Rossmann-Fold Methyltransferases: taking a "*β*-turn" around Their cofactor, S-Adenosylmethionine

Bhanu Pratap Singh Chouhan, Shayida Maimaiti, Madhuri Gade, and Paola Laurino*

Okinawa Institute of Science and Technology Graduate University, 1919-1 Tancha, Onna-son, Okinawa 904-0412, Japan.

Supporting Information

ABSTRACT: Methyltransferases (MTases) are superfamilies of enzymes that catalyze the transfer of a methyl group from S-adenosyl Methionine (SAM) - a nucleoside-based cofactor to a wide variety of substrates such as DNA, RNA, protein, small molecules and lipids. Depending upon their structural features, the MTases can be further classified into different classes – wherein, we consider exclusively the largest class of MTases which are the Rossmann-fold MTases. It has been shown that the nucleoside-cofactor binding Rossmann enzymes particularly the nicotinamide adenine dinucleotide (NAD)-, flavin adenine dinucleotide (FAD)-, and SAM-binding MTases enzymes share common binding motifs which includes: a Gly-rich loop region that interacts with the cofactor, and a highly conserved acidic residue (Asp/Glu) that interacts with the ribose moiety of the cofactor. Here, we observe that the Gly-rich loop region of the Rossmann MTases adapts a specific type-II' β-turn in close proximity to the cofactor (< 4Å) and it appears to be a key feature of these superfamilies. Additionally, we demonstrate that the conservation of this β-turn could play a critical role in the enzyme-cofactor interaction, thereby shedding new light on the structural conformation of Gly-rich loop region from Rossmann MTases.

S-Adenosyl L-Methionine (SAM or AdoMet) Methyltransferases (MTases) are a large group of enzymes implicated in essential cellular activities including regulation of genes, proteins, chromatin, metabolic pathways as well as signal transduction[1]. Members of this group have been categorized into five distinct structural folds (class I-V) that bind SAM and catalyze the transfer of methyl group to different substrates[1]. A large majority of the known MTases belong to class-I 'Rossmann-fold'[1, 2] enzymes which in turn may be part of a multi-domain protein architecture (Figure 1A). The other minor classes include MetH reactivation domain (class-II), Precorrin-4 MTases (class-III), SPOUT family of RNA MTases (class-IV) and, SET-domain protein MTases (class-V)[1].

The α/β class Rossmann-fold[3] is one of the most common and widely distributed folds in nature[4,5,6,7]and therefore considered to be one of the most ancient protein architectures[8,9]. Topologically, the Rossmann fold consists of two tandem repeats constituted by six β-strands intercalated by four α-helices. The overall fold gives rise to a characteristic 'β-α-β' architecture with parallel hydrogen bonded β-strands constituting core of the β-sheet architecture[3,10]. The observed order for β-strands within the sheet architecture is '3-2-1-4-5-6' but the last strand may experience modification and be extended to include up to seven β-strands[1]. This can be seen in case of the Rossmann-fold Methyltransferases where the β-strand order is – '3-2-1-4-5-7-6' and the last beta-strand ('7') is oriented in an anti-parallel manner to the others (Figure 1A). In either case, the α/β units are arranged to give rise to a sandwich topology (defined by CATH as 3.40.50 in version v4.0.0). An intriguing feature of ancient proteins architectures is their dependence on cofactors to perform their activity[11,12]. The Rossmann enzymes are characterized by the use of nucleoside-containing cofactors[4,13,14] and the presence of a shared nucleoside-binding structural motif that has been described previously for nicotinamide adenine dinucleotide (NAD)-, flavin adenine dinucleotide (FAD)-, and SAM-utilizing Rossmann enzymes[15,16,17,18,19]. Specifically, the motif consists of: i). a Gly-rich loop region (β1 loop) between the β1 strand and the α1 helix that interacts with the cofactor and ii). an acidic residue (Asp/Glu) located at the tip of the β2 strand that forms bidentate interactions with the hydroxyls of the ribose moiety from the cofactor[5,20,21,22,23,24]. Furthermore, this structural motif (from NAD-, FAD-, and SAM-binding Rossmann Enzymes) has recently been identified as an ancient signature characteristic of divergent evolution (in contrast to convergent evolution)[25]. Because, the Gly-rich motif for the MTases seemed like the least conserved at the sequence level among the Rossmann enzymes we made an attempt to identify any structural features in this loop region that uniquely characterized the MTase Rossmann enzymes. Also, the proximity and interaction(s) of this loop region to the cofactor makes it a solid candidate for further investigation that could aid in our understanding of loop engineering for this class of enzymes. We performed a systematic study of the β1 loop region (Gly-rich loop) from all the available high-resolution crystal structures of Rossmann MTases in the Protein Data Bank (< 2.5Å) and explored the possibility of identifying potential conserved structural features (such as an α-turn, β-turn, γ-turn, δ-turn, or a π-turn). We then expanded our focus towards other non-MTase Rossmann enzymes and non-Rossmann proteins which harbor other nucleoside-based cofactors to confirm that this structural motif was unique for the MTase enzymes.

We started our analysis by creating a dataset of 199 high-resolution Rossmann MTases structures (<2.5 Å) extracted from PDB (Protein Data Bank). Herein, we observed that the Gly-rich loop region (β1-loop) for >90% of the structures in this dataset (178/199) showed the type-II' β-turn conformation[26] and this β-turn type-II' is constituted by four-residues turn and characterized by torsion angle values of φi+1= 60° and ψi+1= -120°, φi+2= 80° and ψi+2= 0° and a H-bond between the carbonyl oxygen of the amino acid residue (i) and the N-H of the amino acid residue (i+3) (Figure 1C, S1). Only, ~5% (12/199) of the structures are noted to deviate from the aforementioned torsion values and ~1% of the structures (3/199) do not adapt any β-turn conformation (Figure S2).

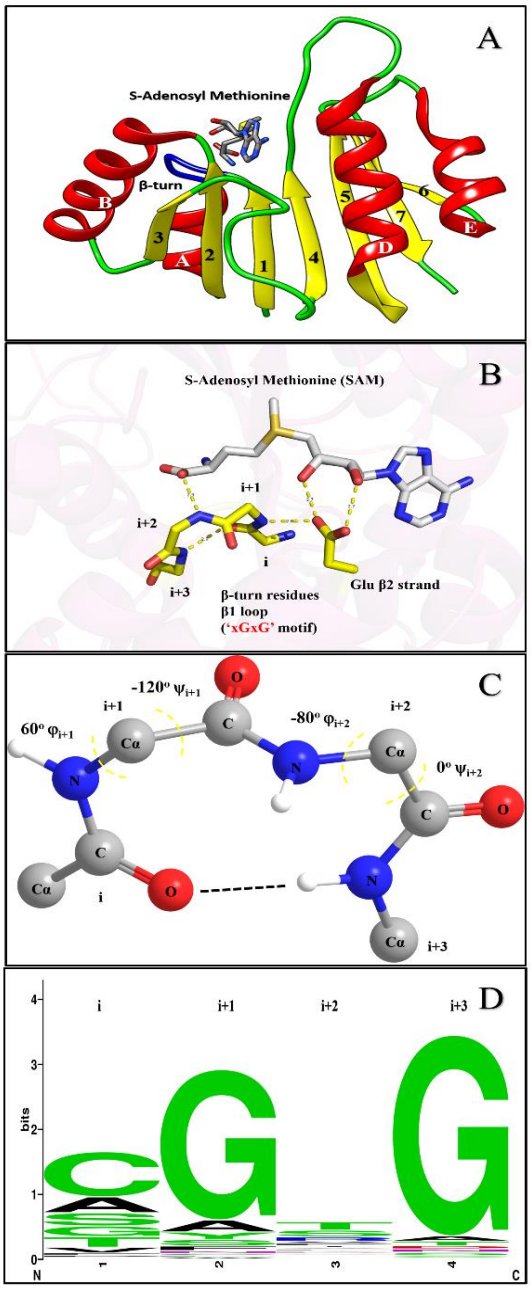

Figure 1. (A) Schematic architecture of the S-Adenosyl Methionine dependent Rossmann-fold Methyltransferase with '3-2-1-4-5-7-6' topology. The β-turn region (blue) with a 'xGxG' motif is located between the β-strand 1 and α-helix 1(or A) and it can form interactions with the cofactor S-Adenosyl Methionine (gray). **(B)** Interactions of S-Adenosyl Methionine with the β-turn from the "Glycine-rich" loop region and the acidic residue from the tip of the β2 strand (PDB: 1G60). **(C)** Three-dimensional ball-stick representation of a Type-II' β-turn with torsion values for positions φi+1= 60º ;ψi+1=-120º and φi+2= 80º;ψi+2= 0º. (D) Weblogo analysis for the Type-II' β-turn from the Rossmann Methyltransferases dataset highlighting the conserved glycine residues at positions i+1 and i+3 (Table S2 and for Weblogo analysis of other MTases superfamilies see Figure S3).

Data set (%)	Type II′ β-turn (%)	Other turns/deviations (%)
Rossmann MTases	94	6
non-MTase Rossmanns[b,c]	3	97
overall data set[b,c]	3	97
SAM-containing proteins[c]	31	69
SAH-containing proteins[c]	25	75

Table 1. Percentage-wise breakdown of type-II' β-turn vs other turn-types/deviations (<4Å distance from the cofactor) in our datasets. (a) Rossmann MTases dataset (Table S2) – for β-turn from the Gly-rich loop region. (b) Non-Mtase Rossmanns - Rossmann fold proteins that are not SAM-dependent MTAses. (c). β-turn composition of the overall dataset. (d,e). SAM/SAH-containing proteins (see Table S1). β-Turns calculated using Promotif[27]and Betaturntool (dunbrack.fccc.edu/betaturn/) to the calculate the occurrence of Type-II' β-turn.

* SAM/SAH containing protein dataset has been omitted, † Type-IV count has been omitted

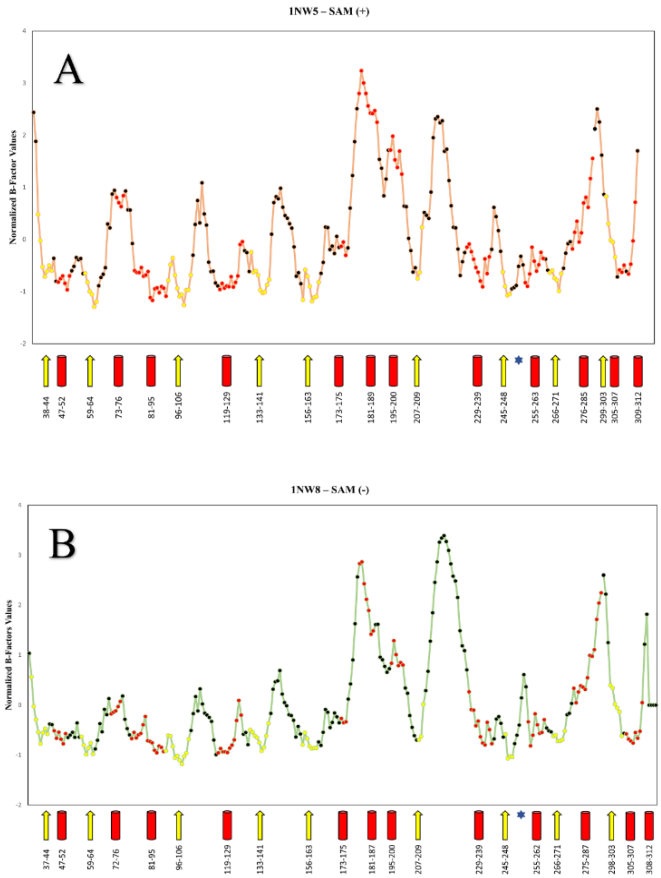

Figure 2. (A) Normalized Crystallographic Cα B-factor of the N6-adenine DNA methyltransferase from *Rhodobacter sphaeroides* (RsrI) bound to S-Adenosyl Methionine (PDB ID = 1NW5). Secondary structure assignment is based on DSSP classification, β-strands on the curve are highlighted in yellow, α-helices in red and loops in black (residue positions are indicated along the X-axis) and the β-turn containing "Glycine-rich" loop region has been highlighted (between β-strand 245-248 and α-helix 255-263). **(B)** S-Adenosyl Methionine absent.

A Weblogo analysis[28] of the type-II' β-turn residues (from Gly-rich loop) in this dataset highlighted a 'xGxG' motif[25] with highly conserved Gly residues at positions i+1 and i+3 (Figure 1, S3). Next, we examined the main chain Cα-atom flexibility of the Gly-rich loop region by performing the normalized B-factor (or temperature factor) analysis for representative structures belonging to different Rossmann Mtases superfamilies (DNA, RNA, protein, fatty acid and metalloid) in apo (- SAM) and holo (+ SAM) form (Figure 2, S4-S7). A comparison of the apo and holo form(s) suggests that generally there is certain amount of constraint in the main chain flexibility of the Gly-rich loop region, especially during the presence of SAM, indicating that this loop region could be pivotal for facilitating interaction between the enzyme and the cofactor SAM. In reference to this, we also observed that in case of some Rossmann MTases in our dataset (PDB ID: 1G60, 1HMY and 1NW5), either the oxygen atom or the carboxylate oxygen atom from the methionine moiety can form a H-bond with main chain nitrogen of the residue at position three (i+2) of the β-turn (Figure 1, S2). Although, this interaction pattern is not consistent across all Rossmann MTases superfamilies, we did observe that in some cases the methionine moiety interacts indirectly with the β-turn mediated by water molecules[24] (Figure S2). Furthermore, we also noted in the majority of the structures (~170/199) that the main chain nitrogen atom from Gly at position (i+1) of the β-turn may be able to form a H-bond with the side chain oxygen atom of the highly conserved acidic residue (Asp/Glu) from the tip of β2 strand (β2 Asp/Glu - 2.6~3.8Å range of distance), thereby stabilizing the side-chain of β2 Asp/Glu and facilitating interaction of the enzyme with the cofactor.

We then performed in-silico Alanine mutagenesis for each of the β-turn residues (from the Gly-rich loop region) from the Rossmann MTases dataset (Table S2)[29]. In this case we observed a wide range of change in the free energy (ΔΔG) values for each position of the β-turn, i.e. positions: i (-3.43 to 6.4), i+1 (-1.15 to 7.94), i+2 (-3.2 to 5.83) and i+3 (-6.33 to 10.15) suggesting that the Ala-mutation at each position could render the enzyme either stabilized or destabilized. This analysis unfortunately did not show a strong trend although, the range of destabilizing ΔΔG values are still higher for positions i+1 and i+3 (Glycines from the "xGxG" motif).

These observations motivated us to experimentally test the importance of the glycine residues in the β-turn. Therefore, we created a library by randomizing the four amino acids of the β-turn of DNA methyltransferase M.HaeIII[30]. Each round of evolution was enriched by transformation, growth, plasmid extraction and selection by digesting with restriction enzyme HaeIII. The methylation activity was gradually changed from non-detectable level in the first-round evolution (R1) to detectable level in the second round (R2), and soon after enriched in the third round (R3). The result demonstrated that the selection pressure is intense as only three rounds of evolution is sufficient for the enrichment of the four-amino acid of glycine rich loop on the β-turn (Figure 3).

Lastly, in order to identify formation of distinct β-turn types (within a range of 4Å of the cofactor) in other nucleoside-based cofactor(s) binding proteins, we created two additional datasets of high resolution protein structures (< 2.5 Å) extracted from the PDB (For dataset details see Supplementary information).

For the non-Mtases Rossmann dataset, we downloaded and analyzed Rossmann fold PDB structures (~1500 structures as classified by CATH DB) that contain other nucleoside-based cofactors [besides SAM (Tables S1 and S3)] to identify different β-turns located within a 4Å distance of the cofactors (results – Table 1, S1). For the overall protein dataset, similarly, we also downloaded and analyzed PDB structures (~8000 structures) that contain other nucleoside-based cofactors (besides SAM – Tables S1 and S3) to identify different β-turns located within a 4Å distance of the cofactors.

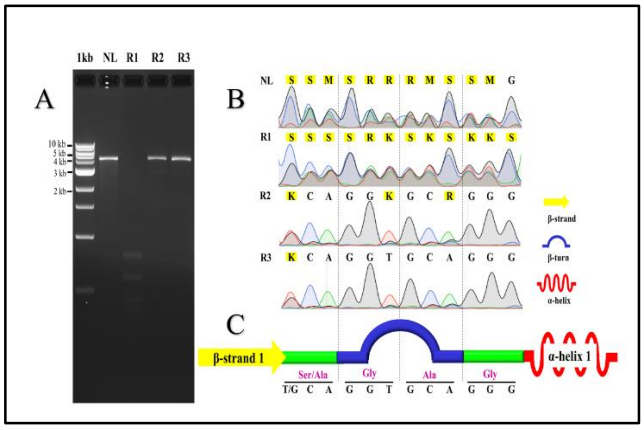

Figure 3. (A) pASK-IBA3+ plasmid containing stabilized M. HaeIII gene undergo random mutagenesis by PCR (Q5 site directed mutagenesis kit) using mutagenic primers on the beta turn sequence. Mutated M. HaeIII gene in pASK-IBA3+ plasmid was transformed into MC 1061 and expressed without induction. After extraction, the plasmid was treated with restriction enzyme HaeIII. NL denotes the Naive library. Round one (R1) showed undetectable level of protection against HaeIII, and whatever survived plasmids re-transformed for the second round (R2) of evolution. The second round indicated increased protection of the plasmid, and the survived plasmid underwent final round of evolution (R3). The third round exhibited complete protection of the plasmid. The plasmids were linearized with restriction enzyme NcoI. **(B)** Sequencing chromatograms of the library pool for the β-turn region between β-strand 1 and α-helix 1. Black line represents the percentage of G (guanine), red line percentage of T (thymine), green line percentage of A (adenine) and blue line percentage of C (cytosine). Nucleotides are highlighted in yellow before it gets enriched. The NL (Naive Library) shows mixed nucleotides signals at all 12 nucleotides positions. R1 (First round of selection) shows no enrichment in any of the 12 positions. R2 demonstrated three codons among four were enriched. R3 shows enriched codons in all 12 positions. **(C)** Schematic representation of the β-turn secondary structure and its surroundings of the M.HaeIII, showing the four amino acid positions where the site-directed mutations were introduced and the original nucleotide sequence.

It has been suggested that the conformational flexibility of the active-site loop region(s) coupled with their destabilization can be considered as key factors contributing towards the acquisition of broader substrate specificity[31]. However, our results suggest that the Gly-rich loop region from Rossmann MTases may not be very flexible [B-factor analysis, (Figure 2, S4-S7)] probably owing to a fixed three-dimensional conformation. Consequently, the Gly-residues placed at positions i+1 and i+3 ("xGxG" motif) of the β-turn are highly conserved [structural analyses (Figure 1) and experimental data (Figure 3)], and mutations of these Gly-residues could be destabilizing for the enzyme (Table S2). Our observations present an interesting scenario wherein on one hand this (Gly-rich) loop region may not be readily evolvable for the MTase Rossmann enzymes. On the other hand, a structured β-turn enables to design loops near to the active site region[32] by introducing motifs (mutations) that could help the loop region adapt a specific conformation with respect to the cofactor.

In summary, we have utilized bioinformatics tools and protein engineering-based approaches to analyze the Gly-rich loop region from the Rossmann MTases, other Rossmann Enzymes as well 40 nucleotide cofactor containing proteins (~8000 proteins) from the Protein Data Bank. This study clearly highlights the unique three-

dimensional geometry adapted by the Gly-rich loop region of the Rossmann MTases in the form of a type-II' β-turn. This structural feature provides important insights into the engineering of key loop regions for Rossmann MTases, especially the ones located in proximity to the cofactor.

ASSOCIATED CONTENT

Supporting Information

Description of materials and methods and supplementary figures provide as PDF supplementary information.

AUTHOR INFORMATION

Corresponding Author

*E-mail: paola.laurino@oist.jp
ORCID: 0000-0002-3725-2645
Funding Information: Financial support by Okinawa Institute of Science and Technology to PL is gratefully acknowledge.
Notes: The authors declare no competing interest.

ABBREVIATIONS

Methyltransferases, MTases; SAM, S-adenosyl Methionine; Nicotinamide adenine dinucleotide, NAD; Flavin adenine dinucleotide FAD, Glycine, Gly; Protein Data Bank, PDB.

REFERENCES

(1) Schubert, H. L., Blumenthal, R. M., and Cheng, X. (2003) Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem Sci*. 28, 329-335.
(2) Struck, A. W., Thompson, M. L., Wong, L. S., and Micklefield, J. (2012) S-adenosyl-methionine-dependent methyltransferases: highly versatile enzymes in biocatalysis, biosynthesis and other biotechnological applications. *Chembiochem*. 13, 2642-2655.
(3) Rao, S. T., and Rossmann, M. G. (1973) Comparison of super-secondary structures in proteins. *J Mol Biol*. 76, 241-256.
(4) Aravind, L., Mazumder, R., Vasudevan, S., and Koonin, E. V. (2002) Trends in protein evolution inferred from sequence and structure analysis. *Curr Opin Struct Biol*. 12, 392-399.
(5) Xie, L., and Bourne, P. E. (2008) Detecting evolutionary relationships across existing fold space, using sequence order-independent profile-profile alignments. *Proc Natl Acad Sci U S A*. 105, 5441-5446.
(6) Edwards, H., Abeln, S., and Deane, C. M. (2013) Exploring fold space preferences of new-born and ancient protein superfamilies. *PLoS Comput Biol*. 9, e1003325.
(7) Bukhari, S. A., and Caetano-Anolles, G. (2013) Origin and evolution of protein fold designs inferred from phylogenomic analysis of CATH domain structures in proteomes. *PLoS Comput Biol*. 9, e1003009.
(8) Caetano-Anolles, G., and Caetano-Anolles, D. (2003) An evolutionarily structured universe of protein architecture. *Genome Res*. 13, 1563-1571.
(9) Caetano-Anolles, G., and Caetano-Anolles, D. (2005) Universal sharing patterns in proteomes and evolution of protein fold architecture and life. *J Mol Evol*. 60, 484-498.
(10) Rossmann, M. G., Moras, D., and Olsen, K. W. (1974) Chemical and biological evolution of nucleotide-binding protein. *Nature*. 250, 194-199.
(11) Caetano-Anolles, G., Kim, H. S., and Mittenthal, J. E. (2007) The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. *Proc Natl Acad Sci U S A*. 104, 9358-9363.
(12) Ma, B. G., Chen, L., Ji, H. F., Chen, Z. H., Yang, F. R., Wang, L., Qu, G., Jiang, Y. Y., Ji, C., and Zhang, H. Y. (2008) Characters of very ancient proteins. *Biochemical and biophysical research communications*. 366, 607-611.
(13) Nath, N., Mitchell, J. B., and Caetano-Anolles, G. (2014) The natural history of biocatalytic mechanisms. *PLoS Comput Biol*. 10, e1003642.
(14) Kim, K. M., and Caetano-Anolles, G. (2011) The proteomic complexity and rise of the primordial ancestor of diversified life. *BMC Evol Biol*. 11, 140.
(15) Gherardini, P. F., Wass, M. N., Helmer-Citterich, M., and Sternberg, M. J. (2007) Convergent evolution of enzyme active sites is not a rare phenomenon. *J. Mol. Biol*. 372, 817-845.
(16) Cheng, H., Schaeffer, R. D., Liao, Y., Kinch, L. N., Pei, J., Shi, S., Kim, B. H., and Grishin, N. V. (2014) ECOD: an evolutionary classification of protein domains. *PLoS Comput Biol*. 10, e1003926.
(17) Efimov, A. V. (1997) Structural trees for protein superfamilies. *Proteins*. 28, 241-260.
(18) Panchenko, A. R., and Madej, T. (2004) Analysis of protein homology by assessing the (dis)similarity in protein loop regions. *Proteins*. 57, 539-547.
(19) Bujnicki, J. M. (1999) 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*. 1, 175-182.
(20) Wierenga, R. K., Terpstra, P., and Hol, W. G. (1986) Prediction of the occurrence of the ADP-binding beta alpha beta-fold in proteins, using an amino acid sequence fingerprint. *J Mol Biol*. 187, 101-107.
(21) Dym, O., and Eisenberg, D. (2001) Sequence-structure analysis of FAD-containing proteins, Protein Sci 10, 1712-1728.
(22) Lupas, A. N., Ponting, C. P., and Russell, R. B. (2001) On the evolution of protein folds: are similar motifs in different protein folds the result of convergence, insertion, or relics of an ancient peptide world? *J Struct Biol*. 134, 191-203.
(23) Gana, R., Rao, S., Huang, H., Wu, C., and Vasudevan, S. (2013) Structural and functional studies of S-adenosyl-L-methionine binding proteins: a ligand-centric approach. *BMC Struct Biol*. 13, 6.
(24) Kozbial, P. Z., and Mushegian, A. R. (2005) Natural history of S-adenosylmethionine-binding proteins. *BMC Struct Biol*. 5, 19.
(25) Laurino, P., Toth-Petroczy, A., Meana-Paneda, R., Lin, W., Truhlar, D. G., and Tawfik, D. S. (2016) An Ancient Fingerprint Indicates the Common Ancestry of Rossmann-Fold Enzymes Utilizing Different Ribose-Based Cofactors. *PLoS Biol*. 14, e1002396.
(26) Hutchinson, E. G., and Thornton, J. M. (1994) A revised set of potentials for beta-turn formation in proteins. *Protein Sci*. 3, 2207-2216.
(27) Hutchinson, E. G., and Thornton, J. M. (1996) PROMOTIF--a program to identify and analyze structural motifs in proteins. *Protein Sci*. 5, 212-220.
(28) Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res*. 14, 1188-1190.
(29) Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L. (2005) The FoldX web server: an online force field. *Nucleic Acids Res*. 33, W382-388.
(30) Rockah-Shmuel, L., and Tawfik, D. S. (2012) Evolutionary transitions to new DNA methyltransferases through target site expansion and shrinkage. *Nucleic Acids Res*. 40, 11627-11637.
(31) Jacquet, P., Hiblot, J., Daude, D., Bergonzi, C., Gotthard, G., Armstrong, N., Chabriere, E., and Elias, M. (2017) Rational engineering of a native hyperthermostable lactonase into a broad spectrum phosphotriesterase. *Scientific reports*. 7, 16745.

(32) Murphy, P. M., Bolduc, J. M., Gallaher, J. L., Stoddard, B. L., and Baker, D. (2009) Alteration of enzyme specificity by computational loop remodeling and design. *Proc Natl Acad Sci U S A.* 106, 9215-9220.

TOC

Insert Table of Contents artwork here