

Metagenome mining and functional analysis reveal oxidized guanine

DNA repair at the Lost City Hydrothermal Field

Payton H. Utzman[¶], Vincent P. Mays[¶], Briggs C. Miller, Mary C. Fairbanks, William J.

Brazelton, Martin P. Horvath*

School of Biological Sciences, University of Utah, 257 S 1400 E, Salt Lake City, Utah, 84112,

USA

*Corresponding author: martin.horvath@utah.edu

[¶]These authors contributed equally to this work.

Abstract

The GO DNA repair system protects against GC → TA mutations by finding and removing oxidized guanine. The system is mechanistically well understood but its origins are unknown. We searched metagenomes and abundantly found the genes encoding GO DNA repair at the Lost City Hydrothermal Field (LCHF). We recombinantly expressed the final enzyme in the system to show MutY homologs function to suppress mutations. Microbes at the LCHF thrive without sunlight, fueled by the products of geochemical transformations of seafloor rocks, under conditions believed to resemble a young Earth. High levels of the reductant H₂ and low levels of O₂ in this environment raise the question, why are resident microbes equipped to repair damage caused by oxidative stress? MutY genes could be assigned to metagenome assembled genomes (MAGs), and thereby associate GO DNA repair with metabolic pathways that generate reactive oxygen, nitrogen and sulfur species. Our results indicate that cell-based life was under evolutionary pressure to cope with oxidized guanine well before O₂ levels rose following the great oxidation event.

Introduction

The Lost City Hydrothermal Field (LCHF) resembles conditions of a younger planet and thus provides a window to study the origin of life on Earth and other planets (1–3). Located 15 kilometers from the Mid-Atlantic Ridge, the LCHF comprises a series of carbonate chimneys at depths ranging from 700–800 meters below the ocean surface (3). The temperature and pH of the LCHF are both elevated, with temperature ranging from 40°C to 90°C and pH ranging from 9 to 11 (1). At this depth, light does not penetrate, magmatic energy sources are unavailable, and

dissolved carbon dioxide is scarce (1). Despite these environmental constraints, archaea and bacteria inhabit the chimneys and hydrothermal fluids venting from the subsurface (1,4). The chemoautotrophic microbes take advantage of chemical compounds generated by seafloor geochemical reactions such as serpentinization, the aqueous alteration of ultramafic rocks (1). Serpentinization produces hydrogen gas and low-molecular-weight hydrocarbons, which fuel modern microbial communities and also would have been needed to fuel self-replicating molecules and the emergence of primitive metabolic pathways as an antecedent to cellular life (1,2,5). Hydrothermal circulation underneath the LCHF depletes seawater oxygen, leading to an anoxic hydrothermal environment very different from the nearby oxygen-rich seawater (1,3,6–8). As such, the subsurface microbial communities may offer a glimpse into how life emerged and existed before the Great Oxidation Event that occurred over two billion years ago.

The unusual environmental conditions of the LCHF present several biochemical challenges to the survival of microbes (4,7,8). For example, high temperatures and alkaline pH conditions present at the LCHF potentiate chemistry to generate DNA-damaging, reactive oxygen species (9). However, it remains unclear whether reactive oxygen species (ROS) are a major threat to resident microbes given that the seafloor underneath the LCHF is largely devoid of molecular oxygen. We reasoned that the prevalence or absence of DNA repair pathways that cope with oxidative damage would provide insight to the question, are ROS a real and present danger to life at the LCHF?

The guanine oxidation (GO) DNA repair system addresses the most common type of DNA damage caused by reactive oxygen species, the 8-oxo-7,8-dihydroguanine (OG) promutagen (**Fig 1**) (10). The OG base differs from guanine by addition of only two atoms, but these change the hydrogen bonding properties so that OG pairs equally well with cytosine and

adenine during DNA replication. The resulting OG:A lesions fuel G:C → T:A transversion mutations if not intercepted by the GO DNA repair system (9). The GO system comprises enzymes encoded by *mutT*, *mutM*, and *mutY*, first discovered through genetic analyses of *Escherichia coli* that demonstrated specific protection from G:C → T:A mutations by these three genes (11–14). Homologs or functional equivalents of these GO system components are found throughout all three kingdoms of life (15–17), underscoring the importance of the system, yet there are several instances where particular bacteria (18,19) or eukaryotes (20) make do without one or more of these genes.

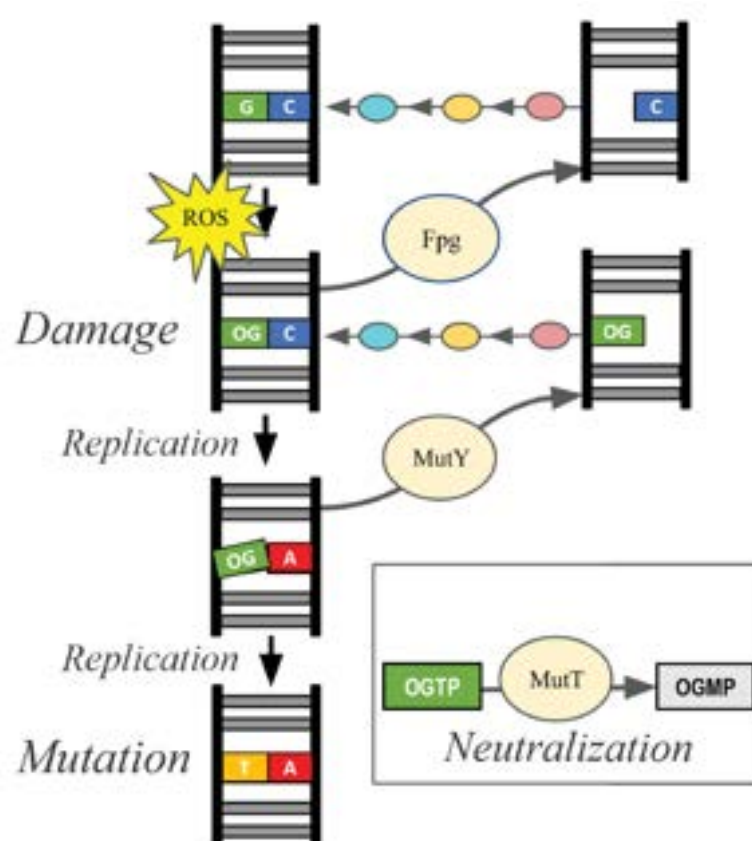


Fig 1. Overview of the GO System.

The gene products of *mutT*, *mutM*, and *mutY* (tan bubbles) prevent or repair oxidized guanine DNA damage caused by ROS. DNA glycosylases Fpg (encoded by *mutM*) and MutY remove OG from OG:C and A from OG:A, respectively, to create AP sites with no base. Additional enzymes (AP nucleases, pink; DNA polymerase, orange; and DNA ligase, teal) cooperate with the GO pathway, process these AP sites and ultimately restore the GC base pair. MutT neutralizes OG nucleotide triphosphates to prevent incorporation of the OG nucleotide during DNA replication, thereby ensuring that OG found in DNA is on the parent strand, not the daughter strand.

Biochemical analyses of gene products have provided a complete mechanistic picture for the GO repair system. MutT hydrolyzes the OG nucleotide triphosphate to sanitize the nucleotide pool, thus limiting incorporation of the promutagen into DNA by DNA polymerase (13,21). The

enzyme encoded by *mutM*, called formamidopyrimidine-DNA glycosylase (Fpg), locates OG:C base pairs and excises the OG base to initiate base excision repair (BER) (12,22). MutY locates OG:A lesions and excises the A base to initiate BER (11,14). Fpg and MutY thus act separately on two different intermediates to prevent G:C → T:A mutations. These DNA glycosylases generate abasic (apurinic/aprimidinic; AP) sites, which are themselves mutagenic if not processed by downstream, general BER enzymes, particularly AP nucleases (e.g. exonuclease III and endonuclease IV), DNA polymerase and DNA ligase (17,23–25) as shown in **Fig 1**.

MutY is the final safeguard of the GO system. If left uncorrected, replication of OG:A lesions results in permanent G:C → T:A transversion mutations as demonstrated by *mutY* loss of function mutants (26,27). Underperformance of the mammalian homolog, MUTYH, leads to early onset cancer in humans, first discovered for a class of colon cancers now recognized as MUTYH Associated Polyposis (28). MutY is made up of two domains that both contribute to DNA binding and biochemical functions. The N-terminal catalytic domain shares structural homology with EndoIII and other members of the Helix-hairpin-Helix (HhH) protein superfamily (17). The C-terminal OG-recognition domain shares structural homology with MutT and other NUDIX hydrolase family members (17,29). Functionally important and highly conserved residues define chemical motifs in both domains (**Fig 2**). These chemical motifs interact with the OG:A lesion and chelate the iron-sulfur cluster cofactor as revealed by x-ray structural analysis (**Fig 2**) (30–33). For example, residues in the N-terminal domain establish the catalytic mechanism for adenine excision (**Fig 2A** and **Fig 2B**) (32,34). Residues found in a beta loop of the C-terminal domain recognize the OG base and thus direct adenine removal from OG:A lesions (**Fig 2C**) (33). Some motifs are shared among other DNA glycosylases, such as the residues that chelate the 4Fe4S iron-sulfur cluster cofactor (**Fig 2D**) (17). Chemical motifs

particular to MutY, especially the OG-recognition residue Ser 308 (**Fig 2C**) and supporting residues in the C-terminal domain, are conserved across organisms and are not found in other DNA glycosylases and therefore can be used to identify MutY genes (17).

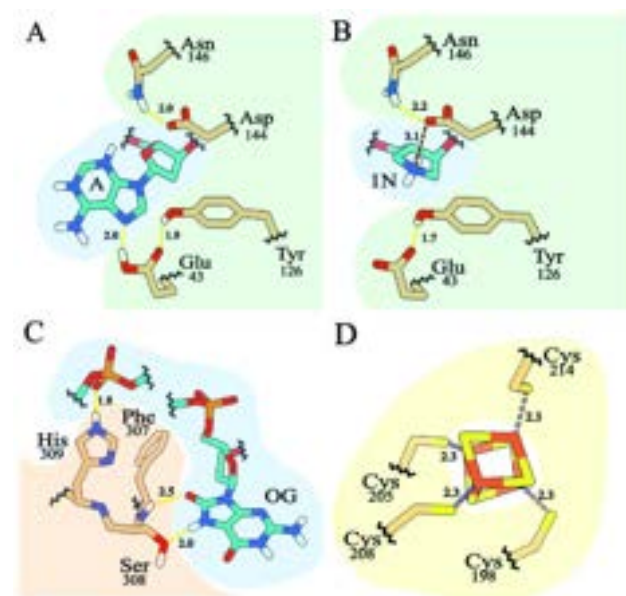


Fig 2. MutY chemical motifs. Panels A-C show interactions between MutY residues and DNA, with DNA highlighted in blue. Panel A was made from PDB ID 3g0q, a structure which describes interaction with the competitive inhibitor fluoro-A (31); panels B-D were made from PDB ID 6u7t, a structure which describes interaction with a transition state analog (1N) of the oxacarbenium ion formed during catalysis of adenine base excision (32). MutY catalytic residues (highlighted in green) interact with an adenine base (panel A) and the 1N moiety which mimics the charge and shape of the transition state (panel B). The OG-recognition residues (highlighted in orange) provide hydrogen bonding interactions with the OG base (panel

C). Four cysteine residues chelate the iron-sulfur metal cofactor (panel D). All of these interactions are important for MutY activity.

Our study investigated whether microbes in the anoxic LCHF environment use the GO DNA repair system to mitigate damage caused by reactive oxygen species. It is important to note that not all organisms have an intact GO repair system; examples are missing one, some or all components. MutY in particular was absent frequently in a survey of 699 bacterial genomes (19), and its absence may indicate relaxed evolutionary selection from oxidized guanine damage (18). We mined for homologous genes within the LCHF microbial community and recombinantly expressed candidate MutY enzymes to characterize function. We found genes encoding GO system components and general base excision repair enzymes at all LCHF sites. MutY homologs from the LCHF suppressed mutations when expressed in *mutY* deficient *E. coli* strains indicating these function similarly to authentic MutY. These Lost City MutY homologs could be assigned

confidently to metagenome assembled genomes (MAG)s, allowing for additional gene inventory analyses that revealed metabolic strategies involving sulfur oxidation and nitrogen reduction. These results have important implications for understanding the repair of oxidative guanine damage in low-oxygen environments, similar to those that existed on a younger Earth, as well as those that may exist on other planets and moons.

Results

Identification of the GO DNA repair system in LCHF microbes

To investigate the potential for LCHF microbes to endure DNA damage caused by ROS despite inhabiting a low oxygen environment, we searched for the GO DNA repair system in metagenomes obtained from LCHF hydrothermal fluids (35). We identified gene homologs for *mutT*, *mutM*, and *mutY*, which constitute the complete GO system (**Fig 3**). The relative abundances of these GO system gene homologs were similar to that of two other DNA repair enzymes that were also frequently found in LCHF metagenomes. Exonuclease III and endonuclease IV work in conjunction with the GO system and perform general functions necessary for all base excision repair pathways, namely the processing of AP sites (24). MutY was underrepresented in each of two samples from a chimney named "Marker 3", indicating that this GO system component is not encoded by some of the LCHF residents (**Fig 3**, pink).

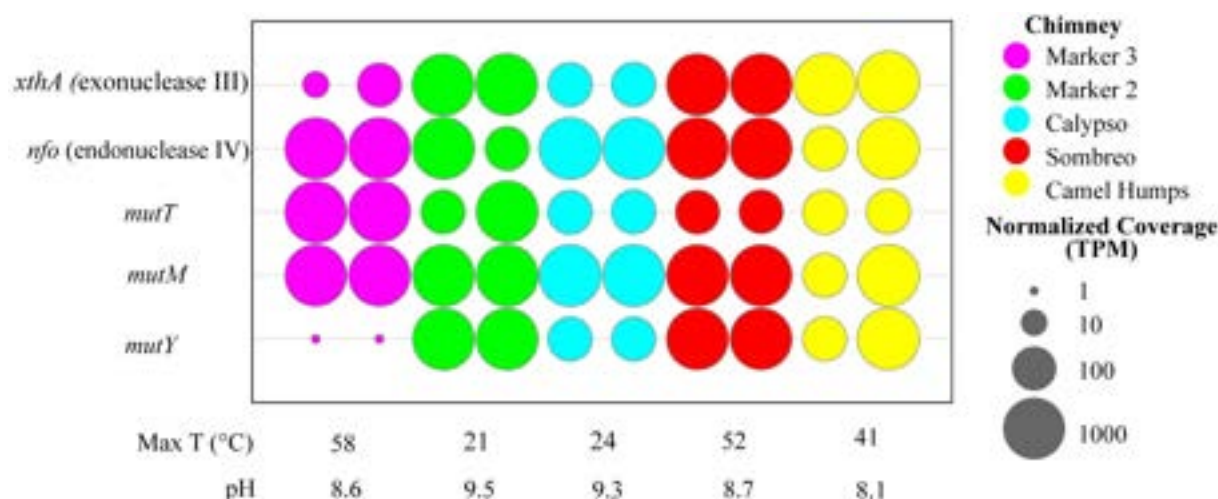


Fig 3. Abundance of GO system gene homologs. Listed on the vertical axis are genes encoding DNA repair enzymes. Genes *xthA* and *nfo* are generally necessary for DNA repair involving base excision repair in bacteria, including the particular GO system investigated here. Together, *mutT*, *mutM* and *mutY* constitute the GO system that deals specifically with oxidized guanine. Across the horizontal axis are the various LCHF sites, coded by color, from which samples were collected in duplicate, along with the reported temperature and pH. The normalized coverage of each gene is reported as a proportional unit suitable for cross-sample comparisons, the transcripts/fragments per million (TPM).

Metagenomic mining for MutY genes

Having determined that GO system gene homologs are abundant at the LCHF, we focused our efforts on the final safeguard of the pathway, MutY. A BLASTP search against the LCHF metagenome with query MutY sequences from *Geobacillus stearothermophilus* (*Gs* MutY) and *E. coli* (*Ec* MutY) preliminarily identified 649 putative MutY candidates on the basis of sequence identity, excluding hits with less than a 30% sequence identity cut-off or E-values exceeding 1E-5 (**Fig 4A**). Structure-guided alignments of these preliminary hits were examined for presence and absence of MutY-defining chemical motifs. We paid particular attention to the chemical motif associated with OG recognition as these residues in the C-terminal domain establish OG:A specificity, which is the hallmark of MutY (29,33,36). This approach

195 authenticated 160 LCHF MutYs (**Fig 4B**). Four representative LCHF MutYs were selected for
196 further analyses described below (red branches in **Fig 4A** and **Fig 4B**).

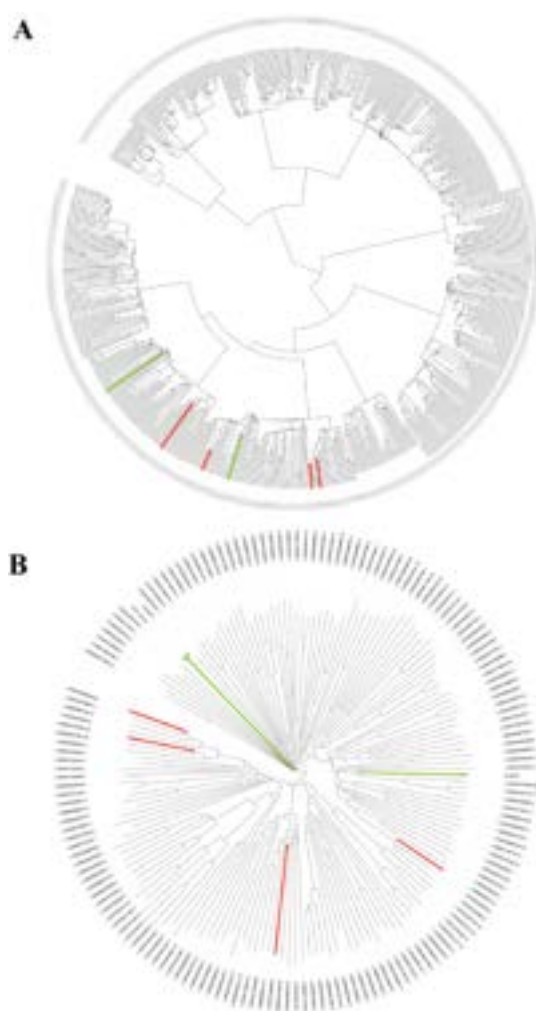


Fig 4. Phylogeny for LCHF MutYs. (A) 649 sequences were identified as LCHF MutY candidates due to sequence similarity to existing MutYs (green branches) and aligned to reconstruct evolutionary relationships. (B) A subset of 160 members contained all necessary MutY-defining chemical motifs. Alignment of these authenticated LCHF MutYs revealed varying evolutionary distances from familiar MutYs and provided a basis for selecting four representative members (red branches).

Fig 5 highlights conservation and diversity for the MutY-defining chemical motifs found in the 160 LCHF MutYs. All of the LCHF MutYs retain the chemical motif to coordinate the iron-sulfur cluster cofactor comprising 4 invariant Cys residues (4 Cs, yellow in **Fig 5**), a feature that is also found in other HhH family members such as EndoIII (16),
210 but which is absent for some “clusterless” MutYs

217 (37). Other invariant and highly conserved motifs make critical interaction with the DNA and
218 provide key catalytic functions for adenine base excision, explaining the high degree of sequence
219 conservation at these positions. For example, all LCHF MutYs use a Glu residue which provides
220 acid base catalysis for the mechanism (first E, green in **Fig 5**). Also, all LCHF MutYs use a Gln
221 (first Q, red in **Fig 5**) and a Tyr at position 88 (first Y, red in **Fig 5**), to wedge between base pairs
222 and thereby distort the DNA for access to the adenine as seen in x-ray crystal structures of *Gs*
223 MutY (30,31). Structures of *Gs* MutY interacting with a transition state analog revealed close

By contrast with these numerous and highly conserved motifs for the N-terminal domain, fewer motifs with greater sequence divergence were found in the C-terminal domain. The MutY ancestor is thought to have resulted from a gene fusion event that attached a MutT-like domain to the C-terminus of a general adenine glycosylase enzyme, and the C-terminal domain of modern *Ec* MutY confers OG:A specificity (29). X-ray crystal structures of *Gs* MutY interacting with OG:A and with G:A highlighted conformational difference for a Ser residue in the C-terminal domain (Ser308 in *Gs* MutY), and mutational analysis showed this Ser residue and its close neighbors (Phe307 and His309 in *Gs* MutY) establish OG specificity (33). Informed by these insights from structure, we eliminated LCHF MutY candidates that lacked a C-terminal domain and its OG-recognition motif. Alignment of the 160 LCHF MutYs that passed this test showed that a second His residue is also well conserved within the H-x-FSH sequence motif (**Fig 5C**). As is evident from the alignment, there are many variations with residues replaced by close analogs at each position. Ser, which makes the key contacts with N7 and O8 of OG and no contact with G, is often replaced by Thr, which can make the same hydrogen bond interactions. Likewise the two His residues are each often replaced by polar residues (e.g. Gln, Asn, Arg or Lys) that can also hydrogen bond to the DNA phosphate backbone as observed for His305 and His309 in *Gs* MutY.

Two other positions with high conservation were revealed for the C-terminal domain in this analysis of the 160 LCHF MutYs. These define a L-xxx-P motif. These residues are replaced by other residues with comparable chemical properties. The Leu position is often another hydrophobic residue such as Met or Phe, and the Pro position is most frequently replaced by Glu, a residue that can present aliphatic methylene groups and thus resemble Pro if the polar group hydrogen bonds with the peptide amide. In the structure of *Gs* MutY, the Pro269 nucleates a

hydrophobic core for the C-terminal domain. The Leu265 makes a strong VDW contact with Tyr89 in the N-terminal domain to support stacking of Tyr88 between bases of the DNA, a molecular contact that suggests communication between the OG-recognition domain and the catalytic domain. Other evolutionary analyses have highlighted the motifs important for DNA contacts, catalysis and OG recognition (17), but the L-xxx-P motif has not been identified previously.

Four representative LCHF MutYs were selected for further analyses. Supporting Information **Table S2** reports the percent identity among these representative LCHF MutYs and the well-studied MutYs from *E. coli* and *Geobacillus stearothermophilus*. LCHF MutY 1 and LCHF MutY2 are most closely related with 65% identity which is almost twice the average in this group. LCHF MutY 3 is most closely related to *Ec* MutY with 48% identity. We examined the representative LCHF MutYs for physical properties as inferred from sequences. **Table 1** reports these physical properties including predicted protein size, isoelectric point (pI), and stability (Tm). Generally the physical characteristics measured for LCHF MutY representatives were comparable to each other and to predicted properties of *Ec* MutY and *Gs* MutY. The predicted Tm for LCHF MutY 3 was above 65°C, distinguishing it as the most stable enzyme (**Table 1**), which may reflect adaptation to a high temperature environment. The isoelectric point predicted for each of the LCHF MutY representatives is 3 pH units above the pI predicted for *Gs* MutY and between 0.1 - 0.5 pH units above the pI predicted for *Ec* MutY, indicating that more numerous positively charged residues have been recruited, possibly as an adaptation to the LCHF environment.

Table 1. Physical Protein Properties.

MutY	Length (residues)	MW (kDa)	pI	Predicted Tm (°C) (N-domain; C-domain)
Gs MutY	372	41.8	5.3	55 - 65 (55 - 65 ; < 55)
Ec MutY	355	39.1	8.6	< 55 (< 55 ; 55 - 65)
LCHF MutY 1	358	39.0	9.1	55 - 65 (55 - 65 ; < 55)
LCHF MutY 2	352	38.3	8.8	< 55 (55 - 65 ; < 55)
LCHF MutY 3	370	42.0	8.7	> 65 (>65 ; 55 - 65)
LCHF MutY 4	376	44.0	9.0	55 - 65 (55 - 65 ; 55 - 65)

Identification of LCHF MutY organisms, gene neighbors, environmental conditions, and metabolic strategies

Our next objective was to identify the organisms from which these LCHF MutY enzymes originated. Each of the four representative LCHF MutY sequences were derived from contiguous DNA sequences (contigs) belonging to a metagenome-assembled genome (MAG) representing a LCHF microbe. The taxonomic classification of these MAGs indicated that LCHF MutY 1 originated from a species of *Marinosulfonomonas*, LCHF MutY 2 from the family *Rhodobacteraceae*, LCHF MutY 3 from the family *Thiotrichaceae*, and LCHF MutY 4 from the family *Flavobacteriaceae* (**Fig 6**). The taxonomic classification of each contig was consistent with the classification of the MAG to which it belonged, supporting the idea that the MutY gene is a long term resident and not a recent arrival through phage infection or some other horizontal gene transfer mechanism. For the remainder of this work we will refer to the MutY-encoding

organisms by the lowest-level classification that was determined for each LCHF MutY (e.g. LCHF MutY 3 will now be referred to as *Thiotrichaceae* MutY).

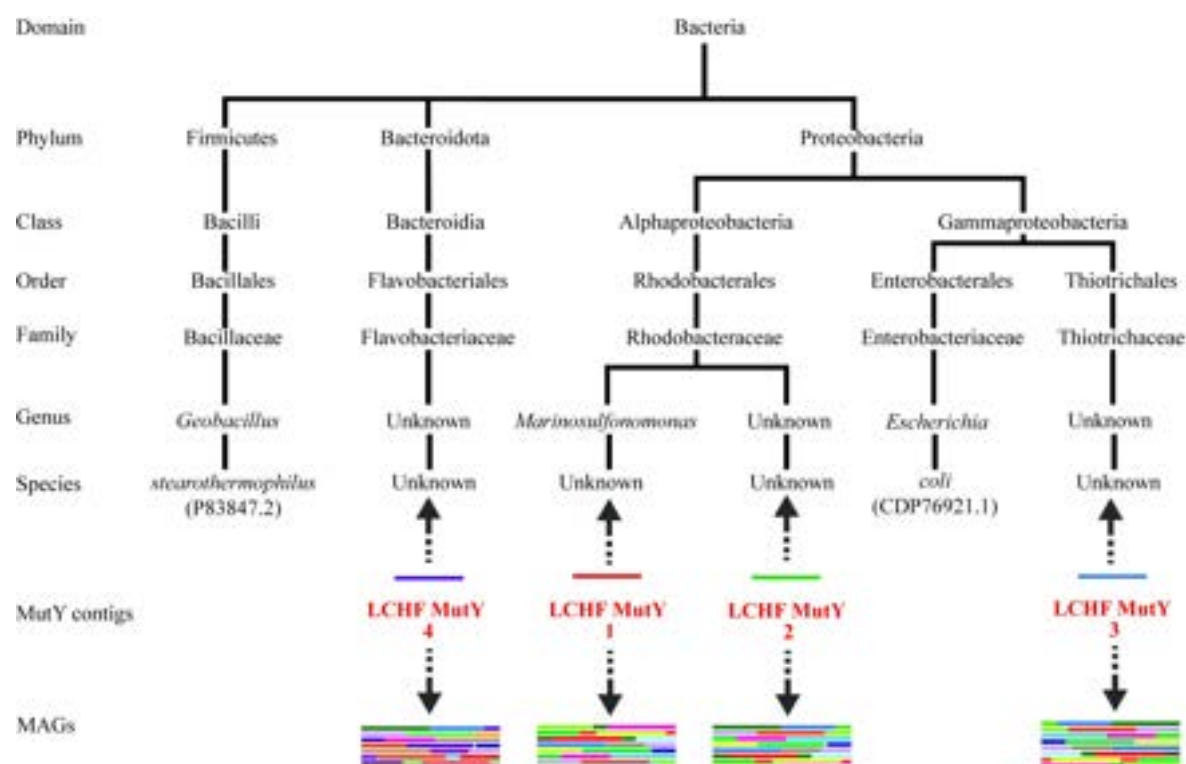


Fig 6. Taxonomic classification. LCHF MutY-encoding contigs were found in several branches of bacteria. The classification places these in relation to MutY of *G. stearothermophilus* and *E. coli* (accession IDs included). LCHF MutYs were mapped to their respective microbes by two methods indicated by the two arrows (see text for details).

The inclusion of MutY contigs in MAGs provided an opportunity to examine gene neighbors for the representative LCHF MutYs. The GO repair genes are located at distant loci in *E. coli* (12), and belong to separate operons (40). However, MutY is the immediate 5'-neighbor to YggX within gammaproteobacteria (40), and homologs of YggX are present outside this lineage, occasionally nearby to MutY (e.g. *Bacillus subtilis*). As gene neighbors, MutY and YggX are part of a SoxRS regulated operon in *E. coli* (40,41). YggX provides oxidative stress protection and iron transport function with a critical Cys residue close to the N-terminus of this

small protein (42,43). A protein matching these features is encoded by a gene partly overlapping with and the nearest 3' neighbor to *Thiotrichaceae* MutY (see Supporting Information **Fig S3**).

To reveal the environmental conditions of these MutY-encoding organisms, we analyzed the sequence coverage of each LCHF MutY contig at each of the sampling sites.

Marinosulfonomonas MutY, *Thiotrichaceae* MutY, and *Flavobacteriaceae* MutY were identified at all sampling sites, ranging from 21°C to 58°C and pH 8.1 to 9.5. *Rhodobacteraceae* MutY was present at all sampling sites excluding Marker 3 and was therefore found in temperatures ranging from 24°C to 52°C and pH 8.1 to 9.5.

We further investigated the metabolic strategies utilized by MutY-encoding microbes by examining the inventory of predicted protein functions in each MAG (**Table 2**, see also Supporting Information **Table S4**). Each LCHF MutY-containing MAG possessed at least two forms of cytochrome oxidase, with the exception of the *Flavobacteriaceae* MAG. The *Flavobacteriaceae* MAG is only 44% complete, however, so no strong conclusions can be made regarding the absence of genes. Cytochrome oxidases commonly provide sources of free radicals and are essential to aerobic metabolism. Predicted proteins indicative of dissimilatory nitrate and nitrite reduction were found in the *Marinosulfonomonas*, *Rhodobacteraceae*, and *Thiotrichaceae* MAGs, suggesting that these organisms may be capable of using nitrate or nitrite as alternative electron acceptors when oxygen is not available. Furthermore, the *Marinosulfonomonas* and *Rhodobacteraceae* MAGs include predicted protein functions associated with the oxidation of reduced sulfur compounds, though it is important to note that the directionality of these reactions cannot be fully determined from bioinformatics alone. These patterns speak to the potential origins of oxidants within the MutY-encoding organisms as discussed below.

354 **Table 2. Metabolic genes identified in LCHF MutY organisms**

			<i>Marinosulfonomonas</i> ^a		<i>Rhodobac-</i> <i>teraceae</i>	<i>Thiotrich-</i> <i>aceae</i>	<i>Flavobac-</i> <i>teriaceae</i>
			MAG 1	MAG 2			
Cytochrome Oxidases	UQCRFS1	K00411	X	X	X	X	
	coxA	K02274			X	X	
	ccoN	K00404	X	X	X		
	cydA	K00425		X			
	cyoB	K02298			X		
Sulfur Oxidation	soxA	K17222	X	X	X		
	soxX	K17223	X	X	X		
	soxB	K17224	X	X	X		
	soxC	K17225	X	X	X		
	soxY	K17226	X		X		
	soxZ	K17227	X		X		
Nitrogen Reduction	narG	K00370	X	X	X		
	narH	K00371	X	X	X		
	nirB	K00362		X	X	X	
	nirD	K00363		X		X	
	nirK	K00368			X		
	norB	K04561			X		
	norC	K02305			X		
	nosZ	K00376			X		
MAG Completeness (%) ^b			88.4	88.2	93.7	66.1	44.3
MAG Contamination (%) ^b			16.4	0.6	1.4	11.8	1.6

^a *Marinosulfonomonas* MutY belongs to two separate MAGs and genes for each are reported separately.

^b Completeness and contamination scores generated by *CheckM* v1.0.5 as described in Brazelton *et al.* 2022 (35).

Predicted protein structures and virtual docking experiments

To assess the likelihood of the LCHF MutY sequences folding into enzymes capable of activity on the OG:A substrate, protein structures were predicted using *Colabfold* (44)(**Fig 7**). These predicted structures were associated with high confidence as indicated by pLDDT scores and PAE profiles (**Table 3** and Supporting Information **Fig S5**). Superpositions revealed that the predicted structures for the LCHF MutYs are each highly comparable with the experimentally determined structure for *Gs* MutY as indicated by visual inspection (**Fig 7**) and by low, pairwise root mean square deviation (RMSD) values (**Table 3**). The whole protein superpositions were dominated by the larger, more structurally conserved N-terminal domain. Breaking the analysis into two separate domains showed that the C-terminal domain, although more plastic, retained core structural features that could be superimposed. The MutY-defining chemical motifs are positioned in locations similar to those seen for the *Gs* MutY reference structure, providing evidence these LCHF MutY enzymes are capable of recognizing OG:A lesions and excising the adenine base. Concisely, the LCHF MutY structure predictions resemble a functional MutY enzyme from a thermophilic bacterium.

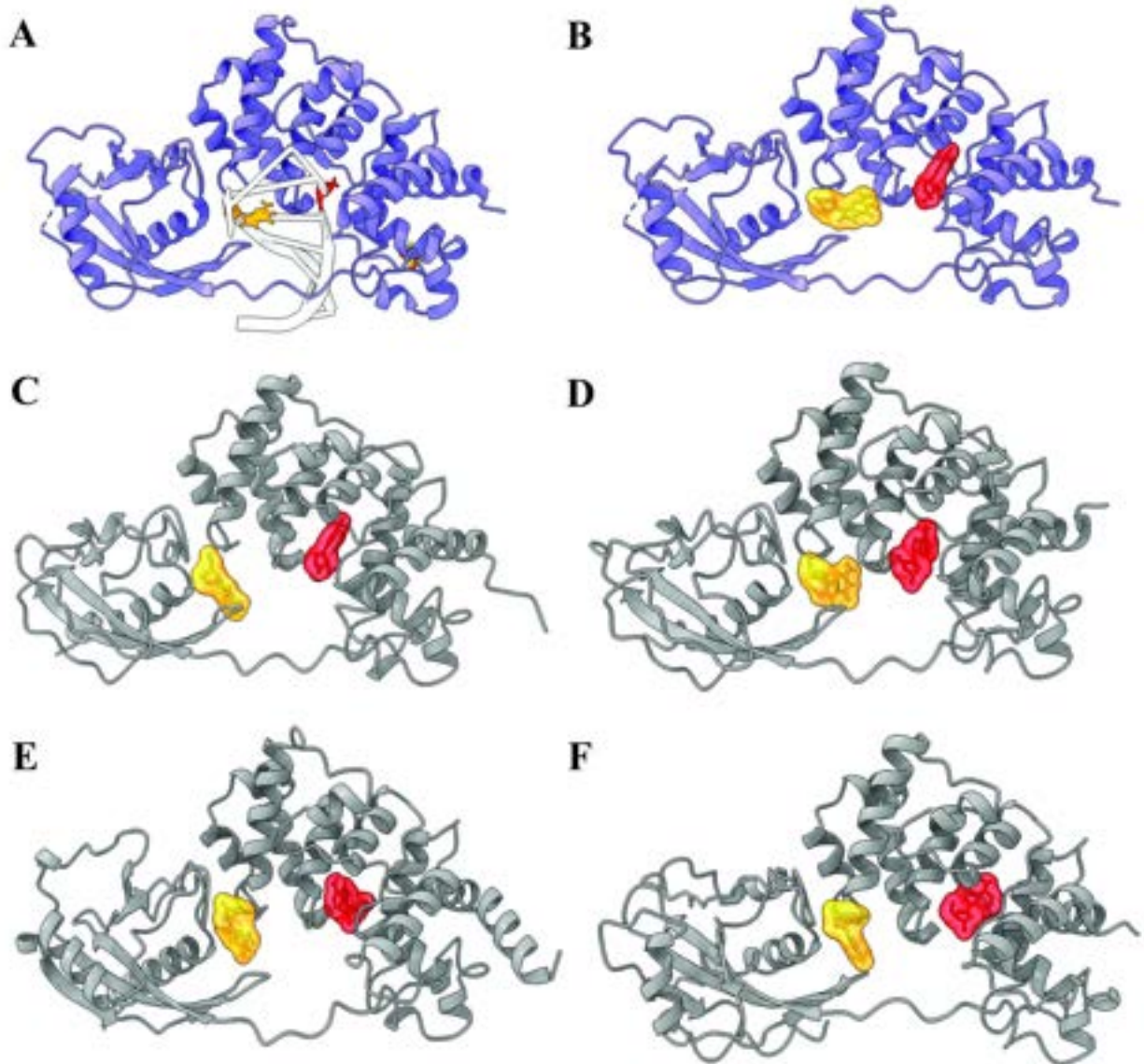


Fig 7. Structure predictions and virtual docking of MutY ligands. (A) The x-ray crystal structure of *Gs* MutY (blue ribbon; PDB ID 3g0q) in complex with DNA (white helix) highlights the positions of the adenosine nucleoside (red) and OG (yellow). (B-F) Virtual docking of ligands. Adenosine and OG were separately docked to identify binding surfaces for these ligands in the structures of *Gs* MutY (B), which served as a positive control, and the four representative LCHF MutYs: *Marinisulfonomonas* MutY (C); *Rhodobacteraceae* MutY (D); *Thiotrichaceae* MutY (E); and *Flavobacteriaceae* MutY (F).

Table 3. Molecular modeling for LCHF MutYs

MutY Model	pLDDT ^d	RMSD (Å) ^a (residues)		Affinity VINA (kcal/ mol) ^b		Energy Amber ^c (kJ/mol)	
		NTD	CTD	A	OG	A	OG
Gs MutY (6u7t) ^e	NA	0.39 (216)	0.54 (116)	-7.3	-7.7	-213 (±38)	-168 (±34)
<i>Marinosulfonomonas</i>	93 (±9)	0.87 (198)	0.90 (64)	-6.8	-7.5	-80 (±37) ^c	-194 (±44)
<i>Rhodobacteraceae</i>	94 (±6)	0.85 (199)	0.98 (51)	-6.8	-7.7	-170 (±37)	-110 (±54) ^c
<i>Thiotrichaceae</i>	92 (±12)	1.0 (195)	1.1 (77)	-7.0	-8.0	-226 (±39)	-205 (±52)
<i>Flavobacteriaceae</i>	92 (±14)	0.84 (199)	1.2 (44)	-7.2	-8.0	-214 (±34)	-197 (±30)

^a Root mean square deviation for the superposition of predicted structures with *Gs* MutY (PDB ID 3g0q) was calculated separately for N-terminal domain (NTD) and C-terminal domain (CTD) by *ChimeraX* (45).

^b Binding affinity for the best outcome from docking adenosine to the enzyme active site and OG to the OG-recognition site as calculated by *Autodock VINA* (46,47).

^c Energy for short-range Coulombic and Lennard-Jones interactions with the ligand as computed by *GROMACS* (48), with the Amber99SB and GAFF force fields (49,50). Energies were averaged over the 100-ns simulation or the time window of the complex, 0 – 26 ns for *Mainosulfonomonas* interaction with A and 0 – 47.5 ns for *Rhodobacteraceae* interaction with OG. Uncertainty is the sample standard deviation.

^d Local distance difference test metric to assess confidence for structures predicted by *Colabfold* (44) averaged over all residues. Uncertainty is the sample standard deviation.

^e Reference superposition values provided by comparing a second structure of *Gs* MutY in complex with its transition state analog (PDB ID 6u7t)

We performed virtual docking experiments to examine the potential for molecular interaction with adenosine and OG ligands. MutY scans DNA looking for the OG:A base pair by sensing the major-groove disposition of the exocyclic amine of the OG base in its *syn* conformation (51,52). After this initial encounter, the enzyme bends the DNA, flips the adenine base from the DNA double helix into the active site pocket, and positions OG in its *anti* conformation as seen in structures of the enzyme complexed to DNA (30,31). Thus, multiple

conformations and orientations for the OG and adenosine ligands were anticipated. The search volume for the adenosine ligand was centered on the active site in the NTD, and the search volume for the OG ligand was defined by the OG-recognition motif found in the CTD. Representative outcomes obtained with *Autodock VINA* are shown in **Fig 7**, and the corresponding binding affinities are reported in **Table 3** and Supporting Information **Table S6**. As anticipated the precise orientation and position for these docked ligands varied, and none exactly match the disposition of the adenine or OG base as presented in the context of double stranded DNA. Nevertheless, binding affinities for the ligand-LCHF MutY complexes ranged from -6.8 to -8.0 kcal/mol, indicating favorable interactions were attainable and similar to the binding affinities measured for *Gs* MutY by the same virtual docking method.

Molecular dynamics simulations

Virtual docking is fast and computationally economical but largely ignores motion and solvent. The reliability of docking improves when complemented with molecular dynamic (MD) simulation (53,54). To further assess stability and dynamic properties of LCHF MutY-ligand complexes derived by docking, we applied MD simulations with the *Amber* force field (49,50), as implemented with *GROMACS* (48). Each protein-ligand complex was solvated in water, charges were balanced with counterions, and the system was equilibrated in preparation for a 100-ns MD simulation (see **Materials and methods**). Supporting Information **Fig S7** and **Movies S8** summarize the resulting trajectories in terms of interaction energy, distance, and structure over time. We focused on mechanistically relevant interactions by tracking distances from the base moiety to the catalytic Glu residue for adenosine complexes, and distances to OG-recognition Ser and Thr residues for OG complexes. MD trajectories for the *Gs* MutY-ligand

complexes (**Fig S7A, Fig S7B, Movie S8A and Movie S8B**) provided a basis of comparison for the LCHF MutY-ligand complexes.

MD analysis revealed dynamic, and in some cases unstable complexes. Relative instability likely reflects the free nature of the ligands, which normally would be presented as part of DNA. As will become evident in later sections, complex instability detected by MD simulation correlates positively with biological activity under mesophile conditions. Even so, many of the MutY-ligand complexes persisted for the entire 100-ns simulation, characterized by favorable binding affinity, extracted as the sum of local Lennard-Jones and Coulombic interactions (**Table 3**). While all ligands were mobile, the MD outcomes separated into two groups distinguished by the degree of ligand movement and persistence of the complex. In the first group the adenosine and OG ligands remained close to the original binding sites for at least 90 ns if not the entire 100-ns MD simulation. This first group with persistently engaged ligands included the complexes with *Gs* MutY (**Fig S7A, Fig S7B, Movie S8A and Movie S8B**), *Thiotrichaceae* MutY (**Fig S7G, Fig S7H, Movie S8G and Movie S8H**) and *Flavobacteriaceae* MutY (**Fig S7I, Fig S7J, Movie S8I and Movie S8J**).

For example, adenosine remains bound to the active site of *Thiotrichaceae* MutY for the entire 100-ns MD simulation. Catalytic Glu46 made contact with N7 of adenosine via a bridging solvent molecule, with this mechanistically relevant interaction observable for the first 11 ns (**Fig S7G and Movie S8G**). Water-mediated interaction of the catalytic Glu and N7 was also observed for *Gs* MutY (**Fig S7A and Movie S8A**), and is comparable to water-bridging interactions described previously in MD simulations of *Gs* MutY complexed to double stranded DNA by others (55). Indeed, such water-mediated interaction was first observed in the crystal structure of

Gs MutY complexed to substrate DNA (30). Thus, our MD analysis captures interactions of functional importance despite lacking a full treatment of DNA.

Similar to observations for adenosine, OG remained bound at the interface of the NTD and CTD in its complex with *Thiotrichaceae* MutY and with *Flavobacteriaceae* MutY, despite notable interdomain hinge motion and flexibility in the CTD. For *Thiotrichaceae* MutY, Ser306 engaged the OG ligand via hydrogen bonds to N1, N2 and O6 of the Watson-Crick-Franklin face during the first 39 ns (**Fig S7H** and **Movie S8H**). Interactions with the Watson-Crick-Franklin face of OG, especially with N2 presented in the major groove, are known to facilitate initial recognition of the OG lesion (51,52). Crystal structures feature the corresponding Ser of *Gs* MutY hydrogen bonded with N7 and O8 of OG (30–32), and similar contacts between Ser305 and N7, O8 and N6 of the Hoogsteen face are observed during the first 13 ns for *Flavobacteriaceae* MutY complexed to OG (**Fig S7J** and **Movie S8J**).

By contrast with these persistently engaged ligands observed in the first group, ligands in the second group disengaged and departed from the original binding site and found new sites within the first 10 ns, as observed for complexes with *Marinosulfonomonas* MutY (**Fig S7C**, **Fig S7D**, **Movie S8C** and **Movie S8D**) and *Rhodobacteraceae* MutY (**Fig S7E**, **Fig S7F**, **Movie S8E** and **Movie S8F**). During the *Marinosulfonomonas* MutY simulation, adenosine slipped out of the active site pocket within 1 ns, remained near the active site entrance until 6.4 ns, when it exited completely and engaged with several different sites on the protein surface (**Fig S7C** and **Movie S8C**). The situation was comparable for adenosine complexed to *Rhodobacteraceae* MutY, but the ligand found a resting place after departing the active site pocket (**Fig S7E** and **Movie S8E**), wedged into a groove with residues Gly126 and Tyr128 on one side and Gln49 and Arg93 on the other side. This alternate adenosine binding site for *Rhodobacteraceae* MutY is

adjacent and partially overlapping with the exosite observed for cytosine in the complex of *Gs* MutY with its OG:C anti-substrate (56). Departure of the base from the active site as observed in our MD simulations was anticipated since crystal structures of MutY in complex with enzyme-generated abasic site (AP) product show no electron density for the base moiety (34), implying that the free base has an escape route.

Binding site departure was also observed for the OG ligand, which disengaged from the CTD of *Marinosulfonomonas* MutY and found new binding sites on the surface of the NTD, as the two domains hinged away from each other (**Fig S7D** and **Movie S8D**). At the outset, OG bound to *Rhodobacteraceae* MutY with OG-specific hydrogen bonds connecting Thr299, N7 and O8 atoms (**Fig S7F** and **Movie S8F**), very comparable to hydrogen bonds seen in the crystal structure of *Gs* MutY bound to DNA with the OG lesion (30–32). However, the FTH loop of *Rhodobacteraceae* MutY pulled away early in the MD simulation at 4.4 ns, thereby breaking these hydrogen bonds. The OG ligand subsequently adopted several novel poses at sites on the NTD and alternatively on the CTD before dissociating completely by 48 ns (**Fig S7F** and **Movie S8F**).

In summary, MD simulations differentiated the LCHF MutYs into two groups based on conformational flexibility and ligand persistence. Ligand persistence was also observed for the complexes with the x-ray crystal structure of *Gs* MutY (PDB ID 6u7t). Kinetically unstable ligand complexes observed for *Marinosulfonomonas* and *Rhodobacteraceae* MutYs prompted further *in vivo* validation to address the open question, which enzyme, if any, could support biological function?

Testing mutation suppression activity of LCHF MutY enzymes by recombinant expression

The *in silico* experiments provided strong evidence that the LCHF MutYs are structurally comparable to authentic MutY enzymes, with affinity for OG:A lesions, albeit with kinetic instability in notable cases, suggesting these may function to prevent mutations. To demonstrate biological function directly, we recombinantly expressed the genes in *E. coli* and measured mutation suppression activity *in vivo*. Three of the representative LCHF MutYs were successfully cloned into the pKK223 expression plasmid as verified by Sanger sequencing. The *Flavobacteriaceae* MutY appeared to be toxic to *E. coli* as only mutant versions of the gene were obtained from multiple cloning attempts, a situation that is reminiscent of *Gs* MutY, which is also apparently toxic to *E. coli* and could not be cloned into pKK223 (33).

To test the mutation suppression activity of *Marinosulfonomonas*, *Rhodobacteraceae*, and *Thiotrichaceae* MutYs, we measured mutation rates with a rifampicin resistance assay (57). Several, independent, single-point mutations within the gene encoding RNA polymerase beta-subunit (*rpoB*) confer antibiotic resistance (58,59). Thus, spontaneous Rif^R mutants arising in overnight cultures can be counted by the colonies that emerge on rifampicin containing plates. Cultures expressing functional MutY delivered by plasmid DNA transformation have low Rif^R frequency compared to the high Rif^R frequency characterizing the reporter strain that lacks *mutY* and *mutM* genes (see **Materials and methods**).

Cultures with an empty plasmid (*null*) and cultures with a plasmid encoding *Ec* MutY showed significant differences in the frequency of Rif^R mutants, with median values of 101 and 12, respectively, indicating the assay was fit for use (significance determined by non-overlap of median 95% confidence intervals) (**Fig 8** and Supporting Information **Table S9**).

Marinosulfonomonas MutY, *Rhodobacteraceae* MutY, and *Thiotrichaceae* MutY each demonstrated significant mutation suppression activity when compared to the *null*. Indeed, *Rhodobacteraceae* MutY showed mutation suppression performance equivalent to that measured for *Ec* MutY, and *Marinosulfonomonas* MutY was apparently better at suppressing mutations than *Ec* MutY, a remarkable outcome given the evolutionary time separating these species (Fig 8). Note that these LCHF MutYs with high mutation suppression function formed unstable complexes as revealed by MD simulation. *Thiotrichaceae* MutY showed partial function. Cultures expressing *Thiotrichaceae* MutY suppressed Rif^R mutants to 50% of the rate observed for *null* cultures, but allowed a mutation rate about 4-fold greater than that measured for cultures expressing *Ec* MutY (Fig 8 and Supporting Information Table S9).

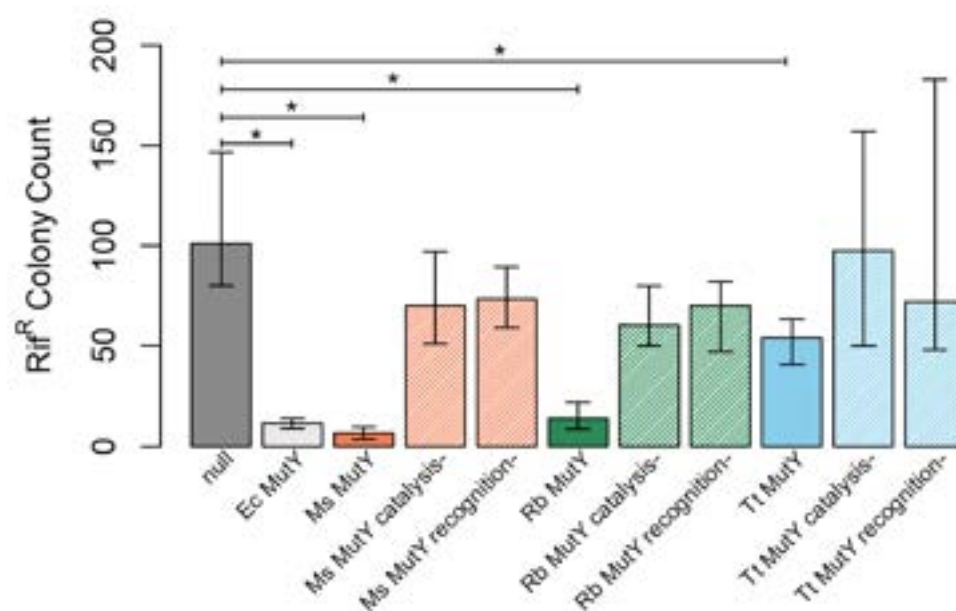


Fig 8. Functional analysis. Bars represent median Rif^R colony counts for *E. coli* cultures expressing MutY, MutY variants, or no MutY (*null*) from a plasmid DNA. Error bars represent 95% confidence intervals as determined by bootstrap sampling (see Supporting Information Table

S9 for tabulated values). *Marinosulfonomonas* (*Ms*) MutY, *Rhodobacteraceae* (*Rb*) MutY, and *Thiotrichaceae* (*Tt*) MutY each suppressed mutations as evidenced by non-overlap of Rif^R confidence intervals compared to *null* cultures. Altered versions of each LCHF MutY tested the importance of residues for catalysis and OG-recognition. Designated *catalysis*- and *recognition*- along the X-axis, these alterations severely impacted mutation suppression function indicating the LCHF MutYs share mechanistic features with the extensively studied enzymes *Ec* MutY and *Gs* MutY.

To investigate the biochemical mechanism employed by LCHF MutY enzymes, we altered residues essential for OG:A recognition and catalysis, then repeated the mutation suppression assay. Two mutants of each LCHF MutY were constructed through site-directed substitution of residues. One set of substitutions was designed to disable the OG-recognition motif by replacing F(S/T)H residues (**Fig 2** and **Fig 5**) with alanine residues (designated *recognition-*); the other set of substitutions was designed to disable catalysis by replacing the active site Asp and Glu residues with structurally similar, but chemically inert Asn and Gln residues (*catalysis-*). For all three LCHF MutYs, these targeted substitutions disabled mutation suppression function *in vivo* as shown by elevated Rif^R frequencies for cultures expressing the *recognition-* and *catalysis-* versions. The mutation frequencies for cultures expressing these site-specific substitution variants were comparable to the Rif^R frequencies measured for *null* cultures as judged by overlapping 95% confidence intervals (**Fig 8** and Supporting Information **Table S9**). These results indicate that the LCHF MutYs suppress mutations by a mechanism that is highly similar to the strategy executed by *Ec* MutY and *Gs* MutY.

Discussion

To gain insight into DNA repair strategies in early Earth-like environments, we investigated the status of the GO DNA repair system within microbes inhabiting the LCHF. Our approach included mining of metagenomic data, bioinformatic comparisons informed by structure and mechanistic understanding, predictive molecular modeling, and functional analyses. The degree to which this approach succeeded was dependent on the assembly of metagenomic sequences into contigs long enough to contain full-length genes (35). Earlier attempts to search for MutY genes within previous LCHF metagenomes with shorter contigs

yielded a number of hits, but these were truncated and therefore missing critical motifs, explaining weak mutation suppression function (unpublished results). The longer contigs utilized in this study allowed us to capture entire MutY genes, bin these MutY-encoding contigs into MAGs to assess associated gene inventories, and thereby infer metabolic strategies for the microbes expressing the GO DNA repair components.

Within the initial set of 649 LCHF MutY candidates identified by sequence identity, 160 genes encoded proteins with all of the chemical motifs known to be important for MutY function. Indeed, leveraging the extensive body of knowledge obtained from crystal structures and mechanistic studies allowed us to select features such as sequence length, presence of MutY motifs, and structural prediction to distinguish LCHF MutYs from other members of the helix-hairpin-helix (HhH) superfamily. Recombinant expression in *E. coli* revealed that LCHF MutY representatives suppress mutations *in vivo* by a mechanism that depends on the catalytic and OG-recognition motifs (**Fig 8**), strongly suggesting these are functional enzymes that actively seek and initiate repair of OG:A lesions within their respective LCHF microbes. Toxicity observed for one LCHF gene encoding *Flavobacteriaceae* MutY, which could not be cloned except with disabling nonsense mutations, underscores the risks and dangers posed by MutY and DNA glycosylases in general, which initiate DNA repair by damaging the DNA further, creating AP sites that are themselves destabilizing (60). The potential for lethal outcomes makes cross-species function observed for *Marinosulfonomonas* MutY and *Rhodobacteraceae* MutY across vast evolutionary time all the more remarkable.

Retained function across evolutionary and species barriers strongly suggests that MutY interacts with the base excision repair apparatus through some well-preserved mechanism that relies on a universal language understood by all organisms. Most critically, the AP sites

generated by MutY should be recognizable to downstream AP nucleases. Protein-protein interactions between AP nucleases and MutY have been discussed as a possible mechanism (61–63), but on its own such a mechanism would rely on coevolution of protein partners. Our results and those reported by others for complementation with the eukaryote homolog MUTYH (64) speaks for a mechanism that is less sensitive to sequence divergence. Therefore, we favor a model where the distorted DNA structure created by MutY signals the location of the AP site for handoff to the BER apparatus, as has been suggested previously (62).

Thiotrichaceae MutY underperformed in our functional evaluation, despite coming from a gammaproteobacterium most closely related to the *E. coli* employed for the bioassay. Lower mutation suppression performance observed for *Thiotrichaceae* MutY may simply be due to differences in conditions for our *in vivo* experiments and more extreme conditions found in the habitat where *Thiotrichaceae* thrives at the LCHF. In support of this adaptation to extreme environments idea, the LCHF enzymes which were predicted to have the highest stability and form the most persistent ligand complexes in MD simulations appeared incompatible with mesophile biology, being either apparently lethal (*Flavobacteriaceae* MutY) or relatively ineffective at suppressing mutations (*Thiotrichaceae* MutY). This pattern of predicted high stability incompatible with mesophile biology extends also to the reference enzyme *Gs* MutY which is from a known thermophile and also appeared lethal in the reporter bacterium, necessitating a chimera approach for evaluation of biological function (33). Adapted for stability at higher temperatures, these enzymes may lack flexibility needed to perform their catalytic duty at lower temperatures, an idea described previously as “corresponding states” of conformational flexibility (65–67). In future work, biochemical characterization of purified LCHF MutY enzymes at high temperatures could address this model directly.

Our metagenomic analysis revealed that gene homologs encoding the GO DNA repair system are abundant in basement microbes inhabiting the LCHF (**Fig 3**). This observation is surprising given that the basement of the LCHF is expected to be anoxic (1,3). The chemical agents commonly thought of for producing oxidized guanine (OG) are ROS derived from molecular oxygen *via* aerobic metabolism. In an anoxic environment, what chemical agents are producing OG and how are these generated? Models of hydrothermal field chemistry predict abiotic production of ROS which the microbial residents may encounter, although these would probably react with cell protective structures before encountering DNA (68). Continual mixing of seawater with the anoxic hydrothermal fluid could provide molecular oxygen at the interface where hydrothermal fluids vent into ambient seawater at the seafloor (1,3). Facultative anaerobes at this interface would inevitably generate ROS (69,70), and therefore benefit from the GO DNA repair system. Intermicrobial competition has driven acquisition of chemical strategies, including ROS, for killing other bacteria (71–73), but there is currently no evidence for such bacterial warfare in basement dwelling microbes of the LCHF.

Another explanation for the source of OG in basement microbes of the LCHF, which is suggested by our gene inventory analysis (**Table 2** and **Supporting Information**), involves reactive sulfur species (RSS) and reactive nitrogen species (RNS). Many of the basement-dwelling microbes within the LCHF appear to metabolize sulfur and nitrogen for energy conservation (4,35), strategies that generate RSS and RNS as metabolic byproducts (74,75). Indeed, mechanisms for the oxidation of guanine by both RSS and RNS have been described, including the formation of 8-oxoguanine (OG) and chemically similar 8-nitroguanine, which templates for adenine in a fashion similar to OG (76–78). The oxidation of guanine by RSS and RNS generated from microbial metabolism would produce OG independent of molecular oxygen

and thereby necessitate the GO DNA repair system for both facultative and obligate anaerobes inhabiting the LCHF.

Whether organisms developed biochemical systems to deal with oxidative damage before or after the Great Oxidation Event (GOE) remains an open question (79,80). It is reasonable to think of these systems arising in response to the selective pressure of oxidative damage from rising O₂ levels. However, it is also possible that these systems were already in place as a coping mechanism for other oxidants and were repurposed to deal with the new source of oxidizing agents when O₂ became readily available. Indeed, obligate anaerobes contain many of the same pathways to deal with oxidative damage as aerobes (79,80). Our discovery of the GO DNA repair system in basement dwellers of the LCHF adds to this body of evidence and supports the hypothesis that oxidative damage repair systems were established before the GOE. We considered the caveat of possible phage-mediated gene transfer – modern microbes adapted to oxygen-rich regions elsewhere may be the source of LCHF MutYs. However, the correspondence of taxonomic assignments based on MutY sequence and based on MAGs and the high degree of sequence diversity seen for LCHF MutY enzymes are inconsistent with expansion of GO DNA repair in the LCHF by horizontal gene transfer. Thus, it seems more likely that LCHF microbes inherited the GO DNA repair system from a common ancestor and retained it through necessity, even in the absence of extrinsic O₂.

Conclusion

Performing empirical studies on how life may have evolved on Earth and other planets is inherently difficult due to time and spatial barriers. Unique sites such as the LCHF serve as representatives of these theorized environments (3,81). By discovering the GO DNA repair

system at the LCHF and validating mutation suppression function by LCHF MutYs, we infer that microbes within the anoxic environment of the LCHF basement are under evolutionary pressure to repair OG lesions. Evolutionary pressure and the source of OG appear to be driven by nitrogen reactive species or sulfur reactive species as supported by the metabolic survey of the MutY-encoding organisms. These results highlight the need for DNA-based life to manage oxidized guanine damage even in anoxic environments. Moreover, this work adds evidence for the more general hypothesis that life established biochemical systems to deal with oxidative damage early, well before the GOE, and should be considered when developing an evolutionary model for early life.

Materials and Methods

Metagenomic sequencing and analysis of LCHF fluid samples

Generation, assembly, and annotation of metagenomes from the Lost City Hydrothermal Field (LCHF) have been described previously (35), and are briefly summarized here. In 2018, the remotely operated vehicle (ROV) Jason collected samples of fluids venting from chimneys at the LCHF, which is located near the Mid-Atlantic Ridge at 30 °N latitude and a depth of ~800 m. Whole-genome community sequences ("metagenomes") were generated from the fluid samples, and assembled metagenomic contigs were binned into metagenome-assembled genomes (MAGs). Potential gene homologs encoding enzymes involved in the GO DNA repair system were identified by conducting KEGG (82,83) orthology assignment using the *BlastKOALA* v2.2 program (84). The selected genes that were identified include: *mutT* (KEGG ID: K03574), *mutM* (K10563), and *mutY* (K03575) along with the genes *xthA* (K01142) and *nfo* (K01151), which

encode exonuclease III and endonuclease IV, respectively.

The relative abundance of each GO repair pathway gene homolog at each LCHF chimney location was calculated as the normalized metagenomic sequence coverage, determined by mapping of reads from each fluid sample against the pooled assembly. Coverages are reported as transcripts (or metagenomic fragments) per million (TPM), which is a proportional unit suitable for comparisons of relative abundances between samples (35,85).

Identification of LCHF MutYs

Candidate MutY genes in the LCHF metagenomes were identified with a *BLASTP* search against predicted protein sequences from the LCHF pooled metagenomic assembly using MutY queries from *Gs* MutY (NCBI Accession ID: P83847.2) and *Ec* MutY (NCBI Accession ID: CDP76921.1). The diversity of these candidates, visualized by aligning the sequences along with *Gs* MutY and *Ec* MutY with *Clustal Omega* (86), and an initial phylogeny was built with *iTOL* (87). LCHF MutY candidate sequences were then aligned by *PROMALS3D* (88), guided by the structure of *Gs* MutY (PDB ID 6u7t). Sequence diversity in the C-terminal domain prevented reliable alignment of this region in this first pass at structure-guided alignment. To overcome this challenge, sequences were split into two parts: one part with all residues before position Val147 in the *Gs* MutY protein which were reliably aligned, and the second part with all residues following position Asn146 which were aligned inconsistently in the first pass. These two parts were separately resubmitted for alignment by *PROMALS3D* guided by the corresponding portions of the crystal structure. For inclusion in this alignment, the C-terminal part was required to pass a minimum length criteria of 160 residues. The resulting alignment was inspected for the MutY-defining chemical motifs described in the text, and a phylogeny was constructed for the

160 authenticated LCHF MutYs with *iTOL* (87). Selection of the four representative LCHF MutYs was guided by this phylogeny and by the completeness of each associated MAG.

Taxonomic classification

Contiguous DNA sequences containing the LCHF MutY representatives were assigned taxonomic classifications using the program *MMseqs2* (89) and the Genome Taxonomy Database (GTDB) as described previously (35). Taxonomic classification of each MAG that included a contig of interest was performed with *GTDB-Tk* v1.5.1 (90). The environmental distributions of MutY-encoding MAGs were inspected for potential signs of contamination from ambient seawater. This possibility was ruled out by the absence of all MutY-encoding taxa reported in this study in the background seawater samples. MAG completeness and contamination scores were generated by *CheckM* v1.0.5 as described previously (35).

Prediction of physical parameters

The theoretical molecular weights and pIs of the LCHF MutY representatives and known MutY sequences were generated with ExPASy (91). The theoretical melting temperature of the representatives was calculated with the Tm predictor from the Institute of Bioinformatics and Structural Biology, National Tsing-Hua University (92).

Molecular modeling

Protein structures for the LCHF MutY representatives were predicted by *Colabfold* with use of *MMseqs2* alignments and relaxed with the Amber force field (44,89,93,94). Predicted structures were superimposed with the crystal structure of *Gs* MutY (PDB ID 3g0q) to generate RMSD (Å) for pruned atom pairs using the MatchMaker tool in *ChimeraX* (45). Initial superpositions were dominated by residues in the N-terminal domain. To fairly compare

structures for the more diverse C-terminal domains, the linker region between domains was identified by inspection, and superposition with *Gs* MutY was repeated with selection of residues in the N-terminal domain and, separately, in the C-terminal domain.

Ligand docking experiments were executed with the program *AutoDock VINA* (46,47). Ligand structures representing adenosine and OG were prepared with the ligand preparation tools implemented with *Autodock Tools* (95,96). Receptor structures were prepared from the structures predicted by *Colabfold* or from the crystal structure of *Gs* MutY (PDB ID 6u7t), each after superposition with PDB ID 3g0q, with the receptor preparation tools as implemented with *Autodock Tools* (95,96). Receptor structures were treated as rigid objects, and ligands included two active torsion angles defined by the C1'-N9 and C4'-C5' bonds. Separate 24 x 24 x 24 Å³ search volumes were defined for adenosine and for OG. The adenosine search volume was centered on the position of atom C1' in the residue A5L:18 in chain C of the *Gs* MutY crystal structure (PDB ID 3g0q), and the OG search volume was centered on the position of atom C1' in residue 8OG:6 in chain B of the same structure.

MD simulations were performed with *GROMACS* version 2022.5 (48), applying the AMBER99SB and GAFF force fields (49,50), with CPU and GPU nodes at the University of Utah's Center for High Performance Computing. We followed steps outlined in the *GROMACS* tutorial "Protein-Ligand Complex" as a guide for our experiments (97). The starting structure for a protein-ligand complex was selected from the binding modes predicted by *Autodock VINA*, choosing the mode with the highest affinity after excluding those that appeared incompatible with the double stranded DNA-enzyme structure. To conserve computational resources, simulation of the complex with adenosine was limited to N-terminal residues as follows: residues 8-220 for *Gs* MutY (PDB ID 6u7t); 6-230 for *Marinosulfonomonas* MutY; 2-220 for

Rhodobacteraceae MutY; 11-223 for *Thiotrichaceae* MutY; and 2-209 for *Flavobacteriaceae* MutY. Simulations with OG omitted the iron-sulfur cluster domain and interdomain linker and thus included limited residues with chain interruptions as follows: residues 29-137, 234-289, 295-360 for *Gs* MutY; 40-142, 239-352 for *Marinosulfonomonas* MutY; 38-139, 233-352 for *Rhodobacteraceae* MutY; 32-140, 238-364 for *Thiotrichaceae* MutY; and 19-127, 234-354 for *Flavobacteriaceae* MutY. Ligand topology files were generated with the *ACPYPE* server (98), applying the general *Amber* force field (49). Each complex was solvated with water molecules with three points of transferable intermolecular potential (TIP3P). Counterions were added to neutralize the net charge of the system. The system was energy minimized by 50000 steepest descent steps and further equilibrated in two phases, NVT followed by NPT, each entailing 100 ps with 2-fs steps. Temperature coupling during NVT and NPT equilibration was accomplished with a modified Berendsen thermostat set to the reference temperature 300 K. Pressure coupling during NPT equilibration was accomplished with the Berendsen algorithm set to the reference pressure 1 bar. The equilibrated system was subjected to a 100-ns MD production run with 2-fs steps, applying a modified Berendsen thermostat (300 K reference temperature) and Parrinello-Rahman barostat (1 bar reference pressure). Short range interaction energies, distances, and structures were extracted from the resulting trajectories with use of *GROMACS* functions and plotted with the *R* package *ggplot2* (99). Figures and movies showing structures were created with *ChimeraX* (45).

Recombinant DNA cloning

Synthetic genes encoding the LCHF MutYs were codon optimized for expression in *E. coli* except that pause sites with rare codons were engineered so as to retain pause sites found in the gene encoding *Ec* MutY. GBlocks gene fragments were ordered from Integrated DNA

Technologies. LCHF MutY genes designed in this way were cloned into the low-expression pKK223 vector by ligation-independent cloning (LIC). PCR reactions with the high-fidelity Phusion polymerase (Agilent) amplified the synthetic gBlock and two overlapping fragments derived from approximate halves of the pKK223 plasmid. PCR products were separated by electrophoresis in 0.8% agarose x1 TAE gels containing 1 µg/mL ethidium bromide. DNA was visualized by long-wavelength UV shadowing to allow dissection of gel bands, and the DNA was purified with the GeneJet gel extraction system (Thermo Scientific), treated with Dpn1 (New England Biolabs) at 37 °C for 45 min, and heat shock transformed directly into DH5α competent cells. Clones were selected on ampicillin media plates. The plasmid DNA was purified from 4-mL overnight cultures by use of the Wizard Plus MiniPrep kit (Promega) according to the manufacturer's instructions. The sequence of the LCHF MutY encoding gene was verified by Sanger sequencing with UpTac and TacTerm primers. Genes encoding site-directed substitution variants were created by amplifying two overlapping fragments of the LCHF MutY-pKK223 plasmid with mutagenic PCR primers followed by similar gel purification and transformation procedures. In our hands the LIC cloning efficiency was close to 95% except for the *Flavobacteriaceae* MutY gene which could not be cloned intact. The pkk223 plasmids containing the LCHF MutY encoding genes can be found on Addgene with ID numbers 210791-210799.

Mutation suppression assay

Mutation rates were measured by the method outlined previously (33,57). The CC104 *mutm::KAN muty::TET* cells (100) were heat-shock transformed with a pKK223 plasmid encoding the *Ec* MutY gene, LCHF MutY genes, or no gene (*null*). Transformants selected from kanamycin-ampicillin-tetracycline (KAT) media plates were diluted prior to inoculation of 2-mL

803 KAT liquid media, and these cultures were grown overnight for 18 hours at 37°C with shaking at
 804 180 rpm. Cultures were kept cold on ice or at +4 °C prior to further processing. Cells were
 805 collected by centrifugation, the media was removed by aspiration, and cells were resuspended in
 806 an equal volume of 0.85% sodium chloride before seeding 100 µL aliquots to kanamycin-
 807 ampicillin-rifampicin (KAR) media plates. Dilutions of the washed cells were also seeded to
 808 KAR plates (10^{-1} dilution) and kanamycin-ampicillin (KA) plates (10^{-7} dilution), and allowed to
 809 overnight at 37°C for 18 hours. The number of Rif^R mutants were counted by counting the
 810 colony forming units (CFU). Statistical analysis was performed in *R* as previously described
 811 (33). Confidence intervals were obtained by bootstrap resampling of 10,000 trials as
 812 implemented in *R* with the *boot* package (101,102).

Acknowledgements

This work was supported by NSF awards to MPH (CHE:CLP- 1905249, 2204229) and to WJB (OCE-1536405), by the NASA Astrobiology Institute Rock-Powered Life team, and by UROP funding from the Office of Undergraduate Research at the University of Utah awarded to PHU. The support and resources from the Center for High Performance Computing at the University of Utah are gratefully acknowledged. There was no additional external funding received for this study. We thank Markel Kolendrianos, Peyton Russelberg, and Sonia Sehgal for technical expertise. We thank the University of Utah undergraduate students enrolled in Molecular Biology of DNA Lab (BIOL3525 fall 2022) for their contribution to measuring Rif^R mutant frequency data. In particular, we would like to thank the following students for their extra efforts and contributions in the lab: Ticker Duncan, Shilpi Kharidia, Jackson Munn, Kenzie Fleming, Alex Ballinger, Andrew Petersen, Brook Miller, Tom Christensen, Madison Haught, Emi Wickens, Spencer Sonntag, Abigail Johnston, Sam Aamodt, Jasmine Jacobo, Alyssa Le, Sam Hendry, Saydra Galloway, Hiroshi Aoki, Peyton Merchant, Kaliece Carter, Annie Joseph, Kathleen Brabb, Natalie Morgan, Sophia Khalaji, Helena Haddadin, Hadlee Young, Brenden Roberts, Mason Hansen, Mackenzie Montzingo, and Quyen Tran. We especially acknowledge Emily Dart who first searched LCHF metagenomes for MutY homologs which provided the impetus for development of this project.

References

1. Kelley DS, Karson JA, Früh-Green GL, Yoerger DR, Shank TM, Butterfield DA, et al. A Serpentinite-Hosted Ecosystem: The Lost City Hydrothermal Field. *Science*. 2005 Mar 4;307(5714):1428–34.
2. Amador ES, Bandfield JL, Brazelton WJ, Kelley D. The Lost City Hydrothermal Field: A Spectroscopic and Astrobiological Analogue for Nili Fossae, Mars. *Astrobiology*. 2017

- Nov;17(11):1138–60.
3. Kelley DS, Karson JA, Blackman DK, Früh-Green GL, Butterfield DA, Lilley MD, et al. An off-axis hydrothermal vent field near the Mid-Atlantic Ridge at 30° N. *Nature*. 2001 Jul;412(6843):145–9.
4. Brazelton WJ, Schrenk MO, Kelley DS, Baross JA. Methane- and Sulfur-Metabolizing Microbial Communities Dominate the Lost City Hydrothermal Field Ecosystem. *Appl Environ Microbiol*. 2006 Sep;72(9):6257–70.
5. Russell MJ, Hall AJ, Martin W. Serpentinization as a source of energy at the origin of life. *Geobiology*. 2010 Dec;8(5):355–71.
6. Proskurowski G, Lilley MD, Seewald JS, Früh-Green GL, Olson EJ, Lupton JE, et al. Abiogenic hydrocarbon production at lost city hydrothermal field. *Science*. 2008 Feb 1;319(5863):604–7.
7. Lang SQ, Butterfield DA, Lilley MD, Paul Johnson H, Hedges JI. Dissolved organic carbon in ridge-axis and ridge-flank hydrothermal systems. *Geochim Cosmochim Acta* [Internet]. 2006 Aug [cited 2023 Nov 28];70(15). Available from: <https://par.nsf.gov/biblio/10088640-dissolved-organic-carbon-ridge-axis-ridge-flank-hydrothermal-systems>
8. Lang SQ, Brazelton WJ. Habitability of the marine serpentinite subsurface: a case study of the Lost City hydrothermal field. *Philos Trans R Soc Math Phys Eng Sci*. 2020 Feb 21;378(2165):20180429.
9. Bruskov VI, Malakhova LV, Masalimov ZK, Chernikov AV. Heat-induced formation of reactive oxygen species and 8-oxoguanine, a biomarker of damage to DNA. *Nucleic Acids Res*. 2002 Mar 15;30(6):1354–63.
10. Kino K, Hirao-Suzuki M, Morikawa M, Sakaga A, Miyazawa H. Generation, repair and replication of guanine oxidation products. *Genes Environ*. 2017 Aug 1;39:21.
11. Nghiem Y, Cabrera M, Cupples CG, Miller JH. The mutY gene: a mutator locus in *Escherichia coli* that generates G.C----T.A transversions. *Proc Natl Acad Sci*. 1988 Apr;85(8):2709–13.
12. Cabrera M, Nghiem Y, Miller JH. mutM, a second mutator locus in *Escherichia coli* that generates G.C----T.A transversions. *J Bacteriol*. 1988 Nov;170(11):5405–7.
13. Maki H, Sekiguchi M. MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis. *Nature*. 1992 Jan;355(6357):273–5.
14. Michaels ML, Cruz C, Grollman AP, Miller JH. Evidence that MutY and MutM combine to prevent mutations by an oxidatively damaged form of guanine in DNA. *Proc Natl Acad Sci U S A*. 1992 Aug 1;89(15):7022–5.
15. Prakash A, Doublé S, Wallace SS. Chapter 4 - The Fpg/Nei Family of DNA Glycosylases: Substrates, Structures, and Search for Damage. In: Doetsch PW, editor. *Progress in Molecular Biology and Translational Science* [Internet]. Academic Press; 2012 [cited 2023 Mar 20]. p. 71–91. (Mechanisms of DNA Repair; vol. 110). Available from: <https://www.sciencedirect.com/science/article/pii/B9780123876652000043>
16. Denver DR, Swenson SL, Lynch M. An evolutionary analysis of the helix-hairpin-helix superfamily of DNA repair glycosylases. *Mol Biol Evol*. 2003 Oct;20(10):1603–11.
17. Trasviña-Arenas CH, Demir M, Lin WJ, David SS. Structure, function and evolution of the Helix-hairpin-Helix DNA glycosylase superfamily: Piecing together the evolutionary puzzle of DNA base damage repair mechanisms. *DNA Repair*. 2021 Dec 1;108:103231.
18. Kuwahara H, Takaki Y, Shimamura S, Yoshida T, Maeda T, Kunieda T, et al. Loss of

- genes for DNA recombination and repair in the reductive genome evolution of
thioautotrophic symbionts of Calyptogenia clams. BMC Evol Biol. 2011 Oct 3;11(1):285.
19. Garcia-Gonzalez A, Rivera-Rivera R, Massey S. The Presence of the DNA Repair Genes
mutM, mutY, mutL, and mutS is Related to Proteome Size in Bacterial Genomes. Front
Genet. 2012;3:3.
20. Jansson K, Blomberg A, Sunnerhagen P, Alm Rosenblad M. Evolutionary loss of 8-oxo-G
repair components among eukaryotes. Genome Integr. 2010 Sep 1;1(1):12.
21. Fowler RG, Schaaper RM. The role of the mutT gene of Escherichia coli in maintaining
replication fidelity. FEMS Microbiol Rev. 1997 Aug;21(1):43–54.
22. Boiteux S, O'Connor TR, Laval J. Formamidopyrimidine-DNA glycosylase of Escherichia
coli: cloning and sequencing of the fpg structural gene and overproduction of the protein.
EMBO J. 1987 Oct;6(10):3177–83.
23. Shida T, Noda M, Sekiguchi J. Cleavage of single- and double-stranded DNAs containing
an abasic residue by Escherichia coli exonuclease III (AP endonuclease VI). Nucleic Acids
Res. 1996 Nov 15;24(22):4572–6.
24. Parikh SS, Mol CD, Tainer JA. Base excision repair enzyme family portrait: integrating the
structure and chemistry of an entire DNA repair pathway. Struct Lond Engl 1993. 1997
Dec 15;5(12):1543–50.
25. Ljungquist S, Lindahl T, Howard-Flanders P. Methyl methane sulfonate-sensitive mutant
of Escherichia coli deficient in an endonuclease specific for apurinic sites in
deoxyribonucleic acid. J Bacteriol. 1976 May;126(2):646–53.
26. Lind PA, Andersson DI. Whole-genome mutational biases in bacteria. Proc Natl Acad Sci.
2008 Nov 18;105(46):17878–83.
27. Foster PL, Lee H, Popodi E, Townes JP, Tang H. Determinants of spontaneous mutation in
the bacterium Escherichia coli as revealed by whole-genome sequencing. Proc Natl Acad
Sci [Internet]. 2015 Nov 3 [cited 2022 Aug 19];112(44). Available from:
<https://pnas.org/doi/full/10.1073/pnas.1512136112>
28. Al-Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, Williams GT, et al.
Inherited variants of MYH associated with somatic G:C→T:A mutations in colorectal
tumors. Nat Genet. 2002;30(2):227–32.
29. Noll DM, Gogos A, Granek JA, Clarke ND. The C-Terminal Domain of the Adenine-DNA
Glycosylase MutY Confers Specificity for 8-Oxoguanine·Adenine Mispairs and May Have
Evolved from MutT, an 8-Oxo-dGTPase. Biochemistry. 1999 May 1;38(20):6374–9.
30. Fromme JC, Banerjee A, Huang SJ, Verdine GL. Structural basis for removal of adenine
mispairs with 8-oxoguanine by MutY adenine DNA glycosylase. Nature. 2004
Feb;427(6975):652–6.
31. Lee S, Verdine GL. Atomic substitution reveals the structural basis for substrate adenine
recognition and removal by adenine DNA glycosylase. Proc Natl Acad Sci. 2009 Nov
3;106(44):18497–502.
32. Woods RD, O'Shea VL, Chu A, Cao S, Richards JL, Horvath MP, et al. Structure and
stereochemistry of the base excision repair glycosylase MutY reveal a mechanism similar
to retaining glycosidases. Nucleic Acids Res. 2016 Jan 29;44(2):801–10.
33. Russelburg LP, O'Shea Murray VL, Demir M, Knutsen KR, Sehgal SL, Cao S, et al.
Structural Basis for Finding OG Lesions and Avoiding Undamaged G by the DNA
Glycosylase MutY. ACS Chem Biol. 2020 Jan 17;15(1):93–102.
34. Demir M, Russelburg LP, Lin WJ, Trasviña-Arenas CH, Huang B, Yuen PK, et al.

- 929 Structural snapshots of base excision by the cancer-associated variant MutY N146S reveal
930 a retaining mechanism. *Nucleic Acids Res.* 2023 Feb 22;51(3):1034–49.
- 931 35. Brazelton WJ, McGonigle JM, Motamedi S, Pendleton HL, Twing KI, Miller BC, et al.
932 Metabolic strategies shared by basement residents of the Lost City hydrothermal field
933 [Internet]. *bioRxiv*; 2022 [cited 2022 Apr 2]. p. 2022.01.25.477282. Available from:
934 <https://www.biorxiv.org/content/10.1101/2022.01.25.477282v1>
- 935 36. Porello SL, Leyes AE, David SS. Single-turnover and pre-steady-state kinetics of the
936 reaction of the adenine glycosylase MutY with mismatch-containing DNA substrates.
937 *Biochemistry.* 1998 Oct 20;37(42):14756–64.
- 938 37. Trasviña-Arenas CH, Lopez-Castillo LM, Sanchez-Sandoval E, Briebe LG. Dispensability
939 of the [4Fe-4S] cluster in novel homologues of adenine glycosylase MutY. *FEBS J.* 2016
940 Feb;283(3):521–40.
- 941 38. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences.
942 *Nucleic Acids Res.* 1990 Oct 25;18(20):6097–100.
- 943 39. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator.
944 *Genome Res.* 2004 Jun;14(6):1188–90.
- 945 40. Gifford CM, Wallace SS. The genes encoding formamidopyrimidine and MutY DNA
946 glycosylases in *Escherichia coli* are transcribed as part of complex operons. *J Bacteriol.*
947 1999 Jul;181(14):4223–36.
- 948 41. Pomposiello PJ, Koutsolioutsou A, Carrasco D, Demple B. SoxRS-regulated expression
949 and genetic analysis of the *yggX* gene of *Escherichia coli*. *J Bacteriol.* 2003
950 Nov;185(22):6624–32.
- 951 42. Gralnick J, Downs D. Protection from superoxide damage associated with an increased
952 level of the YggX protein in *Salmonella enterica*. *Proc Natl Acad Sci U S A.* 2001 Jul
953 3;98(14):8030–5.
- 954 43. Gralnick JA, Downs DM. The YggX Protein of *Salmonella enterica* Is Involved in Fe(II)
955 Trafficking and Minimizes the DNA Damage Caused by Hydroxyl Radicals: RESIDUE
956 CYS-7 IS ESSENTIAL FOR YggX FUNCTION*. *J Biol Chem.* 2003 Jun
957 6;278(23):20708–15.
- 958 44. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold:
959 making protein folding accessible to all. *Nat Methods.* 2022 Jun;19(6):679–82.
- 960 45. Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, et al. UCSF
961 ChimeraX: Tools for structure building and analysis. *Protein Sci Publ Protein Soc.* 2023
962 Nov;32(11):e4792.
- 963 46. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a
964 new scoring function, efficient optimization and multithreading. *J Comput Chem.* 2010 Jan
965 30;31(2):455–61.
- 966 47. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: New Docking
967 Methods, Expanded Force Field, and Python Bindings. *J Chem Inf Model.* 2021 Aug
968 23;61(8):3891–8.
- 969 48. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High
970 performance molecular simulations through multi-level parallelism from laptops to
971 supercomputers. *SoftwareX.* 2015 Sep 1;1–2:19–25.
- 972 49. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a
973 general amber force field. *J Comput Chem.* 2004 Jul 15;25(9):1157–74.
- 974 50. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of

- multiple Amber force fields and development of improved protein backbone parameters. *Proteins*. 2006 Nov 15;65(3):712–25.
51. Manlove AH, McKibbin PL, Doyle EL, Majumdar C, Hamm ML, David SS. Structure–Activity Relationships Reveal Key Features of 8-Oxoguanine: A Mismatch Detection by the MutY Glycosylase. *ACS Chem Biol*. 2017 Sep 15;12(9):2335–44.
52. Lee AJ, Majumdar C, Kathe SD, Van Ostrand RP, Vickery HR, Averill AM, et al. Detection of OG:A Lesion Mispairs by MutY Relies on a Single His Residue and the 2-Amino Group of 8-Oxoguanine. *J Am Chem Soc*. 2020 Aug 5;142(31):13283–7.
53. Liu X, Shi D, Zhou S, Liu H, Liu H, Yao X. Molecular dynamics simulations and novel drug discovery. *Expert Opin Drug Discov*. 2018 Jan;13(1):23–37.
54. Guterres H, Im W. Improving Protein-Ligand Docking Results with High-Throughput Molecular Dynamics Simulations. *J Chem Inf Model*. 2020 Apr 27;60(4):2189–98.
55. Nikkel DJ, Wetmore SD. Distinctive Formation of a DNA-Protein Cross-Link during the Repair of DNA Oxidative Damage: Insights into Human Disease from MD Simulations and QM/MM Calculations. *J Am Chem Soc*. 2023 Jun 21;145(24):13114–25.
56. Wang L, Lee SJ, Verdine GL. Structural Basis for Avoidance of Promutagenic DNA Repair by MutY Adenine DNA Glycosylase. *J Biol Chem*. 2015 Jul 10;290(28):17096–105.
57. Majumdar C, Nuñez NN, Raetz AG, Khuu C, David SS. Chapter Three - Cellular Assays for Studying the Fe–S Cluster Containing Base Excision Repair Glycosylase MUTYH and Homologs. In: David SS, editor. *Methods in Enzymology* [Internet]. Academic Press; 2018 [cited 2021 Jun 22]. p. 69–99. (Fe-S Cluster Enzymes Part B; vol. 599). Available from: <https://www.sciencedirect.com/science/article/pii/S0076687917303816>
58. Garibyan L, Huang T, Kim M, Wolff E, Nguyen A, Nguyen T, et al. Use of the rpoB gene to determine the specificity of base substitution mutations on the Escherichia coli chromosome. *DNA Repair*. 2003 May 13;2(5):593–608.
59. Wu EY, Hilliker AK. Identification of Rifampicin Resistance Mutations in *Escherichia coli*, Including an Unusual Deletion Mutation. *J Mol Microbiol Biotechnol*. 2017;27(6):356–62.
60. Loeb LA, Preston BD. Mutagenesis by apurinic/apyrimidinic sites. *Annu Rev Genet*. 1986;20:201–30.
61. Michaels ML, Tchou J, Grollman AP, Miller JH. A repair system for 8-oxo-7,8-dihydrodeoxyguanine. *Biochemistry*. 1992 Nov 17;31(45):10964–8.
62. Pope MA, Porello SL, David SS. Escherichia coli apurinic-apyrimidinic endonucleases enhance the turnover of the adenine glycosylase MutY with G:A substrates. *J Biol Chem*. 2002 Jun 21;277(25):22605–15.
63. Lu AL, Lee CY, Li L, Li X. Physical and functional interactions between Escherichia coli MutY and endonuclease VIII. *Biochem J*. 2006 Jan 1;393(Pt 1):381–7.
64. Komine K, Shimodaira H, Takao M, Soeda H, Zhang X, Takahashi M, et al. Functional Complementation Assay for 47 MUTYH Variants in a MutY-Disrupted Escherichia Coli Strain. *Hum Mutat*. 2015 Jul;36(7):704–11.
65. Jaenicke R. Protein stability and molecular adaptation to extreme conditions. *Eur J Biochem*. 1991 Dec 18;202(3):715–28.
66. Závodszky P, Kardos J, Svingor Á, Petsko GA. Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proc Natl Acad Sci*. 1998 Jun 23;95(13):7406–11.

67. Dong Y wei, Liao M ling, Meng X liang, Somero GN. Structural flexibility and protein adaptation to temperature: Molecular dynamics analysis of malate dehydrogenases of marine molluscs. *Proc Natl Acad Sci*. 2018 Feb 6;115(6):1274–9.
68. Shaw TJ, Luther GW, Rosas R, Oldham VE, Coffey NR, Ferry JL, et al. Fe-catalyzed sulfide oxidation in hydrothermal plumes is a source of reactive oxygen species to the ocean. *Proc Natl Acad Sci U S A*. 2021 Oct 5;118(40):e2026654118.
69. Messner KR, Imlay JA. The identification of primary sites of superoxide and hydrogen peroxide formation in the aerobic respiratory chain and sulfite reductase complex of *Escherichia coli*. *J Biol Chem*. 1999 Apr 9;274(15):10119–28.
70. Imlay JA. The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. *Nat Rev Microbiol*. 2013 Jul;11(7):443.
71. Price RJ, Lee JS. INHIBITION OF PSEUDOMONAS SPECIES BY HYDROGEN PEROXIDE PRODUCING LACTOBACILLI. *J Food Prot*. 1970 Jan 1;33(1):13–8.
72. Juven BJ, Pierson MD. Antibacterial Effects of Hydrogen Peroxide and Methods for Its Detection and Quantitation†. *J Food Prot*. 1996 Nov 1;59(11):1233–41.
73. Bucci V, Nadell CD, Xavier JB. The Evolution of Bacteriocin Production in Bacterial Biofilms. *Am Nat*. 2011 Dec;178(6):E162–73.
74. Rodionov DA, Dubchak IL, Arkin AP, Alm EJ, Gelfand MS. Dissimilatory Metabolism of Nitrogen Oxides in Bacteria: Comparative Reconstruction of Transcriptional Networks. *PLOS Comput Biol*. 2005 Oct 28;1(5):e55.
75. Han S, Li Y, Gao H. Generation and Physiology of Hydrogen Sulfide and Reactive Sulfur Species in Bacteria. *Antioxid Basel Switz*. 2022 Dec 17;11(12):2487.
76. Suzuki N, Yasui M, Geacintov NE, Shafirovich V, Shibutani S. Miscoding events during DNA synthesis past the nitration-damaged base 8-nitroguanine. *Biochemistry*. 2005 Jun 28;44(25):9238–45.
77. Joyner-Matos J, Predmore BL, Stein JR, Leeuwenburgh C, Julian D. Hydrogen Sulfide Induces Oxidative Damage to RNA and DNA in a Sulfide-Tolerant Marine Invertebrate. *Physiol Biochem Zool PBZ*. 2010;83(2):356–65.
78. Bhamra I, Compagnone-Post P, O’Neil IA, Iwanejko LA, Bates AD, Cosstick R. Base-pairing preferences, physicochemical properties and mutational behaviour of the DNA lesion 8-nitroguanine. *Nucleic Acids Res*. 2012 Nov;40(21):11126–38.
79. Ślesak I, Ślesak H, Kruk J. Oxygen and Hydrogen Peroxide in the Early Evolution of Life on Earth: In silico Comparative Analysis of Biochemical Pathways. *Astrobiology*. 2012 Aug;12(8):775–84.
80. Lu Z, Imlay JA. When anaerobes encounter oxygen: mechanisms of oxygen toxicity, tolerance and defence. *Nat Rev Microbiol*. 2021 Dec;19(12):774–85.
81. Shock EL, Schulte MD. Organic synthesis during fluid mixing in hydrothermal systems. *J Geophys Res Planets*. 1998;103(E12):28513–27.
82. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000 Jan 1;28(1):27–30.
83. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res*. 2023 Jan 6;51(D1):D587–92.
84. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol*. 2016 Feb 22;428(4):726–31.

85. Thornton CN, Tanner WD, VanDerslice JA, Brazelton WJ. Localized effect of treated wastewater effluent on the resistome of an urban watershed. *GigaScience*. 2020 Nov 19;9(11):giaa125.
86. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011 Jan;7(1):539.
87. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. 2021 Jul 2;49(W1):W293–6.
88. Pei J, Kim BH, Grishin NV. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res*. 2008 Apr;36(7):2295–300.
89. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol*. 2017 Nov;35(11):1026–8.
90. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinforma Oxf Engl*. 2019 Nov 15;36(6):1925–7.
91. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*. 2003 Jul 1;31(13):3784–8.
92. Ku T, Lu P, Chan C, Wang T, Lai S, Lyu P, et al. Predicting melting temperature directly from protein sequences. *Comput Biol Chem*. 2009 Dec;33(6):445–50.
93. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021 Aug;596(7873):583–9.
94. Eastman P, Swails J, Chodera JD, McGibbon RT, Zhao Y, Beauchamp KA, et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol*. 2017 Jul;13(7):e1005659.
95. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem*. 2009 Dec;30(16):2785–91.
96. Forli S, Huey R, Pique ME, Sanner M, Goodsell DS, Olson AJ. Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nat Protoc*. 2016 May;11(5):905–19.
97. Lemkul JA. From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1.0]. *Living J Comput Mol Sci*. 2019;1(1):5068–5068.
98. Kagami L, Wilter A, Diaz A, Vranken W. The ACPYPE web server for small-molecule MD topology generation. *Bioinformatics*. 2023 Jun 1;39(6):btad350.
99. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. New York, NY: Springer International Publishing; 2016 [cited 2023 Nov 27]. (Use R!). Available from: <http://link.springer.com/10.1007/978-3-319-24277-4>
100. Cupples CG, Miller JH. A set of lacZ mutations in Escherichia coli that allow rapid detection of each of the six base substitutions. *Proc Natl Acad Sci U S A*. 1989 Jul;86(14):5345–9.
101. Canty A, support) BR (author of parallel. boot: Bootstrap Functions (Originally by Angelo Canty for S) [Internet]. 2022 [cited 2023 Nov 27]. Available from: <https://cran.r-project.org/web/packages/boot/>
102. Davison AC, Hinkley DV. Bootstrap Methods and their Application [Internet]. Cambridge:

- 1113 Cambridge University Press; 1997 [cited 2023 Nov 27]. (Cambridge Series in Statistical
1114 and Probabilistic Mathematics). Available from:
1115 [https://www.cambridge.org/core/books/bootstrap-methods-and-their-](https://www.cambridge.org/core/books/bootstrap-methods-and-their-application/ED2FD043579F27952363566DC09CBD6A)
1116 [application/ED2FD043579F27952363566DC09CBD6A](https://www.cambridge.org/core/books/bootstrap-methods-and-their-application/ED2FD043579F27952363566DC09CBD6A)
1117 103. Osborne MJ, Siddiqui N, Landgraf D, Pomposiello PJ, Gehring K. The solution structure of
1118 the oxidative stress-related protein YggX from Escherichia coli. Protein Sci Publ Protein
1119 Soc. 2005 Jun;14(6):1673–8.
1120
1121

45

c_000001556689_3 18 RDPYAILVSIIMLQTRVETVIN-RYGSFLERFPTIADLAASLDDVLAWSGLGYRRPRLLHRLATVVVDTHGGEIPDTLEELLELPGIGP TSAAVGSIAGFIPALSI 130
c_000001834452_1 1 -----WISIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000004255004_2 35 RDPYAILVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000000605438_3 31 QSAYEIYLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000000581237_15 31 QSAYEIYLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000000525979_2 31 QNSYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000000609738_1 1 RDPYAILVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000000583727_6 27 RDPYAILVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000000580740_2 27 RDPYAILVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_00000013678_1 27 RDPYAILVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000000136725_3 27 RDPYAILVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000001684786_4 24 PEYKRVWVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000002917826_23 31 PEYKRVWVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000002040695_2 31 PEYKRVWVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000003263657_32 41 NDYKRVWVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000005989041_2 36 KDPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000002391082_2 36 KDPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_00000339553_2 43 SDPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_00000396707_2 31 GNPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000002047076_1 16 DDPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000006007075_1 31 KREYTVLVSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000002094036_1 7 KREYTVLVSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000002762689_1 31 KREYTVLVSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000005516980_2 31 KREYTVLVSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000003920004_1 31 PNPYTVLVSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000004008511_2 31 PNPYTVLVSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000004481347_1 31 PNPYTVLVSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000003787733_1 31 PNPYTVLVSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000000754657_2 31 HDLYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000000990943_3 31 NDLYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000004474996_2 31 NDLYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000001463500_11 27 NDLYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000001286181_5 24 NDYKRVWVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000001293628_3 31 REYKRVWVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000001535696_2 31 RDPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000001614067_2 31 SNPYTVLVSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000001765289_1 31 DDPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000001961666_1 18 VDAYKRVWVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000002018097_4 25 NEPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000002843512_36 22 DDYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000002930199_1 22 NNPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000003033975_2 22 RPYKRVWVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000003888107_2 31 HDPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000004013286_6 44 RDPYAILVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000004852258_1 6 SDPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000005254087_1 26 RDPYAILVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000005063677_1 29 QDPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000003283462_3 17 TGPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000002687221_1 31 SNPYTVLVSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000003347358_19 24 DEAYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000002701302_1 24 SDYKRVWVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000000582753_3 32 KSYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000001713769_5 24 NDYKRVWVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000004369364_1 31 RDPYAILVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000006057486_30 28 KDPIKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000005494072_10 24 KDPIKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000001742634_3 31 KDPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000002826998_2 31 NNPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000003195439_1 24 SDPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000004887214_1 21 RPYKRVWVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000005037037_2 29 KNPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000001059964_1 19 YEPYQWVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000002498472_1 17 YSPYQWVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000004615912_3 17 KNPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000004766858_2 31 YHPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000002747260_18 36 DDPYAILVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000005371561_38 36 DDPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_00000169465_2 32 VTPYKRVWVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000003254110_11 32 VTPYKRVWVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000004750284_20 32 VTPYKRVWVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000000214782_15 31 ATPYKRVWVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_00000217847_8 31 ITPYKRVWVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000004054979_1 18 ISPYKRVWVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000001626673_3 41 PNPYTVLVSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000000339186_3 1 -DPSYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000003029168_2 32 RDPYAILVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000001863436_1 48 GDPIYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000002830137_4 41 RDLHYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000004612302_3 35 RDPYAILVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000001933926_1 21 TDYKRVWVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000001365634_3 31 MDYKRVWVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000005849454_9 30 RDPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000001169194_1 41 RDPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000002786947_2 21 RDPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000001279808_52 28 RSPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000000358065_2 31 QKPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000003745941_1 22 AALPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000004187032_4 31 KHYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000006067315_3 31 QTEYTVLVSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000005543774_1 1 KKYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000006211484_1 23 KKYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_00000218976_3 32 KKYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000002992548_1 31 KKYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000004456751_1 1 -----YVLVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000002655634_1 21 KREYTVLVSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_000004232403_2 21 NDYKRVWVSIIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
c_0000007147_1 28 RDPYKIVLSIMLQTRVETVIN-YFEAFIKRFPFISIEVLAASQEEVLSVWAGLGYRRSRLLKKAARVQSEMGGVLPDSATLRLPGIGP TAAAIASISFQEPVIVG 107
Consensus_aa: .s.eY.lE1SEIMLQTPpV.hhhs.YE.p.hpppPaSpLa.tbppllp.hpWGLGYsRt+nhhptp.l.p.h.t.hppshp.l.LPGIG.TAAtLhtHtEsp..shIDEN

Conservation: 95 9 75 87 85 9 5 9 8 9 5 225
sp_p17802_MUTY_ECOL 141 VKRVLARCYAVSGWMPG-----KKEVENKWLSL-----VQE-VTP-----AVGVERPNAIMDLGAMICTRSPKCS-----LCPQNGCAIAAANSWALYPGKPKS----- 225
sp_p83847_MUTY_GEOS 147 VMKVLRSFLVYDDISA-----KPSRKRFQEI-----VRE-IMA-----YENPAGFNAELIEGALVCTPRRPSCL-----LCPQNGCAIAAANSWALYPGKPKS----- 231
VERVMSRLFDISTPLP-----GALKPALKR-----AAS-LTP-----KLRPDGYAIVMDLGATICTARAPTCG-----VCPINCCCHARRAGTAELPKRTPK----- 232
VERVLSRLKIDGNPK-----SGTARKFOQA-----ADD-ALD-----QENPDGHNALMDLGRSICTPRSPDCQ-----RCLASGCGARRAGTARWTRPK----- 233
VERVSRIFQITQIP-----LSRPLTLRQL-----AAG-LTP-----SKRPDGYAIVMDLGLSVCTPRPKPCN-----VCPILHICLANISGTPSPFKRTPR----- 233
TARYFARLALASDLH-----SGEGRQLWSL-----AER-LVP-----RDRPGEFNALIDLGLATVCGV-TPRCQ-----ECPVRGSRSLSELEPARRAS----- 231
VERVTRRLAIGTFLP-----AAKPQVRGA-----VAA-LTP-----AERPDGFAIVMDLGATICTPRRPSCI-----LCPQNGCAIAAANSWALYPGKPKS----- 243
VERVSRIFQITQIP-----LSRPLTLRQL-----AAG-LTP-----SKRPDGYAIVMDLGLSVCTPRPKPCN-----VCPILHICLANISGTPSPFKRTPR----- 190
VERVLSRAYAEAPLP-----GSRPEIRLL-----TQA-LTP-----PDRPDGFAIVMDLGATICTPRPKACA-----LCPMWRPARRSLGTGSEFPRKIRK----- 252
IERVMARHHSPLP-----SAKPELLAC-----AAA-LTP-----DPRAGDYAIVMDLGATICTPKSASCG-----VCPMSFACILRANTAAVLPKRTPK----- 235
VERVMARHHSPLP-----SAKPELLAC-----ATA-LTP-----DLRAGDYAIVMDLGATICTPKSASCG-----VCPMSFACILRANTAAVLPKRTPK----- 235
VNRVMSRLRLHRLFT-----YNGKVLST-----LIR-WLD-----RSRPGDGNALMDLGLSVCTPRPKPCN-----NCPILKLSGTPSGSPSPKPK----- 223

c_000003607531_2 251 VERVTRVFGIFKPLP-----EARPIIKDCAAK--LAP-KTT-----RGRPGDYAIVMDLGATVCRPKPLCD--LCPWEKCYIANQONFVDQVPRKAPK-- 236
c_000003872363_4 259 VERVMARLFNVITPLP-----AAKPELTAT--AAA-LTP-----VDRAGDYAIVMDLGATVCTPKSPGCG--ICPCTEYCAALPEGDAANLPLRPK-- 231
c_000004820107_1 256 IERIMARLYAITEPLP-----DSKANLKNLAA--GLS-EER-----KDRPGDYAIVMDLGATVCTPKSPKCS--LCPVNDSCIAKRGIAESLPIKKGK-- 240
c_000005774797_1 250 INRVLSRVYINDDAIN--RPSARLARI--AET-LAH-----PKRSGDYAIVMDLGATVCKPKPNPNC--TCPICQCKAHSGLHVDLPIKAPQ-- 222
c_000005867021_3 253 VERVMARLYNIHSLP--AVKPMLLTH--ATA-LTP-----EFRAGDYAIVMDLGATVCTPKSPGCG--ICPWSFDCIAHAKGTAALQPKTKP-- 235
c_00000546210_2 144 VKRVLRSRFFGKGMWSG--ESKYSKELMDL--SAK-SLP-----VDNPFYITVIMDLGATVCLPKPNPNC--NCPVINKYSYINLQVDEIPSKGLA-- 228
c_000004551008_5 144 VKRVLRSRFFGKGMWSG--ESKYSKELMDL--SAK-SLP-----VDNPFYITVIMDLGATVCLPKPNPNC--NCPVINKYSYINLQVDEIPSKGLA-- 228
c_000002511148_4 139 VKRVLARHFPIEGKLE--TSSKSKELMDL--SED-LLP-----DNKIEVITVIMDLGATVCLKSSPACN--LCPINQDCLALKEDLVTVLPKTKP-- 223
c_000000577378_2 127 VKRVLARHFPIEGKLE--TSSKSKELMDL--SED-LLP-----DNKIEVITVIMDLGATVCLKSSPACN--LCPINQDCLALKEDLVTVLPKTKP-- 223
c_00000589175_2 150 VKRVLTRVTVGWPWG--KADVLKELMDL--AEQ-TTP-----ASRNDYITVIMDLGATVCTRSKPCSC--DCPLADCLAFQGGQVTVPTSKPK-- 234
c_000000754627_3 147 VKRVLARFFMVGMWG--VKGVENQELMDL--SEQ-LTP-----KXNVTFVIMDLGATVCLSRSPKSC--ACPLNSRCKAHSGLHVDLPIKAPQ-- 231
c_000000811118_1 138 VKRVLTRVLSGSKSPMN--KYKTEKSLKEF--ANT-LTP-----NSKHKEYSVIMDLGATVCPKPNPNC--ECPVDEDCANQKRGIDQELPIKTK-- 222
c_000000899687_2 142 VKRVLRSRFLCNKNGT--TSAFENKLWGE--AED-LLP-----VRRPGDFNIVMDLGATVCLPASPIQ--CQPLRTICKAFKLNKKNVDEIPSKGNK-- 226
c_000001176522_24 161 VKRVLRSRFLGVGWPWG--ERQVEAHLWTL--AEC-LLP--QSEVMPDYITVIMDLGATVCTRTKPKCE--ACQPEADCAVAYQKRGVDELPAKPR-- 247
c_000001345122_1 117 VKRVLTRCDFAVTSWSG--NKKTEKVLMDL--ADE-FTT-----QERFDYITVIMDLGATVCTRSKPKCD--CQPEVADCAKMRKQVIFVPSNPK-- 201
c_000001515736_6 131 VKRVLTRVLAFDGDLA--QSRNRELWMEY--AQO-LCPTT--NLHEAMPRTVIMDLGASVCTRSKPTCL--VCPDSECAARAGNPNENPVTRK-- 219
c_000001682161_6 138 VKRVLTRCDFAVTSWSG--NKKTEKVLMDL--ADE-FTT-----QERFDYITVIMDLGATVCTRSKPKCE--CQPEVADCAKMRKQVIFVPSNPK-- 201
c_000001923643_29 141 VKRVLTRVLAFDADLA--VARNERELMDL--AQO-LCPTT--DLQAMPRTVIMDLGATVCTRSKPTCL--VCPDSECAARAGNPNENPVTRK-- 219
c_000002030994_5 139 VKRVLRSRFTVVGWPWG--KADVLKELMDL--AEQ-TTP-----KTRNDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000002038721_4 141 VKRVLRSRVLGNKSPIN--HSKTKELKKEF--AAE-LLP-----KNDFTYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000002523527_15 156 VKRVLARFFVGIHGYPG--ERALEENMML--AEQALPAG--RHQADHVAITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000002608528_2 139 VKRVLTRVTVGWPWG--KSGILKELMDL--AEQ-TTP-----QTRNDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000002961510_2 142 VKRVLRSRFLCNKNGT--TSASENKLWGE--AED-LLP-----VRRPGDFNIVMDLGATVCLPASPIQ--CQPLRTICKAFKLNKKNVDEIPSKGNK-- 226
c_000002979152_8 147 VKRVLARVHAIEGMPG--KSAVLKELWKE--AEL-HLP-----AERNADYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000003248034_1 159 VLIRVTRVLEIKGNVA--STAQQKLTAA--ARL-LVES--VGPHSAGTVMIMDLGATVCTPQSPSC--LCPINRMSAAKRGVDELPAKPR-- 246
c_000003294679_3 144 VKRVLTRVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000003333364_1 144 VKRVLARLFVDPAIK--SSQAQSFWEET--ARS-LTP-----TGQAREFVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000003140247_5 142 VKRVLTRVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000003492925_4 134 VKRVLRSRFLCNKNGT--TSASENKLWGE--AED-LLP-----VRRPGDFNIVMDLGATVCLPASPIQ--CQPLRTICKAFKLNKKNVDEIPSKGNK-- 226
c_000003766664_5 154 VKRVLRSRFLCNKNGT--TSASENKLWGE--AED-LLP-----VRRPGDFNIVMDLGATVCLPASPIQ--CQPLRTICKAFKLNKKNVDEIPSKGNK-- 226
c_000003948186_72 141 VKRVLRSRVRHVGWAG--RASVNKLWQL--AER-YTP-----KRRVRYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000005057120_4 140 VKRVLTRVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000005590109_5 140 VKRVLTRVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000006063368_2 139 VKRVLTRVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000000467631_2 119 VKRVLARVHAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000000788235_1 140 VKRVLTRVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000001187176_11 141 VKRVLARVHAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000001660697_64 141 VKRVLARVHAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000003378864_2 152 VKRVLRSRFLCNKNGT--TSASENKLWGE--AED-LLP-----VRRPGDFNIVMDLGATVCLPASPIQ--CQPLRTICKAFKLNKKNVDEIPSKGNK-- 226
c_000003910742_3 141 VKRVLTRVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000004116181_2 107 VKRVLARVHAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000006223903_1 142 VKRVLRSRVRHVGWAG--RASVNKLWQL--AER-YTP-----KRRVRYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000001029068_2 114 VKRVLARVHAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000001128125_11 144 VKRVLRSRVRHVGWAG--RASVNKLWQL--AER-YTP-----KRRVRYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000001595844_4 140 VKRVLARVHAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000001719155_2 144 VKRVLRSRFLCNKNGT--TSASENKLWGE--AED-LLP-----VRRPGDFNIVMDLGATVCLPASPIQ--CQPLRTICKAFKLNKKNVDEIPSKGNK-- 226
c_000002363038_6 149 VKRVLRSRVRHVGWAG--RASVNKLWQL--AER-YTP-----KRRVRYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000002561400_1 105 VKRVLTRVFGVHWPWG--KADVLKELMDL--AEQ-TTP-----KTRNDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000002566035_2 141 VKRVLTRVFGVHWPWG--KADVLKELMDL--AEQ-TTP-----KTRNDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000002598045_3 139 VKRVLARVHAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000004188575_41 141 VKRVLARVHAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_00000240049_1 139 VKRVLARVHAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000002139456_2 139 VKRVLARVHAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000003535614_2 141 VKRVLARVHAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000002657784_4 138 VKRVLARVHAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000003800129_15 143 VKRVLRSRVLGNKSPIN--HSKTKELKKEF--AAE-LLP-----KNDFTYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000003716781_1 130 VKRVLARVHAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000001556689_3 139 VKRVLARVHAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000001834452_1 108 VCRVARSRLGKSPG--SSALRGQLEAM--GET-LHAG--VSAGLAGSINIVMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000004255004_2 148 ARRVYARLLSLTN--LRAINTE--AEQ-MYS--HSRPGDFNIVMDLGATVCLPASPIQ--CQPLRTICKAFKLNKKNVDEIPSKGNK-- 226
c_00000605438_3 141 VKRVLRSRVRHVGWAG--RASVNKLWQL--AER-YTP-----KRRVRYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000000581237_15 141 VKRVLRSRVRHVGWAG--RASVNKLWQL--AER-YTP-----KRRVRYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000002529579_2 141 VKRVLRSRVRHVGWAG--RASVNKLWQL--AER-YTP-----KRRVRYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_00000097838_1 114 VARVLTLLGIERPSK--QKQVTEGLWSV--TAD-LVHVAIDLKQSPRNASAFNIVMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_00000583727_6 140 VARVLTLLGIERPSK--QKQVTEGLWSV--TAD-LVHVAIDLKQSPRNASAFNIVMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000005807640_2 142 VARVLTLLGIERPSK--QKQVTEGLWSV--TAD-LVHVAIDLKQSPRNASAFNIVMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_00000134878_1 124 VNRVRSRFLGVGWPWG--KADVLKELMDL--AEQ-TTP-----KTRNDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000005136725_3 139 VNRVRSRFLGVGWPWG--KADVLKELMDL--AEQ-TTP-----KTRNDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000001684786_4 137 VERVTRVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000002971826_23 147 VERVTRVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000002040695_2 150 IARVLSRVLGPGSKE--ALTTSRLQEP--LSL-FID--FKRIGDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000003263657_32 154 IARVLSRVLGPGSKE--ALTTSRLQEP--LSL-FID--FKRIGDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000005989041_2 151 IERVTRVFGICTPFP--EAKTIVKDCAAK--LAP-KTT-----RGRPGDYAIVMDLGATVCRPKPLCD--LCPWEKCYIANQONFVDQVPRKAPK-- 236
c_000002391082_2 149 ITRVLSRFLRIEDPR--RTAEIAELIA--GEA-LIA--RGEAGDFNIVMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000003839553_2 156 ITRVLSRFLRIEDPR--RTAEIAELIA--GEA-LIA--RGEAGDFNIVMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000003996707_2 142 ITRVLSRFLRIEDPR--RTAEIAELIA--GEA-LIA--RGEAGDFNIVMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000002044706_1 129 ITRVLSRFLRIEDPR--RTAEIAELIA--GEA-LIA--RGEAGDFNIVMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000006007075_1 141 VERIIRKILNLAKEKE--ISKENIIRK--KKI-LGM--SDRSDDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000002094036_4 120 IERVLARVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000002762689_1 144 IERVLARVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_00000516980_2 141 IERVLARVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000003920004_1 150 IERVLARVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000004008511_2 144 IERVLARVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000004481347_1 151 IERVLARVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000003787733_3 143 IERVLARVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000000754657_2 143 IERVLARVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000000990943_3 143 IERVLARVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000004474996_2 142 IERVLARVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000001463500_11 140 ITRVLSRFLRIEDPR--RTAEIAELIA--GEA-LIA--RGEAGDFNIVMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000001286181_5 137 VKRVLRSRVLGNKSPIN--HSKTKELKKEF--AAE-LLP-----KNDFTYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000001293628_3 143 VKRVLRSRVLGNKSPIN--HSKTKELKKEF--AAE-LLP-----KNDFTYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000001535696_8 137 VKRVLRSRVLGNKSPIN--HSKTKELKKEF--AAE-LLP-----KNDFTYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000001614067_2 150 VKRVLRSRVLGNKSPIN--HSKTKELKKEF--AAE-LLP-----KNDFTYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000001765289_1 116 AKRVSRLGKIKRL--TSWNLRLNKT--LSN-LTP--EHTPGDFNIVMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000001961666_1 131 VERVTRVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000002018097_4 141 VKRVLRSRVLGNKSPIN--HSKTKELKKEF--AAE-LLP-----KNDFTYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000002843512_36 135 VKRVLRSRVLGNKSPIN--HSKTKELKKEF--AAE-LLP-----KNDFTYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000002930199_1 135 VKRVLRSRVLGNKSPIN--HSKTKELKKEF--AAE-LLP-----KNDFTYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000003037395_2 131 IERVLARVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000003886107_2 132 IERVLARVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000004013286_6 157 SVRVLRSRVLGNKSPIN--HSKTKELKKEF--AAE-LLP-----KNDFTYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000004852258_1 151 HSRIARVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000005254087_1 145 VKRVLRSRVLGNKSPIN--HSKTKELKKEF--AAE-LLP-----KNDFTYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000005603677_1 142 VKRVLRSRVLGNKSPIN--HSKTKELKKEF--AAE-LLP-----KNDFTYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000003283462_3 140 IERVLARVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000002687221_1 144 IERVLARVLAIEGMPG--TMMQVNLWKI--AED-NTP-----SNRVDYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000003347358_19 166 VKRVLRSRVLGNKSPIN--HSKTKELKKEF--AAE-LLP-----KNDFTYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000002701031_2 137 VKRVLRSRVLGNKSPIN--HSKTKELKKEF--AAE-LLP-----KNDFTYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_00000582753_3 143 VKRVLRSRVLGNKSPIN--HSKTKELKKEF--AAE-LLP-----KNDFTYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000001713769_5 137 VKRVLRSRVLGNKSPIN--HSKTKELKKEF--AAE-LLP-----KNDFTYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000004369364_1 140 VKRVLRSRVLGNKSPIN--HSKTKELKKEF--AAE-LLP-----KNDFTYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000006057486_30 141 VKRVLRSRVLGNKSPIN--HSKTKELKKEF--AAE-LLP-----KNDFTYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000005494072_10 137 VKRVLRSRVLGNKSPIN--HSKTKELKKEF--AAE-LLP-----KNDFTYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223
c_000001742634_3 140 VKRVLRSRVLGNKSPIN--HSKTKELKKEF--AAE-LLP-----KNDFTYITVIMDLGATVCTRSKPKCE--DCPLADCLAFQGGQVTVPTSKPK-- 223

c_000002826998_2 140 VFRVLSRYFGIETPM-----TSEGGKQFTHL--ANE-LLL-----KKKPAIYN--AIMDFGAIQCTPKSPDCK--HCFLEKKKQALQENKISKLPVSKKK-- 224
c_000003159439_6 137 VYRVLYRYFGIRTSIN-----STKGIEKFKQL--AQE-LID-----TKDPATFN--AIMDFGAIQCTPKSPNPN--NCPINISCTIALQKGLITLPIKDKK-- 221
c_000004887214_1 142 VYRVLYRYFGIRTSIN-----TSKAKKEFKFV--AFE-LID-----SKNPGGLFN--AIMDFGAIQCTPKSPNPN--ICPINISCTIALQKGLITLPIKDKK-- 226
c_000005037037_2 140 VYRVLSRYFFGIETPQ-----TVKKEKFEQEI--ADV-LIS-----KKSADYN--AIMDFGAIQCTPKSPNPN--KCPISANDYALQKGLITLPIKDKK-- 224
c_000001059964_1 132 MIRVLSRYFFGIETPQ-----LAKTRKRLWLQ--ARE-WLP-----EGEARRFN--AIMDFGAIQCTPKSPNPN--LCPNPNHCRALQKGLITLPIKDKK-- 216
c_000002498472_1 141 VERVFARIFNIDLIPG-----SPEARQNLWEK--AEZ-LIP-----KGHARNFN--AIMDFGAIQCTPKSPNPN--SCPIVTHCRALQKGLITLPIKDKK-- 225
c_000004615912_3 133 VYRVLSRYFFGIETPQ-----ISKNKKIFEVL--AQE-LID-----ASQAGKYN--AIMDFGAIQCTPKSPNPN--LCPNPNHCRALQKGLITLPIKDKK-- 214
c_000002746658_2 143 IERVLCRIEDIEGPPK-----STPVQKQLWSL--AAE-WLP-----HGHAERFN--AIMDFGAIQCTPKSPNPN--LCPANAMVAFHGTGDTITPAPVR-- 227
c_000002747260_18 149 VERVMARLFQIETPLP-----GAKPELITF--AGR-LTP-----DIRPGDHA--AIMDFGAIQCTPKSPNPN--ICPVMFGCTARAAGTQADLPKTKP-- 231
c_000005371561_38 149 VERVMARLFQIETPLP-----AAKPVKLAQ--AAA-LTP-----ATRPDGEHA--AIMDFGAIQCTPKSPNPN--ICPAPACARAAGTQADLPKTKP-- 231
c_00000164166_2 145 VKRVLYRYYSIGWMPG-----OKKVENQLWEV--AEK-WTPTN-----SEGGRCANYT--VIMMELGAMICTRSKPKCD--ECPLQDPCYALQAGQADYDQKSKPK-- 233
c_000003254110_11 145 VKRVLYRYYSIGWMPG-----OKKVENQLWEV--AEK-WTPTN-----PEGGRCANYT--VIMMELGAMICTRSKPKCD--ECPLQDPCYALQAGQADYDQKSKPK-- 233
c_000004750284_20 145 VKRVLYRYYSIGWMPG-----OKKVENQLWEV--AEK-WTPTN-----PEGGRCANYT--VIMMELGAMICTRSKPKCD--ECPLQDPCYALQAGQADYDQKSKPK-- 233
c_00000141782_15 145 VKRVLYRYYSIGWMPG-----OKKVENQLWEV--AEK-WTPTN-----PEGGRCANYT--VIMMELGAMICTRSKPKCD--ECPLQDPCYALQAGQADYDQKSKPK-- 233
c_000002717847_8 144 VKRVLYRYYSIGWMPG-----KEGLFWNL--SAQ-CLD-----EDDPYSYQ--GIMDYGATLCTPKAPLC--ECPLQDPCYALQAGQADYDQKSKPK-- 229
c_000004054799_1 131 VKRVLYRYYSIGWMPG-----ENEGKQKGLWL--SEE-LTP-----NKESFEYT--GIMDYGATLCTPKAPLC--ECPLQDPCYALQAGQADYDQKSKPK-- 229
c_000006126673_1 156 IERVLYRYHAINTPLP-----KAKKEIKERAT--IYT-DNL-----PERPGDYA--AIMDYGATLCTPKAPLC--ECPLQDPCYALQAGQADYDQKSKPK-- 240
c_00000339186_3 113 VYRVLYRYHAINTPLP-----DKATREYLSWL--AET-LVA-----KAENSSALN--SIMELGATVCTQRRPDCA--ACPVARSCTARRERATEIIPAKGPE-- 198
c_000003029168_2 145 VAVRVLSRLFAVGDDLK-----DTALQKGLWEI--TED-LVP-----NKRPGDFN--SIMELGATVCTQRRPDCA--ACPVARSCTARRERATEIIPAKGPE-- 229
c_000001863436_1 161 VAVRVLSRLFAVGDDLK-----SGDNQRRFWL--AGA-LID-----RGRPGDFN--SIMELGATVCTQRRPDCA--ACPVARSCTARRERATEIIPAKGPE-- 245
c_000002830137_4 144 TTVVLSRLFAVGDDLK-----QRRVQQLWMAF--AES-LIP-----RKRVRGFN--SIMELGATVCTQRRPDCA--ACPVARSCTARRERATEIIPAKGPE-- 238
c_000004612302_3 138 TARVLSRLFAVGDDLK-----SSEGRGLWLSL--AER-VVP-----RTRAGEFN--SIMELGATVCTQRRPDCA--ACPVARSCTARRERATEIIPAKGPE-- 231
c_000001933926_1 144 IERVMSRLHRLTEPLP-----AAKPEIAEV--AQR-LTP-----RQRAGEFN--SIMELGATVCTQRRPDCA--ACPVARSCTARRERATEIIPAKGPE-- 216
c_00000136334_1 146 VERVMARLFQIETPLP-----GAKKYLKKEK--AAE-LTP-----SERSGDYA--SIMELGATVCTQRRPDCA--ACPVARSCTARRERATEIIPAKGPE-- 228
c_000005849454_9 143 VKRVLYRYYSIGWMPG-----LALAEKAAALGEAMT--DLP-LGK-----LSAPGEIN--SIMELGATVCTQRRPDCA--ACPVARSCTARRERATEIIPAKGPE-- 233
c_000001169194_1 154 VYRVLYRYYSIGWMPG-----RARGNLNKT--AAE-LVD-----KDRPGDFN--SIMELGATVCTQRRPDCA--ACPVARSCTARRERATEIIPAKGPE-- 233
c_000002786947_2 154 VYRVLYRYYSIGWMPG-----GVRLNMT--AAE-LVD-----KDRPGDFN--SIMELGATVCTQRRPDCA--ACPVARSCTARRERATEIIPAKGPE-- 233
c_000001279808_52 142 ISKLIITVWMP-----KGVV--AES-LVA-----AASCHIVN--SIMELGATVCTQRRPDCA--ACPVARSCTARRERATEIIPAKGPE-- 219
c_00000338065_2 144 IERVLYRYYSIGWMPG-----TKKENLHKL--KVV-FGI-----TRSDCYV--SIMELGATVCTQRRPDCA--ACPVARSCTARRERATEIIPAKGPE-- 223
c_000003745941_1 133 VIRILICRLTADNTVYT-----DNTQAMKVFPTL--AEV-LLD-----TKTPQDYN--SIMELGATVCTQRRPDCA--ACPVARSCTARRERATEIIPAKGPE-- 217
c_000004187032_4 144 IERVLYRYYSIGWMPG-----IQNNLILK--KSI-FGI-----SSRADSYA--SIMELGATVCTQRRPDCA--ACPVARSCTARRERATEIIPAKGPE-- 223
c_000006067315_3 144 IERVLYRYYSIGWMPG-----ISKENLILK--KSI-FGI-----SSRADSYA--SIMELGATVCTQRRPDCA--ACPVARSCTARRERATEIIPAKGPE-- 223
c_000005543774_1 141 VKRVLYRYYSIGWMPG-----PAN-LFK-----TKRNDGIA--SIMELGATVCTQRRPDCA--ACPVARSCTARRERATEIIPAKGPE-- 185
c_000006211484_1 136 VKRVLYRYYSIGWMPG-----AVN-LFK-----TKRNDGIA--SIMELGATVCTQRRPDCA--ACPVARSCTARRERATEIIPAKGPE-- 207
c_000002718976_3 145 VYRVLSRLFAVGDDLK-----INFDKFIKN--KKK-FTI-----TKRNDGIA--SIMELGATVCTQRRPDCA--ACPVARSCTARRERATEIIPAKGPE-- 226
c_000002992548_1 146 IKRVYSRLF--NANL--EKNEKRILKI--TKL-QLY-----TKRNDGIA--SIMELGATVCTQRRPDCA--ACPVARSCTARRERATEIIPAKGPE-- 225
c_0000040456751_1 141 VKRVLSRLFAVGDDLK-----QNDINEIKKI--TYK-LPK-----IDRNFADIA--SIMELGATVCTQRRPDCA--ACPVARSCTARRERATEIIPAKGPE-- 191
c_000002655634_1 141 VKRVLSRLFAVGDDLK-----QNFENEIKKI--TKN-LPK-----VNRNADIA--SIMELGATVCTQRRPDCA--ACPVARSCTARRERATEIIPAKGPE-- 222
c_000004232403_2 140 HVRVLSRLFAVGDDLK-----HNNRNINNY--LQK-LVA-----IDRPFGEIN--SIMELGATVCTQRRPDCA--ACPVARSCTARRERATEIIPAKGPE-- 222
6u7t_after_N146.pdb 141 VYRVLYRYYSIGWMPG-----KPSRKRFEQI--VRE-LMA-----YENPGATN--SIMELGATVCTQRRPDCA--ACPVARSCTARRERATEIIPAKGPE-- 225
Consensus aa: 1.VRhsRhh.lp.....ppph.h.....s.....hhs.....ppsp@sAhM-lGthlC.ppP.C.....CPL...C.t.....hP.....

Conservation: 5 5
sp_P17802_MUTY_ECOL 226 QTLPE--RYGYF--LLIQ--HEDEVLLAQRPSS-GIWMGLVCTF--QFADDESILRQ--WLAQRLA--A 281
sp_P83847_MUTY_GEOS 232 TAVKQ--VPLAV--AVLAD--DEGRVLLIRKSDT--GLANLWETP--SCETDGDAG--MVGEOYGLQ--V 294
c_000003652391_2 233 QPKPT--RYGIY--YLARD--AHGAMILLERRDPK--GLLGGLWMP--TSDWNPARG--DQPPFA--A 286
c_0000031207_6 224 RTEQ--QWMAA--AVVV--CGDQVVLWIGDG--ELLAHGRTT--LARISSGLD--DADRMIIG--KEFKRLGLD--GA 286
c_00000430975_4 234 RKQKR--RYGVV--YMIER--RDGALSILQRPEN--GIFGGMTEIP--GSLMKAAPL--SRH--SHHAQAPIT--A 293
c_000001007148_4 232 RDVTD--VFELL--VVMR--RGRGLLLIQHPDG--ERWGGMWTEP--ALSRDRALRS--RFDERLIE--SEVADWGD--V 297
c_000001092848_3 244 AAKPS--RRGAA--FVAVR--ADDAVLLERRPES--GLLGGLWMP--TSWGSAGSD--GEQGP--SAASFP--A 302
c_000001151029_1 191 TKRKR--RYGVV--YMIER--RDGALSILQRPEN--GIFGGMTEIP--GSLMKAAPL--SRH--SHHAQAPIT--A 250
c_000001797282_1 253 AKGAL--RRGAA--FVAVR--SGDEAVLLRTPPE--GLLGGLWMP--GSAMEPDYV--VAAAL--LDAPLD--A 312
c_000001803648_25 236 PVKPT--RYGIA--YIARR--GDGAWLLERRDPK--GLLGGLWMP--GAQWGDQAV--DKPPVS--A 289
c_000002078955_4 236 PVKPT--RYGIA--YIARR--GDGAWLLERRDPK--GLLGGLWMP--GAQWGDQAV--DKPPVS--A 289
c_000002106160_4 224 QSLFW--GVVLT--GVV--KRDQSLILIKRAGY--KHNNLWELP--GGRTLTQNN--PHQLEE--LLKNKFKLE--V 286
c_000003607531_2 237 KVXAV--RKGVV--FLAAM--DGSGLFLRRRPN--GLLGGLWMP--STVWLSRL--SPK--EICEAAPK--T 296
c_000003872363_4 232 PAKPI--RYGTA--YIARR--HDGAWLLERRDPN--GLLGGLWMP--GSEWGEAAV--ENPPFS--A 285
c_000004820107_1 241 KPKPK--KRGHV--FWISD--DSGNLFIERRPN--EMLGGLWMP--TTEWDIRP--IKLP--Q 291
c_000005774797_1 223 KSKKK--WNAFA--YVIYN--ADGNIGFVRDOK--ALLGGWALP--TSDWSTSPV--FQP--PI--L 275
c_000005867021_3 236 PVKPT--RYGIA--YIARR--PDGAWLLERRDPK--GLLGGLWMP--GAQWGDQAV--DKPPVS--A 289
c_000004546210_2 236 KSKPT--ESMPT--LMIVP--DGRKFLERRKPK--GIWGLWSPF--QTGLADPQI--WCEQPTFN--I 286
c_000004551008_5 229 KSKPT--ESMPT--LMIVP--DGRKFLERRKPK--GIWGLWSPF--QTGLADPQI--WCEQPTFN--I 286
c_000002511148_4 224 KSKPT--KKVWV--LLPQG--PSGEVLLERRKPK--GIWGLWSPF--EAKKNELEL--ELSRKPTD--C 281
c_00000577378_2 212 KSKPT--KKVWV--LLPQG--PSGEVLLERRKPK--GIWGLWSPF--EAKKNELEL--ELSRKPTD--C 269
c_000005598175_2 235 KVMPE--KQAIM--VILGN--SKQEVFMQRPPV--GIWGLWSPF--QFEKRAFAEE--WLKSNVGLK--I 292
c_00000754627_3 232 KAVPK--KSCHQ--LIIL--HNKVVLYQRKPK--GIWGLWSPF--EFNEYSELT--FLAQQGLK--I 286
c_0000081118_1 223 KEKIE--KKIDW--ILIK--TKNVLVLRKPK--GIWGLWSPF--EKDFFSKKI--SKI--L 273
c_00000899687_2 227 ISSKK--TEVSA--GIII--KNKKYIYQQRSKN--ALMGGLWSPF--GGKREKES--PEECLKR--EIKEELRVN--V 289
c_00000176522_24 248 KVMFP--RKVCW--QVSL--QCGVVLVQNPAP--GLMGGLWSPF--EGNR--FLDGNEMVAT--LK 261
c_000001345122_1 202 KILPT--KEKFF--LLIVT--EAFQVALLVKEPK--GIWGLWSPF--EFDVQIEDL--FLDGNEMVAT--LK 261
c_000001515736_6 220 IKRSS--ESWML--LLAVD--ARRRVLMQRRPK--GIWGLWSPF--VFEDYALQS--YQAQWSDG--GT 282
c_000001682161_6 223 KILPT--KEKFF--LLIVT--EAFQVALLVKEPK--GIWGLWSPF--EFDVQIEDL--FLDGNEMVAT--LK 261
c_000001923643_29 230 IKRSS--ESWML--LLAVD--ARRRVLMQRRPK--GIWGLWSPF--VFEDYALQS--YQAQWSDG--GT 282
c_000002030994_5 224 KEMPE--KHTVM--LILGN--DQGEVLMQRRPK--GIWGLWSPF--QFDTALAE--WLDSDFGMS--L 281
c_000002038721_4 226 SKQKE--KNINW--LLIA--SKDKVLLKRVNPK--GIWGLWSPF--EANSINYEK--FNL--L 275
c_000002523527_15 251 AALPE--RSTVM--LIVRH--GRDVLVQLRPPS--GIWGLWSPF--EMPVDTVPFD--SEAAEESA--LELARAYS--P 315
c_000002608528_2 224 KVMPE--KQVVM--LILGN--KNQEVFMQRPPV--GIWGLWSPF--QFKNHENVLE--EDKNNEVYS--F 281
c_000002961510_2 227 VPSKK--TEVSA--GIII--KNKKYIYQQRSKN--ALMGGLWSPF--GGKREKES--PEECLKR--EIKEELRVN--V 289
c_000002979152_6 232 KDKRI--KQSVW--LIVSN--ENADVLLERRSPS--GIWGLWSPF--EFNEYSELT--FLAQQGLK--I 286
c_000003248034_1 247 KSKKR--VMSYV--LACE--QCGVLMQRRPK--GIWGLWSPF--EFGSPFAN--ADL--ARWS--L 301
c_000003294679_3 229 KTNPL--KTNM--LFLRF--EDRYLLERRSPS--GIWGLWSPF--EISKDCSVD--WYQSRGAY--P 285
c_00000333364_1 225 KPAVP--VDVAS--GVIV--HDGYLFIQRRPKD--GVWGLWSPF--GGTVEKGT--PAQCVR--EFOEELEND--V 287
c_000003402697_5 220 TEKPKGRGHR--LEVGI--ACIV--HEGKYLVLQRRPK--GSFGEWSPF--GKREKGEN--FRECVR--EIQEEVGD--V 287
c_000003492925_4 212 VVPT--REENI--MLYV--YNNKLSLITQRG--KFLHGLWSPF--NTEMLP--KAGVGR--LVKTEVGLT--A 252
c_000003766664_5 239 RPKPK--IALVAC--VIRD--EQGAVLLTQRRPK--GLFGGLWSPF--LKEVDTPS--KAGVGR--LVKTEVGLT--A 301
c_000003948186_72 226 NNTPT--RHTQM--LLIEN--TAGEVNLEKRPAP--GIWGLWSPF--ELATQONAIT--YCEQELQIA--V 283
c_000005057120_4 225 KVLAV--KQLIF--LMLQD--QCTFLLEKRPAS--GIWGLWSPF--EFQSFALIS--WCLN-DIA--M 281
c_000005590109_5 225 KSLPV--KKITF--LMQDQ--QCGVLLLEKRPAP--GIWGLWSPF--EYESLANIQS--CFLEK-GIP--I 281
c_000006063368_2 224 KTKIT--KHVHW--LIPYS--ELGKFLVLRKPK--GIWGLWSPF--EKEKEMLK--EKSIFQIK--D 281
c_000006467631_2 204 KTLPV--KNIME--LMLQD--ELGKFLVLRKPK--GIWGLWSPF--EFTLSEIQS--FCQON-NFA--I 260
c_00000788235_1 225 KTLPE--KKAFF--LILLN--INHEIFILKRPSPS--GIWGLWSPF--QFDYTLAQ--WYEQYFSTH--L 282
c_000001187176_11 226 KRLPV--RQCLM--PLLAV--PQGDVLLQRRPKD--GIWGLWSPF--QLDSREQLDA--LVIAQ-GWQ--A 282
c_000001660697_64 231 KALPE--RAATM--VIALH--GETVLLQRRPK--GIWGLWSPF--LVDVTRVQALAYGT--LADDTVTRVQALAYGT--V 296
c_000003378864_2 238 KPVPH--HHIGV--GVIMK--DDQVLLQLRPPS--GLLGGLWSPF--GGKQPAET--TEETVRRREIREGLVD--V 300
c_000003910742_3 226 KVTPE--KNAMV--LMLN--NKDEVFMQRPPV--GIWGLWSPF--QFDYDLQAA--WHENYGFPP--C 283
c_00000416181_2 192 KQKAT--RCVYV--ILIVN--ARQEILLQRRSPV--GIWGLWSPF--EVEDTEAIIH--WCSSR-KIT--V 248
c_000006223903_1 226 KRPKR--REVMV--LMINK--SRGVVLLERRSPS--GIWGLWSPF--ECPLVREALP--WCMRVKID--I 283
c_000001029068_2 192 KRPKR--KTAM--LIFDN--GKGVYLQRRPK--GIWGLWSPF--ECSATKAIT--KTANHFQI--A 256
c_000001128125_11 234 KDKPV--KRAFF--LILLN--SERQVLLQRRPK--GIWGLWSPF--EFDSLQFKE--SIESQTVT--I 290
c_000001595844_4 225 IKRKT--RYFYV--FVFR--KGSYYIYQRRPK--GIWGLWSPF--NIESKPTLN--FQKYPRK--F 287
c_000001719155_2 229 APSK--LAVSA--GVLF--RRNRVYIYQRRPK--GIWGLWSPF--GGKFSGES--PEQCLR--EIKEELVGL--I 291
c_000002363038_6 234 KSKPV--KQKWL--LILQK--SAEQVLYYQRRPK--GIWGLWSPF--EFDSDRAVD--FLNGEHRH--DA 292
c_000002561400_1 190 KSKPV--KQKWL--LILQK--SAEQVLYYQRRPK--GIWGLWSPF--EFDSDRAVD--FLNGEHRH--DA 292
c_000002566035_2 226 AVLPK--RDTVF--AIMEN--NNGEILLQRRPK--GIWGLWSPF--EFSSPLRIE--MIKRYGYFN--I 283
c_000002598045_3 224 KSKLT--KQVW--LLPQG--PSGEILLQRRPK--GIWGLWSPF--ETEKKAEL--ALSRNFDN--I 281
c_000004188575_41 226 KILPT--RQSHW--LILLT--DRRVLLSRKPK--GIWGLWSPF--SFDDTHL--EQHIDT--K 279
c_00000240049_1 224 KILPV--REKRL--LIIRN--KQGHVLEKRPSPS--GIWGLWSPF--ELTLDKSAE--STERNWLS--V 281
c_000002139456_2 224 KILPV--REKRL--LIIRN--KQGHVLEKRPSPS--GIWGLWSPF--ELTLDKSAE--STERNWLS--V 281
c_000003535614_2 226 KILPI--QKRL--LIIRN--EQGAYLEKRPSPS--GIWGLWSPF--ELALAPIT--EVKNTWII--V 283
c_000002657784_4 223 KQKPF--KTVFW--LVVMN--KNGVLLKRRNPT--GVWGLWSPF--ESENIDQLK--ECLTMFKR--K 280
c_000003800129_15 221 TPKKK--VFLNY--EYIK--RDGCIYMIKRMML--GFNGLWSPF--YKIVE--TREVAAELRLF--FAELELEL-- 262
c_000003716781_1 216 PTPID--VELEM--FLVH--DHGRVLLRRDEG--GMAGGLWSPF--TREVAAELRLF--FAELELEL-- 272

c_000001556689_3	216	LVKA	VTEYA	AVIE	DNGRLLFLFRGQRP	SLVSDMWEFF	TLDRLADTSSRSLSITEKPASSHRAQEAQLSR	YIKEQLGWS	293				
c_000001834452_1	196	INWKE	VHLLY	GVAS	CPGVLLEERKS	GMWQGLWEFF	SVYPDQKEE	AWREFQQR	253				
c_000004255004_2	226	TKA	VQWPL	TLAK	WRSRILLHRRPDK	GLIASLWEFF	TPENLPAEL	IDETL	274				
c_00000605438_3	219	IKKPI	RKRAL	IYYA	KNDKYALAAQNEE	RLISGLWGF	QEEDEF		262				
c_00000581237_15	219	KKKPI	RKRRL	IYYT	KDDKYALAAQNEE	RLISGLWGF	QEEDEF		262				
c_000005259579_2	219	IKKPI	RKRRL	VIYH	KANKYALAAQNEE	RLISGLWGF	QEEDEF		262				
c_000006097838_1	208	APVSK	RRFIF	FLIA	WHGNVVFQRRPSG	EVNAGWEFF	NMLVDDTQ	VQPAD	267				
c_000005583727_6	234	AATIN	RRFIF	FLIA	WHGNVVFQRRPSG	EVNAGWEFF	NMLVDDAE	LKPAA	293				
c_000005807640_2	236	AATTK	RRFIF	FLIT	WHGNVVFQRRPSG	EVNAGWEFF	NMLV DAA	ATADQ	294				
c_0000013478_1	207	ETKPH	YDVAV	GLIIV	DKKKLLITRKEE	GLLGLWEFF	GKKIKKEK	IESAIKR	269				
c_000005136725_3	222	HTKPH	YDVAV	GLIIV	DKKKLLITRKEE	GLLGLWEFF	GKKIKMET	IKSAIKR	284				
c_000001684786_4	222	RKTLA	VELEA	GLIIR	RGKRCLLERSSEF	DFLEGLWEFF	LSLPGD	GGIAA	279				
c_00000291826_23	232	RKTLA	VELEA	GLIIR	RGKRCLLERSSEF	DFLEGLWEFF	LARPS	GGIAA	289				
c_000002040695_2	233	TKGPC	RWGHF	FFLVHIPSSGSPV	SLLEKEHEG	PLLEGLWEFF	TQSMVHKKT	EAH	290				
c_000003263657_32	239	KKERP	HHDHVA	GVIK	LDLDFLLIQRPDL	GLLGLWEFF	GGRRESSES	LTGVTIR	302				
c_000005989041_2	237	KVKPV	RKGIV	FLAAM	SDGSLFLRRRPET	GLLGLWEFF	STNMLERRI	SPE	296				
c_000002391082_2	234	KKQKP	HYQVAC	GVIC	KGDRLLLAQRPSA	GMGLWEFF	GKKQEEGET	LQCCLOQ	297				
c_000003839553_2	241	KKKKP	HYQVAA	GVIS	KGRQLLLAQRPAE	VMLGGLWEFF	GKKQEEGET	LEECILR	304				
c_000003996707_2	231	KPTRP	HYEVTA	GVIV	KGSKLLAVQRPSPD	GMGLWEFF	GKKREPSES	LEECILR	294				
c_000002044706_1	214	TKVTF	VQRVA	LFLFK	QDGSFLVRRPAD	GLLAGMWEFF	GRELGDGEAPE	RLASTLCG	270				
c_000006007075_1	224	KKKID	KYFLI	KIYK	KKKKILLIRNTKF	NFLKNLIFP	MEETPKP		267				
c_000002094036_4	200	KKIIN	KFYLA	TLYK	NQNKILLIRNTKF	NFLKNLIFP	MREITYP		244				
c_000002762689_1	224	KKNIN	KYLLI	KVYK	DKNKYLLIRNTKF	NFLKNLIFP	MQELSKP		267				
c_000005516980_2	224	KKNIN	KYFLL	KVYK	KDKRYLLIRNTKF	NFLKNLIFP	MEELFQP		267				
c_000003920004_1	233	KKSRMK	KYTRAY	LIIN	GSEELLVRRASK	GMLPSLWEFF	NDKMWTEKK	LLVRD	292				
c_000004008511_2	227	KSTKKR	KYSRAY	IFYN	EKNEILLIRKRPK	GMLASLWEFF	NDNMVINKK	SLVTD	286				
c_000004481347_1	234	LEQKKI	KFTRAY	LIIMN	KKNEVLLVRRASK	GMLASLWEFF	NDNMVINKK	SLVTD	286				
c_000003787733_3	226	KIIPH	KEIVA	GLIIV	QKQKFLITRKEEN	ALLGLWEFF	SATQPMET	PIGALRR	293				
c_000000754657_2	226	KIIPH	KEIVA	GLIIV	QKQKFLITRKEEN	ALLGLWEFF	GGEIASSST	PVGAIR	288				
c_000000990943_3	226	KVIPH	KDIVT	GLIIV	RGKKFLITRKEEN	ALLGLWEFF	GGEIASSST	PVGAIR	288				
c_000004474996_2	225	KVIPH	KEFVA	GMIC	QKQKFLITRKEEN	ALLGLWEFF	GMEIVSSET	PIGALRR	287				
c_000001463500_11	213	SIKNI	KHFIA	ALIV	YNKFLIMGRPVN	SMGLWEFF	NTQIDLKIS	NTQILFN	275				
c_000001286181_5	218	VTKTQ	ESFHW	FIYQ	KNNKVMCLKNPDE	GIWNLWEFF	KRELFOQ		262				
c_000001293628_3	223	PRVRE	ASYAV	LVTIN	PVGEILLVRRPSPD	GLLGLWEFF	AVEMAEERNPWGELR	VGDSLHGRCASLE	301				
c_000001535696_8	222	TKVRN	RYFNY	LIPIA	ADGDSRKLITLQKRGK	DIWNLWEFF	LIEETKREID	LEEVRK	297				
c_000001614067_2	233	KEKEE	RYGLF	FYLON	KGAVLLETNKKSS	GLANMDL	SIGWYEEN	RFKKSPK	291				
c_000001765289_1	198	KKRKP	HYTIVA	GLIIV	RDNFTFFQRRPEK	AMLGGLWEFF	GKKVEEGES	LEAALKR	261				
c_000001961666_1	213	KIIRK	GKCY	ILIKRL	NDKSKFLIRNPAK	GLLGLWEFF	TYGWLHSDH	DDYVK	273				
c_000002018097_4	221	KPIPT	KEYVA	VFIS	YNDHIFINQRPED	GLLGLWEFF	MLIEVFNSN	KIKALKT	283				
c_000002843512_36	213	VVPT	REQNI	MLYF	YNDKLSLTQRTG	KFLHGLWEFF	ALIEVPL		253				
c_000002930199_1	220	RKPKI	RYFNY	LFIK	KNQQLFQRRGTN	DIWNLWEFF	LIESKEKLN	REKIMA	280				
c_000003033795_2	214	KKIPI	YDIAV	SVVE	YNKILLITRKNK	NFLPGLWEFF	GKKIEKKET	ATQAIR	276				
c_000003888107_2	215	KIITP	IKVST	CIIR	KQKLLILLQRPDL	TMLGGLWEFF	GKKIQNNEE	RKGAIVR	277				
c_000004013286_6	242	PKTQA	VREVA	VIVR	KGKRVLLVRRPSPD	GLLGLWEFF	RPECFSRKSKE	KIRGLMVR	306				
c_000004852258_1	227	KIITP	KTIAA	ALVN	HGDNIFTTRKPLK	GLLGLWEFF	NIELVNGEI	PEDLKI	264				
c_000005254087_1	230	KTRPHYDVAGITWQNG		VSPQ	EGRFLLAQRPDL	GLLGLWEFF	GKQRPED	LPQALRR	299				
c_000005063677_1	207	LKIKK	RYFYV	FWRG	TNGQYLLRKRQEK	DIWGLYEFF	MLIELKPLK	LAHWELEE	289				
c_000003283462_3	213	KNIIPH	YDVVV	GLIIV	KEGRFLITRKEEN	KHLGLWEFF	GKKIENNER	SEALAIR	275				
c_00000687221_1	227	VSRPH	HNVAV	GLIIV	KDNRLILLSKRNAS	GLLGLWEFF	GKKIRSGES	GSSCVVR	289				
c_000003347358_19	244	VVPT	REENI	LVRV	YNDKLSLTQRTG	KFLHGLWEFF	SVEVPH		284				
c_000002701031_2	222	LKIRK	RYFEF	LIIN	ENNEKLLIQRPDL	GMGLYEFF	LIENSVKKE	RDIENSKS	285				
c_00000582753_3	230	KKSTN	RYFTY	YYIS	DSNHLITKRGVKG	GIWGLYEFF	LEENLIAFE	NEVLCK	291				
c_000001713769_5	222	NAPKI	LIHFN	LVL	DSDDHMLCKRIN	GIWNLWEFF	MIESKKEEN	KTVQLSN	287				
c_000004369364_1	225	LKIRK	RYFYV	LVFI	EKDLFLITRKEEN	DIWGLYEFF	NIEGMEELN	ESIEIREA	287				
c_000006057486_30	226	LKIKK	RHFNF	LIIN	NQKTYLITRKEEN	DIWNLWEFF	LIESKHEIDTN	LIKSSIF	291				
c_000005494072_10	222	LKIKK	RYFNY	LVLN	KKNKILLIRKRNK	GIWNLWEFF	LIEVETITIK	TLVEHYL	289				
c_000001742634_3	225	LKIKK	RFYNY	LILIT	KDNKTLIRKRNK	DIWNLWEFF	LIEVETITIK	TLVEHYL	289				
c_000002826998_2	225	LKIKK	RYFNY	LVLN	KKNKILLIRKRNK	GIWNLWEFF	LIEVETITIK	TLVEHYL	289				
c_000003159439_6	222	LKIKK	RYFNY	LVLN	KKNKILLIRKRNK	GIWNLWEFF	LIEVETITIK	TLVEHYL	289				
c_000004887214_1	227	LKIKK	RYFNY	LILIT	KNEETLVKRGK	GIWNLWEFF	LIEVETITIK	TLVEHYL	289				
c_000005037037_2	225	VKVKK	RFYFY	LVLN	INDTYITRKEEN	DIWNLWEFF	MIEVETITIK	TLVEHYL	289				
c_000001059964_1	217	KKVKI	LYRGK	AVIL	FDGGLICKRNPQ	GLLGLWEFF	SVELGRKDN	LRALQKE	279				
c_000002498472_1	226	KKIIP	IDMAT	GLII	HNGMLPIQRRPLD	DWVGSYEFF	GGRMELKET	PEQTVVR	288				
c_000004615912_3	215	IRKKK	RYFHY	FISHF	EKKILLIRKRNK	DIWNLWEFF	MLIELKPLK	LAHWELEE	289				
c_000004766858_2	228	PEITR	IKKIL	ALVQ	NGDCMLVRRKRPD	GLLGLWEFF	TWELPSSENS	AAAWLQ	291				
c_000002747260_18	232	KPKPT	RHGIA	YLGR	ADGAMLLERRPEK	GLLGLWEFF	GSDMAEAT	EAAPPD	285				
c_000005371561_38	232	KPKPT	RHGIA	YLGR	ADGAMLLERRPEK	GLLGLWEFF	GSDMAESP	EAAPPD	285				
c_00000169465_2	234	KALPE	KSTYM	MVAQ	FNSQVYLQRPST	GLWGLYEFF	EVSSIEEGE	QKARGVN	289				
c_000003254110_11	234	KALPE	KATFM	MVAQ	FNSQVYLQRPST	GLWGLYEFF	EVSSIEEGE	QKARGVN	289				
c_000004750284_20	230	KVTPV	QRTYM	LIPM	FGQVFLQRPSPD	GIWGLYEFF	EAQSAEEGIS	LQAQRTG	285				
c_000000141782_15	235	REKPL	RRTRM	LILVD	GGVFLVRRPSPS	GIWGLWEFF	ECPEVDVES	YCRHELGE	291				
c_000002717847_8	224	KKD	VYLEF	SLYK	DGPKYFLAQTEAL	GFWKLMWPF	VKVIN		263				
c_000004054799_1	215	QKPT	VKLN	FILPH	TKNEILMHKQAS	EYEWELWPI	DGDEVKIK		259				
c_000006126673_1	241	KARPQ	RYGYV	YIWD	DGNNILVGRPEK	GLLGLWEFF	TSSWETDIE	TVEH	301				
c_00000339186_3	199	IRYED	KQIDV	YLIT	RGRPYLVQRPSPD	GLWGLWEFF	NSDSGNKYS	APK	297				
c_000003029168_2	230	VPLQSRKE	VGVIVWAG	KPIN	ADSLITVQRPSPD	GLWAGWEFF	HDEQTDPE	HNAALR	248				
c_000001863436_1	246	ARQKP	VAARTAVTSSGA		EGRRFLVRRPAD	GLLGLWEFF	TVEAAGDAV	PTQAQVS	315				
c_000002830137_4	239	QTEH	VVEAA	VVIR	RGSRVLLRREDR	ERWAGLWEFF	RFPFDPDATN	KSRLETGPE	303				
c_000004612302_3	232	VRTD	LYQLL	LIVR	RDRKLLIQHPER	ERWAGLWEFF	ALSTRGPLLKS	RIDEGLIA	297				
c_000001933926_1	217	VARPT	RHGVV	FWATR	GDGCVLLRREDR	GLLGLWEFF	STEMRECPW	RIRSAV	276				
c_000003136334_1	227	KKRPT	RYATV	FWLLD	GRGNVLLRREDR	GLLGLWEFF	STDWLEENT	PFKENS	286				
c_000005849454_9	234	KQWQ	LEMFV	LVRH	VQVYVLLQKRNK	GWSPLYEFF	SAICESPATANE	ASSAHSTAH	299				
c_000001169194_1	234	VVARE	ISFLL	VILQT	EEGEVMLVRRPSPD	GLLAGWAF	EQLAKPLD	CAATSRD	301				
c_000002786947_2	234	APARE	IFPFL	VILQT	EEGEVMLVRRPSPD	GLLAGWAF	EQLAKPLD	CAATSRD	301				
c_000001279808_52	230	KKTKRKRHR	IEVGI	ACIW	REGKYLVRQRPSPD	KSPFEGSWEFF	GKKREKES	FRGCVKR	287				
c_000000358065_2	224	KKIIN	KFYLA	TLYK	HDDQVLLITRKEEN	NFLKNLIFP	MKEISQS		268				
c_000003745941_1	218	KIKQE	TLIDRA	WVEQ	GGKLLIRKRNK	RRLSGLEL	TLDMGY	SAT	268				
c_000004187032_4	224	KKKNE	KYFIL	KVYK	KNQKYLILVRRPSPD	NFLKNLIFP	MEELSKP		267				
c_000006067315_3	224	KFNKI	KYFEA	NIYQ	NQNKYLLIRKRNK	RFLKNLIFP	MEELSKP		267				
c_000005543774_1	186	TESKT	YDIFC	YLK	NKQKIALTKNNDL	GFLKFNLP	IKATSKK		230				
c_000006211484_1	208	TSKKN	YDIFC	YLK	NKQKIALTKNNDL	GFLKFNLP	IKATSKK		230				
c_000002718976_3	227	IKSKN	YDIFC	YLK	NKQKIALTKNNDL	GFLKFNLP	IKATSKK		230				
c_000002992548_1	226	KKEQK	YNYVC	YLK	KKKEIALTKNNDL	SFLSNFNP	TKIETSNKK		273				
c_000000456751_1	192	IKKEK	INIVC	YLK	KKKEIALTKNNDL	SFLSNFNP	TKIETSNKK		273				
c_000002655634_1	223	TKKKK	FNIVC	YLK	KKKEIALTKNNDL	SFLSNFNP	TKIETSNKK		273				
c_000004232403_2	223	KKIPT	RNLAA	ALIE	YGDYLLITRKEEN	GLLGLWEFF	NIELANGKS	PEKSLKE	285				
6u7t_after_N146.pdb	226	TVKQ	VPLAV	AVLAD	DEGRVLIRKRDST	GLLGLWEFF	SCETDGADG	KEKLEQ	283				
Consensus_aa:					h	hh	phh	pp	shh	shh	hp	p	

Conservation:	282	DNLTLQTAFT	SHFH	LDIVPMLEF	SSFTG	CMDEGNA	LWYNLAQPPSVG	LAAPVERLLQQLR	TGAPV	350
sp_p17802_MUTY_ECOL	295	ELTEPVSFE	SHLV	WQTVTFGR	VHGG	FVEEY	RLAPEDELKAYA	FVPSHQVWREYK	EWASGVRRPD	366
sp_p83847_MUTY_GEOS	287	DWRLARGEVR	THFH	LILIRVAEL	PDDI	TPAVG	FLLSKHAFRPSD	LPTV		338
c_000003652391_2	287	ELIGHADRFQ	IAITRK	LFIHLVRY	QGE	IPAGV	ETIPIDGDR	LPAHRRKIAAAR	ELLLEVFR	352
c_000000031207_6	294	DKMHPGRVY	SHFH	LELKIMTINS	EVT	PEGT	FWCRKEEVKAQA	LPSLMKIMTHMY	KA	354
c_000000430975_4	288	EFQWMDDEFQ	GVTRF	LHLQPIVAMW	RSGSW	EGPVA	EWVGIDELATRP	LSVAGRRVATRE	RILRHNGAQA	370
c_000001007148_4	303	PWPC	GTVTIT	THFH	LALSIVYARV	ADMP	APAGS	WWCAPGLAGEA	LPTVMCKVVEAI	372
c_000001092848_3	301	DKWQPGLVY	SHFH	LELKI						273
c_000001151029_1	313	RWRLPGLVR	SHFH	LELTVFARV	ALATP	APATG	RFTPRSLADDEP	LPGLMRKVLAAHAF	DPKPEPEKKRP	385
c_000001797282_1	229	DWQDPGAEVR	SHFH	LRLSLRISV	GNA	KPTAG	HFADADFDPE	LPTVMRRKAYRIAR	TFDFGDANG	358
c_000001803648_25	280	DWQVLAENVR	THFH	LRLSLRVAT	GNA	PPATG	DFISKNADFPET	LPTVMRRKAYRIAR	TQFDND	355
c_000002078955_4	287	TKNKIGEIR	SHFH	ITLSPGHCR	INGSK	FEVNRPF	RWIRLINEINQYA	FPKANHKLFINMQ	E	351

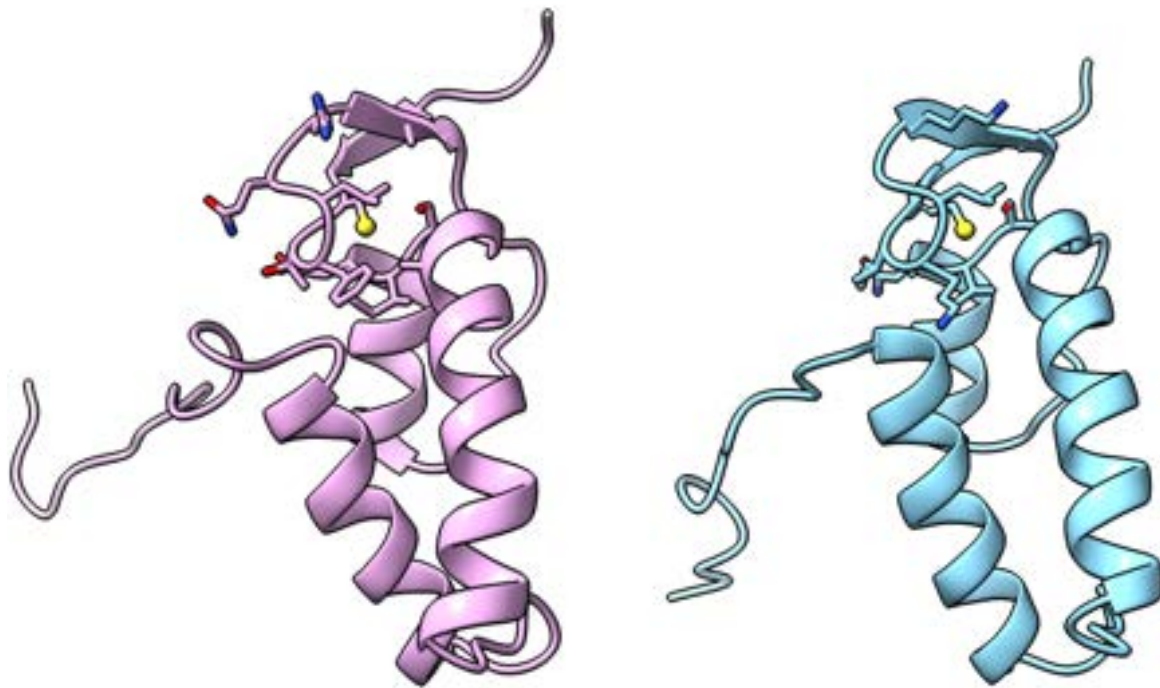
c_000002826998_2 290 QIKKISLPPFRNLTIRK---IIATFFFEIDF---MDKFS-----NKNNNY-----IKIERKNLSKFA--FPKIIDLYLKDNS--LNLAMK----- 358
c_000003159439_6 289 TSLYSDNIKAHKLSHQH--IFSKEWIVET---KQNIL-----NSILIEDLEKYP--VPKLIQNFIEKFD--NIKLNPAK-- 354
c_000004887214_1 294 LQLLTTPKVVVVKLSHQH--LYIKFWEIRV--PTFRS-----ATISWQELKLPL--FPVIVFKFIREFL--ETSMNPFDTPL 361
c_000005037037_2 288 KVKGRSKNFTIRLTHQK--LNAIFIEIDL--KTK-----INRQF-----INTDINNLSKFA--FP-- 336
c_000001059964_1 280 QLGKEIHTLEIRLTHFR--TRVSFFTATV--HVLPE-----SSQKRV--RFVTKSEVE----- 326
c_000002498472_1 289 EIVEKITNTVHTTHYK--VTLHCFLRL--KKGSDT-----DPVLHAAQDF--NWVPFKALQEYA--FPAGHRKLIITYL-- 355
c_000004615912_3 285 QTKK---YQELTHQK--ISAYFWEIN--LNVAP-----VNIYDF--ILIDSKNLAKFA--FPKIIALYLENNS--LNL-- 345
c_000004766858_2 292 RIGPELTFMEIDTRFH--ATLHCRLCTV--AEIPS-----VLPDPA--QWLTLGEISALA--LPRVFQRLKRLL--DEVSA-- 359
c_000002747260_18 286 FWRDPGAEVRTHTHFH--LRLALRVADL--PASA-----TFPTG--QFIPHHAFRPGQ--LPTVMRKAYDLAS--ATFLGN-- 352
c_000005371561_38 287 NWRVTPGEVRTHTHFH--LVLELRADL--FEDC-----TMRG--QFLAPGFRPSD--LPTAMRKAFDLAR--DG-- 349
c_000000169465_2 290 DETRTLEGFRTHTHFH--LDITPVAVV--NSAPS-----KRVAETAF--RWFSLNEPIEVG--LAAPTTHIIQQLM--R-- 355
c_000003254110_11 290 EETKTLETFRTHTHFH--LDITPVAVV--NSPPT-----KRVADTAF--RWFSLGEPIEVG--LAAPTTHIIQQLI--G-- 355
c_000004750284_20 286 ERIQQLEFRTHTHFH--LDITPLIAVV--NSTPQ-----KRVAENES--RWFLDLDEPIEVG--LAAPTTHIIKTILA--K-- 351
c_000000141782_15 292 TRPQAMPGLRTHTHFH--LDIQPVRLRL--TAAPA-----RLMDGDRH--VWYNVESPDARG--LAAPVARLLQGLA--LEDTRA-- 362
c_000002717847_8 264 ---KFNANKKIALSHRH--IHFREFKQIAN--LPT-----GMKG--EWFSSKDDLANIA--LPKPIDSKLLQDG-- 319
c_000004054799_1 260 NQKYSKIKVNPISHLN--LDMKIKTFKI--DKKCD-----LDSN--L-- 296
c_000006126673_1 302 KPYKQLFFIMVPTHFS--LKLYLCKTSE--FICDN-----VPEEH--KWVFISTLEDIG--FPSVFEKVKKII-- 362
c_000000339186_3 248 -NQPPFASARISITTYR--IQLNAHLIHA-----KRGKG--MWCTSDERSLP--FSSAHLKLLGKLP--HRLDK-- 308
c_000003029168_2 299 QKLTKLVTITITMTR--TTLHVFEGLM--DLETS-----TMQEASSR--LWVRLDGLHKHP--FSSPQHEIITKLD--KEAALQLAGLV 374
c_000001863436_1 316 TPRVELGEVVRHSHIR--MTITAEHRVA--SSEQVRL--RSDDEVGQEA--RWVSAREVLDLP--MSAAMRKIEGLYR--ELSGAI-- 391
c_000002830137_4 304 EPVRQLTTIQRGVTRFQ--ITLQCHLAQC--VSGRK--RGPHL--SWRMPEHEHLPL--LSVTARKIGRLVA--KLAD-- 369
c_000004612302_3 298 EVEEWLQEFRUGVTRFR--LHLQPVVASW--RAGHW-----IGPVVA--EWEIDEILTARP--LSVAARRVADGVQ--AGQRGGASG-- 370
c_000001933926_1 277 AWRELPGAVRISHTFH--LELQIMTVET--NEEVEG----- 309
c_000003136334_1 289 DWQMLSGKVRITTHFH--LEL----- 308
c_000005849454_9 300 KLGAYQQQVRITTHHR--IRAHVYAAEG-----Q-----APTHLEDPTQVP--LTGLARKVLDRI--SA-- 354
c_000001169194_1 302 GVASALAEVQRTHLQ--ARYRPWVVPV--AKL-----LNNGN--VMPTPGEPVDFP--IPVAQHVKLDALA--ECGATLGSSSES 373
c_000002786947_2 302 GVASALAEVQRTHLQ--ARYRPWVVPV--AKL-----LTGEGS--MMPTPGEPVDFP--IPAAQHVKLDALA--ECRATPDLSLES 373
c_000001279808_52 288 SVRSHFYEELEETE--LLAFHRCQI--QAGEP--QALENQL--KWSPEDFDSVG--FLAKNHKALEKL--SMRV-- 356
c_000000358065_2 269 SLSPKPDHKLKKNMNM--MNISIDFSNN--KKK-----LNKG--LWVEKTKLENHM--IPTPTKKIFASVK--HNL-- 330
c_000003745941_1 269 RGNTPFIVTRKGLNSQ--ITEPIWRLSD--NSEA-----DHAEM--EWIDLEQLIKVT--LSGPHRGRWINELL--TDFNRQK-- 336
c_000004187032_4 268 --KFNENFNKKNMNM--MNIKIQLKLN--LKK-----FPSS--YWFDKRRLDNYM--LPTPTKKVFNKYLE--KN-- 326
c_000006067315_3 269 FNTSLNKNKKNMMD--MKILNKNK--LKK-----LNKS--YLVDKRRLDNYM--LPTPTKKVFNKYLE--KN-- 330
c_000005543774_1 231 KKNWFLCNKYKNSNKK--LNINLYYKFS--SKI-----PSKY--NWYSL--KKNKEF--IPTPTKKIFKQIT--HLY-- 291
c_000006211484_1 253 KKNWFLCNKYKNSNKK--LNINLYYKFT--SKI-----PSEY--NWYSL--KKNKEF--IPSTFK-- 303
c_000002718976_3 273 QNWIFLCNKYKNSNKK--LNINLYYKFS--KKI-----PRSF--NWYSL--KKNKEF--IPSTFKRFRQVS--TLF-- 333
c_000002992548_1 274 NGWYILCNKYKNSNKK--MNINLYYKFT--KNK-----PKKY--MWYPIDRRSNEF--IPSTFK-- 325
c_00000456751_1 241 RGWYILCNKYKNSNKK--MNINLYYKFT--KKI-----PKKF--NWYSIDRRSNEF--IPSTFKKIFKKIA--KVYS-- 303
c_000002655634_1 272 REWNYILCNKYKNSNKK--MNINLYYKFT--KNK-----PKKF--DWYSVDKSSQEF--MPSFTKKII-- 326
c_000004232403_2 286 EVGNIGLIQHTTHFH--MNITLFCCLL--KYDQ-----KIDSSA--KWVRLFELDQYA--FSRGNHKLFLQYQ--KENV-- 351
c_000001933926_1 284 ELTEPIVSFEIASHLV--WQLTVFPGRLL--VHGG-----PVEEPP--RLAPEDELKAYA--FPVSHQRVWREYK--EWAS-- 349
Consensus_aa:hp+.ho+bp...l.hp.h.h.....@h..pph.phs..hs..p+hhp.h.....

1139
1140

S2 Table. Percent identity matrix.

	Gs MutY	LCHF MutY 4	LCHF MutY 1	LCHF MutY 2	Ec MutY	LCHF MutY 3
Gs MutY	100	37.5	31.2	31.2	38.0	33.3
LCHF MutY 4	37.5	100	34.4	33.4	33.3	33.7
LCHF MutY 1	31.2	34.4	100	65.5	37.0	34.8
LCHF MutY 2	31.2	33.4	65.5	100	39.9	34.8
Ec MutY	38.0	33.3	37.0	39.9	100	48.3
LCHF MutY 3	33.3	33.7	34.8	34.8	48.3	100

We were interested in determining how similar the amino sequences of LCHF MutY representatives were to existing MutY enzymes. We visualized this in the form of a percent identity matrix that was generated by *Clustal Omega*.



S3 Fig. MutY gene neighbors. Structural homology for *Ec* YggX (left) and nearest neighbor to *Thiotrichaceae* MutY (right). The structure solution NMR structure for *Ec* YggX (PDB ID 1yhd) (103) is superimposable to the structure predicted by *Colabfold* for the nearest neighbor to *Thiotrichaceae* MutY, with RMSD of 0.99 Å for 69 pruned pairs selected from 88 possible pairs. The Cys residue critical for function is highlighted with a yellow sphere for the sulfhydryl group.

1157 **S4 Table. Metabolic gene identification**

			<i>Marinosulfonomonas</i> MAG 1	MAG 2 ^a	<i>Rhodobac-</i> <i>teraceae</i>	<i>Thiotrich-</i> <i>aceae</i>	<i>Flavobac-</i> <i>teriaceae</i>
Cytochrome Oxidases	UQCRFS1	K00411	X	X	X	X	
	coxA	K02274			X	X	
	ccoN	K00404	X	X	X		
	cydA	K00425		X			
	cyoB	K02298			X		
Aerobic CODH	coxS ^b	K03518					
	coxM ^b	K03519					
	coxL ^b	K03520					
Methanogenesis	mcrA ^b	K00399					
	mcrB ^b	K00401					
	mcrG ^b	K00402					
Methane Oxidation or Nitrification	pmoA ^b	K10944					
	pmoB ^b	K10945					
	pmoC ^b	K10946					
Sulfur Oxidation	soxA	K17222	X	X	X		
	soxX	K17223	X	X	X		
	soxB	K17224	X	X	X		
	soxC	K17225	X	X	X		
	soxY	K17226	X		X		
	soxZ	K17227	X		X		
General Nitrogen Metabolism	narG	K00370	X	X	X		
	narH	K00371	X	X	X		
Dissimilatory Nitrate Reduction	nirB	K00362		X	X	X	
	nirD	K00363		X		X	
	nrfA ^b	K03385					
Denitrification	nirK	K00368			X		
	norB	K04561			X		
	norC	K02305			X		
	nosZ	K00376			X		
MAG Completeness (%) ^c			88.4	88.2	93.7	66.1	44.3
MAG Contamination (%) ^c			16.4	0.6	1.4	11.8	1.6

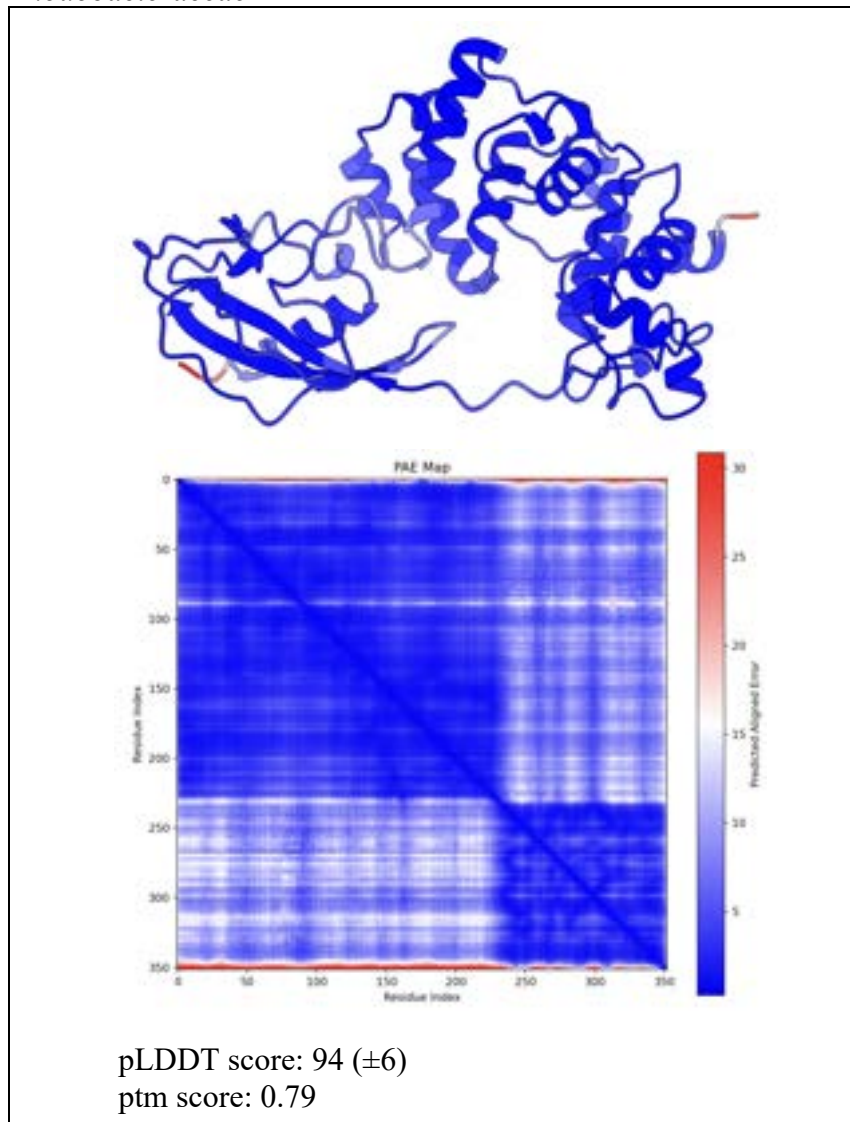
^a *Marinosulfonomonas* MutY contig belongs to two separate MAGs and each are reported separately as MAG 1 and MAG 2, respectively.

^b KEGG ID gene not identified in any MAG and not reported in Table 2.

^c Completeness and contamination scores generated by CheckM v1.0.5 as described in Brazelton et al 2022 (35).

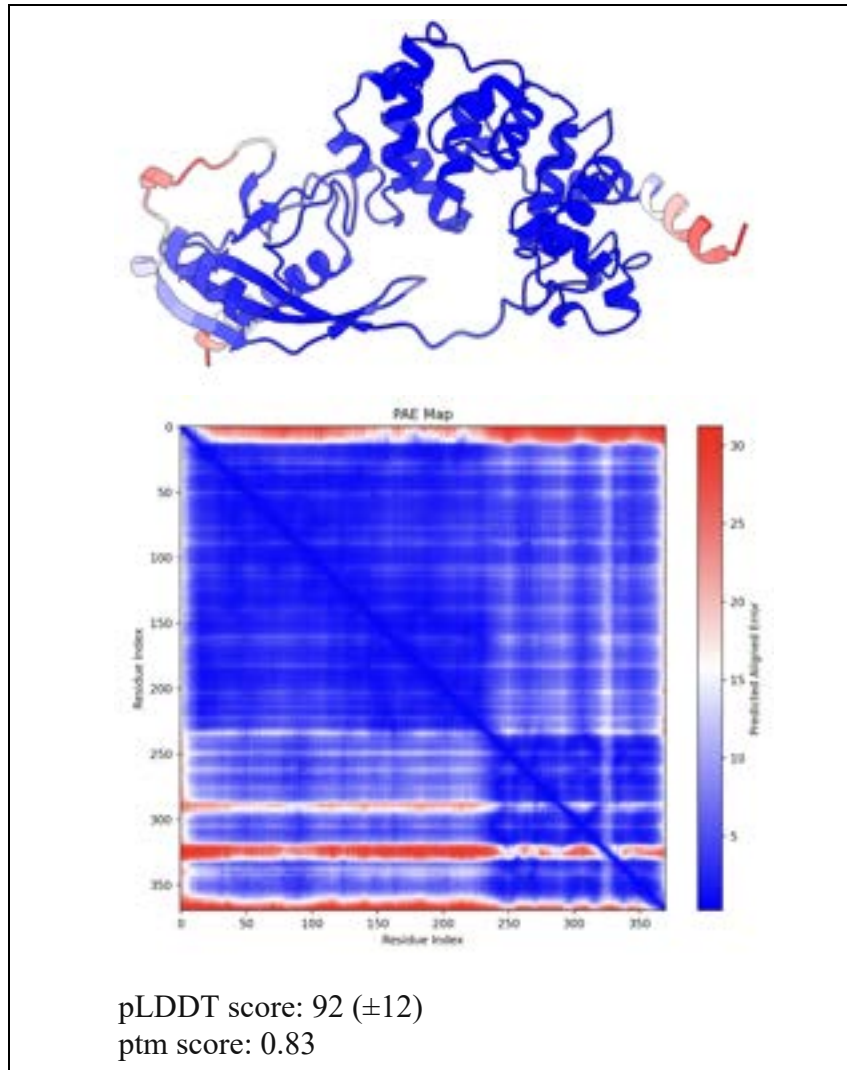
A KEGG ID analysis was used to identify the potential metabolic strategies of the MutY encoding organisms at the LCHF. The full metabolic KEGG ID search is shown above.

1171 *Rhodobacteraceae* MutY Structure Prediction



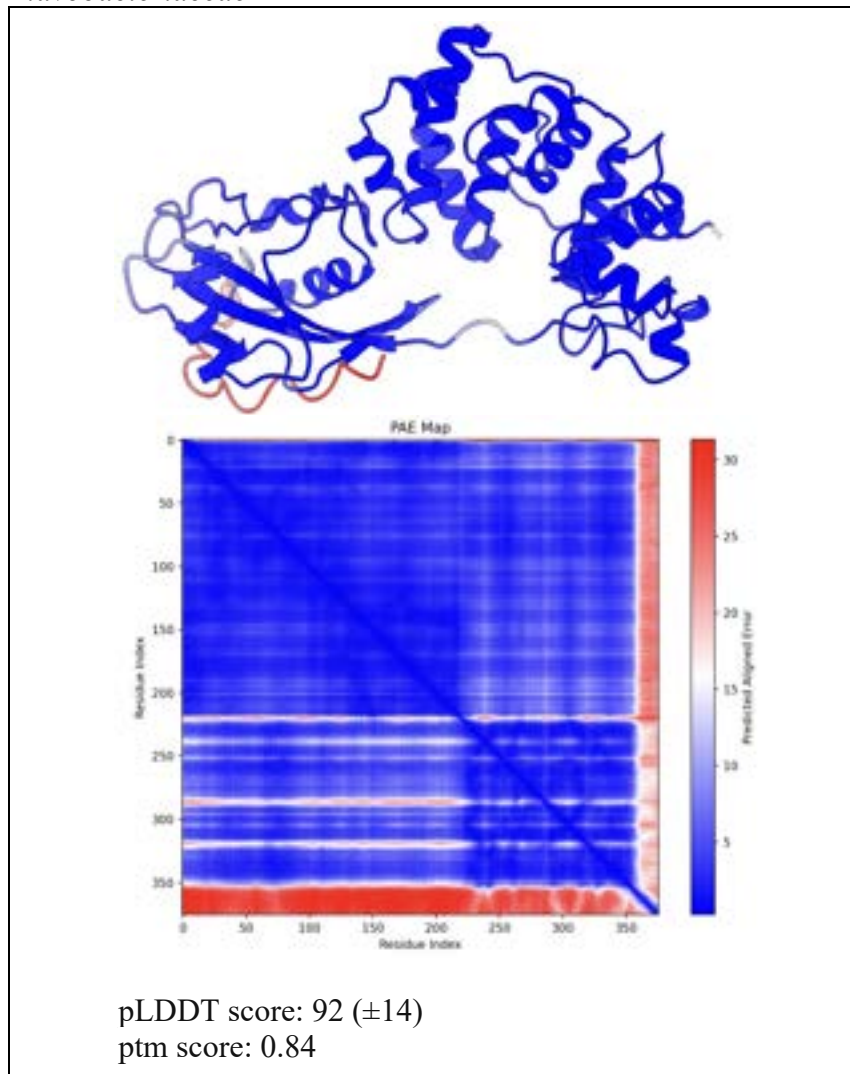
1172

1173 *Thiotrichaceae* MutY Structure Prediction



1174

1175 *Flavobacteriaceae* MutY Structure Prediction



1176

S6 Table. Ligand binding affinity (kcal / mol)

Adenosine	Gs MutY (3g0q)	<i>Marinosulfonomonas</i> MutY	<i>Rhodobacteraceae</i> MutY	<i>Thiotrichaceae</i> MutY	<i>Flavobacteriaceae</i> MutY
Mode 1	-7.3	-6.8	-6.8	-7.0	-7.2
Mode 2	-6.9	-6.3	-6.5	-6.9	-7.1
Mode 3	-6.8	-6.2	-6.3	-6.3	-7.0
Mode 4	-6.7	-6.2	-6.3	-6.3	-6.9
Mode 5	-6.7	-6.2	-6.1	-6.3	-6.8
Mode 6	-6.7	-6.0	-6.0	-6.3	-6.6
Mode 7	-6.6	-5.9	-6.0	-6.2	6.5
Mode 8	-6.5	-5.8	-6.0	-6.2	-6.4
Mode 9	-6.5	-5.7	-5.8	-6.1	-6.3

OG	Gs MutY (3g0q)	<i>Marinosulfonomonas</i> MutY	<i>Rhodobacteraceae</i> MutY	<i>Thiotrichaceae</i> MutY	<i>Flavobacteriaceae</i> MutY
Mode 1	-7.7	-7.5	-7.7	-8.0	-8.0
Mode 2	-6.9	-6.8	-7.6	-7.5	-7.5
Mode 3	-6.9	-6.8	-7.5	-7.3	-7.4
Mode 4	-6.8	-6.8	-7.3	-7.0	-7.3
Mode 5	-6.7	-6.7	-7.1	-7.0	-7.2
Mode 6	-6.6	-6.7	-7.1	-6.8	-7.2
Mode 7	-6.5	-6.4	-7.1	-6.8	-7.2
Mode 8	-6.5	-6.2	-7.0	-6.6	-7.1
Mode 9	-6.3	-6.2	-6.9	-6.6	-7.0

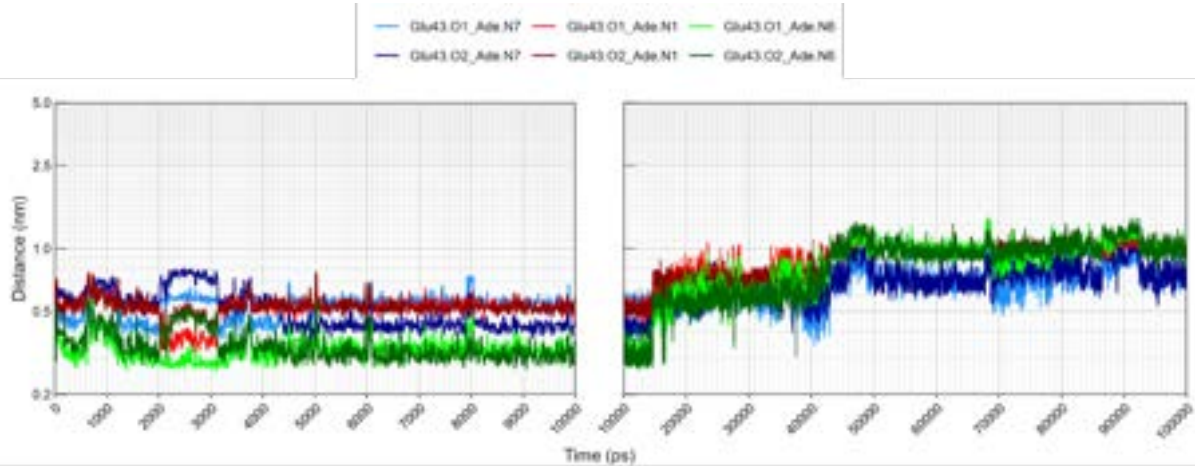
*Binding affinities are reported for the binding modes generated by *AutoDock VINA*. Each mode represents a predicted ligand pose, which differs by a combination of position, orientation, and rotamer conformation. The receptor structure was obtained from PDB ID 3g0q for *Gs* MutY and through structure prediction for the LCHF *Marinosulfonomonas* MutY, *Rhodobacteraceae* MutY, *Thiotrichaceae* MutY, and *Flavobacteriaceae* MutY. The binding mode representing the starting complex for molecular dynamics analysis is highlighted.

S7 Fig. Molecular dynamics. Molecular dynamics simulations were calculated by *GROMACS* with the Amber99SB and GAFF force fields for MutY complexed to adenosine and OG. For each MD simulation, short range interaction energies, distances between the ligand and functionally relevant residues, and representative structures are shown. Note that the Y-axis is logarithmic for distance. The adenosine and OG ligands are shown with all atoms wrapped in transparent surfaces. For the adenosine complexes, the protein structure was truncated so as to focus on the NTD (residues 8 - 220 in *Gs* MutY, and corresponding residues for the LCHF MutYs). Catalytic residues are shown: Glu43 and Asp144 in the *Gs* MutY protein and corresponding residues in the LCHF MutYs. The distance versus time plot for the adenosine complex, tracks potential contacts between the catalytic Glu (atoms OE1 and OE2) and the hydrogen bond donors and acceptors on adenosine (atoms N1, N6 and N7). For the OG complex, the iron-sulfur cluster domain and inter-domain linker were omitted so as to focus on the OG-recognition site found at the interface between NTD (residues 29-137 in *Gs* MutY) and CTD (residues 234-360 in *Gs* MutY). Residues that interact with OG are shown: Thr49, and Ser308 in the *Gs* MutY protein and corresponding residues in the LCHF MutYs. The distance versus time plot for the OG complex tracks potential contacts between the critical Ser/Thr residues and the hydrogen bond donors and acceptors on OG (atoms N1, N2, O6, N7 and O8). The total short range interaction energy (black trace) is the sum of short range Lennard-Jones (salmon trace) and Coulombic (sky blue) interaction energies. (A) Molecular dynamic simulation for *Gs* MutY NTD complexed with adenosine. The ligand complex persisted for the entire 100,000 ps, with changes in location and orientation evident at 16,000 ps and 42,000 ps in the distance plot. Hydrogen bonds between catalytic Glu43 and the Hoogsteen face of the adenine base were observed during the first 16,000 ps. These consistently involved direct contact with N6, as evidenced by close distance (green traces) and inspection of structures. N7 was also engaged (blue traces), with relevance for catalysis, via bridging water molecules (O red and H white). (B) Molecular dynamic simulation for *Gs* MutY complexed with OG. The ligand complex was stable for 92,000 ps, with the OG ligand bound to a cleft between the NTD (white) and CTD (gray). The functionally relevant hydrogen bond between the amide N of Ser308 and atom O8 of OG was frequently observed (not shown), sometimes accompanied by a second OG-specific hydrogen bond between the hydroxyl oxygen of Ser308 and atom N7 of OG (sky blue trace in the distance plot). (C) Molecular dynamic simulation for *Marinosulfonomonas* MutY NTD complexed with adenosine. In the first 3,000 ps, the adenine base approached closely catalytic Glu49 (green traces), often directly hydrogen bonded and occasionally bridged by a solvent molecule. However, the complex was relatively unstable, and the ligand departed the active site and found a new binding site by 8,000 ps. Favorable VDW interactions characterize both binding sites, but favorable Coulombic interactions are diminished substantially at the second site. (D) Molecular dynamic simulation for *Marinosulfonomonas* MutY complexed with OG. The initial ligand complex was unstable with a hinge-like motion creating new contacts between the NTD (white) and CTD (gray). After nearly escaping at ~4,000 ps, the OG ligand found several alternate sites on the NTD and CTD. (E) Molecular dynamic simulation for *Rhodobacteraceae* MutY NTD complexed with adenosine. The complex was relatively unstable. The adenine base initially hydrogen bonded with catalytic Glu45 during the first 3,800 ps but then changed orientation and drifted to a new site distinct and different from its original docking site. Note, catalytic E43 is not visible in the 10,000-ps representative structure as the new position of adenosine blocks its view. (F) Molecular dynamic simulation for *Rhodobacteraceae* MutY complexed with OG. The ligand complex was unstable and dissociated completely within 48 ns. Functionally relevant

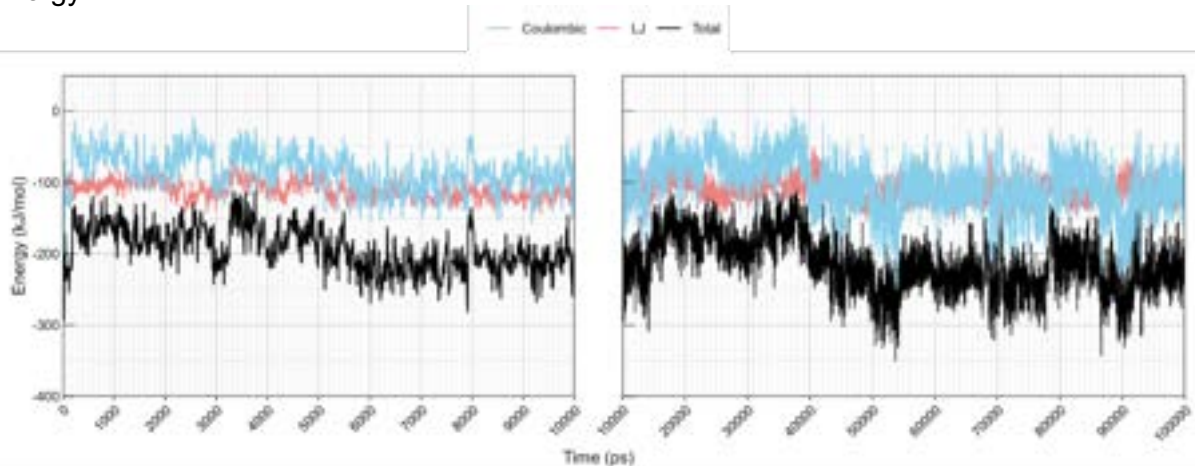
hydrogen bonds between Thr299 and OG observed for the initial starting structure were lost as the ligand moved to new positions on the NTD and CTD before dissociation. (G) Molecular dynamic simulation for *Thiotrichaceae* MutY NTD complexed with adenosine. Note that Ser replaces active site Tyr for this LCHF MutY, as is also the case for *Ec* MutY. The complex was relatively stable with the ligand persisting in the active site throughout the simulation. Hydrogen bonds between catalytic Glu46 and the Hoogsteen face of adenine were evident by close distance to N7 (blue traces) and N6 (green traces) and by inspection of structures. Water frequently bridged N7 to Glu46 as seen in the representative structure at 10,000 ps. (H) Molecular dynamic simulation for *Thiotrichaceae* MutY complexed with OG. The ligand complex was stable for the entire 100,000-ps simulation with the OG ligand bound to a cleft between the NTD (white) and CTD (gray). Hydrogen bonds between Ser306 and OG were frequently observed. (I) Molecular dynamic simulation for *Flavobacteriaceae* MutY NTD complexed with adenosine. The ligand persisted in the active site throughout the simulation, with the ligand periodically finding new orientations as evident in different distance traces vying for close approach to catalytic Glu43. For example, N7 of the adenine base was very close to Glu43 (blue trace) during the first 2,700 ps, suggesting catalytic engagement, but slipped out of reach at later time points. Water frequently bridged contacts between Glu43 and the adenine base. (J) Molecular dynamic simulation for *Flavobacteriaceae* MutY complexed with OG. The ligand complex was relatively stable. Functionally relevant hydrogen bonds between Ser305 and the Hoogsten face of OG can be inferred from recurring close distances up until 13,000 ps when the ligand adopts a new pose at the NTD-CTD interface.

(A) *Geobacillus stearothermophilus* MutY NTD complexed with adenosine

Distance Adenosine - Glu43

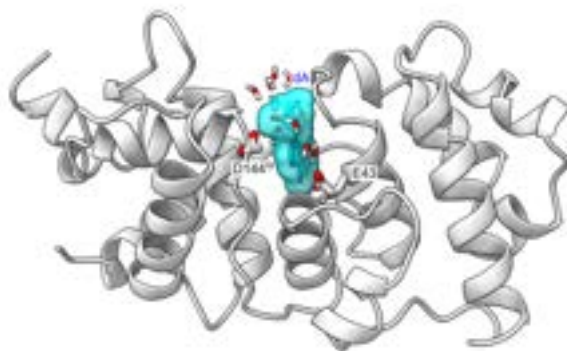


Energy



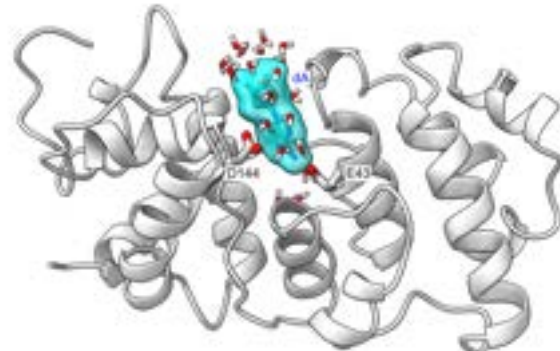
Adenosine and *G. stearothermophilus* MutY NTD

0,000 ps



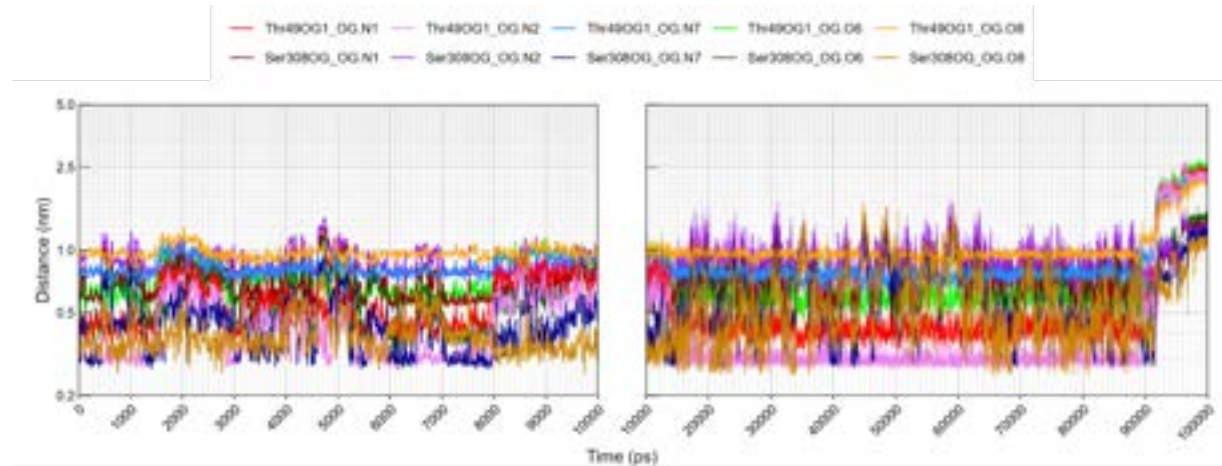
Adenosine and *G. stearothermophilus* MutY NTD

10,000 ps

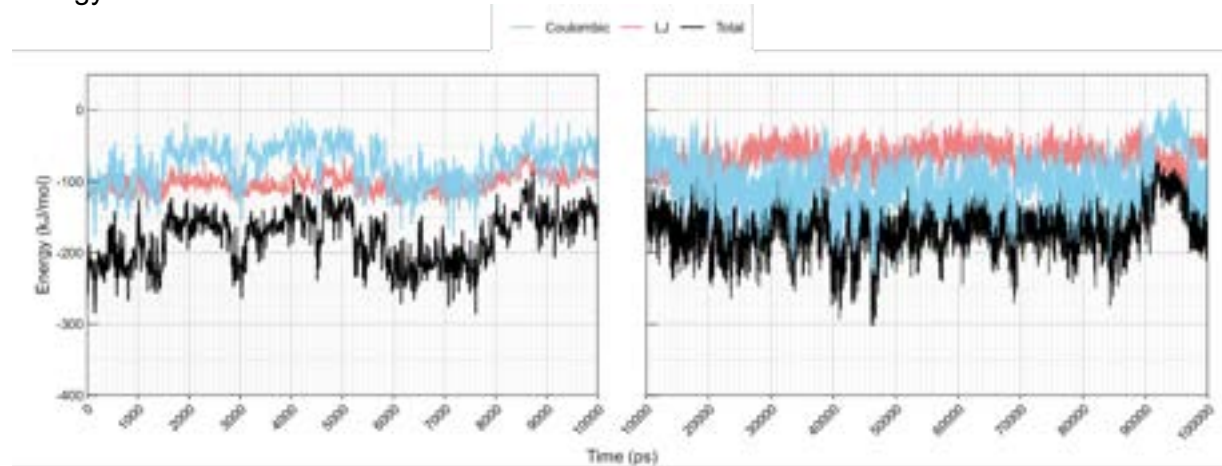


(B) *Geobacillus stearothermophilus* MutY complexed with OG

Distance OG - Thr49 and Ser308

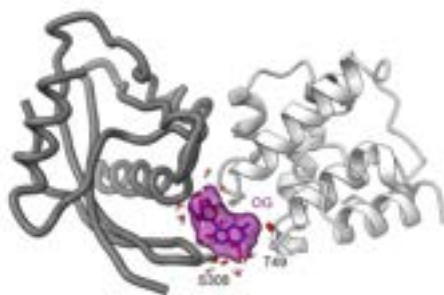


Energy



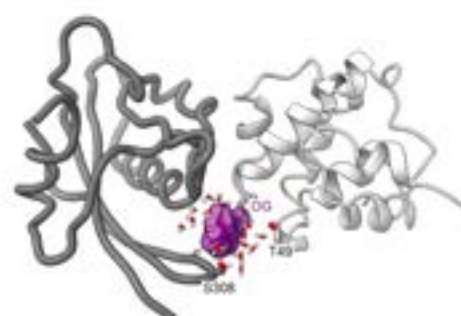
OG and *Geobacillus stearothermophilus* MutY

0,000 ps



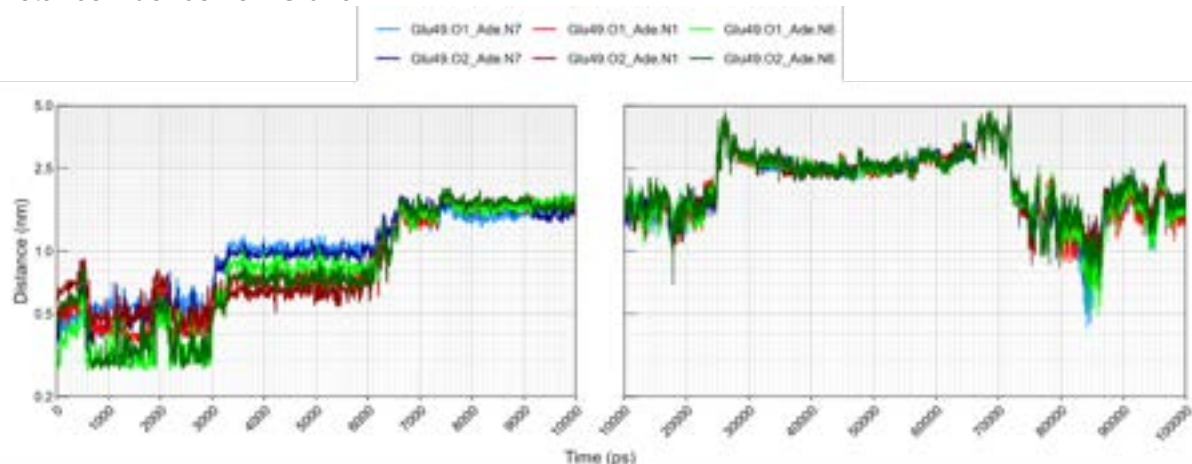
OG and *Geobacillus stearothermophilus* MutY

10,000 ps

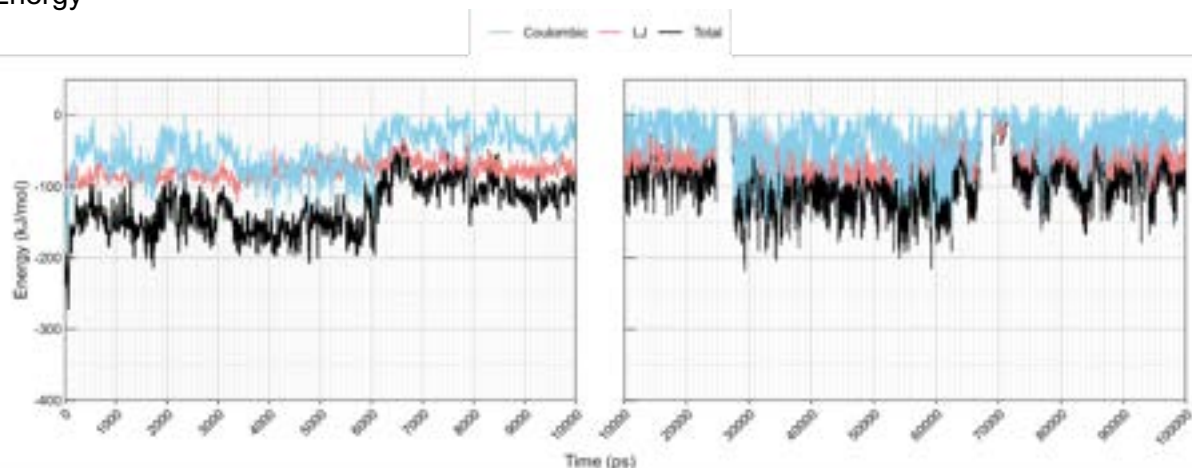


(C) *Marinosulfonomonas* MutY NTD complexed with adenosine

Distance Adenosine - Glu49

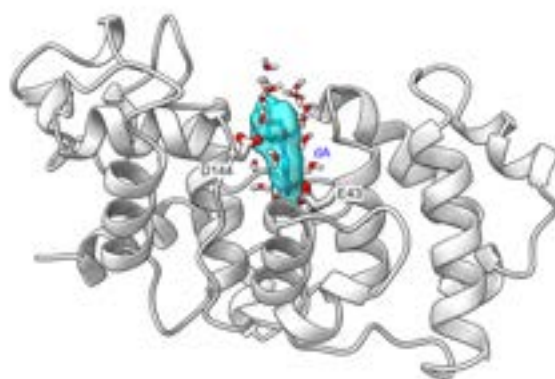


Energy



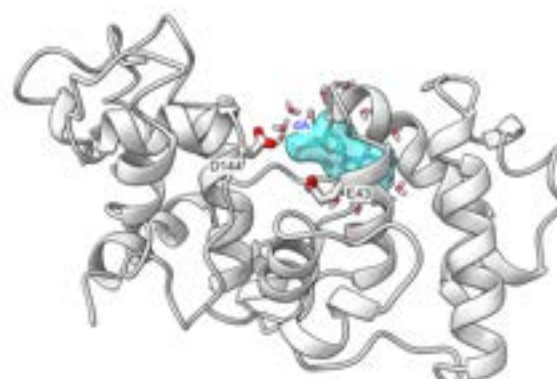
Adenosine and *Marinosulfonomonas* MutY NTD

0,000 ps



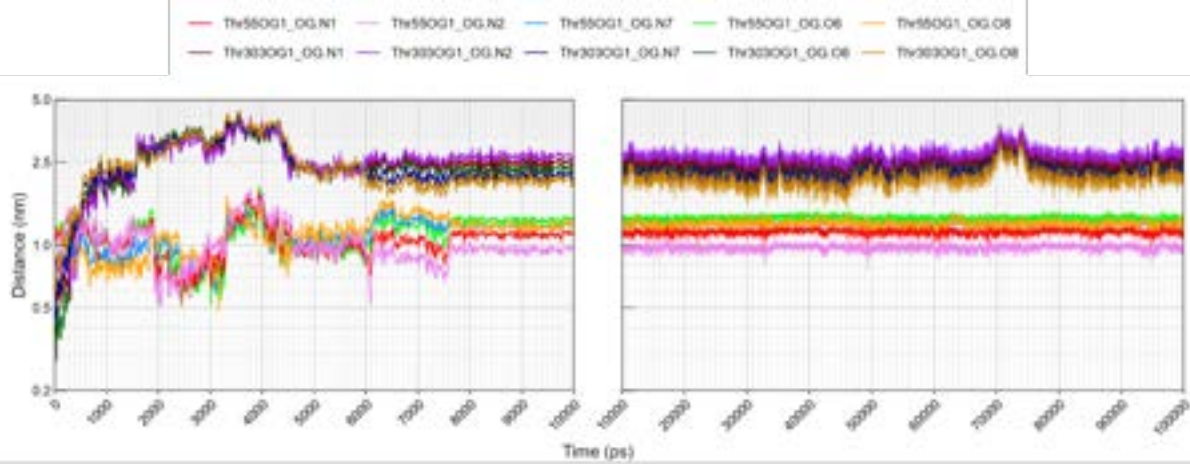
Adenosine and *Marinosulfonomonas* MutY NTD

10,000 ps

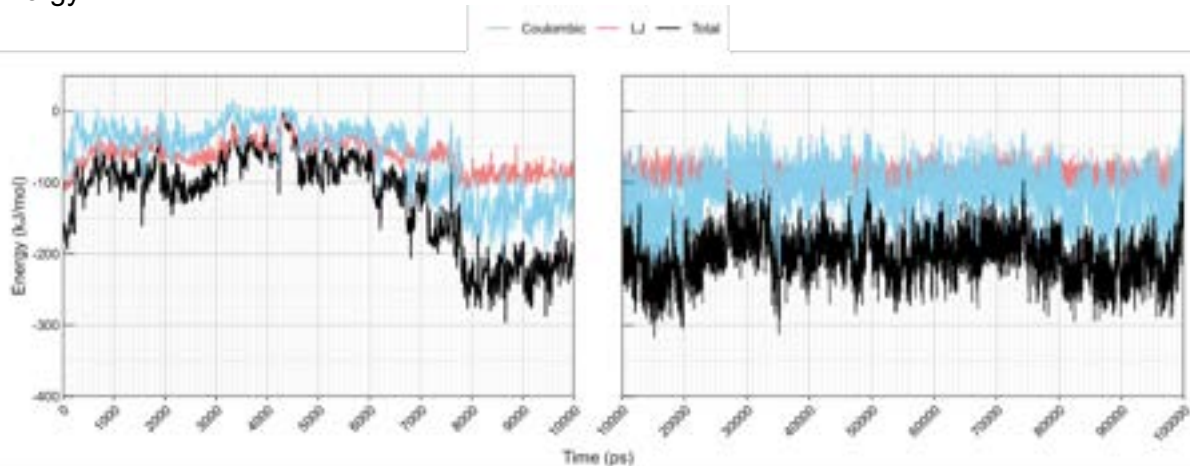


(D) *Marinosulfonomonas* MutY complexed with OG

Distance OG - Thr55 and Thr303

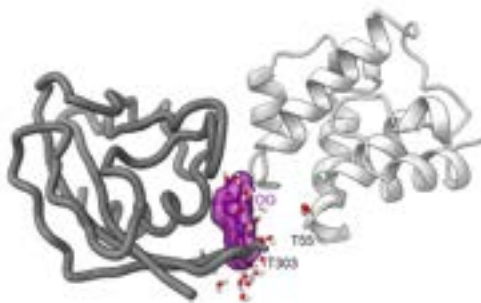


Energy



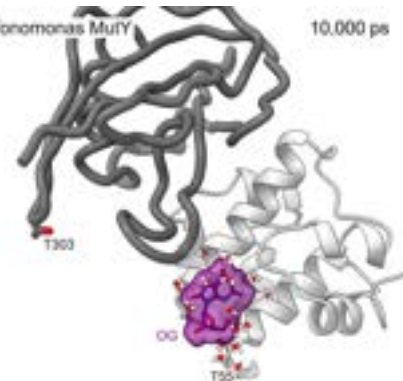
OG and *Marinosulfonomonas* MutY

0,000 ps



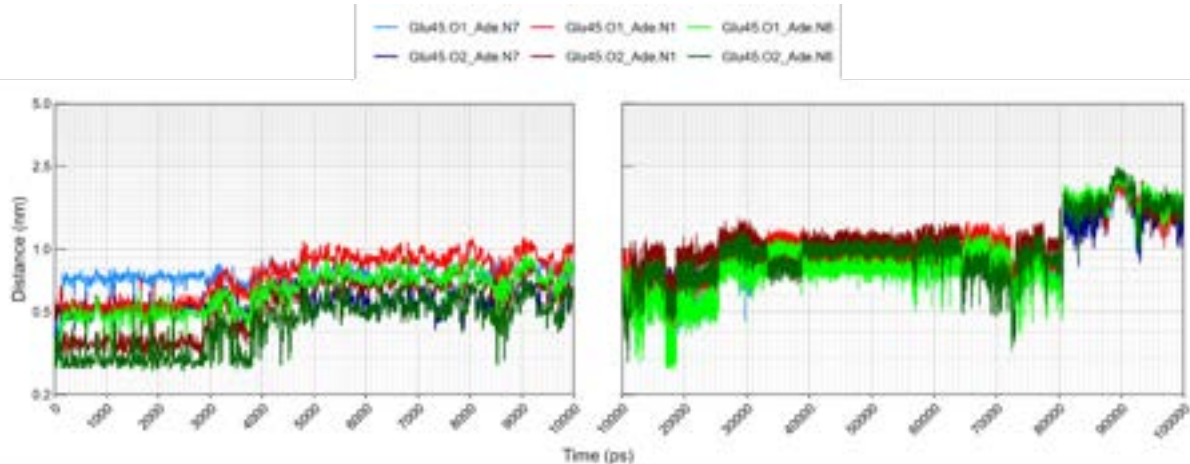
OG and *Marinosulfonomonas* MutY

10,000 ps

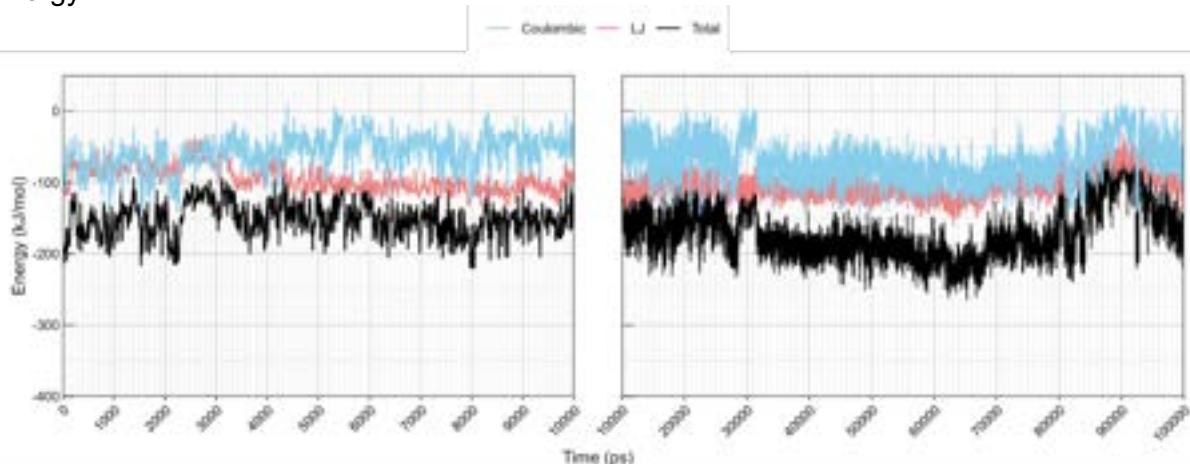


(E) *Rhodobacteraceae* MutY NTD complexed with adenosine

Distance Adenosine - Glu45

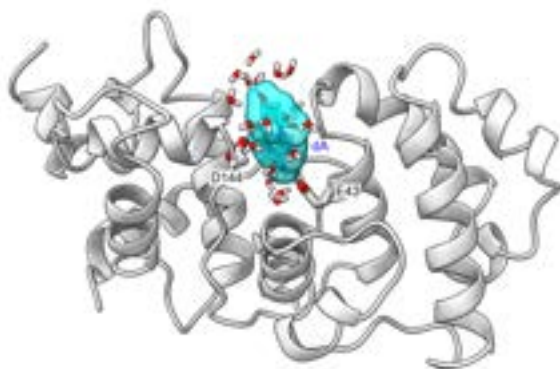


Energy



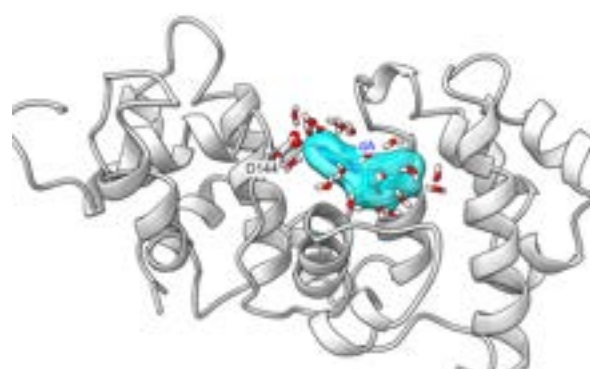
Adenosine and *Rhodobacteraceae* MutY NTD

0,000 ps



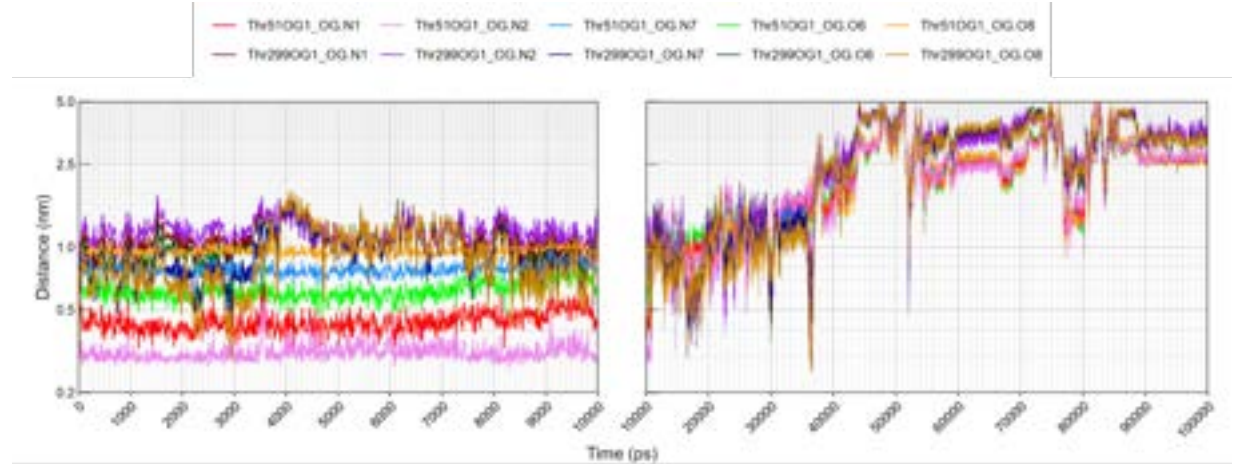
Adenosine and *Rhodobacteraceae* MutY NTD

10,000 ps

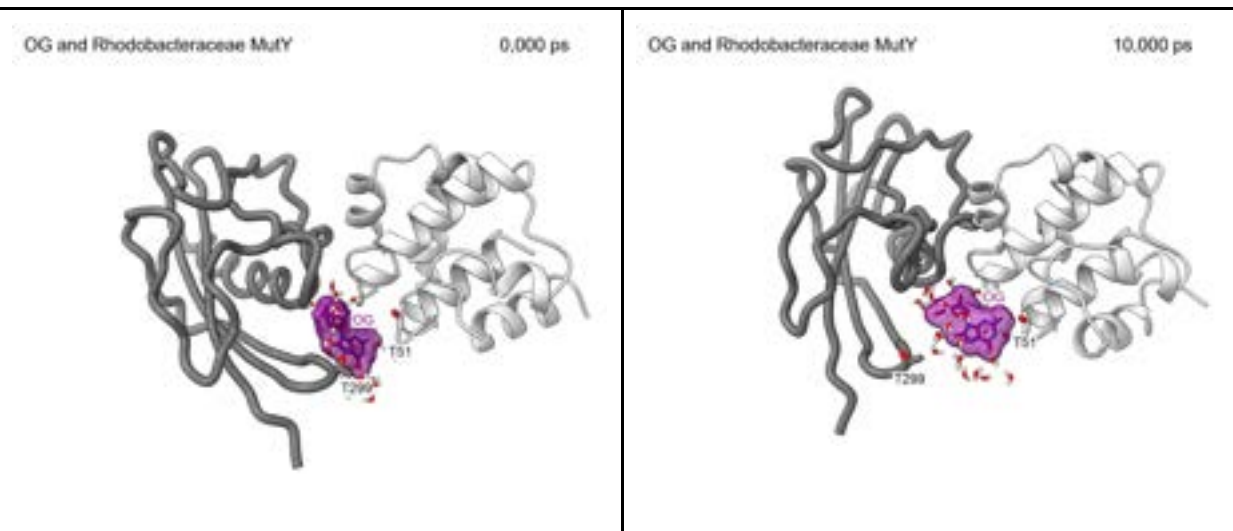
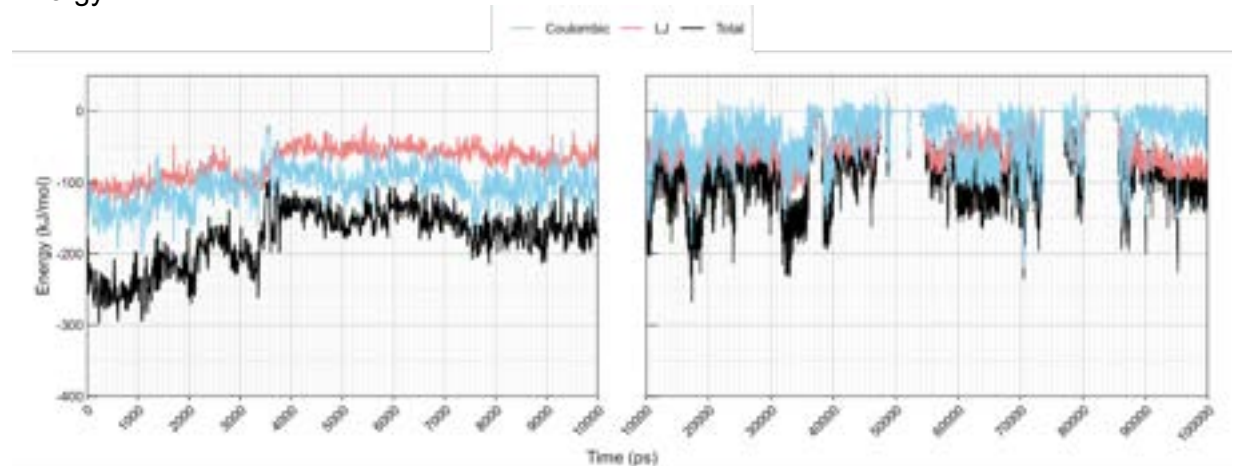


(F) *Rhodobacteraceae* MutY complexed with OG

Distance OG - Thr51 and Thr299

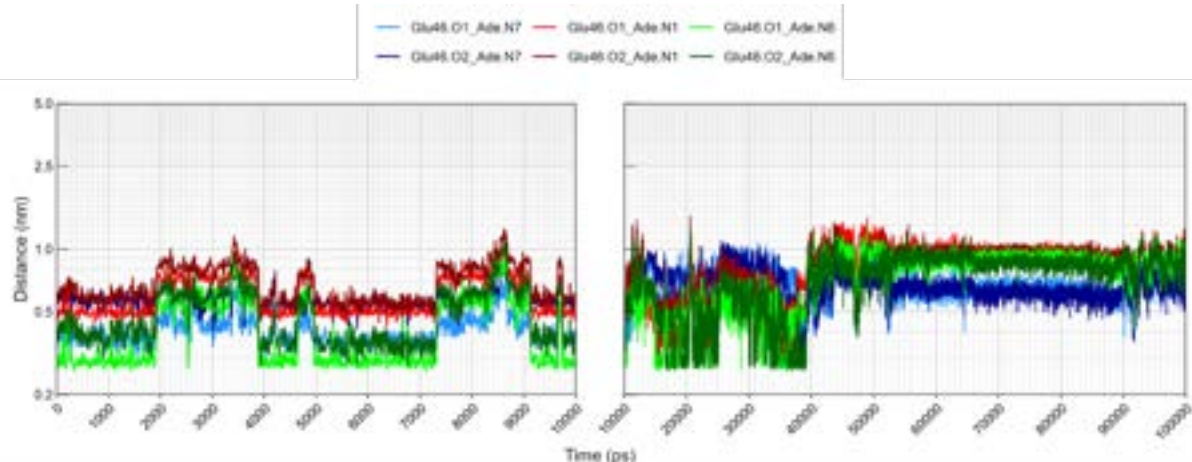


Energy

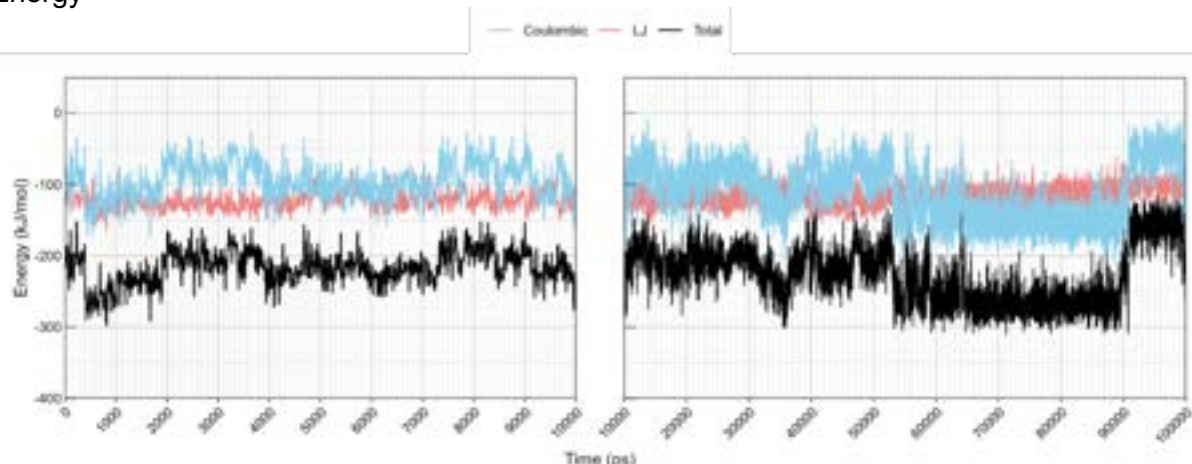


(G) *Thiotrichaceae* MutY NTD complexed with adenosine

Distance Adenosine - Glu46

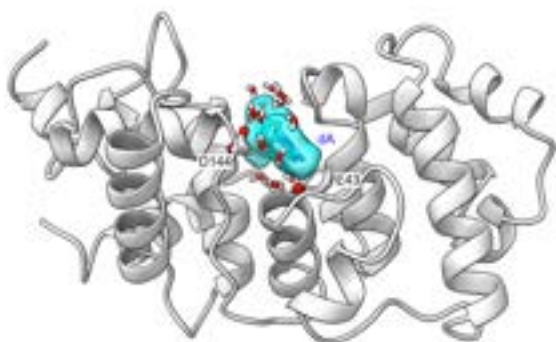


Energy



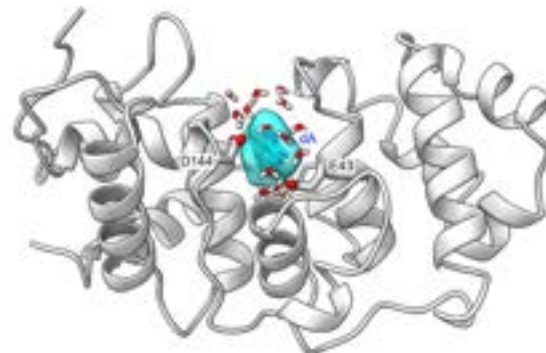
Adenosine and *Thiotrichaceae* MutY NTD

0,000 ps



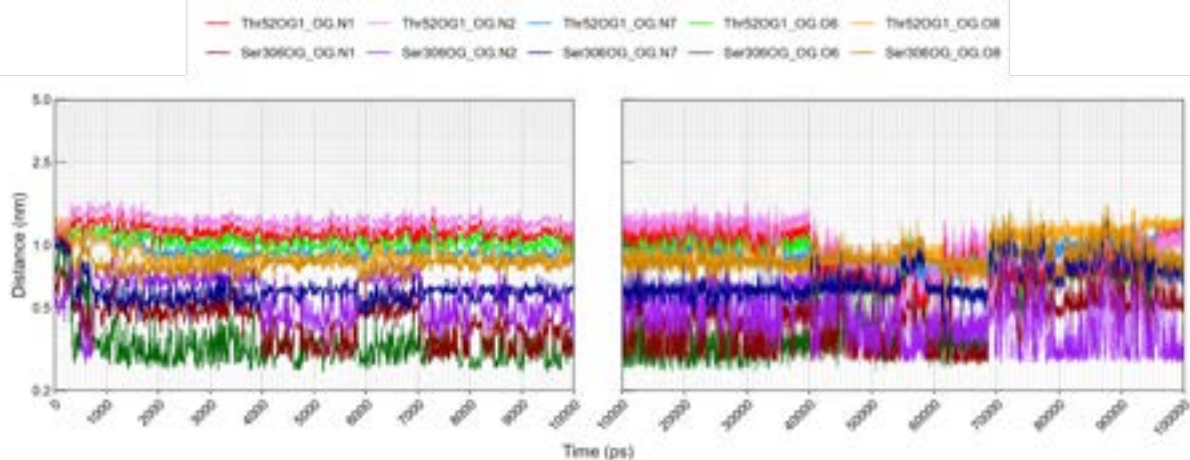
Adenosine and *Thiotrichaceae* MutY NTD

10,000 ps

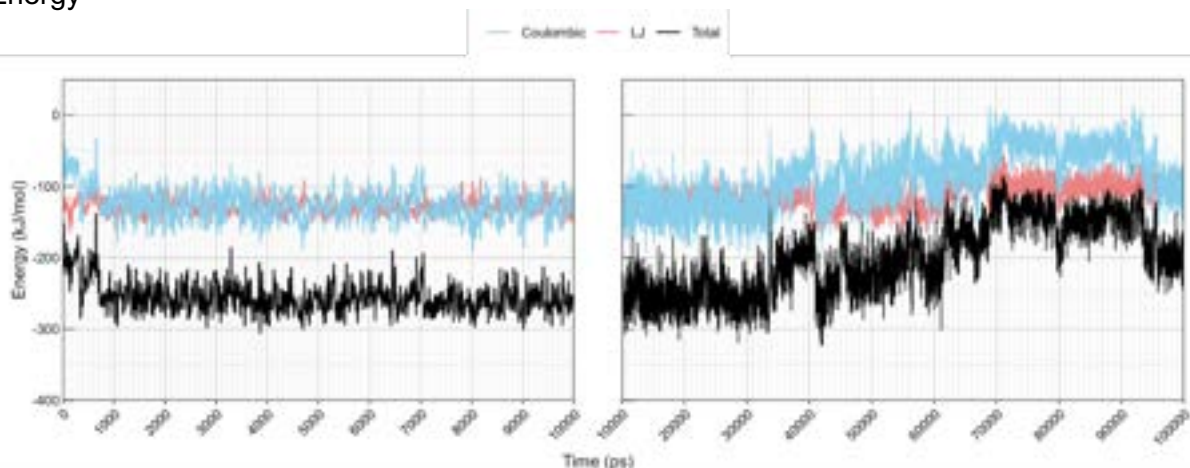


(H) *Thiotrichaceae* MutY complexed with OG

Distance OG - Thr52 and Ser306

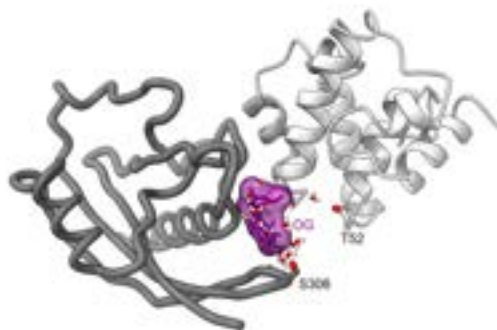


Energy



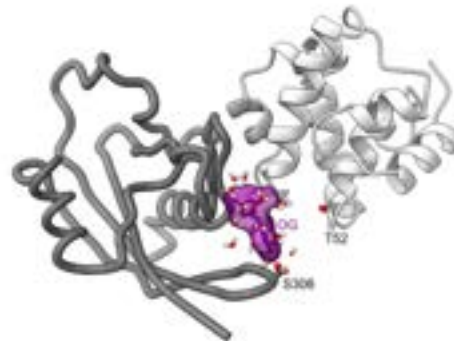
OG and *Thiotrichaceae* MutY

0,000 ps



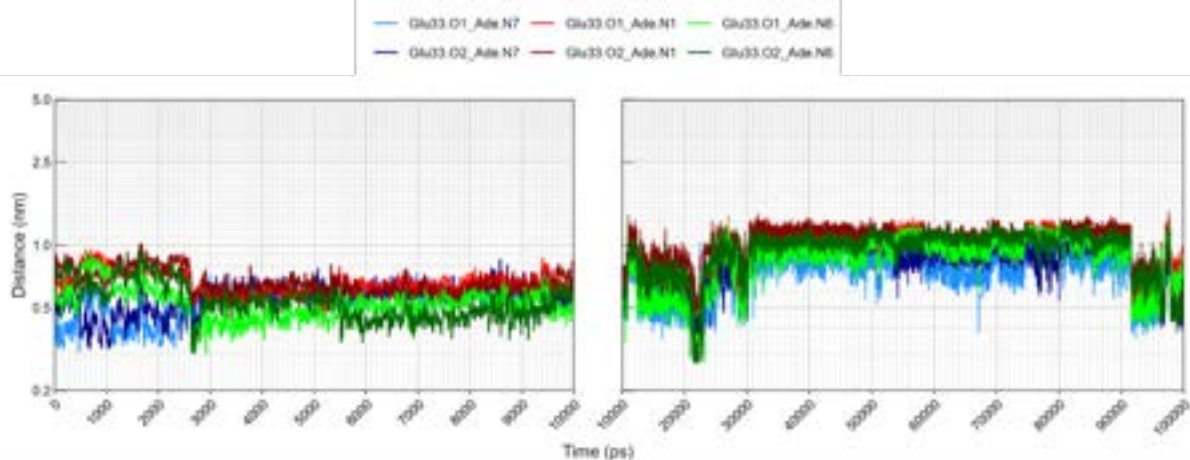
OG and *Thiotrichaceae* MutY

10,000 ps

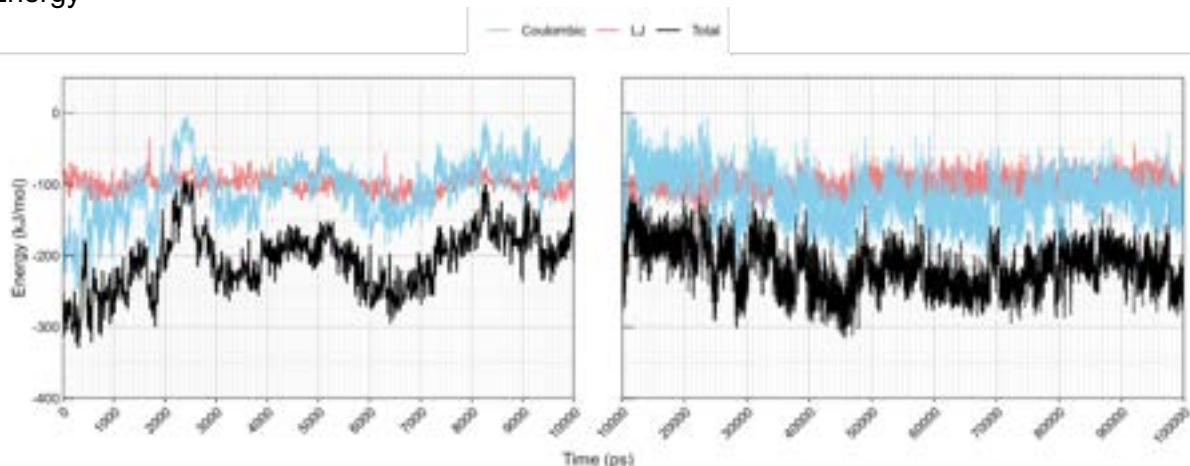


(I) *Flavobacteriaceae* MutY NTD complexed with adenosine

Distance Adenosine - Glu33

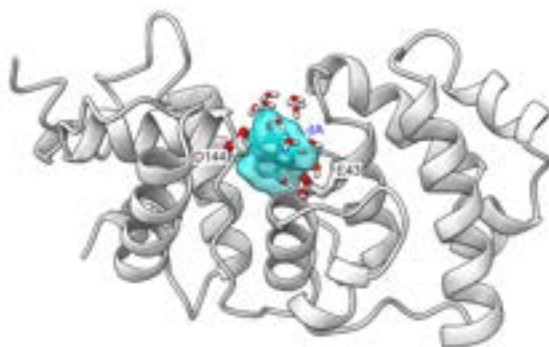


Energy



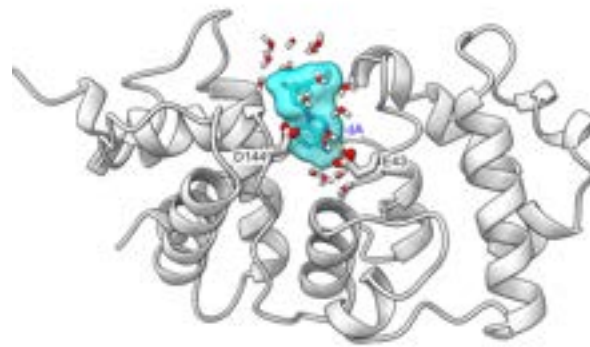
Adenosine and *Flavobacteriaceae* MutY NTD

0,000 ps



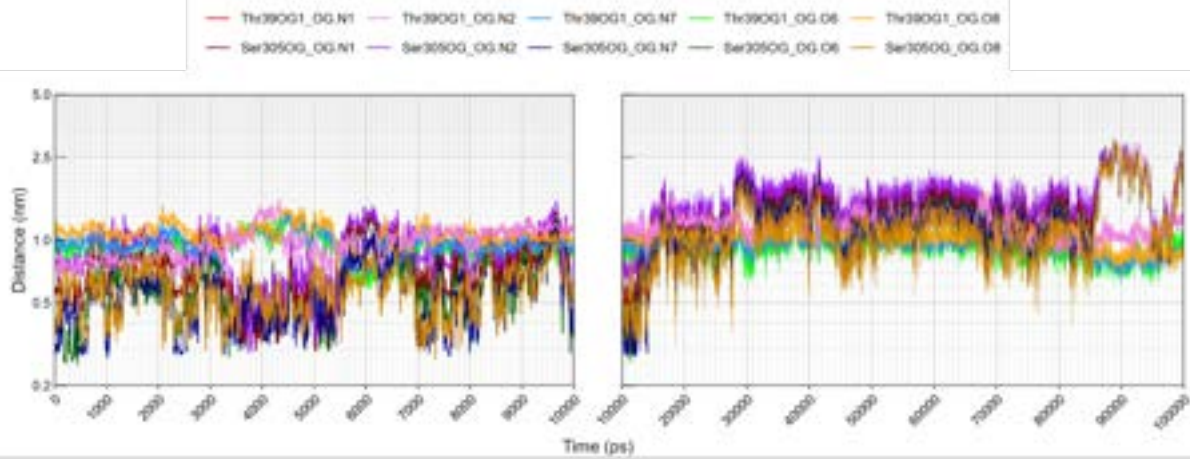
Adenosine and *Flavobacteriaceae* MutY NTD

10,000 ps

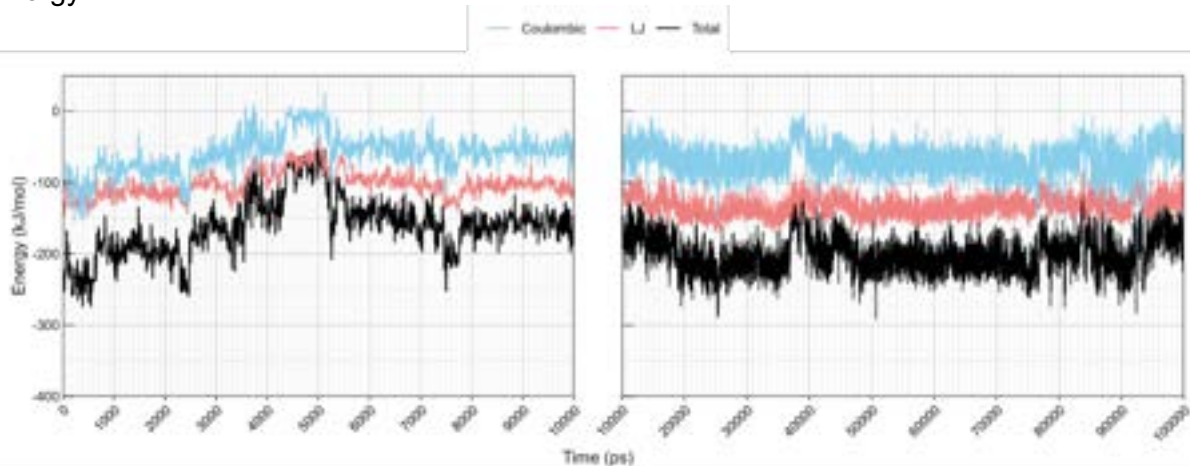


(J) *Flavobacteriaceae* MutY complexed with OG

Distance OG - Thr39 and Ser305

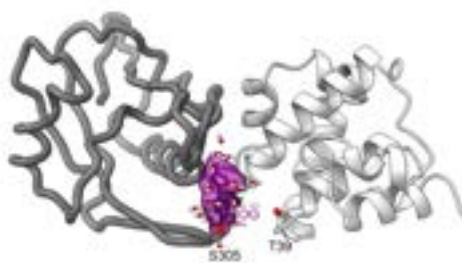


Energy



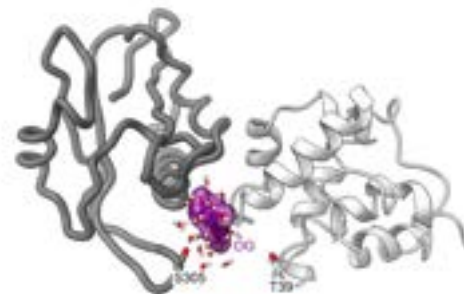
OG and *Flavobacteriaceae* MutY

0,000 ps



OG and *Flavobacteriaceae* MutY

10,000 ps



S8 Movies. Molecular animations. Structures for each MD trajectory were sampled at 200-ps intervals from 0 – 10,000 ps and at 1,000-ps intervals from 10,000 – 100,000 ps and movies were recorded with *ChimeraX*. Residues belonging to the NTD are depicted with a traditional ribbon style cartoon and colored light gray. Residues belonging to the CTD are depicted with a licorice cartoon style and colored dark gray. Solvent molecules that are within 4 Å of both the ligand and protein are shown (O, red; H, white). Each movie highlights particular features and events with time paused, the scene rotating about the y axis, and a brief caption. Links to view and download each movie are provided.

(A) Molecular animation for *Gs* MutY NTD complexed with adenosine. Adenosine remains within the active site pocket throughout the entire 100,000-ps simulation. At ~45,000 ps adenosine rotates within the active site to place its sugar in close proximity to the catalytic Glu43 residue, demonstrating a limited degree of flexibility within the active site pocket. Link to view movie: <https://youtu.be/CO-ljoafn1Q>

(B) Molecular animation for *Gs* MutY complexed with OG. OG remains wedged between NTD and CTD for most of the trajectory. Interactions with the Hoogsteen and Watson Crick face of OG relevant for OG recognition are highlighted at pauses. A new pose emerges at 90,000 ps just prior to departure of the ligand from the NTD-CTD interface. Link to view movie: <https://youtu.be/sVujlyoViRU>

(C) Molecular animation for *Marinosulfonomonas* MutY NTD complexed with adenosine. Adenosine slips back toward the entrance of the active site pocket within 1ns and exits completely by ~6,000-7,000 ps. It then settles in a pocket on the surface of the protein defined by the loop containing Ser24 and a helix from Ala58 to His65. It remains there until ~25,000 ps when it begins to move freely in the solvent, engaging, disengaging, then re-engaging with the surface of the protein for the remainder of the 100,000-ps simulation. Link to view movie: <https://youtu.be/8RAHtBEzjTA>

(D) Molecular animation for *Marinosulfonomonas* MutY complexed with OG. The two domains adopt a different disposition with a new inter-domain interface early in the simulation. The OG ligand finds two new binding sites on the NTD, each distinct from the original binding site, and persists complexed with the NTD until the end of the 100,000-ps simulation. Link to view movie: <https://youtu.be/IUCuN82XJ-U>

(E) Molecular animation for *Rhodobacteraceae* MutY NTD complexed with adenosine. Similar to the *Marinosulfonomonas* MutY NTD simulation, adenosine is completely outside the active site pocket relatively early in the simulation by ~5,000 ps. It then settles on the surface of the protein and wedges into a groove with residues Gly126 and Tyr128 on one side and Gln49 and Arg93 on the other side, and remains at this binding site for the rest of the 100,000-ps simulation. Link to view movie: https://youtu.be/qvgICZj6k_o

(F) Molecular animation for *Rhodobacteraceae* MutY complexed with OG. The animation features a highly dynamic OG-MutY complex that dissociates completely by 48,000 ps. The OG

ligand disengages from functionally relevant interactions at the NTD-CTD interface to find a new site on the NTD by 4,400 ps, nearly escapes at 13,000 ps, and samples several alternate sites on the NTD or the CTD or at a new site at the NTD-CTD interface prior to exiting this region and exploring new sites on the surface of the NTD. The molecular animation is discontinued at 48,000 ps with the complex dissociated. The NTD-CTD structure remains intact for the remainder of the 100,000-ps simulation but the OG ligand did not rebind (not shown). Link to view movie: <https://youtu.be/M9N8OXkGre4>

(G) Molecular animation for *Thiotrichaceae* MutY NTD complexed with adenosine. Adenosine remains in the active site pocket for the entire 100,000-ps simulation. Similarly to the *Gs* MutY-adenosine simulation, the ligand rotates within the active site pocket at ~43,000 ps to place its sugar within close proximity of the active site Glu46 residue, demonstrating limited flexibility within the active site pocket. Link to view movie: https://youtu.be/WyWPjDp_pqI

(H) Molecular animation for *Thiotrichaceae* MutY complexed with OG. The complex persists for the entire 100,00-ps simulation. The initial complex features interaction of Ser305 with the Watson-Crick-Franklin face of the OG base. This pose persists until transition to a new pose at ~69,000 ps with the deoxyribose sugar closer to Ser305 and the base wedged between two helices that converge at the NTD-CTD interface. Link to view movie: <https://youtu.be/B64q0FWSrBs>

(I) Molecular animation for *Flavobacteriaceae* MutY NTD complexed with adenosine. Adenosine remains within the active site pocket for the entire 100,000 ps, It starts with its sugar facing the catalytic Glu33 residue. At ~3,000 ps it rotates to bring the base portion deeper within the active site. It remains in this general orientation for the remainder of the 100,000 ps with the sugar engaging Glu33 in the ~60,000 – 80,000-ps time window. Link to view movie: <https://youtu.be/X7az3V8-wMY>

(J) Molecular animation for *Flavobacteriaceae* MutY complexed with OG. During the first 9,800 ps, Ser305 makes hydrogen bonds with the Hoogsteen face of OG in a manner relevant for recognition. At 10,000 ps, a new pose emerges with the base wedged between helices in the NTD and CTD and thus removed from the FSH recognition loop. The complex with this new pose persists for the remainder of the 100,000-ps simulation. Link to view movie: <https://youtu.be/FJuTrgaGxNQ>

S9 Table. Rifampicin resistance assay.

MutY Expressed	Median (ci.low^a, ci.high^a)	Mutation Frequency^b (per 10⁸ cells)	Fold Change of EcMutY (ci.low^a, ci.high^a)	n
<i>null</i>	101 (81, 145)	20.26	8.78 (6.44, 13.3)	90
<i>E. coli</i> MutY	12 (9, 14)	6.70	1.00 (NA, NA)	79
<i>Marinosulfonomonus</i> MutY	7 (4, 10)	4.13	0.57 (0.27, 0.91)	30
<i>Marinosulfonomonus</i> MutY (catalysis-)	70 (51, 97)	30.77	6.09 (4.32, 9.30)	29
<i>Marinosulfonomonus</i> MutY (recognition-)	74 (59, 90)	11.53	6.39 (4.69, 8.80)	40
<i>Rhodobacteraceae</i> MutY	14 (8, 22)	3.48	1.22 (0.71, 2.00)	30
<i>Rhodobacteraceae</i> MutY (catalysis-)	61 (50, 80)	41.37	5.26 (4.00, 8.00)	38
<i>Rhodobacteraceae</i> MutY (recognition-)	70 (50, 82)	37.65	6.09 (4.12, 8.75)	30
<i>Thiotrichaceae</i> MutY	54 (42, 63)	14.77	4.70 (3.23, 6.44)	47
<i>Thiotrichaceae</i> MutY (catalysis-)	98 (50, 157)	28.89	8.48 (4.46, 14.28)	48
<i>Thiotrichaceae</i> MutY (recognition-)	72 (48, 183)	10.75	6.26 (3.77, 18.00)	20

^a Confidence intervals (95%) determined by a bootstrap method, see **Materials and methods** for details.

^b Mutation frequency reported as median number of resistant colonies per 10⁸ viable colonies. Fold change was calculated by dividing Rif^R frequency by the frequency measured for cultures expressing *Ec* MutY.