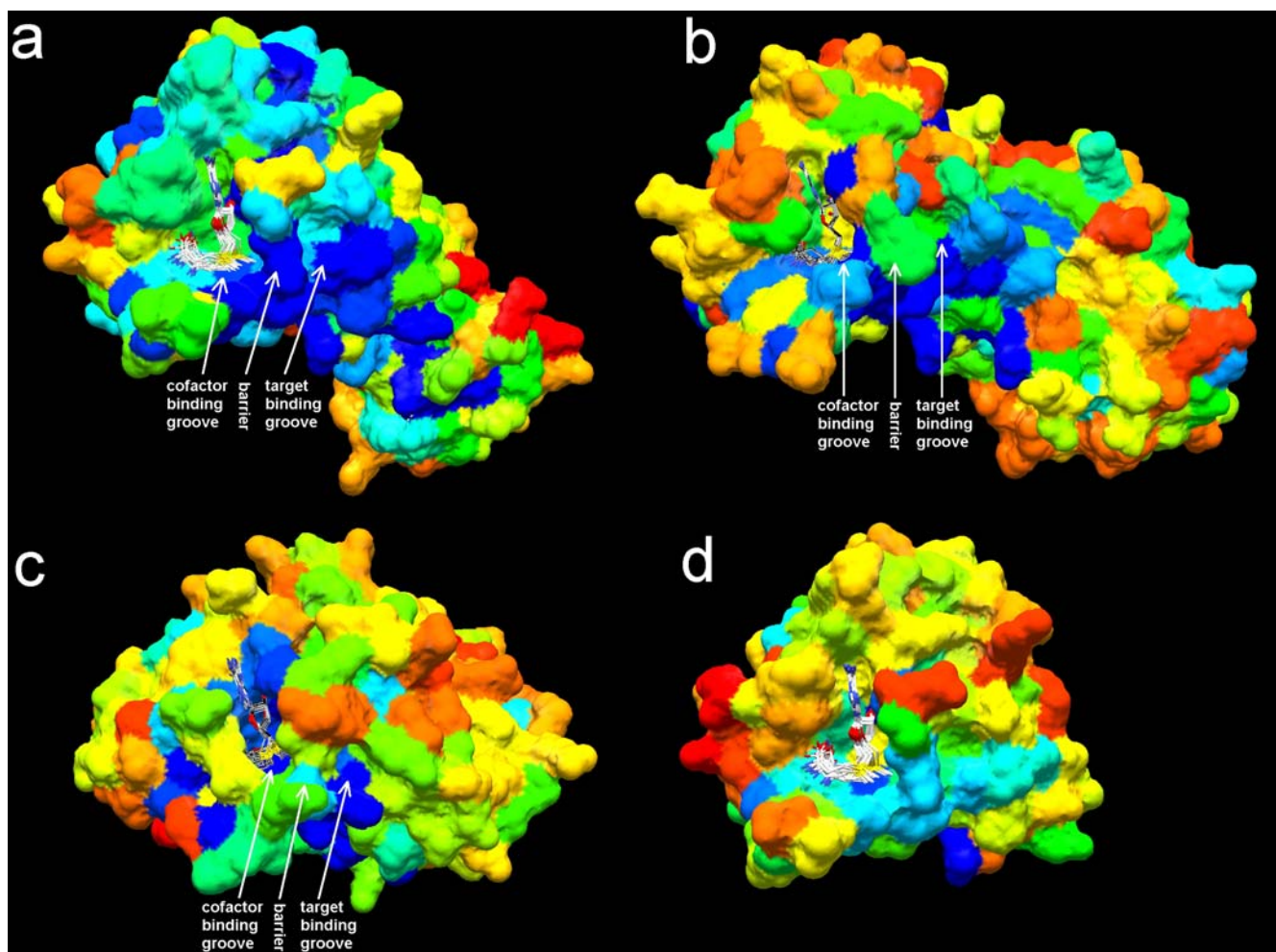


Figure 3
20 best docking solutions obtained during the refined (local) docking procedure, mapped onto the surface of the SPOUT superfamily members (monomers) colored by sequence conservation. The protein surface is colored according to the relative sequence conservation among its orthologs (blue – strongly conserved, to cyan – moderately conserved, to red – variable). The orthologous sets for each of the three structures are non-overlapping. a) 1gz0 (53 orthologous sequences), b) 1ipa (54 sequences), c) 1k3r (30 sequences). d) the SPOUT domain of 1gz0, colored by sequence conservation computed for all the three aligned families. Notably, when all three families are considered, the invariant and nearly-invariant residues disappear from the predicted target-binding groove, while a few of them still remain in the cofactor-binding groove (including Gly predicted to interact directly with AdoMet). The invariant side chain at the "bottom" of each monomer comes from the Arg residue, which interacts with the "barrier" structure of the second monomer in the dimer (shown in detail in Figure 4).

structures along with the results of our docking simulations allowed to analyze the evolutionary conservation and variability in the structural context and thereby infer sequence-structure-function relationships in the SPOUT superfamily.

In all three structures of SPOUT superfamily members, two grooves (one deeper, one more shallow) can be iden-

tified on the protein surface, separated by a barrier (Figure 3). The deep groove corresponds to the predicted cofactor-binding site, while the shallow groove is lined up with subfamily-specific residues, which are not conserved on the superfamily level. The barrier is formed by a G-loop (196-GAEGEG-201 in 1gz0, 215-GPEHEG-220 in 1ipa, 218-GGPYKG-223 in 1k3r). In all low-energy docking solutions, the methyl group of the cofactor is invariably

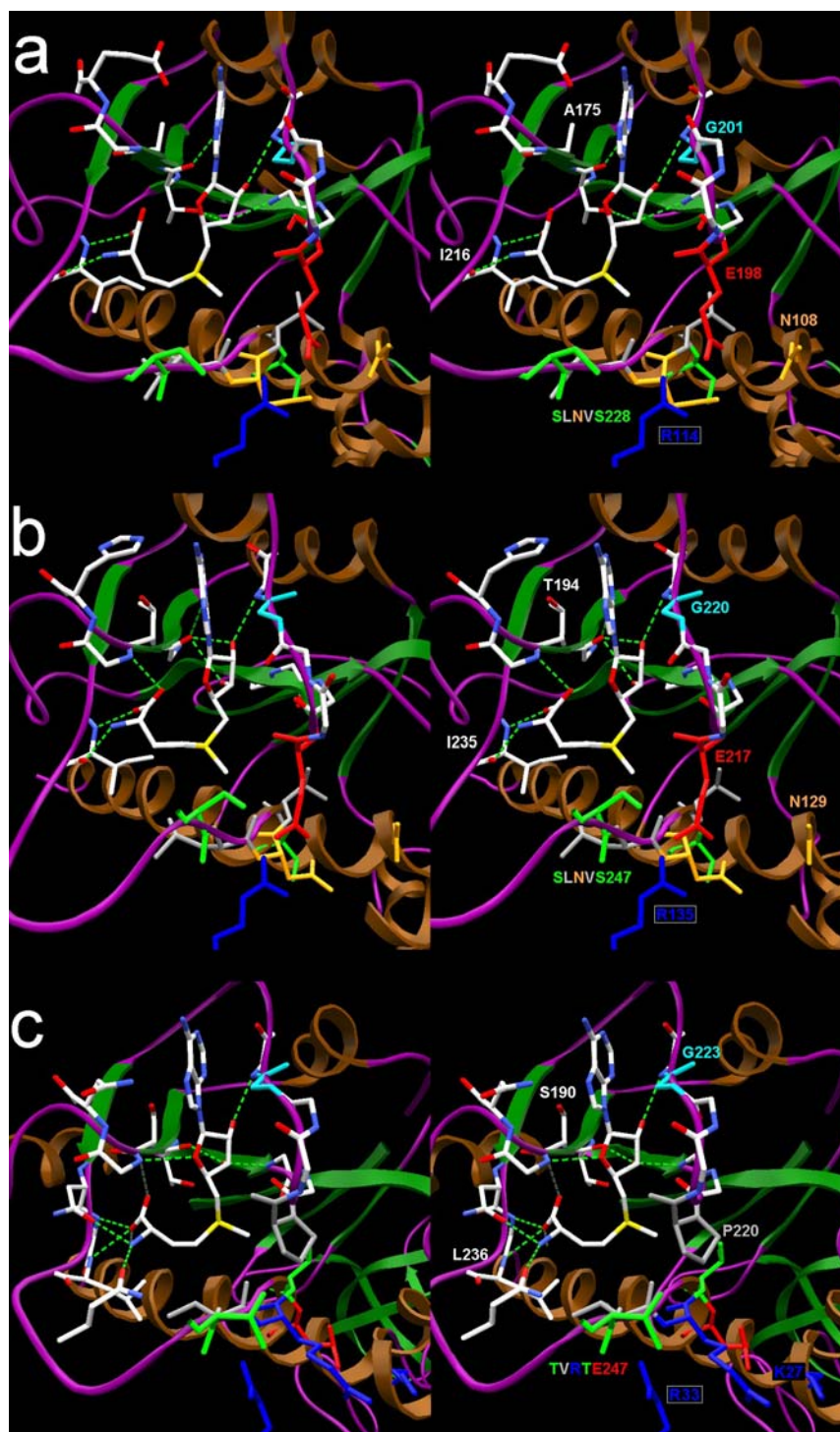


Figure 4

Lowest-energy docking solutions obtained for a) Igz0, b) Iipa, c) Ik3r. AdoMet and selected important residues are shown in the wireframe representation and labeled. The label for the invariant Arg side chain, provided by the second monomer, is boxed. The rest of the protein is shown in a schematic representation (brown helices, green strands and purple loops). Selected sidechains are colored according to their physicochemical properties (Arg and Lys – blue; Glu – red; Thr and Ser – green; aliphatic (Pro, Val, Leu) – gray; Gly – cyan). For the residues that bind the cofactor and for the cofactor itself, the following color scheme is used: C – white, O – red, N – blue, S – yellow). Predicted hydrogen bonds are shown as green broken lines.

Table 2: AutoDock Docking Results. Min and max energy fields correspond to the best LGA – global grid run. Rmsd values are taken from the clustering experiment; distances measured with the minimum-energy conformation as a reference.

structure	min docking energy [kcal/mol]	max docking energy [kcal/mol]	rmsd for top 50 clusters [Å]
1gz0	-14.18	-12.55	1.70
1ipa	-14.00	-12.65	1.82
1k3r	-15.01	-13.56	0.90

Table 3: Predicted function/contacts of conserved residues. Homologous residues making similar contacts in different structures are in the same rows. Interacting functional groups are indicated in parentheses ("b" denotes backbone). Contacts absent from the lowest-energy solution, but present in other "top 20" solutions are shown in italics.

Structure:	1gz0	1ipa	1k3r
a) AdoMet-binding			
adenine	A175 G201	T194 G220	S190 G223
ribose O2	M202 (bNH)	L221 (bNH)	L224 (bNH)
ribose O3	G196 (bNH)	<i>G215 (bNH)</i>	G218 (bNH)
	T174 (OH)	T193 (bCO)	T189 (bCO)
methionine COO⁻	I216 (bNH)	I235 (bNH)	L236 (bNH) T235 (bNH)
methionine NH₃⁺	I216 (bCO)	I235 (bCO)	N234 (OD1) Q239 (OE1)
b) target binding & catalysis (precise role unknown)	N108 E198 S224 N226 S228 R114 (2nd subunit)	N129 E217 S243 N245 S247 R135 (2nd subunit)	K27 P220 T243 R245 T246 E247 R33 (2nd subunit)

directed towards the barrier, and the shallow groove behind it. In the light of our docking simulation, this arrangement of protein surface features and residue conservation patterns strongly suggest that the shallow groove corresponds to the target-binding/catalytic site, which evolved to interact with different substrates. We hypothesize that the unbound structures (1gz0, 1ipa, and 1k3r) represent a "closed", catalytically inactive conformation, and that a structural change is required to open the barrier between the two grooves and thereby allow the substrate to carry out an enzyme-assisted nucleophilic attack on the methyl group of AdoMet. It seems that a moderate conformational change of the G-loop would be sufficient to bring the substrate close to the methyl group donor. It is tempting to speculate that such change could be induced by substrate binding. Nevertheless, modeling of such conformational transitions is at the brink of the available methodology and at the same time beyond the scope of the present study. Moreover, the target specificity is known only for the RlmB MTase/1gz0 (2'-hydroxyl of

G2251 in 23S rRNA) [20], while it remains to be determined for the other two proteins (1ipa and 1k3r), therefore we did not attempt to dock the substrate to the predicted catalytic pocket.

The list of conserved residues mapping to the vicinity of the predicted catalytic pocket (Figure 4) is shown in Table 3b. On the one hand, it is evident that 1gz0 and 1ipa have very similar active sites, which confirms earlier prediction that 1ipa is an RNA:ribose 2'-O-MTase [17]. On the other hand, the predicted active site of 1k3r is different, which suggests this protein and its orthologs may be involved in a different type of methylation.

It is noteworthy that without the knowledge of the ligand-binding site, confident predictions of the catalytic sites in the SPOUT superfamily could not be made. For the conserved residues in the 1k3r structure no functional predictions have been made even though its relationship to SPOUT MTases has been unambiguously identified [18].

In the 1ipa structure, three conserved residues (N129, E217, and N245) were hypothesized to "form the AdoMet-binding site and the catalytic site". Our analysis suggests that they all involved in target binding and catalysis and not in AdoMet binding. Michel et al [16] analyzed the conserved residues in the light of their structure (1gz0) and suggested that, by analogy to the classis AdoMet-binding site, E198 could bind the ribose moiety, while N108, R114, S224, and N226 could participate in recognition of the adenine ring or the α -amino and α -carboxyl group of the methionine. According to our model, all these residues may be important for substrate-binding and catalysis rather than AdoMet-binding. The presented docking solutions and predicted functionally important residues will facilitate experimental studies and aid the elucidation of the catalytic mechanism of the SPOUT-superfamily MTases

Conclusions

To explore potential cofactor-binding sites on the surface of the SPOUT-superfamily MTases, and to establish possible spatial relationships between the methyl group donor and the predicted catalytic sites of *E. coli* RlmB (1gz0), hypothetical MTase from *T. thermophilus* (1ipa), and *M. thermoautotrophicus* MT1 protein (1k3r), we simulated docking of AdoMet using AutoDock. The results show unequivocally that the preferred cofactor-binding site is in the groove in the knotted region and suggest a preferable ligand conformation, with its methyl group directed towards the putative catalytic site. Hence the computational docking procedure has not only identified the generic cofactor-binding cleft, but also generated protein-ligand complexes in a biologically relevant conformation. It is intriguing, but probably purely coincidental, that the cofactor-binding site of unrelated MTases from the SET superfamily is also localized on a knot.

We consider it significant that the three independent docking simulations identified a homologous region of the protein structure as the most likely cofactor-binding site, even in the absence of significant sequence similarity between the structures of 1ipa and 1gz0 and the structure of 1k3r. Our results reinforce the sequence-based prediction of Anantharaman et al. [13] that the cofactor-binding site is localized in the C-terminal part of the SPOUT domain. Our analysis has identified conserved protein-ligand contacts, which may be typical for the entire SPOUT superfamily and suggested which of the subfamily-specific residues may be important for binding of specific substrates and catalysis of particular variants of the methyltransfer reaction. The results of our study will be useful in designing future experiments to analyze the proposed interactions between the SPOUT MTases, AdoMet and the substrates.

Methods

Preparation of the ligand and target molecules for docking

Structures of Protein Data Bank (PDB) entries 1gz0, 1ipa and 1k3r were used for docking simulations of AdoMet. The protein targets and the ligand were prepared for docking using Autodock 3.05 [19,21] and AutoDockTools 1.1-alpha [22]. All "heteroatoms", including water molecules and ions were removed from the original files. The positions of polar hydrogens and charges were assigned using the Kollman algorithm [23]. Atomic solvation parameters and fragmental volumes were determined using the Addsol program. AutoTors was used to define AdoMet torsion angles. Flexible torsions were enabled on all bonds except the adenine ring. The ribose C1 atom was chosen as the ligand root. The carbon atoms of the adenine ring were designated as aromatic, the other carbons were designated as aliphatic. This affects both the torsions definition stage and subsequent force fields definition stage. After picking the root, the AutoTors procedure builds up the tree describing degrees of freedom for all atoms in the ligand molecule. The tree is then traversed to expand all branches defining existing torsions. Leaving out the rigid adenine bonds we obtained six flexible torsions total. Polar hydrogen charges of the Gasteiger-type [24] were assigned and the non-polar hydrogens were merged with the carbons. The protein sidechains were not allowed to change their conformation in any docking simulations described here.

Docking simulations

In the preliminary, global docking experiment, mass-centered grid maps were generated with 0.375 Å spacing by the AutoGrid program for the whole protein target. Lenard-Jones parameters 12-10 and 12-6 (supplied with the program package) were used for modeling H-bonds and Van der Waals interactions, respectively. The distance-dependent dielectric permittivity of Mehler and Solmajer [25] was used for the calculations of the electrostatic grid maps. The Genetic algorithm (GA) and Lamarckian genetic algorithm with the pseudo-Solis and Wets modification (LGA/pSW) methods were used with default parameters (Table 1). Random starting positions on the entire protein surface, random orientations and torsions were used for all structures. For all simulations the populations in the genetic algorithm was 50. Each simulation comprised $2,5 \times 10^6$ energy evaluations. Each docking experiment consisted of a series of 100 simulations.

In the refined, local experiment, the $20 \text{ \AA} \times 20 \text{ \AA} \times 20 \text{ \AA}$ grid was generated, with the center corresponding to the lowest-energy solution from the global search. Only the LGA/pSW method was used. The analysis was carried out until the total number of clusters for a given docking experiment reached 50. Other parameters of the docking

Table 1: Docking Parameters

Translation step	2.0 Å
Quaternion step	50.0°
Torsion step	50.0°
Translation reduction factor	1/cycle
Quaternion reduction factor	1/cycle
Torsion reduction factor	1/cycle
Elitism	1
Rate of gene mutation	0.02
Rate of crossover	0.8
No of generations for picking worst individual	10
Mean of Cauchy distribution	0
Variance of Cauchy distribution	1
No of iterations of Solis and Wets local search	300
No of consecutive successes before changing rho	4
No of consecutive failures before changing rho	4
Size of local search space to sample	1
Lower bound on rho	0.01

simulations were identical to that of the global search (see above and Table 1).

Clustering of the docking solutions

Docked conformations of the ligand were sorted in the order of increasing binding energy. The lowest-energy solution was used as a reference, and all other conformations were binned using a rms (root-mean-square) deviation threshold of 0.5 Å. For the whole-molecule docking procedure, clustering was carried out without taking any ligand conformation as a reference.

Sequence and structure analysis

Orthologs of the SPOUT MTases 1gz0, 1ipa, and 1k3r were identified and aligned using PSI-BLAST [26]. Multiple sequence alignments were edited manually to maximize the sequence conservation and to minimize the number of insertions and deletions in helices and strands in the reference structures. For closely related sequences (>90% identity) only single representatives were retained. The final alignments included only the orthologs of each of the structurally characterized protein (1gz0-53, 1ipa-54, and 1k3r-30) and were mutually exclusive (i.e. no sequence appeared in more than one alignment). The alignment of all sequences was guided by structural superposition of the SPOUT domain of 1gz0, 1ipa, and 1k3r. For each residue in the reference structure, the % sequence identity was calculated from the corresponding column in the alignment, and translated into temperature factors (using a reverse linear relationship, with 100% identity corresponding to the temperature factor of 00.00 and 10 % or less to 99.99). Protein surface analysis, including mapping the distribution of residue conservation and electrostatic potential, was carried out using SwissP-DB-Viewer [27].

Authors' contributions

MK performed the docking analysis and drafted the Methods section. JS participated in sequence analyses, mapping of the residue conservation onto the protein surface and identification of protein-ligand interactions. MF contributed to analyses of the crystallographic data. JD contributed to sequence analyses. JMB conceived of the project, coordinated its execution, analyzed and interpreted all data, prepared all figures and the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank Dr. Miroslaw Cygler for providing the 1gz0 structure and preprint of the manuscript before publication and Dr. Andrzej Joachimiak for releasing the 1k3r structure to the public (in PDB) before publication of its analysis. We are grateful to Dr. Osnat Herzberg for providing us with the unpublished, uncomplexed structure of another member of the SPOUT superfamily (not analyzed in this work). We would also like to thank Dr. Iddo Friedberg for critical reading of our manuscript. This analysis was supported by the EMBO/HHMI Young Investigator Award to JMB.

References

- Chiang PK, Gordon RK, Tal J, Zeng GC, Doctor BP, Pardhasaradhi K and McCann PP **S-adenosylmethionine and methylation**. *FASEB J* 1996, **10**:471-480
- Banerjee RV, Harder SR, Ragsdale SW and Matthews RG **Mechanism of reductive activation of cobalamin-dependent methionine synthase: an electron paramagnetic resonance spectroelectrochemical study**. *Biochemistry* 1990, **29**:1129-1135
- Cermakian N and Cedergren R **Modified nucleosides always were: an evolutionary model**. In *Modification and editing of RNA* (Edited by: Grosjean H, Benne R) Washington, DC: ASM Press 1998, 535-542
- Cheng X and Blumenthal RM *S-adenosylmethionine-dependent methyltransferases: structures and functions* Singapore: World Scientific Inc 1999,
- Bujnicki JM **Comparison of protein structures reveals monophyletic origin of the AdoMet-dependent methyltransferase family and mechanistic convergence rather than recent differentiation of N4-cytosine and N6-adenine DNA methylation**. In *Silico Biol* 1999, **1**:1-8 [<http://www.bioinfo.de/isb/1999/01/0016>]

6. Fauman EB, Blumenthal RM and Cheng X **Structure and evolution of AdoMet-dependent MTases.** In *S-Adenosylmethionine-dependent methyltransferases: structures and function* (Edited by: Cheng X, Blumenthal RM) Singapore: World Scientific Inc 1999, 1-38
7. Bugl H, Fauman EB, Staker BL, Zheng F, Kushner SR, Saper MA, Bardwell JC and Jakob U **RNA methylation under heat shock control.** *Mol Cell* 2000, **6**:349-360
8. Drennan CL, Huang S, Drummond JT, Matthews RG and Lidwig ML **How a protein binds B12: A 3.0 Å X-ray structure of B12-binding domains of methionine synthase.** *Science* 1994, **266**:1669-1674
9. Schubert HL, Wilson KS, Raux E, Woodcock SC and Warren MJ **The X-ray structure of a cobalamin biosynthetic enzyme, cobalt-precorrin-4 methyltransferase.** *Nat Struct Biol* 1998, **5**:585-592
10. Dixon M, Fauman EB and Ludwig ML **The black sheep of the family: AdoMet-dependent methyltransferases that do not fit the consensus structural fold.** In *S-Adenosylmethionine-dependent methyltransferases: structures and functions* (Edited by: Cheng X, Blumenthal RM) Singapore: World Scientific Inc 1999, 39-54
11. Jacobs SA, Harp JM, Devarakonda S, Kim Y, Rastinejad F and Khorasanizadeh S **The active site of the SET domain is constructed on a knot.** *Nat Struct Biol* 2002, **9**:833-838
12. Trievel R, Beach B, Dirk L, Houtz R and Hurley J **Structure and Catalytic Mechanism of a SET Domain Protein Methyltransferase.** *Cell* 2002, **111**:91
13. Anantharaman V, Koonin EV and Aravind L **SPOUT: a class of methyltransferases that includes SpoU and TrmD RNA methylase superfamilies, and novel superfamilies of predicted prokaryotic RNA methylases.** *J Mol Microbiol Biotechnol* 2002, **4**:71-75
14. Cavaille J, Chetouani F and Bachellerie JP **The yeast *Saccharomyces cerevisiae* YDL112w ORF encodes the putative 2'-O-ribose methyltransferase catalyzing the formation of Gm18 in tRNAs.** *RNA* 1999, **5**:66-81
15. Bystrom AS and Bjork GR **Chromosomal location and cloning of the gene (trmD) responsible for the synthesis of tRNA (mIG) methyltransferase in *Escherichia coli* K-12.** *Mol Gen Genet* 1982, **188**:440-446
16. Michel G, Sauve V, Larocque R, Li Y, Matte A and Cygler M **The structure of the RlmB 23S rRNA methyltransferase reveals a new methyltransferase fold with a unique knot.** *Structure* 2002, **10**:1303-1315
17. Nureki O, Shirouzu M, Hashimoto K, Ishitani R, Terada T, Tamakoshi M, Oshima T, Chijimatsu M, Takio K and Vassylyev DG **An enzyme with a deep trefoil knot for the active-site architecture.** *Acta Crystallogr D Biol Crystallogr* 2002, **58**:1129-1137
18. Zarembinski T, Kim Y, Peterson K, Christendat D, Dharamsi A, Arrowsmith CH, Edwards AM and Joachimiak A **Deep trefoil knot implicated in RNA binding found in an archaeobacterial protein.** *Proteins* 2003, **50**:177-183
19. Goodsell DS, Morris GM and Olson AJ **Automated docking of flexible ligands: applications of AutoDock.** *J Mol Recognit* 1996, **9**:1-5
20. Lovgren JM and Wikstrom PM **The rlmB gene is essential for formation of Gm2251 in 23S rRNA but not for ribosome maturation in *Escherichia coli*.** *J Bacteriol* 2001, **183**:6957-6960
21. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK and Olson AJ **Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function.** *J Comput Chem* 1998, **19**:1662
22. Sanner MF, Duncan BS, Carillo CJ and Olson AJ **Integrating computation and visualization for biomolecular analysis: an example using Python and AVS.** *Pac Symp Biocomput* 1999, 401-412
23. Chiche L, Gregoret LM, Cohen FE and Kollman PA **Protein model structure evaluation using the solvation free energy of folding.** *Proc Natl Acad Sci U S A* 1990, **87**:3240-3243
24. Gasteiger J and Marsili M **Iterative partial equalization of orbital electronegativity – a rapid access to atomic charges.** *Tetrahedron* 1980, **36**:3219-3228
25. Mehler EL and Solmajer T **Electrostatic effects in proteins: comparison of dielectric and charge models.** *Protein Eng* 1991, **4**:903-910
26. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402
27. Guex N and Peitsch MC **SWISS-MODEL and the Swiss-Pdb-Viewer: an environment for comparative protein modeling.** *Electrophoresis* 1997, **18**:2714-2723

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

