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 H_2 and low levels of O_2 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 O_2 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

41 dissolved carbon dioxide is scarce (1). Despite these environmental constraints, archaea and
42 bacteria inhabit the chimneys and hydrothermal fluids venting from the subsurface (1,4). The
43 chemoautotrophic microbes take advantage of chemical compounds generated by subseafloor
44 geochemical reactions such as serpentinization, the aqueous alteration of ultramafic rocks (1).
45 Serpentinization produces hydrogen gas and low-molecular-weight hydrocarbons, which fuel
46 modern microbial communities and also would have been needed to fuel self-replicating
47 molecules and the emergence of primitive metabolic pathways as an antecedent to cellular life
48 (1,2,5). Hydrothermal circulation underneath the LCHF depletes seawater oxygen, leading to an
49 anoxic hydrothermal environment very different from the nearby oxygen-rich seawater (1,3,6–8).
50 As such, the subsurface microbial communities may offer a glimpse into how life emerged and
51 existed before the Great Oxidation Event that occurred over two billion years ago.
52 The unusual environmental conditions of the LCHF present several biochemical
53 challenges to the survival of microbes (4,7,8). For example, high temperatures and alkaline pH
54 conditions present at the LCHF potentiate chemistry to generate DNA-damaging, reactive
55 oxygen species (9). However, it remains unclear whether reactive oxygen species (ROS) are a
56 major threat to resident microbes given that the subseafloor underneath the LCHF is largely
57 devoid of molecular oxygen. We reasoned that the prevalence or absence of DNA repair
58 pathways that cope with oxidative damage would provide insight to the question, are ROS a real
59 and present danger to life at the LCHF?
60 The guanine oxidation (GO) DNA repair system addresses the most common type of
61 DNA damage caused by reactive oxygen species, the 8-oxo-7,8-dihydroguanine (OG)
62 promutagen (**Fig 1**) (10). The OG base differs from guanine by addition of only two atoms, but
63 these change the hydrogen bonding properties so that OG pairs equally well with cytosine and

64 adenine during DNA replication. The resulting OG:A lesions fuel G:C → T:A transversion

65 mutations if not intercepted by the GO DNA repair system (9). The GO system comprises

66 enzymes encoded by *mutT*, *mutM*, and *mutY*, first discovered through genetic analyses of

67 *Escherichia coli* that demonstrated specific protection from G:C → T:A mutations by these

68 three genes (11–14). Homologs or functional equivalents of these GO system components are

69 found throughout all three kingdoms of life (15–17), underscoring the importance of the system,

70 yet there are several instances where particular bacteria (18,19) or eukaryotes (20) make do

71 without one or more of these genes.

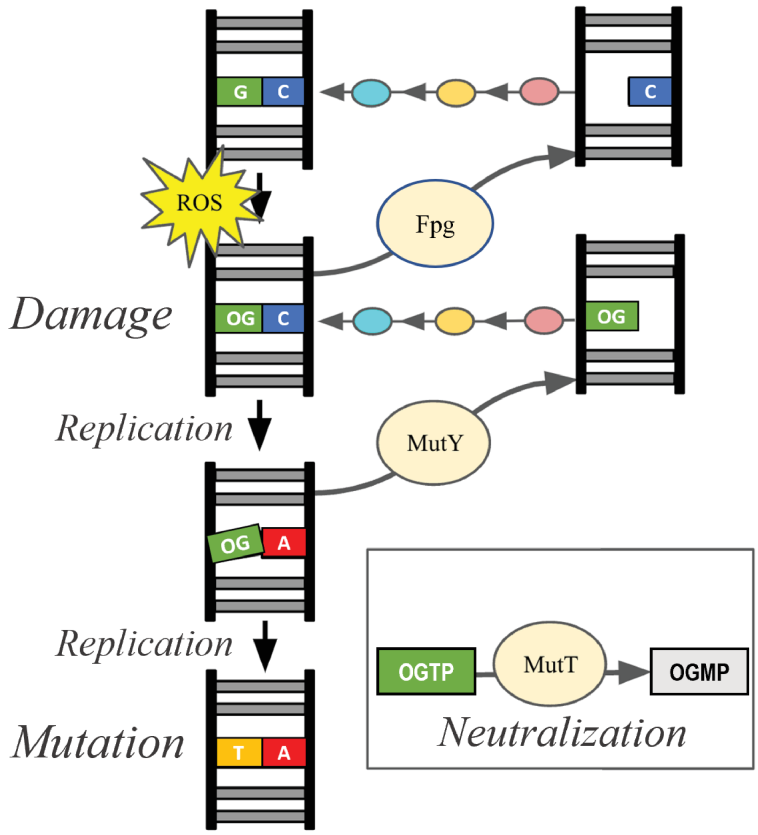


Fig 1. Overview of the GO System. The gene products of *mutT*, *mutM*, and *mutY* (tan bubbles) prevent or repair oxidized guanine DNA damage caused by ROS. DNA glycosylases Fpg (encoded by *mutM*) and MutY remove OG from OG:C and A from OG:A, respectively, to create AP sites with no base. Additional 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.

98 Biochemical analyses of gene products have provided a complete mechanistic picture for

99 the GO repair system. MutT hydrolyzes the OG nucleotide triphosphate to sanitize the nucleotide

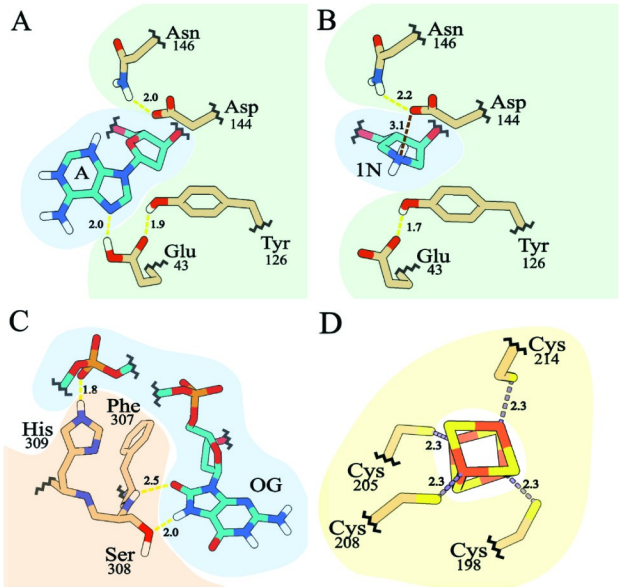
100 pool, thus limiting incorporation of the promutagen into DNA by DNA polymerase (13,21). The

101 enzyme encoded by *mutM*, called formamidopyrimidine-DNA glycosylase (Fpg), locates OG:C
102 base pairs and excises the OG base to initiate base excision repair (BER) (12,22). MutY locates
103 OG:A lesions and excises the A base to initiate BER (11,14). Fpg and MutY thus act separately
104 on two different intermediates to prevent G:C → T:A mutations. These DNA glycosylases
105 generate abasic (apurinic/apyrimidinic; AP) sites, which are themselves mutagenic if not
106 processed by downstream, general BER enzymes, particularly AP nucleases (e.g. exonuclease III
107 and endonuclease IV), DNA polymerase and DNA ligase (17,23–25) as shown in **Fig 1**.
108 MutY is the final safeguard of the GO system. If left uncorrected, replication of OG:A
109 lesions results in permanent G:C → T:A transversion mutations as demonstrated by *mutY* loss of
110 function mutants (26,27). Underperformance of the mammalian homolog, MUTYH, leads to
111 early onset cancer in humans, first discovered for a class of colon cancers now recognized as
112 MUTYH Associated Polyposis (28). MutY is made up of two domains that both contribute to
113 DNA binding and biochemical functions. The N-terminal catalytic domain shares structural
114 homology with EndoIII and other members of the Helix-hairpin-Helix (HhH) protein
115 superfamily (17). The C-terminal OG-recognition domain shares structural homology with MutT
116 and other NUDIX hydrolase family members (17,29). Functionally important and highly
117 conserved residues define chemical motifs in both domains (**Fig 2**). These chemical motifs
118 interact with the OG:A lesion and chelate the iron-sulfur cluster cofactor as revealed by x-ray
119 structural analysis (**Fig 2**) (30–33). For example, residues in the N-terminal domain establish the
120 catalytic mechanism for adenine excision (**Fig 2A** and **Fig 2B**) (32,34). Residues found in a beta
121 loop of the C-terminal domain recognize the OG base and thus direct adenine removal from
122 OG:A lesions (**Fig 2C**) (33). Some motifs are shared among other DNA glycosylases, such as the
123 residues that chelate the 4Fe4S iron-sulfur cluster cofactor (**Fig 2D**) (17). Chemical motifs

124 particular to MutY, especially the OG-recognition residue Ser 308 (**Fig 2C**) and supporting

125 residues in the C-terminal domain, are conserved across organisms and are not found in other

126 DNA glycosylases and therefore can be used to identify MutY genes (17).



143 **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

144 C). Four cysteine residues chelate the iron-sulfur metal cofactor (panel D). All of these

145 interactions are important for MutY activity.

146

147 Our study investigated whether microbes in the anoxic LCHF environment use the GO

148 DNA repair system to mitigate damage caused by reactive oxygen species. It is important to note

149 that not all organisms have an intact GO repair system; examples are missing one, some or all

150 components. MutY in particular was absent frequently in a survey of 699 bacterial genomes (19),

151 and its absence may indicate relaxed evolutionary selection from oxidized guanine damage (18).

152 We mined for homologous genes within the LCHF microbial community and recombinantly

153 expressed candidate MutY enzymes to characterize function. We found genes encoding GO

154 system components and general base excision repair enzymes at all LCHF sites. MutY homologs

155 from the LCHF suppressed mutations when expressed in *mutY* deficient *E. coli* strains indicating

156 these function similarly to authentic MutY. These Lost City MutY homologs could be assigned

157 confidently to metagenome assembled genomes (MAG)s, allowing for additional gene inventory
158 analyses that revealed metabolic strategies involving sulfur oxidation and nitrogen reduction.
159 These results have important implications for understanding the repair of oxidative guanine
160 damage in low-oxygen environments, similar to those that existed on a younger Earth, as well as
161 those that may exist on other planets and moons.
162

163 Results

164 Identification of the GO DNA repair system in LCHF microbes

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

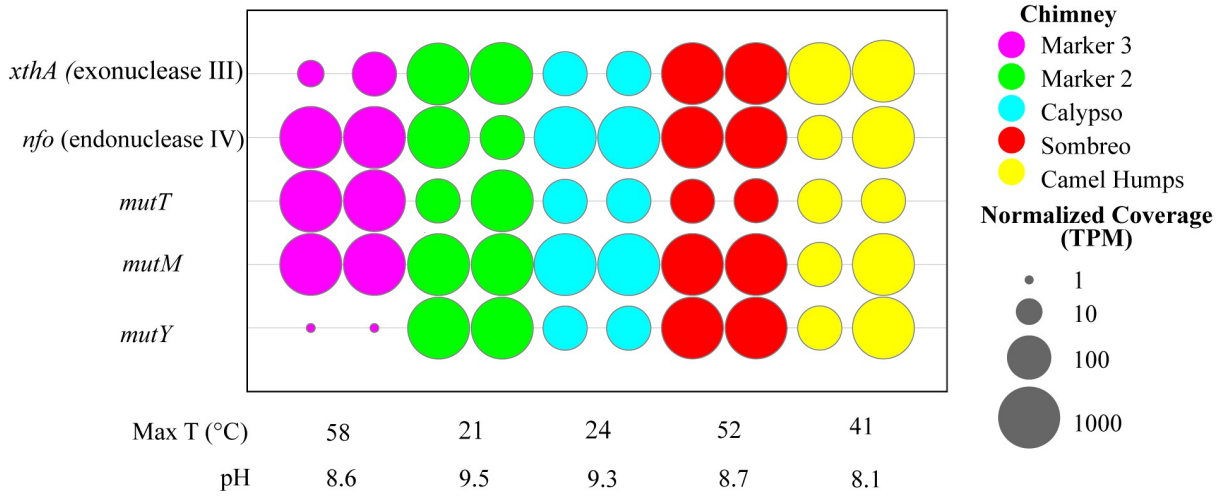


Fig 3. Abundance of GO system gene homologs. Listed on the vertical axis are genes encoding DNA repair enzymes. Genes *xthA* and *nfo* are generally necessary for DNA repair involving base excision repair in bacteria, including the particular GO system investigated here. Together, *mutT*, *mutM* and *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**).

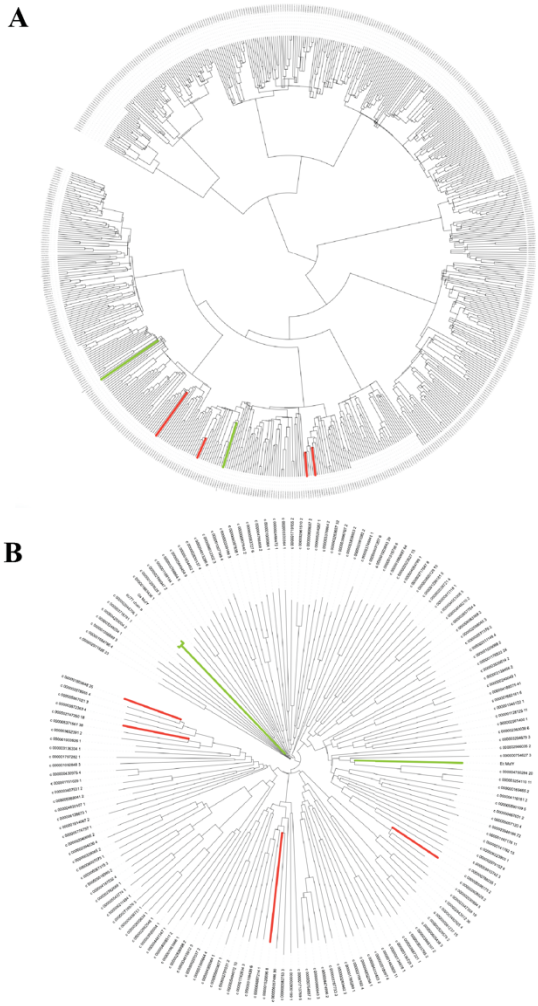


Fig 4. Phylogeny for LCHF MutYs. (A) 649 sequences were identified as LCHF MutY candidates due to sequence similarity to existing MutYs (green branches) and aligned to reconstruct evolutionary relationships. (B) A subset of 160 members contained all necessary MutY-defining chemical motifs. Alignment 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), 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

224 contact with Tyr126 (second Y, green in **Fig 5**), Asp144 (D, green in **Fig 5**), and supported by
225 Asn146 (N, green in **Fig 5**) indicating these chemical motifs stabilize the transition state during
226 catalysis (32–34). For the LCHF MutYs, the Asn residue is invariant, the catalytic Asp is nearly
227 invariant, replaced by chemically similar Glu for five LCHF MutYs, and the catalytic Tyr
228 residue is often replaced by Ser and sometimes by Thr or Asn. The residue found between the
229 catalytic Asp and Asn is always a small residue, most often Gly (G, red in **Fig 5**) but sometimes
230 Ala, Val or Thr.

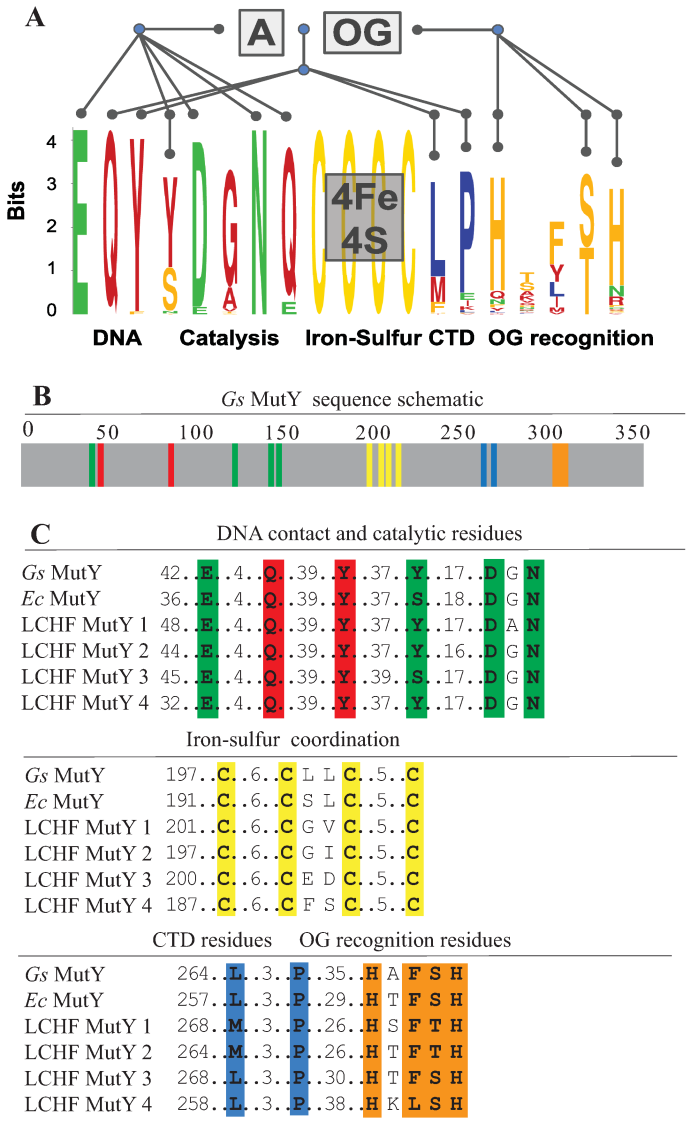


Fig 5. LCHF MutY chemical motifs. (A) Conservation and diversity of MutY-defining chemical motifs are depicted with a sequence logo for the 160 LCHF MutYs. These motifs are associated with biochemical functions including DNA binding, enzyme catalysis, attachment of the iron-sulfur cofactor, and recognition of the damaged OG base. Sequence logo generated by Weblogo (38,39). See **Supporting Information** for a complete alignment (**Data S1**) and a percent identity matrix for the representative LCHF MutYs (**Table S2**). (B) Chemical motifs located in MutY as shown with color-coded positions for *Gs* MutY. (C) Alignment for select chemical motifs highlights conservation among the four representative LCHF MutYs, *Ec* MutY and *Gs* MutY

261 By contrast with these numerous and highly conserved motifs for the N-terminal domain,
262 fewer motifs with greater sequence divergence were found in the C-terminal domain. The MutY
263 ancestor is thought to have resulted from a gene fusion event that attached a MutT-like domain to
264 the C-terminus of a general adenine glycosylase enzyme, and the C-terminal domain of modern
265 *Ec* MutY confers OG:A specificity (29). X-ray crystal structures of *Gs* MutY interacting with
266 OG:A and with G:A highlighted conformational difference for a Ser residue in the C-terminal
267 domain (Ser308 in *Gs* MutY), and mutational analysis showed this Ser residue and its close
268 neighbors (Phe307 and His309 in *Gs* MutY) establish OG specificity (33). Informed by these
269 insights from structure, we eliminated LCHF MutY candidates that lacked a C-terminal domain
270 and its OG-recognition motif. Alignment of the 160 LCHF MutYs that passed this test showed
271 that a second His residue is also well conserved within the H-x-FSH sequence motif (**Fig 5C**).
272 As is evident from the alignment, there are many variations with residues replaced by close
273 analogs at each position. Ser, which makes the key contacts with N7 and O8 of OG and no
274 contact with G, is often replaced by Thr, which can make the same hydrogen bond interactions.
275 Likewise the two His residues are each often replaced by polar residues (e.g. Gln, Asn, Arg or
276 Lys) that can also hydrogen bond to the DNA phosphate backbone as observed for His305 and
277 His309 in *Gs* MutY.
278 Two other positions with high conservation were revealed for the C-terminal domain in
279 this analysis of the 160 LCHF MutYs. These define a L-xxx-P motif. These residues are replaced
280 by other residues with comparable chemical properties. The Leu position is often another
281 hydrophobic residue such as Met or Phe, and the Pro position is most frequently replaced by Glu,
282 a residue that can present aliphatic methylene groups and thus resemble Pro if the polar group
283 hydrogen bonds with the peptide amide. In the structure of *Gs* MutY, the Pro269 nucleates a

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

290 Four representative LCHF MutYs were selected for further analyses. Supporting
291 Information **Table S2** reports the percent identity among these representative LCHF MutYs and
292 the well-studied MutYs from *E. coli* and *Geobacillus stearothermophilus*. LCHF MutY 1 and
293 LCHF MutY2 are most closely related with 65% identity which is almost twice the average in
294 this group. LCHF MutY 3 is most closely related to *Ec* MutY with 48% identity. We examined
295 the representative LCHF MutYs for physical properties as inferred from sequences. **Table 1**
296 reports these physical properties including predicted protein size, isoelectric point (pI), and
297 stability (Tm). Generally the physical characteristics measured for LCHF MutY representatives
298 were comparable to each other and to predicted properties of *Ec* MutY and *Gs* MutY. The
299 predicted Tm for LCHF MutY 3 was above 65°C, distinguishing it as the most stable enzyme
300 (**Table 1**), which may reflect adaptation to a high temperature environment. The isoelectric point
301 predicted for each of the LCHF MutY representatives is 3 pH units above the pI predicted for *Gs*
302 MutY and between 0.1 - 0.5 pH units above the pI predicted for *Ec* MutY, indicating that more
303 numerous positively charged residues have been recruited, possibly as an adaptation to the LCHF
304 environment.
305
306

307 **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)

308

309 Identification of LCHF MutY organisms, gene neighbors, 310 environmental conditions, and metabolic strategies

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

321 organisms by the lowest-level classification that was determined for each LCHF MutY (e.g.

322 LCHF MutY 3 will now be referred to as *Thiotrichaceae* MutY).

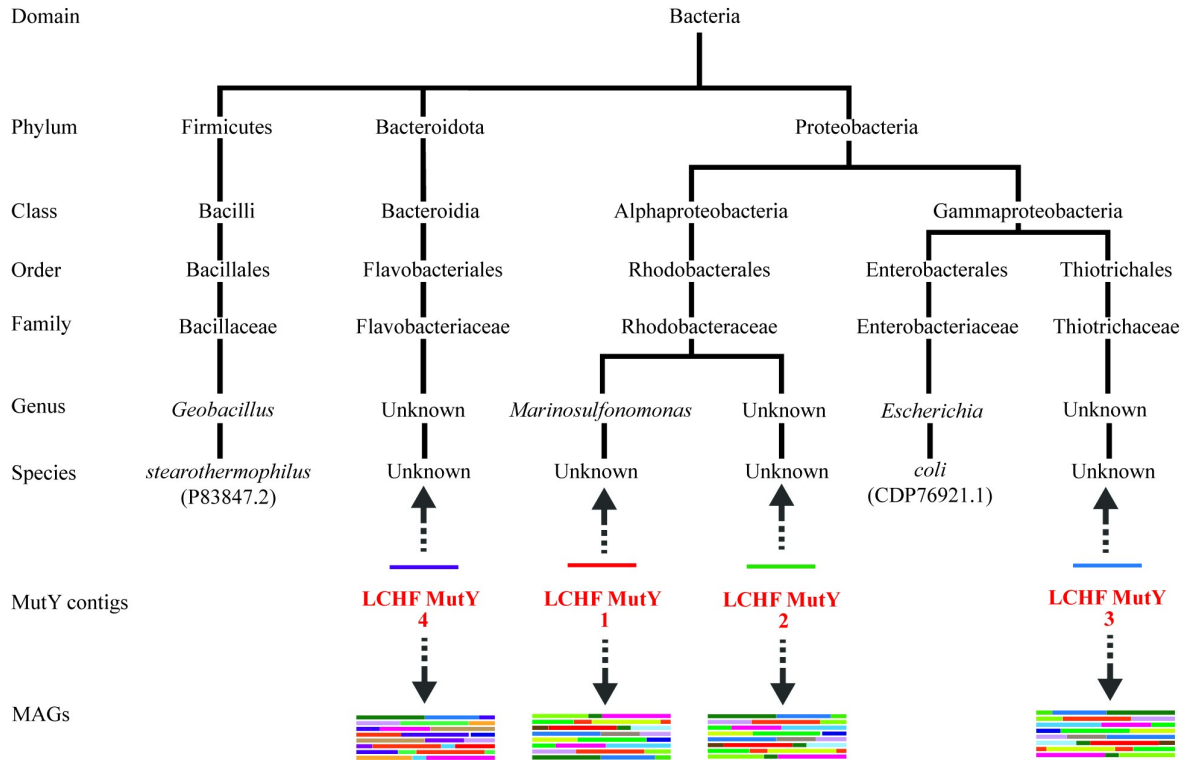


Fig 6. Taxonomic classification. LCHF MutY-encoding contigs were found in several branches of bacteria. The classification places these in relation to MutY of *G. stearothermophilus* and *E. coli* (accession IDs included). LCHF MutYs were mapped to their respective microbes by two methods indicated by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

323

324 The inclusion of MutY contigs in MAGs provided an opportunity to examine gene

325 neighbors for the representative LCHF MutYs. The GO repair genes are located at distant loci in

326 *E. coli* (12), and belong to separate operons (40). However, MutY is the immediate 5'-neighbor

327 to YggX within gammaproteobacteria (40), and homologs of YggX are present outside this

328 lineage, occasionally nearby to MutY (e.g. *Bacillus subtilis*). As gene neighbors, MutY and

329 YggX are part of a SoxRS regulated operon in *E. coli* (40,41). YggX provides oxidative stress

330 protection and iron transport function with a critical Cys residue close to the N-terminus of this

331 small protein (42,43). A protein matching these features is encoded by a gene partly overlapping
332 with and the nearest 3' neighbor to *Thiotrichaceae* MutY (see Supporting Information **Fig S3**).
333 To reveal the environmental conditions of these MutY-encoding organisms, we analyzed
334 the sequence coverage of each LCHF MutY contig at each of the sampling sites.
335 *Marinosulfonomonas* MutY, *Thiotrichaceae* MutY, and *Flavobacteriaceae* MutY were identified
336 at all sampling sites, ranging from 21°C to 58°C and pH 8.1 to 9.5. *Rhodobacteraceae* MutY was
337 present at all sampling sites excluding Marker 3 and was therefore found in temperatures ranging
338 from 24°C to 52°C and pH 8.1 to 9.5.
339 We further investigated the metabolic strategies utilized by MutY-encoding microbes by
340 examining the inventory of predicted protein functions in each MAG (**Table 2,** see also
341 Supporting Information **Table S4**). Each LCHF MutY-containing MAG possessed at least two
342 forms of cytochrome oxidase, with the exception of the *Flavobacteriaceae* MAG. The
343 *Flavobacteriaceae* MAG is only 44% complete, however, so no strong conclusions can be made
344 regarding the absence of genes. Cytochrome oxidases commonly provide sources of free radicals
345 and are essential to aerobic metabolism. Predicted proteins indicative of dissimilatory nitrate and
346 nitrite reduction were found in the *Marinosulfonomonas, Rhodobacteraceae,* and *Thiotrichaceae*
347 MAGs*,* suggesting that these organisms may be capable of using nitrate or nitrite as alternative
348 electron acceptors when oxygen is not available. Furthermore, the *Marinosulfonomonas* and
349 *Rhodobacteraceae* MAGs include predicted protein functions associated with the oxidation of
350 reduced sulfur compounds, though it is important to note that the directionality of these reactions
351 cannot be fully determined from bioinformatics alone. These patterns speak to the potential
352 origins of oxidants within the MutY-encoding organisms as discussed below.
353

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

			Marinosulfonomonas [a]		*Rhodobac-teraceae*	*Thiotrich-aceae*	*Flavobac-teriaceae*
			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

355
356 [a] *Marinosulfonomonas* MutY belongs to two separate MAGs and genes for each are reported separately.
357 [b] Completeness and contamination scores generated by *CheckM* v1.0.5 as described in Brazelton *et al.* 2022 (35).

358 ## Predicted protein structures and virtual docking experiments

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

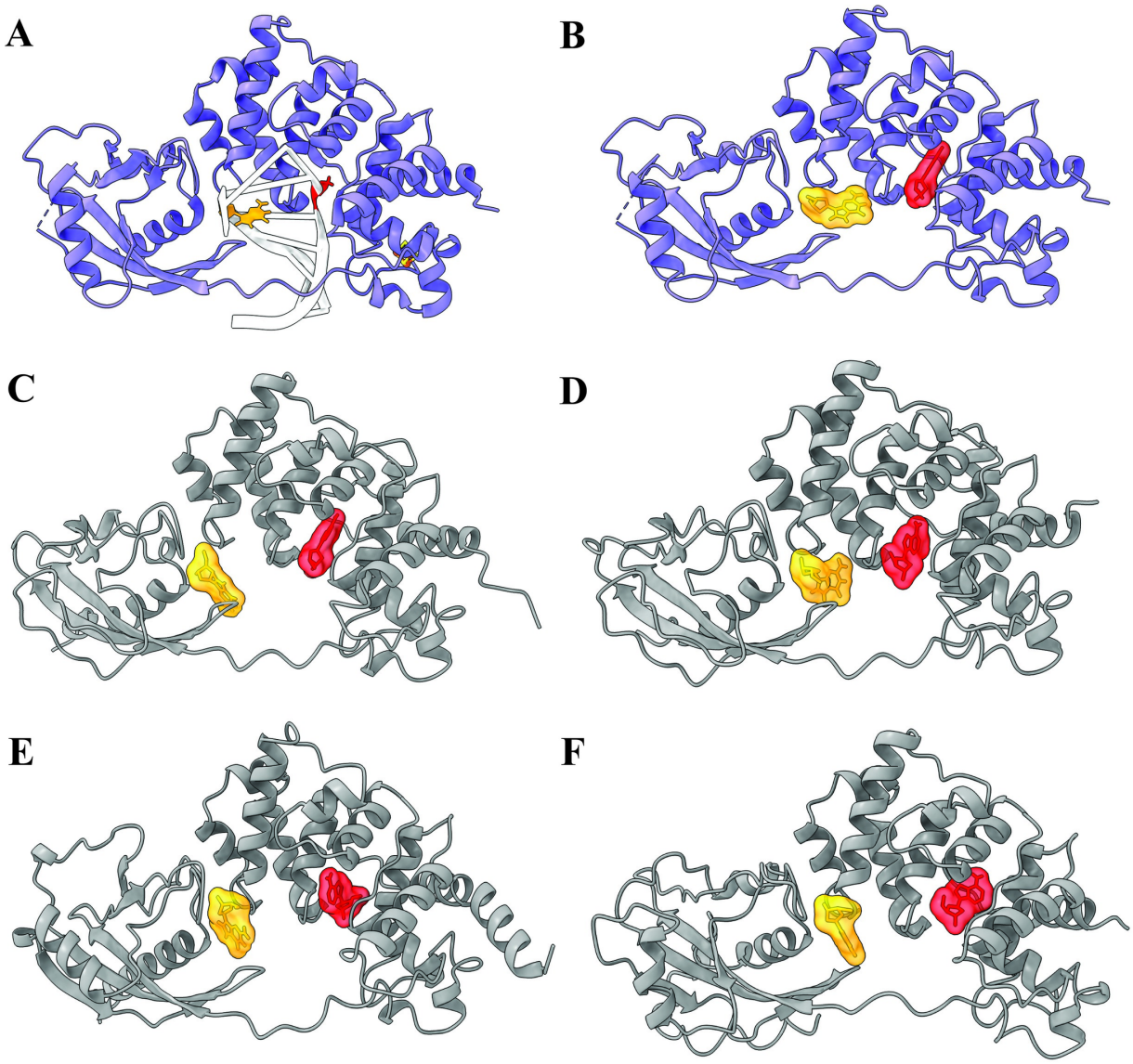


Fig 7. **Structure predictions and virtual docking of MutY ligands**. (A) The x-ray crystal structure of *Gs* MutY (blue ribbon; PDB ID 3g0q) in complex with DNA (white helix) highlights the positions of the adenosine nucleoside (red) and OG (yellow). (B-F) Virtual docking of ligands. Adenosine and OG 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: *Marinosulfonomonas* MutY (C); *Rhodobacteraceae* MutY (D); *Thiotrichaceae* MutY (E); and *Flavobacteriaceae* MutY (F).

382 **Table 3. Molecular modeling for LCHF MutYs**

		RMSD (Å) [a] (residues)		Affinity VINA (kcal/ mol) [b]		Energy Amber [c] (kJ/mol)	
MutY Model	**pLDDT [d]**	**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)
Marinosulfonomonas	93 (±9)	0.87 (198)	0.90 (64)	-6.8	-7.5	-80 (±37) [c]	-194 (±44)
Rhodobacteraceae	94 (±6)	0.85 (199)	0.98 (51)	-6.8	-7.7	-170 (±37)	-110 (±54) [c]
Thiotrichaceae	92 (±12)	1.0 (195)	1.1 (77)	-7.0	-8.0	-226 (±39)	-205 (±52)
Flavobacteriaceae	92 (±14)	0.84 (199)	1.2 (44)	-7.2	-8.0	-214 (±34)	-197 (±30)

383 [a] Root mean square deviation for the superposition of predicted structures with *Gs* MutY (PDB ID 3g0q) was
384 calculated separately for N-terminal domain (NTD) and C-terminal domain (CTD) by *ChimeraX* (45).
385
386 [b] Binding affinity for the best outcome from docking adenosine to the enzyme active site and OG to the OG-
387 recognition site as calculated by *Autodock VINA* (46,47).
388
389 [c] Energy for short-range Coulombic and Lennard-Jones interactions with the ligand as computed by *GROMACS*
390 (48), with the Amber99SB and GAFF force fields (49,50). Energies were averaged over the 100-ns simulation or the
391 time window of the complex, 0 – 26 ns for *Mainosulfonomonas* interaction with A and 0 – 47.5 ns for
392 *Rhodobacteraceae* interaction with OG. Uncertainty is the sample standard deviation.
393
394 [d] Local distance difference test metric to assess confidence for structures predicted by *Colabfold* (44) averaged over
395 all residues. Uncertainty is the sample standard deviation.
396
397 [e] Reference superposition values provided by comparing a second structure of *Gs* MutY in complex with its
398 transition state analog (PDB ID 6u7t)
399

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

406 conformations and orientations for the OG and adenosine ligands were anticipated. The search
407 volume for the adenosine ligand was centered on the active site in the NTD, and the search
408 volume for the OG ligand was defined by the OG-recognition motif found in the CTD.
409 Representative outcomes obtained with *Autodock VINA* are shown in **Fig 7,** and the
410 corresponding binding affinities are reported in **Table 3** and Supporting Information **Table S6**.
411 As anticipated the precise orientation and position for these docked ligands varied, and none
412 exactly match the disposition of the adenine or OG base as presented in the context of double
413 stranded DNA. Nevertheless, binding affinities for the ligand-LCHF MutY complexes ranged
414 from -6.8 to -8.0 kcal/mol, indicating favorable interactions were attainable and similar to the
415 binding affinities measured for *Gs* MutY by the same virtual docking method.
416

417 Molecular dynamics simulations

418 Virtual docking is fast and computationally economical but largely ignores motion and
419 solvent. The reliability of docking improves when complemented with molecular dynamic (MD)
420 simulation (53,54). To further assess stability and dynamic properties of LCHF MutY-ligand
421 complexes derived by docking, we applied MD simulations with the *Amber* force field (49,50),
422 as implemented with *GROMACS* (48). Each protein-ligand complex was solvated in water,
423 charges were balanced with counterions, and the system was equilibrated in preparation for a
424 100-ns MD simulation (see **Materials and methods**). Supporting Information **Fig S7** and
425 **Movies S8** summarize the resulting trajectories in terms of interaction energy, distance, and
426 structure over time. We focused on mechanistically relevant interactions by tracking distances
427 from the base moiety to the catalytic Glu residue for adenosine complexes, and distances to OG-
428 recognition Ser and Thr residues for OG complexes. MD trajectories for the *Gs* MutY-ligand

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

431 MD analysis revealed dynamic, and in some cases unstable complexes. Relative
432 instability likely reflects the free nature of the ligands, which normally would be presented as
433 part of DNA. As will become evident in later sections, complex instability detected by MD
434 simulation correlates positively with biological activity under mesophile conditions. Even so,
435 many of the MutY-ligand complexes persisted for the entire 100-ns simulation, characterized by
436 favorable binding affinity, extracted as the sum of local Lennard-Jones and Coulombic
437 interactions (**Table 3**). While all ligands were mobile, the MD outcomes separated into two
438 groups distinguished by the degree of ligand movement and persistence of the complex. In the
439 first group the adenosine and OG ligands remained close to the original binding sites for at least
440 90 ns if not the entire 100-ns MD simulation. This first group with persistently engaged ligands
441 included the complexes with *Gs* MutY (**Fig S7A, Fig S7B, Movie S8A** and **Movie S8B**),
442 *Thiotrichaceae* MutY (**Fig S7G**, **Fig S7H, Movie S8G** and **Movie S8H**) and *Flavobacteriaceae*
443 MutY (**Fig S7I**, **Fig S7J, Movie S8I** and **Movie S8J**).

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

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

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

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

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

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

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

495 **Testing mutation suppression activity of LCHF MutY enzymes by**

496 **recombinant expression**

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

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

514 Cultures with an empty plasmid (*null*) and cultures with a plasmid encoding *Ec* MutY
515 showed significant differences in the frequency of Rif^R mutants, with median values of 101 and
516 12, respectively, indicating the assay was fit for use (significance determined by non-overlap of
517 median 95% confidence intervals) (**Fig 8** and Supporting Information **Table S9**).

518 *Marinosulfonomonas* MutY, *Rhodobacteraceae* MutY, and *Thiotrichaceae* MutY each
519 demonstrated significant mutation suppression activity when compared to the *null*. Indeed,
520 *Rhodobacteraceae* MutY showed mutation suppression performance equivalent to that measured
521 for *Ec* MutY, and *Marinosulfonomonas* MutY was apparently better at suppressing mutations
522 than *Ec* MutY, a remarkable outcome given the evolutionary time separating these species (**Fig**
523 **8**). Note that these LCHF MutYs with high mutation suppression function formed unstable
524 complexes as revealed by MD simulation. *Thiotrichaceae* MutY showed partial function.
525 Cultures expressing *Thiotrichaceae* MutY suppressed Rif^R mutants to 50% of the rate observed
526 for *null* cultures, but allowed a mutation rate about 4-fold greater than that measured for cultures
527 expressing *Ec* MutY (**Fig 8** and Supporting Information **Table S9**).

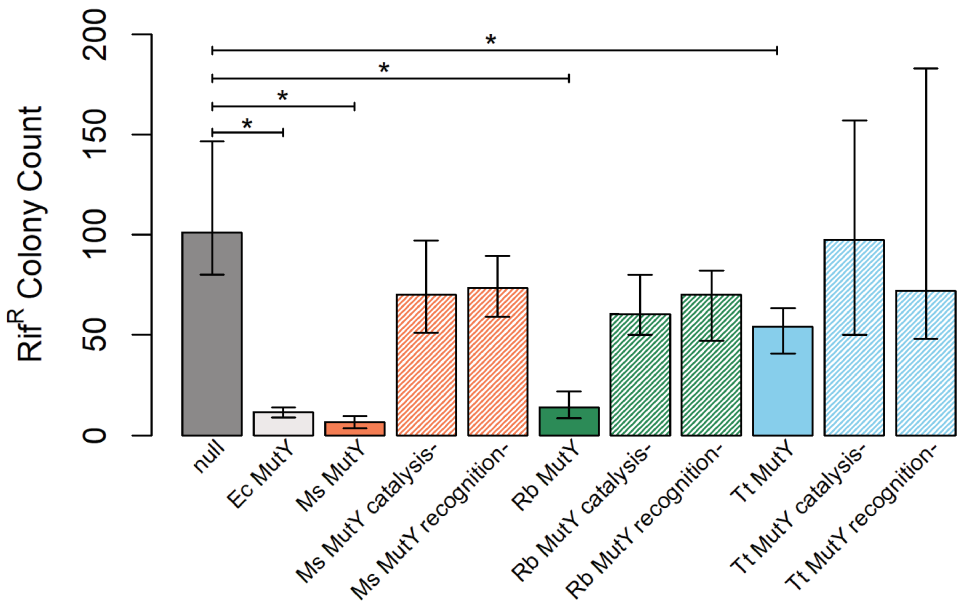


Fig 8. **Functional analysis.** Bars represent median Rif^R colony counts for *E. coli* cultures expressing MutY, MutY variants, or no MutY (*null*) from a plasmid DNA. Error bars represent 95% confidence intervals as determined by bootstrap sampling (see Supporting Information **Table**
545 **S9** for tabulated values). *Marinosulfonomonas* (*Ms*) MutY, *Rhodobacteraceae* (*Rb*) MutY, and
546 *Thiotrichaceae* (*Tt*) MutY each suppressed mutations as evidenced by non-overlap of Rif^R
547 confidence intervals compared to *null* cultures. Altered versions of each LCHF MutY tested the
548 importance of residues for catalysis and OG-recognition. Designated *catalysis*- and *recognition*-
549 along the X-axis, these alterations severely impacted mutation suppression function indicating
550 the LCHF MutYs share mechanistic features with the extensively studied enzymes *Ec* MutY and
551 *Gs* MutY.
552

553 To investigate the biochemical mechanism employed by LCHF MutY enzymes, we
554 altered residues essential for OG:A recognition and catalysis, then repeated the mutation
555 suppression assay. Two mutants of each LCHF MutY were constructed through site-directed
556 substitution of residues. One set of substitutions was designed to disable the OG-recognition
557 motif by replacing F(S/T)H residues (**Fig 2** and **Fig 5**) with alanine residues (designated
558 *recognition*-); the other set of substitutions was designed to disable catalysis by replacing the
559 active site Asp and Glu residues with structurally similar, but chemically inert Asn and Gln
560 residues (*catalysis*-). For all three LCHF MutYs, these targeted substitutions disabled mutation
561 suppression function *in vivo* as shown by elevated Rif^R frequencies for cultures expressing the
562 *recognition*- and *catalysis*- versions. The mutation frequencies for cultures expressing these site-
563 specific substitution variants were comparable to the Rif^R frequencies measured for *null* cultures
564 as judged by overlapping 95% confidence intervals (**Fig 8** and Supporting Information **Table**
565 **S9**). These results indicate that the LCHF MutYs suppress mutations by a mechanism that is
566 highly similar to the strategy executed by *Ec* MutY and *Gs* MutY.
567

568 # Discussion

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

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

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

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

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

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

622 Our metagenomic analysis revealed that gene homologs encoding the GO DNA repair
623 system are abundant in basement microbes inhabiting the LCHF (**Fig 3**). This observation is
624 surprising given that the basement of the LCHF is expected to be anoxic (1,3). The chemical
625 agents commonly thought of for producing oxidized guanine (OG) are ROS derived from
626 molecular oxygen *via* aerobic metabolism. In an anoxic environment, what chemical agents are
627 producing OG and how are these generated? Models of hydrothermal field chemistry predict
628 abiotic production of ROS which the microbial residents may encounter, although these would
629 probably react with cell protective structures before encountering DNA (68). Continual mixing
630 of seawater with the anoxic hydrothermal fluid could provide molecular oxygen at the interface
631 where hydrothermal fluids vent into ambient seawater at the seafloor (1,3). Facultative anaerobes
632 at this interface would inevitably generate ROS (69,70), and therefore benefit from the GO DNA
633 repair system. Intermicrobial competition has driven acquisition of chemical strategies, including
634 ROS, for killing other bacteria (71–73), but there is currently no evidence for such bacterial
635 warfare in basement dwelling microbes of the LCHF.
636 Another explanation for the source of OG in basement microbes of the LCHF, which is
637 suggested by our gene inventory analysis (**Table 2** and **Supporting Information**), involves
638 reactive sulfur species (RSS) and reactive nitrogen species (RNS). Many of the basement-
639 dwelling microbes within the LCHF appear to metabolize sulfur and nitrogen for energy
640 conservation (4,35), strategies that generate RSS and RNS as metabolic byproducts (74,75).
641 Indeed, mechanisms for the oxidation of guanine by both RSS and RNS have been described,
642 including the formation of 8-oxoguanine (OG) and chemically similar 8-nitroguanine, which
643 templates for adenine in a fashion similar to OG (76–78). The oxidation of guanine by RSS and
644 RNS generated from microbial metabolism would produce OG independent of molecular oxygen

645 and thereby necessitate the GO DNA repair system for both facultative and obligate anaerobes
646 inhabiting the LCHF.
647 Whether organisms developed biochemical systems to deal with oxidative damage before
648 or after the Great Oxidation Event (GOE) remains an open question (79,80). It is reasonable to
649 think of these systems arising in response to the selective pressure of oxidative damage from
650 rising O_2 levels. However, it is also possible that these systems were already in place as a coping
651 mechanism for other oxidants and were repurposed to deal with the new source of oxidizing
652 agents when O_2 became readily available. Indeed, obligate anaerobes contain many of the same
653 pathways to deal with oxidative damage as aerobes (79,80). Our discovery of the GO DNA
654 repair system in basement dwellers of the LCHF adds to this body of evidence and supports the
655 hypothesis that oxidative damage repair systems were established before the GOE. We
656 considered the caveat of possible phage-mediated gene transfer – modern microbes adapted to
657 oxygen-rich regions elsewhere may be the source of LCHF MutYs. However, the
658 correspondence of taxonomic assignments based on MutY sequence and based on MAGs and the
659 high degree of sequence diversity seen for LCHF MutY enzymes are inconsistent with expansion
660 of GO DNA repair in the LCHF by horizontal gene transfer. Thus, it seems more likely that
661 LCHF microbes inherited the GO DNA repair system from a common ancestor and retained it
662 through necessity, even in the absence of extrinsic O_2.
663

664 # Conclusion

665 Performing empirical studies on how life may have evolved on Earth and other planets is
666 inherently difficult due to time and spatial barriers. Unique sites such as the LCHF serve as
667 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 ºN 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

690 encode exonuclease III and endonuclease IV, respectively.

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

696 Identification of LCHF MutYs

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

712 160 authenticated LCHF MutYs with *iTOL* (87). Selection of the four representative LCHF
713 MutYs was guided by this phylogeny and by the completeness of each associated MAG.

714 Taxonomic classification

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

723 Prediction of physical parameters

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

728 Molecular modeling

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

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

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

748 MD simulations were performed with *GROMACS* version 2022.5 (48), applying the
749 AMBER99SB and GAFF force fields (49,50), with CPU and GPU nodes at the University of
750 Utah's Center for High Performance Computing. We followed steps outlined in the *GROMACS*
751 tutorial "Protein-Ligand Complex" as a guide for our experiments (97). The starting structure for
752 a protein-ligand complex was selected from the binding modes predicted by *Autodock VINA*,
753 choosing the mode with the highest affinity after excluding those that appeared incompatible
754 with the double stranded DNA-enzyme structure. To conserve computational resources,
755 simulation of the complex with adenosine was limited to N-terminal residues as follows: residues
756 8-220 for *Gs* MutY (PDB ID 6u7t); 6-230 for *Marinosulfonomonas* MutY; 2-220 for

757 *Rhodobacteraceae* MutY; 11-223 for *Thiotrichaceae* MutY; and 2-209 for *Flavobacteriaceae*
758 MutY. Simulations with OG omitted the iron-sulfur cluster domain and interdomain linker and
759 thus included limited residues with chain interruptions as follows: residues 29-137, 234-289,
760 295-360 for *Gs* MutY; 40-142, 239-352 for *Marinosulfonomonas* MutY; 38-139, 233-352 for
761 *Rhodobacteraceae* MutY; 32-140, 238-364 for *Thiotrichaceae* MutY; and 19-127, 234-354 for
762 *Flavobacteriaceae* MutY. Ligand topology files were generated with the *ACPYPE* server (98),
763 applying the general *Amber* force field (49). Each complex was solvated with water molecules
764 with three points of transferable intermolecular potential (TIP3P). Counterions were added to
765 neutralize the net charge of the system. The system was energy minimized by 50000 steepest
766 descent steps and further equilibrated in two phases, NVT followed by NPT, each entailing 100
767 ps with 2-fs steps. Temperature coupling during NVT and NPT equilibration was accomplished
768 with a modified Berendsen thermostat set to the reference temperature 300 K. Pressure coupling
769 during NPT equilibration was accomplished with the Berendsen algorithm set to the reference
770 pressure 1 bar. The equilibrated system was subjected to a 100-ns MD production run with 2-fs
771 steps, applying a modified Berendsen thermostat (300 K reference temperature) and Parrinello-
772 Rahman barostat (1 bar reference pressure). Short range interaction energies, distances, and
773 structures were extracted from the resulting trajectories with use of *GROMACS* functions and
774 plotted with the *R* package *ggplot2* (99). Figures and movies showing structures were created
775 with *ChimeraX* (45).

776 Recombinant DNA cloning

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

780 Technologies. LCHF MutY genes designed in this way were cloned into the low-expression
781 pKK223 vector by ligation-independent cloning (LIC). PCR reactions with the high-fidelity
782 Phusion polymerase (Agilent) amplified the synthetic gBlock and two overlapping fragments
783 derived from approximate halves of the pKK223 plasmid. PCR products were separated by
784 electrophoresis in 0.8% agarose x1 TAE gels containing 1 μg/mL ethidium bromide. DNA was
785 visualized by long-wavelength UV shadowing to allow dissection of gel bands, and the DNA
786 was purified with the GeneJet gel extraction system (Thermo Scientific), treated with Dpn1
787 (New England Biolabs) at 37 °C for 45 min, and heat shock transformed directly into DH5α
788 competent cells. Clones were selected on ampicillin media plates. The plasmid DNA was
789 purified from 4-mL overnight cultures by use of the Wizard Plus MiniPrep kit (Promega)
790 according to the manufacturer's instructions. The sequence of the LCHF MutY encoding gene
791 was verified by Sanger sequencing with UpTac and TacTerm primers. Genes encoding site-
792 directed substitution variants were created by amplifying two overlapping fragments of the
793 LCHF MutY-pKK223 plasmid with mutagenic PCR primers followed by similar gel purification
794 and transformation procedures. In our hands the LIC cloning efficiency was close to 95% except
795 for the *Flavobacteriaceae* MutY gene which could not be cloned intact. The pkk223 plasmids
796 containing the LCHF MutY encoding genes can be found on Addgene with ID numbers 210791-
797 210799.

798 Mutation suppression assay

799 Mutation rates were measured by the method outlined previously (33,57). The CC104
800 *mutm::KAN muty::TET* cells (100) were heat-shock transformed with a pKK223 plasmid
801 encoding the *Ec* MutY gene, LCHF MutY genes, or no gene (*null*). Transformants selected from
802 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: Tieker 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, Doublié 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

883 genes for DNA recombination and repair in the reductive genome evolution of
884 thioautotrophic symbionts of Calyptogena clams. BMC Evol Biol. 2011 Oct 3;11(1):285.
885 19. Garcia-Gonzalez A, Rivera-Rivera R, Massey S. The Presence of the DNA Repair Genes
886 mutM, mutY, mutL, and mutS is Related to Proteome Size in Bacterial Genomes. Front
887 Genet. 2012;3:3.
888 20. Jansson K, Blomberg A, Sunnerhagen P, Alm Rosenblad M. Evolutionary loss of 8-oxo-G
889 repair components among eukaryotes. Genome Integr. 2010 Sep 1;1(1):12.
890 21. Fowler RG, Schaaper RM. The role of the mutT gene of Escherichia coli in maintaining
891 replication fidelity. FEMS Microbiol Rev. 1997 Aug;21(1):43–54.
892 22. Boiteux S, O'Connor TR, Laval J. Formamidopyrimidine-DNA glycosylase of Escherichia
893 coli: cloning and sequencing of the fpg structural gene and overproduction of the protein.
894 EMBO J. 1987 Oct;6(10):3177–83.
895 23. Shida T, Noda M, Sekiguchi J. Cleavage of single- and double-stranded DNAs containing
896 an abasic residue by Escherichia coli exonuclease III (AP endonuclease VI). Nucleic Acids
897 Res. 1996 Nov 15;24(22):4572–6.
898 24. Parikh SS, Mol CD, Tainer JA. Base excision repair enzyme family portrait: integrating the
899 structure and chemistry of an entire DNA repair pathway. Struct Lond Engl 1993. 1997
900 Dec 15;5(12):1543–50.
901 25. Ljungquist S, Lindahl T, Howard-Flanders P. Methyl methane sulfonate-sensitive mutant
902 of Escherichia coli deficient in an endonuclease specific for apurinic sites in
903 deoxyribonucleic acid. J Bacteriol. 1976 May;126(2):646–53.
904 26. Lind PA, Andersson DI. Whole-genome mutational biases in bacteria. Proc Natl Acad Sci.
905 2008 Nov 18;105(46):17878–83.
906 27. Foster PL, Lee H, Popodi E, Townes JP, Tang H. Determinants of spontaneous mutation in
907 the bacterium *Escherichia coli* as revealed by whole-genome sequencing. Proc Natl Acad
908 Sci [Internet]. 2015 Nov 3 [cited 2022 Aug 19];112(44). Available from:
909 https://pnas.org/doi/full/10.1073/pnas.1512136112
910 28. Al-Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, Williams GT, et al.
911 Inherited variants of MYH associated with somatic G:C→T:A mutations in colorectal
912 tumors. Nat Genet. 2002;30(2):227–32.
913 29. Noll DM, Gogos A, Granek JA, Clarke ND. The C-Terminal Domain of the Adenine-DNA
914 Glycosylase MutY Confers Specificity for 8-Oxoguanine·Adenine Mispairs and May Have
915 Evolved from MutT, an 8-Oxo-dGTPase. Biochemistry. 1999 May 1;38(20):6374–9.
916 30. Fromme JC, Banerjee A, Huang SJ, Verdine GL. Structural basis for removal of adenine
917 mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. Nature. 2004
918 Feb;427(6975):652–6.
919 31. Lee S, Verdine GL. Atomic substitution reveals the structural basis for substrate adenine
920 recognition and removal by adenine DNA glycosylase. Proc Natl Acad Sci. 2009 Nov
921 3;106(44):18497–502.
922 32. Woods RD, O'Shea VL, Chu A, Cao S, Richards JL, Horvath MP, et al. Structure and
923 stereochemistry of the base excision repair glycosylase MutY reveal a mechanism similar
924 to retaining glycosidases. Nucleic Acids Res. 2016 Jan 29;44(2):801–10.
925 33. Russelburg LP, O'Shea Murray VL, Demir M, Knutsen KR, Sehgal SL, Cao S, et al.
926 Structural Basis for Finding OG Lesions and Avoiding Undamaged G by the DNA
927 Glycosylase MutY. ACS Chem Biol. 2020 Jan 17;15(1):93–102.
928 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, Brieba 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

975 multiple Amber force fields and development of improved protein backbone parameters.
976 Proteins. 2006 Nov 15;65(3):712–25.
977 51. Manlove AH, McKibbin PL, Doyle EL, Majumdar C, Hamm ML, David SS. Structure–
978 Activity Relationships Reveal Key Features of 8-Oxoguanine: A Mismatch Detection by
979 the MutY Glycosylase. ACS Chem Biol. 2017 Sep 15;12(9):2335–44.
980 52. Lee AJ, Majumdar C, Kathe SD, Van Ostrand RP, Vickery HR, Averill AM, et al.
981 Detection of OG:A Lesion Mispairs by MutY Relies on a Single His Residue and the 2-
982 Amino Group of 8-Oxoguanine. J Am Chem Soc. 2020 Aug 5;142(31):13283–7.
983 53. Liu X, Shi D, Zhou S, Liu H, Liu H, Yao X. Molecular dynamics simulations and novel
984 drug discovery. Expert Opin Drug Discov. 2018 Jan;13(1):23–37.
985 54. Guterres H, Im W. Improving Protein-Ligand Docking Results with High-Throughput
986 Molecular Dynamics Simulations. J Chem Inf Model. 2020 Apr 27;60(4):2189–98.
987 55. Nikkel DJ, Wetmore SD. Distinctive Formation of a DNA-Protein Cross-Link during the
988 Repair of DNA Oxidative Damage: Insights into Human Disease from MD Simulations
989 and QM/MM Calculations. J Am Chem Soc. 2023 Jun 21;145(24):13114–25.
990 56. Wang L, Lee SJ, Verdine GL. Structural Basis for Avoidance of Promutagenic DNA
991 Repair by MutY Adenine DNA Glycosylase. J Biol Chem. 2015 Jul 10;290(28):17096–
992 105.
993 57. Majumdar C, Nuñez NN, Raetz AG, Khuu C, David SS. Chapter Three - Cellular Assays
994 for Studying the Fe–S Cluster Containing Base Excision Repair Glycosylase MUTYH and
995 Homologs. In: David SS, editor. Methods in Enzymology [Internet]. Academic Press; 2018
996 [cited 2021 Jun 22]. p. 69–99. (Fe-S Cluster Enzymes Part B; vol. 599). Available from:
997 https://www.sciencedirect.com/science/article/pii/S0076687917303816
998 58. Garibyan L, Huang T, Kim M, Wolff E, Nguyen A, Nguyen T, et al. Use of the rpoB gene
999 to determine the specificity of base substitution mutations on the Escherichia coli
1000 chromosome. DNA Repair. 2003 May 13;2(5):593–608.
1001 59. Wu EY, Hilliker AK. Identification of Rifampicin Resistance Mutations in ***Escherichia***
1002 ***coli***, Including an Unusual Deletion Mutation. J Mol Microbiol Biotechnol.
1003 2017;27(6):356–62.
1004 60. Loeb LA, Preston BD. Mutagenesis by apurinic/apyrimidinic sites. Annu Rev Genet.
1005 1986;20:201–30.
1006 61. Michaels ML, Tchou J, Grollman AP, Miller JH. A repair system for 8-oxo-7,8-
1007 dihydrodeoxyguanine. Biochemistry. 1992 Nov 17;31(45):10964–8.
1008 62. Pope MA, Porello SL, David SS. Escherichia coli apurinic-apyrimidinic endonucleases
1009 enhance the turnover of the adenine glycosylase MutY with G:A substrates. J Biol Chem.
1010 2002 Jun 21;277(25):22605–15.
1011 63. Lu AL, Lee CY, Li L, Li X. Physical and functional interactions between Escherichia coli
1012 MutY and endonuclease VIII. Biochem J. 2006 Jan 1;393(Pt 1):381–7.
1013 64. Komine K, Shimodaira H, Takao M, Soeda H, Zhang X, Takahashi M, et al. Functional
1014 Complementation Assay for 47 MUTYH Variants in a MutY-Disrupted Escherichia Coli
1015 Strain. Hum Mutat. 2015 Jul;36(7):704–11.
1016 65. Jaenicke R. Protein stability and molecular adaptation to extreme conditions. Eur J
1017 Biochem. 1991 Dec 18;202(3):715–28.
1018 66. Závodszky P, Kardos J, Svingor Á, Petsko GA. Adjustment of conformational flexibility is
1019 a key event in the thermal adaptation of proteins. Proc Natl Acad Sci. 1998 Jun
1020 23;95(13):7406–11.

1021 67. Dong Y wei, Liao M ling, Meng X liang, Somero GN. Structural flexibility and protein
1022 adaptation to temperature: Molecular dynamics analysis of malate dehydrogenases of
1023 marine molluscs. Proc Natl Acad Sci. 2018 Feb 6;115(6):1274–9.
1024 68. Shaw TJ, Luther GW, Rosas R, Oldham VE, Coffey NR, Ferry JL, et al. Fe-catalyzed
1025 sulfide oxidation in hydrothermal plumes is a source of reactive oxygen species to the
1026 ocean. Proc Natl Acad Sci U S A. 2021 Oct 5;118(40):e2026654118.
1027 69. Messner KR, Imlay JA. The identification of primary sites of superoxide and hydrogen
1028 peroxide formation in the aerobic respiratory chain and sulfite reductase complex of
1029 Escherichia coli. J Biol Chem. 1999 Apr 9;274(15):10119–28.
1030 70. Imlay JA. The molecular mechanisms and physiological consequences of oxidative stress:
1031 lessons from a model bacterium. Nat Rev Microbiol. 2013 Jul;11(7):443.
1032 71. Price RJ, Lee JS. INHIBITION OF PSEUDOMONAS SPECIES BY HYDROGEN
1033 PEROXIDE PRODUCING LACTOBACILLI1. J Food Prot. 1970 Jan 1;33(1):13–8.
1034 72. Juven BJ, Pierson MD. Antibacterial Effects of Hydrogen Peroxide and Methods for Its
1035 Detection and Quantitation†. J Food Prot. 1996 Nov 1;59(11):1233–41.
1036 73. Bucci V, Nadell CD, Xavier JB. The Evolution of Bacteriocin Production in Bacterial
1037 Biofilms. Am Nat. 2011 Dec;178(6):E162–73.
1038 74. Rodionov DA, Dubchak IL, Arkin AP, Alm EJ, Gelfand MS. Dissimilatory Metabolism of
1039 Nitrogen Oxides in Bacteria: Comparative Reconstruction of Transcriptional Networks.
1040 PLOS Comput Biol. 2005 Oct 28;1(5):e55.
1041 75. Han S, Li Y, Gao H. Generation and Physiology of Hydrogen Sulfide and Reactive Sulfur
1042 Species in Bacteria. Antioxid Basel Switz. 2022 Dec 17;11(12):2487.
1043 76. Suzuki N, Yasui M, Geacintov NE, Shafirovich V, Shibutani S. Miscoding events during
1044 DNA synthesis past the nitration-damaged base 8-nitroguanine. Biochemistry. 2005 Jun
1045 28;44(25):9238–45.
1046 77. Joyner-Matos J, Predmore BL, Stein JR, Leeuwenburgh C, Julian D. Hydrogen Sulfide
1047 Induces Oxidative Damage to RNA and DNA in a Sulfide-Tolerant Marine Invertebrate.
1048 Physiol Biochem Zool PBZ. 2010;83(2):356–65.
1049 78. Bhamra I, Compagnone-Post P, O'Neil IA, Iwanejko LA, Bates AD, Cosstick R. Base-
1050 pairing preferences, physicochemical properties and mutational behaviour of the DNA
1051 lesion 8-nitroguanine. Nucleic Acids Res. 2012 Nov;40(21):11126–38.
1052 79. Ślesak I, Ślesak H, Kruk J. Oxygen and Hydrogen Peroxide in the Early Evolution of Life
1053 on Earth: In silico Comparative Analysis of Biochemical Pathways. Astrobiology. 2012
1054 Aug;12(8):775–84.
1055 80. Lu Z, Imlay JA. When anaerobes encounter oxygen: mechanisms of oxygen toxicity,
1056 tolerance and defence. Nat Rev Microbiol. 2021 Dec;19(12):774–85.
1057 81. Shock EL, Schulte MD. Organic synthesis during fluid mixing in hydrothermal systems. J
1058 Geophys Res Planets. 1998;103(E12):28513–27.
1059 82. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids
1060 Res. 2000 Jan 1;28(1):27–30.
1061 83. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for
1062 taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023 Jan
1063 6;51(D1):D587–92.
1064 84. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for
1065 Functional Characterization of Genome and Metagenome Sequences. J Mol Biol. 2016 Feb
1066 22;428(4):726–31.

1067 85. Thornton CN, Tanner WD, VanDerslice JA, Brazelton WJ. Localized effect of treated
1068 wastewater effluent on the resistome of an urban watershed. GigaScience. 2020 Nov
1069 19;9(11):giaa125.
1070 86. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation
1071 of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol.
1072 2011 Jan;7(1):539.
1073 87. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree
1074 display and annotation. Nucleic Acids Res. 2021 Jul 2;49(W1):W293–6.
1075 88. Pei J, Kim BH, Grishin NV. PROMALS3D: a tool for multiple protein sequence and
1076 structure alignments. Nucleic Acids Res. 2008 Apr;36(7):2295–300.
1077 89. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the
1078 analysis of massive data sets. Nat Biotechnol. 2017 Nov;35(11):1026–8.
1079 90. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify
1080 genomes with the Genome Taxonomy Database. Bioinforma Oxf Engl. 2019 Nov
1081 15;36(6):1925–7.
1082 91. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the
1083 proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 2003
1084 Jul 1;31(13):3784–8.
1085 92. Ku T, Lu P, Chan C, Wang T, Lai S, Lyu P, et al. Predicting melting temperature directly
1086 from protein sequences. Comput Biol Chem. 2009 Dec;33(6):445–50.
1087 93. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate
1088 protein structure prediction with AlphaFold. Nature. 2021 Aug;596(7873):583–9.
1089 94. Eastman P, Swails J, Chodera JD, McGibbon RT, Zhao Y, Beauchamp KA, et al. OpenMM
1090 7: Rapid development of high performance algorithms for molecular dynamics. PLoS
1091 Comput Biol. 2017 Jul;13(7):e1005659.
1092 95. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4
1093 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput
1094 Chem. 2009 Dec;30(16):2785–91.
1095 96. Forli S, Huey R, Pique ME, Sanner M, Goodsell DS, Olson AJ. Computational protein-
1096 ligand docking and virtual drug screening with the AutoDock suite. Nat Protoc. 2016
1097 May;11(5):905–19.
1098 97. Lemkul JA. From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the
1099 GROMACS-2018 Molecular Simulation Package [Article v1.0]. Living J Comput Mol Sci.
1100 2019;1(1):5068–5068.
1101 98. Kagami L, Wilter A, Diaz A, Vranken W. The ACPYPE web server for small-molecule
1102 MD topology generation. Bioinformatics. 2023 Jun 1;39(6):btad350.
1103 99. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. New York, NY:
1104 Springer International Publishing; 2016 [cited 2023 Nov 27]. (Use R!). Available from:
1105 http://link.springer.com/10.1007/978-3-319-24277-4
1106 100. Cupples CG, Miller JH. A set of lacZ mutations in Escherichia coli that allow rapid
1107 detection of each of the six base substitutions. Proc Natl Acad Sci U S A. 1989
1108 Jul;86(14):5345–9.
1109 101. Canty A, support) BR (author of parallel. boot: Bootstrap Functions (Originally by Angelo
1110 Canty for S) [Internet]. 2022 [cited 2023 Nov 27]. Available from: https://cran.r-
1111 project.org/web/packages/boot/
1112 102. Davison AC, Hinkley DV. Bootstrap Methods and their Application [Internet]. Cambridge:

Cambridge University Press; 1997 [cited 2023 Nov 27]. (Cambridge Series in Statistical and Probabilistic Mathematics). Available from: https://www.cambridge.org/core/books/bootstrap-methods-and-their-application/ED2FD043579F27952363566DC09CBD6A

103. Osborne MJ, Siddiqui N, Landgraf D, Pomposiello PJ, Gehring K. The solution structure of the oxidative stress-related protein YggX from Escherichia coli. Protein Sci Publ Protein Soc. 2005 Jun;14(6):1673–8.

Supporting Information

Data and code availability

Primary data and code for analysis of data have been deposited with GitHub: https://github.com/paytonutzman/Lost-City-MutY-Discovery. These include FASTA files for the discovered MutY enzymes, predicted structures, virtual docking outcomes, MD trajectories and scripts to analyze these trajectories, rifampicin resistance frequency data, and the *R* code to report statistics, median and 95% confidence intervals.

S1 Dataset. Alignment of Lost City MutY homologs

Chemical motifs are highlighted in columns. Alignment was generated by *Promals3D* guided by the structure of *Gs* MutY. It was necessary to align sequences in the first block including up to N146 separately from the second block and third block because otherwise the C-terminal domain residues were aligned inconsistently. The homologs flagged with dark red highlighting were eliminated because of missing chemical motifs. The homolog flagged with light pink highlighting required manual adjustment so as to align the H-x-FSH motif.

```
Conservation:                    9 5 579 67999 5  5   65  7         55      6   9 98989 9   5      5        8        65969 67  66 5          859
sp_P17802_MUTY_ECOL     28  KTPYKVWLSEVMLQQTQVATVIP-YFERFMARFPTVTDLANAPLDEVLHLWTGLGYYARARNLHKAAQQVATLHGGKFPETFEEVAALPGVGRSTAGAILSLSLGKHFPILDGN  140
sp_P83847_MUTY_GEOS     34  RDPYKVWVSEVMLQQTRVETVIP-YFEQFIDRFPTLEALADADEDEVLKAWEGLGYYSRVRNLHAAVKEVKTRYGGKVPDDPDEFSRLKGVGPYTVGAVLSLAYGVPEPAVDGN  146
c_000003652391_2        37  PDPYTVWLSEVMLQQTTVAAVTG-YFRRFTTRWPTVQALAAADDGDVMGAWAGLGYYARARNLLKCARVVCETHAGAFPNTYAGLIALPGIGPYTAAAIAAIAFDRAEVVVDGN  149
c_000000031207_6        26  EDPYAIWLSEVMLQQTRVTTVIP-YWHRFLEKYPDPKSLARAPLEQVLELWAGLGYYSRGRNLHAAACQIVNDHDGIFPRDPELIGKLPGIGEYTTAAICSIAYGEPLAVIDGN  138
c_000000430975_4        38  PNVYHVWLSEIMLQQTKVATVLS-YYLNFLHKWPTLNDLATATRQNVLQAWAGLGYYARARNLHACAQIVMEKYGGTFPKTEKELLRLPGIGHYTAAAIVAIAHGQPSVVVDGN  150
c_000001007148_4        35  RDPYVVWVREIMLQQTTVTAVVP-YLDRFLKRFPTVESLAGAKSAEVLRVWEGLGYYSRARNLSRGAKLIVEEWGGQWRNTATELQSLPGVGPYTAGAIASFAFDQSAGIVEAN  147
c_000001092848_3        48  PDPYRVWLSEIMLQQTTVGAVKP-YCERFLERWPTVEALARADLQEVMKAWAGLGYYSRARNLKKCAEEVAHRHGGVFPDTEAGLKALPGIGDYTAAAVAAIAFGRRAAVVDGN  160
c_000001151029_1         1  ------WLSEIMLQQTKVATVLS-YYSNFLHKWPTLADLAMATRHNVLQAWAGLGYYARARNLHACSQIVMEKYGGTFPKTEKELLKLPGIGHYTAAAIVAIAHGQPSVVIDGN  107
c_000001797282_1        58  PDPYRVWLSEVMLQQTTIAAVRP-YFERFLTRFPDIFALAEAPEEAVMSAWAGLGYYSRARNLHACARTVAA-AGGRFPDTAEGLRKLPGIGAYTAGAIAAIAFDRQEAAVDGN  169
c_000001803648_25       40  PDPYRIWMSEVMLQQTTVAAVAK-YFHAFTTRWPTIIALASANDDDVMGQWAGLGYYARARNLLKCARVVVADYGGNFPDTRDELLKLPGIGPYTAAAIASIAFGRSETVMDAN  152
c_000002078955_4        40  AEPYRVWMSEVMLQQTTVAAVAK-YFHAFTKRWPTVDALAAAKDQDVMGQWAGLGYYARARNLLKCARMVVSDYNSEFPETRDELLKLPGIGPYTAAAIASIAFGRPECVMDGN  152
c_000002106160_4        28  KEPYKIWLSEIMLQQTQVKTVIP-FYNSWIEKYPDFESVAGARLDSLLKSWEGLGYYNRCHNFHKAVKIIVNKYNSSLPINIKDFMVLPGVGSYTSAAVYSIVFLHPIPVVDGN  140
c_000003607531_2        38  KVPYKIWLSEIMLQQTTVTTVGP-YFQHFLKTWPTIDELANASLDDVLKVWQGLGYYSRARNLHKCAKLICNNFDGRFPNTEEQLEQLPGIGPYTAAAILAIAFDKPAIVVDGN  150
c_000003872363_4        36  PDPYHIWLSEVMLQQTTVAAVVK-YFNTFTKRWPNIQALATAQDADVMAEWAGLGYYARARNLLKCARMIVNDHGGVFPNDRAALLALPGIGPYTAAAISAIAFGRSACVVDGN  148
c_000004820107_1        43  PNPYYVWLSEIMLQQTTVVTVGP-YFTKFVERWPAIHDLAKADRDDVMHEWAGLGYYARARNLHKCAQIVSKELNGIFPEEQKDLEKLPGIGGYTSAAIRAIAFSKPANVVDGN  155
c_000005774797_1        29  PNPYHVLLSEFMLQQTTVATVKT-YFEQFTKKWPSVEHFSHATDDEIMAQWAGLGYYARARNLVKTIHTIH--NQGHFPASAYELQKLPGIGPYTSAAIAAIAFNEPILPRDGN  139
c_000005867021_3        40  PDPYRIWLSEVMLQQTTVVAVAK-YFHVFTKRWPNITALANAKDEDVMGEWAGLGYYARARNLLKCARMIVADFGGVFPDTRDDLLKLPGIGPYTAAAIASIAFGRSECVMDGN  152
c_000004546210_2        31  PSKYKTWISEIMLQQTQVNTVEP-YFHRFMERFPTVEELASSSPDEVLQIWSGLGYYARARNIHRSAQLIIKQHKGELPSDYESLLHLPGIGESTAGAILSLSGIEPKPILDGN  143
c_000004551008_5        31  PSKYKTWISEIMLQQTQVATVEP-YFIQFIQRFPEVTELASSSLDEVLQLWSGLGYYARARNIHRSARIIVNQHHSELPSDYESLLTLPGIGKSTAGAILSLSGIEPKPILDAN  143
c_000002511148_4        26  KTPYRVWVSEIMLQQTQVSTAIP-YFEKFIKEIPDVKSLSEASLDLVLSLWSGLGYYARARNLHKSAKIINLEHGGKIPSTRDELMALPGIGRSTAGAILALGFNKKAPILDGN  138
c_000000577378_2        14  KTPYRVWVSEIMLQQTQVSTVIP-YFKKFVREIPNIKSLSEASLDLVLSLWSGLGYYARARNLHKSAKIINLDHGGKIPSTGGELMALPGIGRSTAGAILALGFNMKAPILDGN  126
c_000000598175_2        37  INPYRVWVSEIMLQQTQVKTVIP-YFEKFMATFPDVQTLAHADQDLVLKHWSGLGYYARARNMHKAAKMICGDFKGQFPDNLEAMQSLSGIGRSTAAAILSIASNQPQAILDGN  149
c_000000754627_3        34  KTPYKVWVSEVMLQQTQVVTVIP-YFEKFMQSFPDIIALANADEDQVLHHWTGLGYYARARNLHKAAKIVRDKYQGQFPSTLEEVMDLPGIGRSTAGAVLSLSLGQHHPILDGN  146
c_000000811118_1        25  KDPYKVWISEIMLQQTQVTTVIP-YFNKFISEYPTIKTLASTSLDEVLSLWSGLGYYTRARNIHKTAKILKKDFDCKLPNEIEALMSLPGIGFSTAGAILSLGFEQSGVILDGN  137
c_000000899687_2        29  NDPYSIWVSEIMLQQTQVKTVIP-YYERWIKTLPTIDKLANAPEQKILKLWEGLGYYSRARNLKKSAKIICKEMNGKLPNTVKNLQNLPGIGRYTAGAISSIAFGLKAPVLDGN  141
c_000001176522_24       50  RTPYRVWVAEVMLQQTQVKTVVP-YFERFMEAFPAVDVLADAEMDAVLSQWAGLGYYARGRNLHKAARQIVAQ--GGFPETLAGWQALPGVGRSTAAAIMAQAYGQRAAILDGN  160
c_000001345122_1         4  ISPYRVWVSEIMLQQTQVKTVIP-YYQRFMKSFPSIEQLASATQDEVLAHWAGLGYYARGRNLHKSAQIIQTEYYGKFPQIFDDIIALPGIGKSTAGAIMSIALKQRMPILDGN  116
c_000001515736_6        18  RDPYRVWLSEIMLQQTQVSTVLG-YYQRFLDAFPDVASLAGAPQDAVLALWSGLGYYSRARNLHKCAQTVMQQWGGAFPQTAEELATLPGIGRSTAGAISSFCFSERVPILDAN  130
c_000001682161_6        25  PTPYRTWVSEIMLQQTQVKTVIP-YYLRFMESFPTIQSLAKASQEQVLAHWAGLGYYARGRNLHKTAQTIVEQHKGQFPNRFEEVVALTGIGRSTAGAILSIAMQQRMPILDGN  137
c_000001923643_29       28  RDPYRVWLSEIMLQQTQVSTVLG-YYQRFLDAFPDVASLAAAPQDAVLALWSGLGYYSRARNLHRCAQAVVQEWGGAFPRAEDLATLPGIGRSTAGAIASFCFSERVPILDAN  140
c_000002030994_5        26  INPYRVWVSEIMLQQTQVKTVIP-YYEKFMTSFPDVQTLASADQDTVLKHWSGLGYYARARNMHKAAQMVCDDFAGVFPDNLESMQSLSGIGRSTAAAILSIASNQQEAILDGN  138
c_000002038721_4        28  KSPYHIWISEIMLQQTQVKTVIP-YFHKFINSFPSIKSLAKSDLDKVLANWSGLGYYTRAKNIHQASKIIHQKYHGSIPKNYEQLIELPGIGESTAGALLTLAFDKPGIILDGN  140
c_000002523527_15       43  RDPYRIWLSEIMLQQTQVAAVIE-YFQRFVAALPTVAALAAASADEVMALWAGLGYYSRARNLHRCAKTVVDEHAGVFPTDPEVLVTLPGIGRSTAAAIAAFSAGVRSPILDGN  155
c_000002608528_2        26  INPYRVWVSEIMLQQTQVKTVIP-YYQKFMTTFPDVQTLANADQDLVLKHWSGLGYYARARNMHKAAKMVCDDFSGILPDNLEDMQSLSGIGRSTAAAILSIASNQSQAILDGN  138
c_000002961510_2        29  KDPYSIWVSEIMLQQTQVKTVIP-YYERWIKTLPTIGKLANAPEQKILKLWEGLGYYSRARNLKKSAQIICKEMNGELPKTVNDLKNLPGVGRYTAGAISSIAFGLKAPVLDGN  141
c_000002979152_8        34  RTSYRVWVSEIMLQQTQVVTVIP-YYERFMQSFPAVEGLAKATQNEVLDHWTGLGYYSRARNLHKAAQFVVEQCNGVLPESLDGLIELPGIGRSTAAAILSLAGNKSEAILDGN  146
c_000003248034_1        46  KNPYHIWLSEVMLQQTRVTTVVP-YYERFLDRFPTLESLATASEQSVLQAWSGLGYYRRAKYLRASAQQIINKWNGQLPKEYDKLRALPGFGAYTAAAVSSIAFGRKHAAVDGN  158
c_000003294679_3        31  KTPYSIWVSEIMLQQTQVKTVIP-YFKKFIVSFPTLRVLAQASLDDVLHHWSGLGYYARAHNLHKTACLVNEKFSAQLPLDMKELCLLPGIGRSTAGAILALSCNQKHPILDGN  143
c_000003333364_1        28  YDPYEVWISEVILQQTQMERGVA-YIERFLERFPTMESLAAAEEDEVLKVWEGLGYYARARNLRKGAQHMVEEHGGTFPADVEAARALPGVGRYTAGAVLSQAYNLEEPIVDAN  140
c_000003402697_5        28  REPYYVWVSEIMSQQTQIDRVADQFFPRFIETFPTVEALAAADWETVYPCWDGLGYYNRGKNLLKAAKVLVDKHEGELPQTVAGLEALPGVGKYTAHAVLAFAFDKKVPAIDTN  141
c_000003492925_4        25  DDPYHIYLSEVMLQQTQVKTVLERYYFPFLKKFPTLKALGDANLDDVLKMWEGLGYYSRAKNLHKTAGLIEV-----LPSDIDELIRLPGIGKNTAHAVATFVYKKPVPIMEAN  133
c_000003766664_5        41  PTPYKVWVSEIMLQQTRVETVIP-YFQRFLSSFPDVHALAAASLDEVLSRWSGLGYYSRARNLHKASQFVVESWDGIFPSDAKMLEEMPGVGRYTAGAIASIAFGMPAPIVDGN  153
c_000003948186_72       28  ITPYRVWLSEIMLQQTQVATVIS-YFERFVQTFPDVGALARAPLDEVLHLWTGLGYYARARNLHKAAGLIVDTHGSELPDSVQQLCELPGIGRSTAGAICAIALRKHAVIMDGN  140
c_000005057120_4        28  ITPYRIWLSEVMLQQTQVATVIP-YFNQFIDYFPDIHKLATAEDDAVLHLWSGLGYYARARNLHKTAKILSKN-KGIFPDDLDSLMELPGIGRSTAGAILSIAFNNSYPILDGN  139
c_000005590109_5        28  ISPYRVWLSEVMLQQTQVTTVIP-YFNHFIQKFPDVHQLANAPLDSVLHLWSGLGYYARARNLHKTAKIISET-DGEFPNDLNLLMQLPGIGRSTAGAISSIAFNKSQPILDGN  139
c_000006063368_2        26  KSPYKVWISEIMLQQTQVTTVIP-YFNKFLLRFPNIEKLAESELDEVMSYWSGLGYYSRARNLYKAALEITCKHKSKLPNSYNELVALPGIGKTTAGAILALGFQKKAAILDGN  138
c_000000467631_2         7  STPYQTWLSEIMLQQTQVTTVIP-YFNRFIQTFPNIATLADAPLDAVLQHWAGLGYYARARNLHKTA-IIINKNGGVFPDDVDGLLELPGIGRSTAGAILSIAFNKSHPILDGN  118
c_000000788235_1        27  ITPYRVWISEIMLQQTQVATVIP-YYQRFITSFPDVTTLANADEDLVLQHWAGLGYYSRARNLHKAAQQVRDEFSATVPADLDSLQSLVGIGRSTAGAILSLAYEQRHAILDGN  139
c_000001187176_11       28  MTPYRVWVSEIMLQQTQVATVIP-YYERFMAALPDVHALAAAPADEVLHLWTGLGYYSRARNLHRAAKRVVEECAGQFPGDVQGLTDLPGIGPSTAGAIASLSMGVRAPILDGN  140
c_000001660697_64       28  GDAYRTWLSEIMLQQTQVSAVLG-YYARFIERFPTVQALAAAPADDVMAAWAGLGYYTRARNLHRCAQIVVAEHGGIFPRDPEVLASLPGIGRSTAAAIAAFSYGVRAAILDGN  140
c_000003378864_2        39  RDPYAIWLSEIMLQQTQVETVLP-YYDRFLTAFPTIADLAAAPLGRVLKMWEGLGYYARCRNLHRAAQKIVSDHNGRFPATFEDVLALPGIGQSTAGAILTFAYDQRHPLLDGN  151
c_000003910742_3        28  PTPYRVWISEIMLQQTQVVTVIP-YYQNFMDSFPDVRTLAHAEEDLVLQHWAGLGYYSRARNLHKSAKLICDEFSGVIPDDIDQIQSLPGIGRSTAGAILSLAYQQQQAILDGN  140
c_000004116181_2         1  -----VWVSEIMLQQTQVATVVP-YFNRFMAQFPDVKTLAQTSLDAVLQHWAGLGYYARARNLHKTAKIIAQQ--GCFPNTLETLMQLPGIGQSTAGAIMSIAFQQSCPILDGN  106
c_000006223903_1        29  RNPYRIWISEIMLQQTQVTTVIP-YYERFMARFPHIVTLASAKLDTVLAEWTGLGYYARARHLHQAAGVIVARHGGVFPDQFHDVVALPGIGRSTAGAILAFSHGQRHAILDGN  141
c_000001029068_2         2  -NVYYVWLSEVVLQQTQVASAVG-YFSNFIKRFPTIGALSAASEDDVLSAWSGLGYYARARNLHKTAQIINRQYSGVFPSTVEQVLALPGIGRSTAGAILSIAYNQACPILDGN  113
c_000001128125_11       36  ISPYRVWVSEIMLQQTQVATVIP-YFKRFMLSFPNVEVLAKATQDEVLSHWAGLGYYARGRNLHKAARQIVEQHDGIFPSQFDDILALPGIGRSTAGAILSILNQRFAILDGN  148
c_000001595844_4        27  SDAYAIWLSEIILQQTRVAQGMP-YYIRFLEAFPTVFDLAKADEQEVMRLWQGLGYYSRARNLHFTAQQVANEFGGKFPDNYADLLALKGIGPYTAAAIASFAFGLPHAVVDGN  139
c_000001719155_2        31  RDPYRILVSEFMLQQTQVKTVIP-YFERWIKSFPTLQKLAEARESTVLKHWEGLGYYSRARNLRKAAQKIQQEYSGKVPESMNEIMKLPGVGRYTAGAVLSIAFGQKVPVLDGN  143
c_000002363038_6        36  INPYRVWVSEIMLQQTQVVTVIP-YFERFMQRFPEVSDLAAASQDEVLSHWAGLGYYARGRNLHKAAQLIVEQHNGVFPKSFDEVLNLPGIGRSTAGAILSISLQQRFPILDGN  148
c_000002561400_1         1  ---------EIMLQQTQVATVIP-YFERFMQRFPKVADLACADQEEVLTYWAGLGYYARGRNLHKAAQLIVEKHGGQFPSDFEEVLALPGIGRSTAGAILSISLEQRFPILDGN  104
c_000002566035_2        28  VSPYKVWVSEIMLQQTQVITVIP-YFKRFMKRFPDVQTLANSSLDEVLHYWTGLGYYARARNLYKTAKIITKEYQGNFPLSLNEIMSLPGIGRSTAGAVLSLSYDKCFPILDGN  140
c_000002598045_3        26  ITPYKVWVSEIMLQQTQVVTAIP-YFKKFIQEFPTVKSLSKATPDQVLKWSGLGYYARARNLHKAAKLICSNHEGKIPSSINELVELPGIGKSTAGAILSLGFKKKAPILDGN  138
c_000004188575_41       28  PTPYQVWVSEIMLQQTQVATVIP-YYQRFMIRFPDNATLAAASQDDVLSYWTGLGYYARARNLHRCAQYVAEIHQGQLPAELDALIALPGIGRSTAGAIMALAYHQRFSILDGN  140
c_000000240049_1        26  KSLYRVWVSEIMLQQTQVTTVIP-YFARFMLQFNDIEALSNATQDDVLLYWAGLGYYSRARNLHKAAQIIQQQHGGCFPRQYNDVLALPGIGPSTAGAILAQSQGQRHAILDGN  138
c_000002139456_2        26  KSLYRVWVSEIMLQQTQVATVIP-YFARFMQQFSDIEALSNATQDEVLLYWAGLGYYSRARNLHKAAQIIQQQHGGCFPQQYDDVLALPGIGPSTAGAILAQALGQRHAILDGN  138
c_000003535614_2        28  RSLYRVWVSEIMLQQTQVATVIP-YYQRFMQRFKNIAALAEATQDDVLLHWAGLGYYSRARNLHKAAQIITASHQGKFPENYHDVLTLPGIGPSTAAAILAQATGQRHAILDGN  140
c_000002657784_4        25  KTPYKVWISEIMLQQTQVVTVIP-FYSAFINRFPNIQELATSSEEEVMSFWSGLGYYSRARNIYKTAQILQKKFESKLPSSLSQLISLPGIGPSTAGAILSLGYKKRAPILDAN  137
c_000003800129_15       30  TTPYRVWISEMMLQQTQVNTAKD-YFLNFIEQYPDLESIKNASEEEILILWKGLGYYRRAKFIYQAKEKIFKDYNGLFPDEFEDIISLPGIGKSTAGAILSIAYGKPFPILDAN  142
c_000003716781_1        17  DDAYSVWISESMLQQTRVETVIP-YFERFLKRFPTVEALAEADIDEVLARWSGLGYYSRARRLHAAAREIQDRHDGEFPQTRDEAEDLPGVGSYTAGAVLSIAYGLDEPLVDGN  129
```

```
c_000001556689_3        18  RDPYAILVSEIMLQQTRVETVIN-RYGSFLERFPTIADLAIASLDDVLAEWSGLGYYRRPRLLHRLATVVVDTHGGEIPDTLEELLELPGIGPYTSAAVGSIAFGIPALSIDGN  130
c_000001834452_1         1  ------WISEIMLQQTRVEVVVP-YFEAFIKRFPSIEVLANAQEEEVLSYWAGLGYYKRSRLLLKAARVIQSEMDGVLPCDSATLRRLPGIGDYTAAAIASISFQEQIPVVDGN  107
c_000004255004_2        35  RDPYAIWVSEIMLQQTRVATVLP-YYYRFMKAFPSVEALDQAPLDAVRNVWSGLGYYQRAANLKKAASVLMNRHNGKLPADYQTLLGLPGVGQYTAGAVLSIAFDRRYPAPDGN  147
c_000000605438_3        31  QSAYEIYLSEIMLQQTQVKTVLERFYFQFLEKFPTLKAVANAPVDDILKAWEGLGYYTRARNLHKTAIVTKGI----LPKSAEALEQLSGIGKSTAHAVACFAFNEPLPILDAN  140
c_000000581237_15       31  QSAYEIYLSEIMLQQTQVKTVLERFYFQFLDKFPTLKDVANAPVDDVLKAWEGLGYYTRARNLHKTAVDTKGI----LPNNAEKLEELSGIGKSTAHAVACFAFDEPLPILDAN  140
c_000002529579_2        31  QNSYKIYLSEIMLQQTQVKTVLERFYFQFLEKFPTLEAVANAPVDDVLKAWEGLGYYTRARNLHKTAIATKGV----LPKRAEALEKLSGIGKSTAHAVACFAFNEPLPILDAN  140
c_000006097838_1         1  RDPYAIWVSEVMLQQTQVKTVIP-YWRRWMEQLPDIASLAAADEDTVIKLWEGLGYYSRARNLQKAARLLCEELSGKFPRDLCGVLALPGVGRYTAGAICSIAYNQPEPLVDGN  113
c_000000583727_6        27  RDPYAIWVSEIMLQQTQVKTVIP-YWQRWMKQLPDIASLAAADEDTVIKLWEGLGYYSRGRNLQKAAKRIYDELAGRFPRDLAGMLALPGVGRYTAGAICSIAYNQPEPLVDGN  139
c_000005807640_2        29  RDPYAIWVSEIMLQQTQVKTVIP-YWQRWMQRLPDIASLAAADEDTVIKLWEGLGYYSRARNLQKAAKRIRDALSDRFPRDLAGALALPGVGRYTGGAICSIAYNQPEPIVDGN  141
c_000000134878_1        12  NDPYKIWISEIMLQQTQVSTVQD-YYLRWIKRFPTIQSVADASIDEVLRNWEGLGYYARARNFYHSCKALSTTT-SCVPKGMDEFQKLKGVGPYIGAAVQSIAFNIPAGVVDGN  123
c_000005136725_3        27  SDPYKIWISEIMLQQTQVSTVRD-YYIRWIKQFPDIQSVADASIDKVLKYWEGLGYYSRVRNFHKSCQMLIRNK-QDIPKGLDEFQKLKGVGPYIASAVQSIAFNIPTGVIDGN  138
c_000001684786_4        24  PEPYRVWVSEIMLQQTRVETVIP-YFERFLDRFPTPAALAEASEDEVLTLWSGLGYYRRARSLLEGAKAVVRDHEGVFPRSYEAALDIPGVGPYTAAAILSIAYGEPWPVLDGN  136
c_000002971826_23       34  PEPYRVWVSEIMLQQTRVDTVIP-YFERFLDRFPTLSDLAAASEDEVLTLWSGLGYYRRARSLLAGAKTVIEHHAGVFPRDYEAALEIPGVGPYTAAAILSVAYGEPWPVLDGN  146
c_000002040695_2        37  ANPYFVFISEVMLQQTTVATVVP-YFKKWIKKWPTLHALAEASLDDVLHTWQGLGYYRARNVHRCAQEVVQSYGGLFPQNVSQLRKLPGIGPYTAGALSAIAFEQNSVAVDGN  149
c_000003263657_32       41  NDPYSVWISEMMLQQTQVTQAEP-YYERFMAQFPSVETLAGATLDDVLKAWEGLGYYARARYLHRAAVQVVDQYGGRIPESLSEISILPGIGPYTAAAILSIAYHRDHAAVDGN  153
c_000005989041_2        38  KDPYKIWLSEIMLQQTTVKAVGP-YFQLFLKTWPTIDKLAKATLDDVLRVWQGLGYYSRARNMHCAKLICNNFGGRFPNTEEQLERLPGIGPYTAAAISAIAFDKPANVVDGN  150
c_000002391082_2        36  DNPYHIWISEIMLQQTRVDQMGA-YFERFVGAFPTVEKLAAASEDQVLKVWEGLGYYARARNMHKAAVQIAGQLGGCIPDTYAGLIELPGIGEYTAAAVSSIAFDRDHPVLDGN  148
c_000003839553_2        43  SDPYHIWISEIMLQQTRVDQMGG-YFERFITAFPTLEALAGASGDEVLKVWEGLGYYARACNMHKAARIAFELEGQIPNTYDGLIDLPGVGEYTAAAVSSIAFDCDHPVLDGN  155
c_000003996707_2        34  GNPYHVWISEIMLQQTRVDQMQP-FFERFTAAFPTVGALAAASQEEVLKAWEGLGYYARARNLHRAAQRVV-EAGGNIPDRYEELVGLPGIGPYTAAAVSSIAFDRDHPVLDGN  145
c_000002044706_1        16  DDPYRILVSEFMLQQTRVTTVLR-YYPAFLERFPTLADLAAADQEDVLAAWSGLGFYRRARNLHACARQVVDRYDGEVPRDPEELGSLPGIGRYTVGAILSSAHNARLPILDGN  128
c_000006007075_1        31  KREYYTLVSEFMLQQTQVSTVIP-YFKNFIKDIPDLNSLAKINEKKLLKYWEGLGYYSRVRNLKKTAKIVVSKYKKKLPTTLEELKTLPGIGDYTASAILSIAFNKPFIPLDGN  143
c_000002094036_4         7  QKEYYTLVSEFMLQQTKVSTVIP-YFNNFLKNIPNMQSLAKINEKKLLKYWQGLGYYSRAKNLKKTAKIIVDNYQGKLPSDFEELKKLPGIGDYTANAISAIVFNKPTIPLDGN  119
c_000002762689_1        31  KKQYYTLVSEFMLQQTQVATVIP-FFNRFIKNIPDLEAALANFDNHKLIKLWEGLGYYSRVRNLKKTAQIVVKNFDKKLPRKYLDLKSLPGIGDYTASAISAIAFNKRIIPLDGN  143
c_000005516980_2        31  KREYYTLVSEFMLQQTQVSTVIP-YFKNFIKNIPNLKMLAKVQNKKLIKLWEGLGYYSRARNLKKTSQVIIKNFKGKVPENFEDLKSLPGIGNYTASAILAIAFNKPYIPLDGN  143
c_000003920004_1        37  PNPYYVLVSEFMLQQTTVNTVIS-RFKDFIKKWPNLKKLSMINENQILQFWSGLGYYARAKNLLNSAKIISLKFNNIVPDNYNDLIDLPGVGDYTAKAVLGIGYNKSVMPVDAN  149
c_000004008511_2        31  PDPYFVFVSEYMLQQTTVGTVKT-RFEEFILKWPSINDLARISEKTILNFWSGLGYYSRARNLLKAAKIIKKEFNSKIPNNYDDLINLPGIGDYTTKAILGIAYNKSVMPLDVN  143
c_000004481347_1        38  TNAYYVLVSEFMLQQTTVNTVTK-RFNEFIKLWPSIDKLSRISENRILRFWSGLGYYNRATNLLKTIKIISRDFKFKVPQHYDQLIELPGVGDYTAKAILGIAYNQPVLPLDAN  150
c_000003787733_3        30  NDLYKIWLSEVMLQQTRVNTVVP-YYLLWVKKYKSIKSVAKADYQNLLKLWEGLGYYSRCKNFYKSCLIIINDYDGKIPLDYLTFRSLPGVGDYIASAVLSIGLNQPHAAIDSN  142
c_000000754657_2        30  HDLYKIWISEVMLQQTRVNTVVP-YYLRWVEKYESIKSVAKADELKLLKLWEGLGYYSRCKNFHKACKIIVNDYDGEVPSDYQKFLLLPGVGEYTASAVFSIGLKQYYAAIDGN  142
c_000000990943_3        30  NDLYKIWISEVMLQQTRVNTVVP-YYLRWVEKYKSIKSVAKADDLKLLKLWEGLGYYSRCRNFHKACKIIVNDYGGEFPSDYQKFRLLPGVGDYTASAVFSIGLKQTYAAIDGN  142
c_000004474996_2        29  NDLYKIWLSEVMLQQTRVETVIP-YYLKWVEKYKTIKSVAEADGQKLLKLWEGLGYYSRCRNFHKACKIVNDFTGEIPSDYQNFRSLPGVGQYTAGAILSIAFNQPYIPALDGN  141
c_000001463500_11       17  NNLYHVYLSEVMLQQTKVKTAIP-YYKRWLKKYPTIESVALSNLDDLLKLWEGLGYYSRCVNFYKACIILYSQYNSQLPNNMTDFLKLPGVGNYISAAVYSIVDKIPAPAVDVN  129
c_000001286181_5        24  NDPYKVWVSEIMLQQTQVSTVIP-FYENFMKSFPDVLTLSKANLDEVLNNWAGLGFYSRARNLHQTAKIITDKYHQVFPNNYEDLVALPGIGEYTAGAILALAFNKKGTILDGN  136
c_000001293628_3        30  REPYRVWVSEIMLQQTRVETAIP-YYNRWMERFPTVDALAEGHEQAVLKAWEGLGYYSRARNLRRSALIVRERLGGVLPQDSRSLKTLPGIGESYSAGAIASIAYGEVVPAVDGN  142
c_000001535696_8        24  RDPYKIWLSEIMLQQTRVAQGAP-YYLKFIKHFATVYDLAKASEEQVLKLWQGLGYYSRARHLHATAKTIVNEYGGNFPNTYKELLKLKGIGDYTASAIASICFDRPEPVVDGN  136
c_000001614067_2        37  SNPYFVYLSEIMLQQTVVKTVIP-YFLKFVRKWPDINALAKAELHEINSYWAGLGYYSRCHNMFNAAKLI----NSTFPNNYDDLINLPGIGDYTAKTILAIAFDRKSNVVDGN  149
c_000001765289_1         3  DDPYKIWLSEIMLQQTQVKTVVP-YYNRWIKRYPSIKSVALADLGAVLKMWEGLGYTRCRNFHTAAKIVVKKFNGIIPNNWENFSSLPGVGDYTAAAVLSIAFNKPYAVMDGN  115
c_000001961666_1        18  VDAYKVWVSEIMLQQTTVNAVIS-RYKKFISRFTNVKVLAKAPLEEVLEEWAGLGYYSRARNLHSSAKIIVNKYEGIIPSEENKLINLPGIGTYTSGAIRAIAFNRPSLAIDVN  130
c_000002018097_4        25  NEPYNIWLSEVMLQQTQVKTVVP-YYTKWIKKFPTLKSVAESDLNSLLKLWEGLGYYSRCRNFHKATQIVVEKYKSVVPNDYKLFRQLPGVGDYIASAVLSIAYNKKYPAVDAN  137
c_000002843512_36       26  DDPYHIYLSEVMLQQTQLKTVLERYYFPFLKFRTLTLGETPLDDVLKMWEGLGYYSRAKNLHKTATLIEE-----LPSNIDELIKLPGIGKNTAHAVATFAYKKPVPIMEAN  134
c_000002930199_1        22  NNPYKIWISEIILQQTQIQTGIR-YYHKFIRHFPDINSLALAEEVDVLNVWQGLGYNRVLNMLYSAKIIVKQHKSTFPTKYDELIKLKGIGPYTAAAISSICKNEKRAVLDGN  134
c_000003033795_2        22  RTPYRVWVSEIMLQQTQVNTVIP-YYKKWIKKYPTLKSFKESNFDDVIKIWEGLGYYSRCHNMFNAAKLI----NSTFPNNYDDLINLPGIGDYTAKTILAIAFDRKSNVVGIDTN  130
c_000003888107_2        23  HDPYKIWISEIMLQQTQVSTVIP-YYNKWINKFPNATKLSKANYNDVLKLWEGLGYYSRCKNIYKTSHLV----QKNFPDSYSELIKLPGIGDYTAKTILAIAFDKNQVGIDTN  131
c_000004013286_6        44  RDPYAIWVSEIMLQQTQVNTVVP-FFERFIESFPTVEKLAAATEAAVLKRWEGLGYYSRARNLHRAAREVTDQYDGQLPRDPATLSRLPGIGQYTAGAILSIAFDCREPIVEAN  156
c_000004852258_1         6  SDPYKIWLSEVMLQQTQVETVIP-YYQKWINHYPTIKSVALENIDNLLKIWEGLGYYSRCRNFHEAIKDIFYNHNGVIPKQIEDFKSLKGVGIVISGAVMSIAFNQTHIAVDGN  118
c_000005254087_1        32  RDPYAVWVAEVMLQQTQIATVIP-YYERWLARFPTVQALAEASLDEVLKLWEGLGYYGRARNLHAAVQMVMNEYDGRIPTTAADLLKLKGIGRYTANAIASIAYNEAVPVLDGN  144
c_000005603677_1        29  QDPYKIWLAEIIMQQTRLEQGLP-YYERFIAAYPSIKELAETEEQKLMKLWQGLGYYARARNLQATAKRIVAEYAGIFPDNYKAILDLKGVGEYTAAAIASAFYADLPYAVVDGN  141
c_000003283462_3        17  TGFYRIWLSEIMLQQTQVKRVSI-FYTKWIKRFPTLESVANAKEEHLLKVWEGLGYNRCRNFYKATKIIIDDYNGKIPNDYQTFRLLPGVGDYTASAVLSIAFGMPYPAIDGN  129
c_000002687221_1        31  SNPYFIWLSEVMLQQTRVSTASP-YYQHWISALPDIHSVAEAPIDNILKLWEGLGYYGRARNFHKACRIIEKHGGVIPSDPLEFSKLPGVGQYTICAAVMSIAFNLPIPAADGN  143
c_000003347358_19       57  DEAYYIYLSEVMLQQTQVKTVLERYYIPFIERFPTLKHLGEAPLDDVLKMWEGLGYNRAKNLHKTATLVDQ-----LPSNIDELIKLPGIGKNTAHAVATFAFHQPVPIMEAN  165
c_000002701031_2        24  SDSYKIWLSEIILQQTRVEQGLP-YFKEFITKFPKIELLAKASEDEVLKLWQGLGYYSRARNMRLTAKAIVEKFNGVFPKDYEKILKLKGIGPYTAAAICSFAYNLPYAVVDGN  136
c_000000582753_3        32  KSSYKIWLSEIILQQTKVSQGLP-YYRSFIESYPRLEDLAHSSEEKVLKLWQGLGYYSRARNLHFTAQYIFNELSNEFPNNYKDLLKLKGVGDYTASAIASFAFDEVVPVLDGN  144
c_000001713769_5        24  NDPYSVWLSEIILQQTRIAQGLP-YYLKFINKYPTINSLAKANENDILILWQGLGYYSRARNLLKTAKFIVDERNGKFPTTYIELIKLKGIGEYTAGAISSICFNERRAVLDGN  136
c_000004369364_1        27  SDPYAIWLSEIILQQTRVAQGLP-YYLKFLEAFPTVQDLAQAPEQDIMLLWQGLGYYSRARNLHFTAKQVSDELNGLFPNSYTGLLSLKGVGPYTAAAIASFAYKLPHAVVDGN  139
c_000006057486_30       28  KDPYLIWLSEIILQQTKIAQGTP-YYLKFVEKFGNVSLLAKADEKQVLKLWQGLGYYSRARNLHKTAIYINNNLKGVFPNNYTDLKKLKGIGDYTASAIASFCFNLPEATVDGN  140
c_000005494072_10       24  KDPYKIWLSEIILQQTKIEQGLP-YYMSFVENYPNIETLAIADEQEVLKLWQGLGYYSRARNLHFTAKSIVKNHNGKFPDNYNELLKLKGVGDYTASAISSICFDEKCAVVDGN  136
c_000001742634_3        27  KKPYFIWLSEIILQQTRTTQGLP-YYIAFIKHFPTIFDLAKTSETKVLKLWQGLGYYSRARNLHFTAYVIDNYNGKFPENYKELIKLKGVGDYTASAIASICYNEKTAVVDGN  139
c_000002826998_2        27  KNPYLIWLSEIILQQTRVEQGLP-YFLKFKEKYPTVTDLANAEEDEIMKLWEGLGYYSRARNLHFSAKYITAELDGIFPKKHTEILKLKGVGTYTAAAISSFAYDLPHAVLDGN  139
c_000003159439_6        24  SNPYHIWISEIILQQTRVDQGLD-YYKRFIATFPTINNLAKAKEEAVLKLWQGLGYYSRARNLHFSAQYIVNELDGKFPSSYKELLKLKGVGDYTASAIASICYNEVTAVVDGN  136
c_000004887214_1        29  REPYKIWLSEIILQQTRVAQGLP-YYQRFIESFPILQNLAAASEQNVLKLWQGLGYYSRARNMHFTAQYVVNELGGEFPCSYKELMKLKGVGDYTASAVASFCYDEVQPVLDGN  141
c_000005037037_2        27  KNPYYIWLSEIILQQTRVEQGTK-YFLAFKKNYLTIFDLAQAKETEVLKLWEGLGYYSRARNLHFTAKYIANDLEGVFPRNYTDLLKLKGVGSYTASAIASFAYDIPKAVVDGN  139
c_000001059964_1        19  YEPYQVWISEIMLQQTQVQTMLP-YYHRWMERLPRLQDVAEASEDLLIKLWEGLGYYRRVHNIKKTAVICIRDYGGTFPGSYRQQIQLPGIGPYTAGAIASIAFNFSEPVLDGN  131
c_000002498472_1        28  YSPYQVWISEIMLQQTQMDRVTL-FFTRWMDTFPDVNTLASASETRVLKLWEGLGYYSRARNILKTAQLLVTEYQGQIPDDPKQLLSFPGIGPYTAGAIASISFNHDVPVVDAN  140
c_000004615912_3        17  KNPYLIWLSEVILQQTRTAQGLP-YYEKFKNRFPTIDLMANAHEDEIMKLWEGLGYYSRARNMHFTAKYITEELGGEFPDTLEGLKKLKGVGDYTAAAIASAFYADLPHAVLDGN  129
c_000004766858_2        30  YHPYEIWVSEIMLQQTQVKTMLP-FYTRWMSELPTVHSVAAASEEMVLKLWEGLGYYSRARNLQKAARVICDEHGGELPADFDHIRALPGIGPYTAGALASIAFNQPRPVVDGN  142
c_000002747260_18       36  PDAYAVWLSEVMLQQTTVAAVRD-YFRRFTTRWPTVADLAAAEDADVMGEWAGLGYYARARNLLKCARVVVAEHGGHFPDTRDGLLKLPGIGPYTAAAIAAIAFDEPAVVVDGN  148
c_000005371561_38       36  PDPYRVWLSEIMLQQTTVAAVRD-YFHRFTARWPTVADLAAAEDGDVMGEWAGLGYYARARNLLKCARVVVSEHGGRFPDDPDALLALPGVGPYTAAAIAAIAFDKPETVMDGN  148
c_000000169465_2        32  VTPYKVWVSEIMLQQTQVTTVIP-YFKRFMASFPTVHDLAKASQDDVLHHWTGLGYYARARNLHKAANTLVEKYNGEFPYSLEEVMDLPGIGRSTAGAILSLSRNMRFPILDGN  144
c_000003254110_11       32  VTPYKVWVSEIMLQQTQVTTVIP-YFERFMASFPTVHDLAKASQDDVLHHWTGLGYYARARNLHKAANRLVDEYNGEFPFSLEEVMDLPGIGRSTAGAILSLSRNMRFAILDGN  144
c_000004750284_20       32  VTPYKVWVSEIMLQQTQVATVIP-YFNRFMASYPSVFELANASQDDVLHHWTGLGYYARARNLHKAAKTLVEDYEGVFPPTLDEVVSLPGIGRSTAGAILSLSRDERYPILDGN  144
c_000000141782_15       37  ATPYRVWISEVMLQQTQVATVIP-YFERFLQRFPDVATLAAADADSVLHLWSGLGYYARARNLHRAARVVVEAHGGEFPRRFEEVQALPGIGRSTAGAILSLGCGQRHAILDGN  149
c_000002717847_8        31  ITPYKVWISEMMLQQTQVVTAQS-YYSKFLKKYPSLTSMKMATEDDILSLWSGLGYYRASFIYQAKEKIFLDFNGRFPATYDELITLPGIGRSTAGAILSIAYQKPFPILDAN  143
c_000004054799_1        18  ISPYRVWISEIMLQQTQVKTVIP-YFYKFIKKYPNLGALLQASEDEILAQWSGLGFYRRAKNIYKACRIISEDFNSKLPTNINDLESLPGIGRSTAGAIMSIAFNEGHPILDAN  130
c_000006126673_1        43  PNPYFVWLSEIMLQQTTVSAVIP-YFLKFTQRWPDVWALANADVDDIMSAWAGLGYYARARNMHKCAKTIVSDYDGIFPQEQAVLKTLAGIGDYTSAAITAIVFDKYATVVDGN  155
c_000000339186_3         1  -DPYSIWVSEIMLQQTQVSTVIA-YWERWMRALPTIDKLARAKPTQVLKLWEGLGYYSRARNLQVAAKEIVKNHDGQFPRDRPAIRALPGIGRYTAGAISSIAFGLPEPIVDGN  112
c_000003029168_2        32  RDPYAIWISEIMLQQTTVESVIP-IYERFMELFPTVDVLAVANPTDVLQAWAGLGYYKRARHLHEAAIEIVNRFGSRIPDHLSDLQSLPGIGRYTAGAIASIAFDQCVPVVDTN  144
c_000001863436_1        48  GDPYAIWVSEVMLQQTQVVTVID-YYTRWLARFPDVETLAAAEIDEVLELWAGLGYYRARLLHRAAQHVVEECGGALPQTAGELEDLPGIGPYTAGAIASIAFGEAAPLVDGN  160
c_000002830137_4        41  RDLYHIWVSETMLQQTQVATVVP-YFQRFVERFPNITSLAAADEEDVLRLWEGLGYYRARANLHRAAQQVVQQHHAQFPADFDTLLSLPGIGRYTAGAILSIALDQRLPILEAN  153
c_000004612302_3        35  RDPYAVWIREIMLQQTTVKAVVP-YLERFLDQFPTIEVLAASETDEVLRVWEGLGYYSRARNLHRAARMIVEEWGGQWRETAVELQALPGVGPYTAGAISAFAFDQPAAIVEAN  147
c_000001933926_1        21  TDPYRVWLSEIMLQQTTVATVTP-YFLRFVERWPTVHDLAAADLDRVLHAWQGLGYYARARNMHACAKRISGDMAGVFPDQENELRRLPGIGPYTAAAIAAIAFDRPATPVDGN  133
c_000003136334_1        33  MDPYRVWLSEIMLQQTTTITVRG-YFEDFVSRWPSIEDLAAAELDAVLHAWQGLGYYARARNLHCAKLITKEYGGKFPEREEALVKLPGIGPYTAAAIAAIAFGHIATPVDGN  145
c_000005849454_9        30  RDPYRTWVSEIMLQQTRVEAVIP-YFEAFLKRFPDVHALATATEEEVLQRWAGLGYYRARAFLHAAAQKVVEDWDGRFPETADTLRAMPGVGATSAAIASLAFGEAVPVVDGN  142
c_000001169194_1        41  TDPYRILVSEIMLQQTRVETVLR-YYESWLKRFPTLGALASANSREVLKAWEGLGYYRARANLHSAARRIRELPGGHFPTSYRDLRKLPGVGQYTAGAVASIGFGEIVPAVDGN  153
c_000002786947_2        41  TDPYRILVSEIMLQQTRVETVLR-YYESWLKQFPTLGTLASADSTEVLKAWEGLGYYRARANLHRAAKHIRNVLGGRFPTSSRDLRQLPGVGQYTAGAVASIGFGEIVPAVDGN  153
c_000001279808_52       28  RSPYYVWISEIMSQQTQIDRVAEKFFPRFIATFPTVEALAAASWDEVYPVWEGLGYNRGKNLLKAAQIIVEKHDGEIPQTAVELEALPGVGKYTAHAVLAFAFDQKVPAIDTN  141
c_000000358065_2        31  QKQYFILISEFMLQQTQVKTVIP-YFENFINKIPNLKSLAKVNNRKLMKCWEGLGYYSRAKNLKKSAKIIDNYKGRLPNDFEELKKLPGIGDYTASAISAIAFNKPIIPLDGN  143
c_000003745941_1        22  AALYPTVVSEFMLQQTQVKTVLP-YFHRWLKRFPDFKSLAKAKEETVLKHWEGLGYYSRARNLHRLAREILALN--SIPEDPESWQQFKGIGAYTAAAITSISFNHPVACVDGN  132
c_000004187032_4        31  KKHYYTLVSEFMLQQTQVVTVIP-YFDRFINKIPNLKKLSTIQDNELIKLWEGLGYYSRARNLKTAQIISKHHGKIPNNYEDLISLPGIGNYTANAILAIAFNKSYIPLDGN  143
c_000006067315_3        31  QTEYFTLVSEFMLQQTQVKTVIP-YFNNFINKIPNLKKLANVNETRLMKCWEGLGYYSRARNLKKTAKIIISEFNSNLPNTIEELKNLPGIGDYTSRAVLAIAFNKPIIPLDGN  143
c_000005543774_1         1  KKLYYRILSEFMLQQTQVKTVIP-YFNNFVKKVPSKLALSKSNEKKILKLWEGLGYYSRAKNLHKTSKILVKKYDSQIPDSFEKLKELPGIGDYTANILLAIYNHPRIGIDGN  113
c_000006211484_1        23  KKQYYRILSEFMLQQTQVKTVIP-YFNNFVKKIPSLKALSNSDEKTVLKLWEGLGYYARAKNLHKTSKILIKKYNGQIPKSFVKLKELPGIGDYTANILLAIYNQPCIALDVN  135
c_000002718976_3        32  KKLYYRLLSEFMLQQTQVKTVIP-YFNKFTKKFKTIKALSKSNEREILKMWEGLGYYRARNLLACSKILVKNYKSRLPRSIIEIKKLPGIGDYTANALLGLVYNEPRIAVDGN  144
c_000002992548_1        33  KEQYYRILSEFMLQQTQVKTVIP-YFKNFINKIPNLKVLSSSNEKTILKLWEGLGYYRARANLHKTAKILTKNYKGKLPKEFDEIKKLPGVGEYTANALLALVYNQPRIPFDGN  145
c_000000456751_1         1  ----YVLVSEFMLQQTQVKTVIP-YFTKFVKKIPNLKLSTSNEKKILKLWEGLGYYRARANLHTAKILIKNCKGKLPKKFAEIIKLPGVGEYTANSLLALVHNKPSIPIDGN  109
c_000002655634_1        28  KKEYYVLVSEFMLQQTQVKTVIP-YFKKFVTKIPNLKTLSTSNEKQVLKLWEGLGYYRVRNLHKTAKILLKQYKAKLPKKFQEIIKLPGVGEYTANSLSALIHNKPCIPIDGN  140
c_000004232403_2        27  NDPYKIWLSEVMLQQTRVETVIP-YYNTWLENYKTIQSVAQANLDDLLKIWEGMGYYSRCKNFYYAAKEICKYKGTIPNQFDVFKSLPGVGDYTAGAVMSIAFNQPHPSIDGN  139
6u7t.before.V147.pdb    28  RDPYKVWVSEVMLQQTRVETVIP-YFEQFIDRFPTLEALADADEDEVLKAWEGLGYSRVRNLHAAVKEVKTRYGGKVPDDPDEFSRLKGVGPYTVGAVLSLAYGVPEPAVDGN  140
Consensus_aa:                .sSY.l@lSElMLQQTpV.hhhs.Y@.p@hpp@PslpsLA.tpbpplphWpGLGYYsRt+Nlhphtp.l.pph.t.hPpshp.l..LPGlG..TAtAlhtht@sp..shlDtN
```

```
Conservation:                  95  9                                                  75 87 85  9    5 9      8     9 5
sp_P17802_MUTY_ECOL     141  VKRVLARCYAVSGWPG------KKEVENKLWSL----SEQ-VTP---------AVGVERFNQAMMDLGAMICTRSKPKCS----LCPLQNGCIAAANNSWALYPGKKPK-----  225
sp_P83847_MUTY_GEOS     147  VMRVLSRLFLVTDDIA------KPSTRKRFEQI----VRE-IMA---------YENPGAFNEALIELGALVCTPRRPSCL----LCPVQAYCQAFAEGVAEELPVKMKK-----  231
c_000003652391_2        150  VERVMARLFDISTPLP--------GAKPALKAR----AAS-LTP---------KLRPGDYAQAVMDLGATICTARAPTCG----VCPLNDCCHARRAGTAAELPKRTPK-----  232
c_000000031207_6        139  VERVLSRYLKIDGNPK------SGTARKRIQQA----ADD-ALD---------QENPGDHNQALMDLGRSICTPRSPDCQ----RCPLASGCGARASGEPTRWPTRRQK-----  223
c_000000430975_4        151  VERVVSRIFQITQPIP--------LSRPTLRQL----AAG-LTP---------SKRPGDYAQAIMDLGSLVCTPRKPKCN----VCPLHDICLANISGTPSDFPKRTPR-----  233
c_000001007148_4        148  TARFYARLLALRSDLH------SGEGRRQLWSL----AER-IVP---------RDRPGEFNQALIDLGATVCGV-TPRCQ----ECPVRGSCRSLELGEVERIPARASR-----  231
c_000001092848_3        161  VERVITRLRAIGTPLP--------AAKPQVRKA----VAA-LVP---------AERPGDFAQAMMDLGATICTPRRPSCI----LCPLQAFCAARGAAAQESFPVKASK-----  243
c_000001151029_1        108  VERVVSRIFQITQPIP--------LSRPTLRQL----AAG-LTP---------SKRPGDYAQAMMDLGSLVCTPRNPKCN----VCPWHDICLANISGTPSDFPKRTPR-----  190
c_000001797282_1        170  VERVLSRAYAVEAPLP--------GSRPEIRRL----TQA-LVP---------PDRPGDFAQALMDLGATICTPKRPACA----LCPWMRPCRARSLGTQESFPRKIKV-----  252
c_000001803648_25       153  IERVMARMHDIHSPLP--------SAKPELLAC----AAA-LTP---------DFRAGDYAQAVMDLGATICTPKSASCG----VCPWSFACLARINGTAAVLPKKTPK-----  235
c_000002078955_4        153  VERVMARLHNIHTPLP--------AAKPELFSK----ATA-LTP---------DLRAGDYAQAVMDLGATICTPKSPTCG----ICPWSCACVALISGTAAELPKKTPK-----  235
c_000002106160_4        141  VNRVMSRLLRLLRFTP--------YNKKRVLST----LIR-WLD---------RSRPGDFNQALMELGSTVCLPKNPKCE----NCPLKLSCGGTLSGSPESYPSPKPK-----  223
```

```
c_000003607531_2      151  VERVITRVFGIFKPLP-----EARPIIKDCAAK----LAP-KTT---------RGRPGDYAQAVMDLGATVCRPRKPLCD----LCPWEKYCIANQQNFVDQVPRKAPK-----  236
c_000003872363_4      149  VERVMARLFNVITPLP--------AAKPELTAH----AAA-LTP---------VDRAGDYAQAVMDLGATICTPKSPGCG----ICPCTEYCAALPEGDAANLPLRPPK-----  231
c_000004820107_1      156  IERIMARLYAITEPLP------DSKANLKNLAA----GLS-EER---------KDRPGDYAQALMDLGATICTPTSPKCS----LCPVNDSCTAKRQGIAESLPIKKKK-----  240
c_000005774797_1      140  INRVLSRVYNIDDAIN--------RPSARLARI----AET-LAH---------PKRSGDVAQAMMDLGAAICKPKNPNCE----TCPICDQCKAHSLGRTHELPVKAPQ-----  222
c_000005867021_3      153  VERVMARLYNIHSPLP--------AVKPMLLTH----ATA-LTP---------EFRAGDYAQAVMDLGATVCTPKSPTCG----ICPWSFDCIAHAKGTAAQLPKKTPK-----  235
c_000004546210_2      144  VKRVLSRFFGIKGWSG------ESKVSKELWDL----SAK-SLP---------VDNFQVYTGGIMDLGATICLSRNPNCS----ICPIQRRCYAHQNHLVDEIPSKKPK-----  228
c_000004551008_5      144  VKRVLSRFFGVKGWSG------ESKVSKELWDL----SAK-SLP---------VDNFEIYTGGIMDLGATVCLPKNPNCS----NCPIVNKCYSYLNQLVDEIPSKKLA-----  228
c_000002511148_4      139  VKRVLARHFKIEGKLE------TSSKRKELWQI----SED-LLP---------DNKIEVYTGAIMDLGATLCIKSSPACN----LCPLNQDCLALKEDLVTVLPSKTKP-----  223
c_000000577378_2      127  VKRVLARYFKIEGKLE------TSSKRKELWQV----SEG-LLP---------DNKIEIYTGAIMDLGATVCTKSNPVCN----LCPLNQDCLALKEDLVTVLPSKTKP-----  211
c_000000598175_2      150  VKRVLTRVFVVDGWPG------KADVLKKLWLL----AEQ-TTP---------ASRNADYTGAIMDLGATLCTRSKPRCE----DCPLKADCLAFQQGNQTVYPTSKPK-----  234
c_000000754627_3      147  VKRVLARFFMVEGWYG------VKKVENQLWHL----SEQ-LTP---------KNNVTEFNQAMMDLGASLCSRSRFDCE----ACPLNSRCGAFNAGKVKEFPHSKPK-----  231
c_000000811118_1      138  VKRVLLRISGSKSPMN------KYKTEKSLKEF----ANI-LLP---------NSKHKEYSQGLMDLGAIICRPNNPNCS----ECPVDEDCAANQKEIQDEIPIKITK-----  222
c_000000899687_2      142  VKRVLSRLFCINKNGT------TSAFENCLWGE----AED-LLP---------VRRPGDFNQALMELGATVCIPASPICQ----QCPLRTICKAFLKNKVNEFPPSKNK-----  226
c_000001176522_24     161  VRRVLCRLFGVEGWPG------ERQVEAHLWTL----AEC-LLPA-------QSEVMPDYTQAQMDLGATLCTRTKPRCE----ACPFQAECVAYRQGRVAELPAPRPR-----  247
c_000001345122_1      117  VKRVLCRFDAVTSWSG------NKKTEKVLWDR----ADE-FTP---------QERFDDYTQAIMDLGATLCTRSKPKCD----VCPVEADCDAKKMDKVIQFPYSNPK-----  201
c_000001515736_6      131  VRRVLTRVLAFDGDLA------QSRNERQLWEY----AQQ-LCPTE-----NLHEAMPRYTQGMMDLGASVCTSRKPTCL----VCPLDSECRAARAGNPENYPVRTRR-----  219
c_000001682161_6      138  VKRVLCRFDAINSWSG------NKQTESQLWQR----AEQ-LTP---------SKRVKDYTQAIMDLGATVCTRSKPSCE----ICPVAENCQARQLNQVAKYPFSKPK-----  222
c_000001923643_29     141  VRRVLTRVLAFDADLA------VARNEERELWDL----AQQ-LCPTE-----DLQQAMPRYTQGLMDLGATICTPRKPSCL----VCPLQPQCRAARFGNPENYPVRTRK-----  229
c_000002030994_5      139  VKRVLSRVFTVVGWPG------KADVLKKFWVL----AEQ-TTP---------KTRNADYTQAIMDLGATLCTRSKPRCE----DCPLKTDCLAFQQGNQTEYPTKKPK-----  223
c_000002038721_4      141  VKRVLLRVLGNKSPIN------HSKTQKELKKF----AAE-LLP---------KNDFRTYSQGIMDLGSMICTPINPTCK----NCPIQNKCKSFQLNLQNMIPVKTSK-----  225
c_000002523527_15     156  VKRVFARVFGIHGYPG------ERAIENRMWAL----AEQALPPAG----RHQADHMVAYTQGLMDLGATICSRGKPACLADAEACPLVADCVARKDNLTAVLPTPKPR-----  250
c_000002608528_2      139  VKRVLTRVFTVEGWPG------KSGILKKLWLL----AEQ-TTP---------QTRNADYTQAIMDLGATLCTRSKPRCE----DCPLRAGCLAFQQGNQTDYPTKKPK-----  223
c_000002961510_2      142  VKRVLSRLFCINENGT------TSASENRLWQQ----AED-LLP---------VRRPGDFNQALMELGATVCIPKSPICQ----QCPMRTLCKAFLKNKITEFPPSKKK-----  226
c_000002979152_8      147  VKRVLARYHAIEGWPG------KSAVLKELWEK----AEL-HLP---------AERNADYTQAMMDLGATLCSRSKPACE----RCPVVDGCAAARKSSWALYPGKKPR-----  231
c_000003248034_1      159  VLRVLTRVLDEKGNVA------STAIQQKLTAA----ARL-LVES------VGPHSAGTWNQALMEVGATVCTPQSPSCD----RCPINRWCSAWKAGTQKTRPVKARP-----  246
c_000003294679_3      144  VRRVFTRHKAIKGWPG------TTMVQNVLWKI----AED-NTP---------SNRVADYTQAIMDLGATICTRSHPRCE----LCPVNKDCIAHSRGQEAEFPHKNIK-----  228
c_000003333364_1      141  VERVLARLFDVDAPIK------SSQAQSFLWET----ARS-LIP---------TGQAREFNQAIMEFGNLVCGR-KPSCM----FCPLTEECEAHYLGIVEHRPVPAER-----  224
c_000003402697_5      142  ISKIITVLWPG----------------QDTVEV----ANE-LVS--------LSQSGYAWNSAMMDLASSLRAGNPVE-------GALGETYFSDEAICRLFKPKRKKPEKKIK  219
c_000003492925_4      134  VKRILCRLHKLKTP----------SDKELWRI----AYN-LVD---------KKNPFDYNQAMMDIGATLCTPKKPDCE----TCPLSDICQGGNDPTLYPTKKKR-----  211
c_000003766664_5      154  VIRVLSRLCDLEEEVG------ETKAKERIWDW----AER-LVD---------PQNASAFNQGMMELGAMLCTPTSPGCD----QCPWREDCRARAAGTVLERPIKKAK-----  238
c_000003948186_72     141  VRKVLSRVHRVAGWAG------RASVNKELWQL----AER-YTP---------KRRVERYTQAIMDLGATVCVRGQPRCH----ECPVAKHCEALRHDCVQSFPQPKAR-----  225
c_000005057120_4      140  VRKVLTRYFAISGWTG------NSKISNELWEI----SSR-YTP---------KIRVADYTQAMMDLGATLCTRSKPQCT----ICPIQADCLAKKQNKTAELPTPKPA-----  224
c_000005590109_5      140  VRRVLARFHVISGWTG------GTKVSNKLWQL----SSD-YTP---------LRIADYTQAIMDLGATICIRSKPKCT----QCPINSACLAKIEGKISELPTPKPK-----  224
c_000006063368_2      139  VKRVLTRHRKIETDLS------KSSTLKKLWKL----SEL-LLP---------DVRIDTYTQAIMDLGAIVCKKKSPLCE----ICPVNKDCLALKANLVDCLPVKRKI-----  223
c_000000467631_2      119  VKRVLARYHAISGWTG------SAKISSELWKI----SAR-YTP---------AERCADYTQAMMDLGATLCTRSKPRCG----ECPISSDCVARIGDKVKLLPTPKPA-----  203
c_000000788235_1      140  VRKVLTRYHAIEGWSG------KGHILKQLWSI----AES-HLP---------DTRYANYTQAMMDLGATVCTRSKPKCL----GCPIHADCSAFQAGTPTAYPTPKPK-----  224
c_000001187176_11     141  VRKVLARYHAVQGWPG------EKPVHDRLWAI----AEA-HTP---------HSRVAHYTQAMMDLGATLCTRSKPSCL----ICPLQAGCQARMLGEPTAYPQARPR-----  225
c_000001660697_64     141  VRKVFARVFGIDGFPG------DKRVEDTLWRI----AEAVLPP---------ADGIQPYTQGLMDMGATVCTRGKPACLTGERACPLESLCEARSTDRVMELPVPRPR-----  230
c_000003378864_2      152  VRKVLSRLYDVSDP----------VGTAAATRTLWAYSTELLN-------QADDAYTHNQAIMELGATLCTPKKTACT----RCPVSVRCDALARDTVPLRPIKRPR-----  237
c_000003910742_3      141  VRKVLTRYHAVEGWTG------KSQVLKQLWLL----AES-HLP---------ATRYANYTQAMMDLGATVCSRTKPDCD----HCPLNSNCAALSKGNPTAYPTPKPK-----  225
c_000004116181_2      107  VRKVLARFNAVQGWPG------NASVNKQLWAI----SAR-YTP---------VERVADYTQAMMDLGATVCTRTRPGCQ----ICPVNENCLARKKSLTEIIPASRPV-----  191
c_000006223903_1      142  VRKVLSRYHAVEGWSG------RAEVARRLWSL----AEA-HTP---------TRRVAAYTQAIMDLGAMVC-RRRPDCA----TCPLAGGCKAHKQNVHDYPASKPT-----  225
c_000001029068_2      114  VRKVLARYHQVDGHYG------QGKTIKQLWKL----SEK-HTP---------NTRNNDYTQAIMDLGSIVCTRLNPVCD----CCPVALRCLAAKKTHTQHLYPASKPK-----  198
c_000001128125_11     149  VRKVLSRYFGIHGWPG------EKKIEADLWHK----ADL-LTP---------PKRFADYTQAIMDLGATTCKRSKPLCG----VCPVQKSCVALKREETHSIPASKPK-----  233
c_000001595844_4      140  VFRVLARYTANFIDIG------TPAGKKEFTLL----ANQ-LLD---------KKQPGKFNQAIMDLGALVCKPTNPDCS----NCPLQIKCQAFSQDTIAQLPVKEKK-----  224
c_000001719155_2      144  VKRVLSRLFLLKENGE------TRKSENILWET----MQQ-LLP---------ETGAGNFNQAIMELGATVCLPKNPLCL----LCPLKRNCQAQKSGKQNLYPPRKRA-----  228
c_000002363038_6      149  VKRVLSRYFAIEGWPG------EKRVENQMWGI----ADN-LTP---------ETDFADYTQAIMDLGATLCKRSKPNCN----YCPVKTDCRALKTDQVASFPNSKPK-----  233
c_000002561400_1      105  VKRVLTRFFGVHGWPG------EKQVENRLWHR----ADE-LTP---------SVRFADYTQAIMDLGATLCKRSKPNCI----TCPVHSGCQALKKGEVHLLPSFKPK-----  189
c_000002566035_2      141  VKRVLTRYHGIKGWPG------EKKIENKLWSL----AEL-YTP---------RKLTKQYTQAIMDLGATVCTRRKPDCD----ICPLTLTCFAREMGKQHDFPGKKPT-----  225
c_000002598045_3      139  VKRVLARHFKIEGNLQ------ASSTTKQLWQI----SES-LLP---------GDRIDIYTQAIMDLGATVCTKSNPLCS----NCPVNQDCSAFKEGLVDTLPLKVKR-----  223
c_000004188575_41     141  VRKVLARFYAIASWPG------NKATEQQMWGQ----ADN-LLP---------DQRIAAYIQAQMDLGATLCTRSKPNCD----CCPLQSNCQAYARGTPTDFPIAKPK-----  225
c_000000240049_1      139  VRKVLARHQAIEGWSG------QSPVSKKLWTI----AEQ-FTP---------ENQLADYTQAIMDLGATVCTRSSAKCG----VCPLSEDCLAYQQQRVADFPGKKPK-----  223
c_000002139456_2      139  VRKVLARYQAVEGWSG------QSQVSKKLWAI----AEQ-LTP---------KEQLADYTQAIMDLGATVCTRSTAKCG----GCPLSEDCLAHQQQTVADFPGKKPK-----  223
c_000003535614_2      141  VRKVLARYQAIEGWPG------KPAIEKQLWVF----AER-YTP---------TEELADYTQAIMDLGATLCTRASPKCG----SCPLNNGCKALANNCVAALPTRKPK-----  225
c_000002657784_4      138  VRKVLARYKKIEGDLS------KALNIKTLWRI----SES-LTP---------EEKVDLYNQAIMDLGALVCTKSSPKCR----ICPVSRDCIAFNKKLIQVLPRKKLT-----  222
c_000003800129_15     143  VRKVVGRLYNQESY----------QEKKFWEL----SKK-ILD---------KSNPFLFQQGIMDIGATICHIHNPQCD----ECPLGRWCLSKKNGSFYTYKKKK-------  220
c_000003716781_1      130  VRRVLARWFAVEGVV------TKAERPREIEGL----AAA-LVG--------GVNDPGSWNQALMELGATVCLPKDPRCD----SCPVMTECRARARGLERAVPELPVR-----  215
c_000001556689_3      131  VARIMCRLAGIADDTR------RSKVRQRLEAM----AAD-AIE---------CHPPGTFNQAIMELGARVCTPRSPRCD----QCPCSSYCEAHELGIEEQIPPRRSK-----  215
c_000001834452_1      108  VCRVAARFLGLKSPLG------SSALRGQARDW----GET-LHAG-----VSAGLAGSLNEGLMELGATVCRPRAPLCG----TCPISEKCVARATNRVTEIPKKAKR-----  195
c_000004255004_2      148  ARRVYARLLSLTN-------------LRAINET----AEQ-MVS---------HSRPGDFNQAVMELGATVCLPGKPRCE----RCPLAYRCQARSVGAFKLQLRRPTR-----  225
c_000000605438_3      141  VKRILYRFFAVKEA----------NDKKLWEM----AYD-LYD---------KDNAYIYNQTMMDIGSAICTHKNPLCT----QCPFVSLCQGKDEPL--LYPTKKKK-----  218
c_000000581237_15     141  VKRILYRFFAMTSC----------NDKKLWEM----SYE-LYD---------RENSYIYNQTMMDIGSAICTHKNPACD----VCPFEPLCQGKESPL--LYPEKKVK-----  218
c_000002529579_2      141  VKRILYRFFAVKEA----------NEKKLWEM----AYD-LYD---------KENAYIYNQTMMDIGSLICTHKNPLCT----QCPFVSLCQGKDEPL--LYPTKKKK-----  218
c_000006097838_1      114  VARVLTRLLGIERPSK------QKQVVEQLWSV----TAD-LVKHAVDLQKTSRRNATAFNQAMMELGALICTPRNPLCH----DCPLKGHCIARKQNKVNRIPVSAKR-----  207
c_000000583727_6      140  VARVLTRLLGIEQSPK------DKPVTDRLWSV----ADE-LVKHTVTLQKPRQRNASAFNQAMMELGALVCTPRNPLCG----DCPLKGRCTARKQNKVETIPALAKR-----  233
c_000005807640_2      142  VARVLTRLLGIEESPK------ARPVADRLWSV----ADE-LVNHAVTLRKPSQQNASAFNQAMMELGALVCTPRNPLCG----DCPLKGRCIARKQNKVETIPALAKR-----  235
c_000000134878_1      124  VNRVVSRFYAYQDPPS-------KNKKKMEKF----MSL-IIN---------DNRPGDINQAIMDLGRYVCKPSTPLCD----ECPIHNGCKALKLGIQSDLPIKIKK-----  206
c_000005136725_3      139  VNRVVSRFYGYRDPPL-------KNKKKIEKF----MNA-IIN---------DKRPGDINQSVMDMGRYICKPRTPLCD----RCPIQYGCKALKLGIQADLPVKIKK-----  221
c_000001684786_4      137  VERVMTRILRLAENPR------QVSTKKRLKEI----AGG-WMP---------ADKASSFNQAMMELGALICSPVSPDCP----SCPVREVCAAAEQGDPENFPRLPPK-----  221
c_000002971826_23     147  VERVLTRILRLADNPR------TAATKKYLKEI----AGG-WMP---------RGKASSFNQAMMELGALICVPVSPNCA----ACPIHDACAAAEHGDPENFPKLPPR-----  231
c_000002040695_2      150  IARVFSRVLQYPGSKE-------ALTTSLRQP----ISL-FLP---------FKRIGDYTQAIMDLGALLCRPCNPQCS----LCPIKTLCKSAHHNTVHLFPHKTKK-----  232
c_000003263657_32     154  VIRVLSRLYRIHEDAT-----KGPAKRRFQEL----ADG-LLQ---------KGQAADFNQAMMELGATVCLPQKPRCT----VCPVQCHCRAFHELSDPSGLPYKPP-----  238
c_000005989041_2      151  IERVITRIFGICTPFP-----EAKTIVKDCAAK----LAP-KTT---------RGRPGDYAQAVMDLGAIVCRPKKPLCD----LCPWKKCCIANQRNFVDQVPQKVSK-----  236
c_000002391082_2      149  VTRVLCRLLRIEEDPR------RTAIEAELIAA----GEA-LLA---------RGEAGDFNQAMMELGARVCTPRNPLCP----TCPVGQWCRAQAELDDPSALPYKVP-----  233
c_000003839553_2      156  VIRVLCRHLRIEGDPR------RAQIKAELIAA----VEA-LLA---------RGQAGDFNQAMMELGARICTPRKPLCE----TCPVGATCRARAELDDPTALPYKAP-----  240
c_000003996707_2      146  VTRVLCRLLRLEGDPR------KAAVKTELIAA----GER-LLA---------RGRAGDFNQAMMELGARICKPTSPACG----QCAVVEMCRAHAELDDPAQLPTKVR-----  230
c_000002044706_1      129  VIRVLTRLFRIPGDPG------TAAVSRRLWQL----ATD-VLP---------QNRPGDFNQAIMDLGATVCKPIAPLCE----QCPLAGLCAVRAAGNAEDFPNAGRR-----  213
c_000006007075_1      144  VERIIKRILNLKTEKE--------ISKENIIKK----KKI-LGM---------SDRSSDYAQALMELGALVCKPKNPYCK----KCPITKNCLSYKKKDFEI---ISKN-----  223
c_000002094036_4      120  IERVLKRILNLKSEVK--------TKKENLHKQ----KKV-FGQ---------TNRSNDYVQALMEIGALICKPKIPHCN----KCPITKNCLSYKKKDFKI---KPKN-----  199
c_000002762689_1      144  IERVLKRYLYLKKENE--------INKENLIKK----KKI-FGY---------SNRASDYAQALMELGALICKPNNPHCE----QCPISNKCIALKKKDFLL---TKIK-----  223
c_000005516980_2      144  IERVLKRYLYLKKENE--------IQKENLIEK----KSV-FGL---------SQRSSDYAQALMELGALICKPSNPLCN----QCPISKKCKSFRKKDFNL---IKNT-----  223
c_000003920004_1      150  IKRMLARLYGLDQSIN--------LINKKITSL----SKF-YES---------KKQSSNLIQAFMDYGSIICVPRNPKCG----ICIISRECIANQRKISNIIPKKIKS-----  232
c_000004008511_2      144  IERILVRLYGLKLPII--------KIKNDLRKK----SDN-FIS---------KNSSSDLIQAFMDFGSLICTPRNPDCV----NCLIQKNCNAFNYNLQDIIPVKLKS-----  226
c_000004481347_1      151  IKRIIARLYGIKTSLL--------LNKKTIEDI----ASK-YQS---------SNKASDLIQSFMDYGSVICLPRNPKCD----ECLIEKFCEARKKNIQHLIPFKNLS-----  233
c_000003787733_3      143  VKRVMARILKIKHITK--------RNNRRIRNR----LIK-WMD---------PERPGDINQALMDLANKICRVDHAYCN----TCPIDKVCMAEKMSIPESYPTKLNK-----  225
c_000000754657_2      143  VRRVMSRLLRIKNITK--------RNKKRIHNT----LIE-WMD---------RERPGDINQALMDLANKVCRVDHAQCN----SCPIDEICMANKMSIPESYPTRIKR-----  225
c_000000990943_3      143  VRRVMARVLRIKNITN--------RNKKRINNT----LIK-WMD---------PVRPGDINQALMDLANSICRVDHAHCS----ICPIDEICMAEKMPIPESYPIKVKR-----  225
c_000004474996_2      142  VIRVMARLLRIKNITK--------RNKNRIYNT----LIK-WMD---------KERPGDINQALMDLANKICCVDHALCS----TCPIDEICMANNMSIPESYPAKLEK-----  224
c_000001463500_11     130  VTRVLSRIAIRNLTP--------YNRRRIFNV----LEK-FID---------KERPGDVNQALMDLGSSICTPINTECI----ICPVSQFCKAKKTLLPFGYPIKINK-----  212
c_000001286181_5      137  VKRVVSRYLNVENNPR--------ELKKSIKPF----LYQ-NSP---------KANYQSFGQGMMDLGSLICVKEKPKCD----LCPIQKTCLSFGQGMMDGFIKTKKNT-------  217
c_000001293628_3      143  VRRVLSRLFDLADP----------TDASLRTR----AGE-LVD---------PDRPGDWNQALMELGATVCTARSPRCS----VCPLTKQCRAETEGTQELRPAPKRR-----  222
c_000001535696_8      137  VYRVLSRYFGVDLPIN------TTEGVKYYKEL----AFK-VMD---------SQNIRDYNQGIMEFGAIQCAPKNPNCF----SCPLSDSCVALQENKIDTLPVKLKK-----  221
c_000001614067_2      150  VERVFSRFYAVEKPIK--------ESKIFIKNI----AEK-HLP---------DNRHGDYAQALMDLGSLICIPKSPRCK----MCPLLAICDVGGTTSAKQYPIKLPK-----  232
c_000001765289_1      116  AKRVMSRILGIKNL--------TSWNLSRINKT----LSN-IIP---------EHTPGNFNQSVMELGATLCRPRSPSCN----KCPLSFGCKA-FKTNKPDYYPKPAA-----  197
c_000001961666_1      131  VERVITRLFKIDNNK---------KVKNKIKKY----TQS-LIP---------QNKPGDFAEALMDLGSLICKRANPTCS----LCPLKLNCKGYISNQIKVSVKNKKL-----  212
c_000002018097_4      138  LKRVISRYLGIKKLTR--------RNTLRIHNQ----LNK-MLT---------NGRPGDINQALMDVGSLICKPNDVICI----KCPLINKCKGYSGYNPIMYPDKNKR-----  220
c_000002843512_36     135  VKRILCRLQKLKTP----------TDKELWKI----AYD-LVN---------KKNPFDYNQAMMDIGATLCTPKNPNCE----ACPLSDICQGGSDPALYPTKKKR-------  212
c_000002930199_1      135  VYRLISRLYNINTAIN------TNKGREEFQSI----ANN-LLP---------NKNTGLYNQAIMDFGSIQCKKYNPKCN----ICPLQKECQSAILGIVNERPVKTLS-----  219
c_000003033795_2      131  LERIGYRILGLKTKTK--------RNKKRVVEY----LEK-NQC---------ANNPGDYNEAMDLGSSLCKASITYCN----QCPLKKICKAYTSSSPILYPTPKVP-----  213
c_000003888107_2      132  LERIGFRLLGLKKRSK--------FNQNKVKKF----LEK-IQD---------KNNPGDFNQALMDLGSNICQSQSVNCR----VCPINDSCRAFLSLNPVDYPAPIKP-----  214
c_000004013286_6      157  SRRVLSRVFAIKGDLS------RQPMKHLLWDL----AKF-VLP---------EKNVGDFNQAMMELGSQCCQKQYPRCM----QCPIKEWCRAFQEGLQEELPFTPIQ-----  241
c_000004852258_1      119  HSRIIARVLGVKNQTS--------RNLNRINNY----LKK-LVR---------EGNPGEINQALMDMGSNICKSNKVICV----NCPFHFSCKAFQSGHPLSYPKKILP-----  201
c_000005254087_1      145  VVRVLSRLTDFAEDVT------TTPAKKQLWQM----AAK-LVP---------ADRPGDYNEAIMELGQTLCLPTKTDCV----ACPLTAVCLARQRGTQMERPVRPPR-----  229
c_000005603677_1      142  VYRVLSRYFGQALPID------SSAGKKWFKAK----AQA-LLW---------EDDPANYNQAMMDFGALHCRPKNPSCG----DCPLAKNCQAKANKQQLELPKKGKK-----  226
c_000003283462_3      130  IKRIMARMLGVKKITN--------FNYKRIKKT----LLN-LIS---------SSRPGDFNQALMDLGSQVCRPNQAHCY----RCPLKLFCKAGNTLNPEKYPRPLSK-----  212
c_000002687221_1      144  AIRVISRMNSIDSPYP--------KSRKKIINF----LSE-HID---------TIRPGYFNQALMDLGRVICTYKSPSCY----ICPVNTYCYSYVNNTVDKFPTKIKK-----  226
c_000003347358_19     166  VKRILCRLEKLRTP----------SDTELWDI----AYT-LVD---------KINPFDYNQAMMDIGATVCIPKKPQCD----ICPLNDICKGQEEPTLYPTKKRR-------  243
c_000002701031_2      137  VIRVLSRYFGIQTPFD------SSDGKKEFQLL----AEE-LLI---------KEKAGENNQAIMEFGALQCTPKSPNCN----VCPLQNTCFAFNNNLILELPKKAKK-----  221
c_000000582753_3      145  VYRFISRLYGVSTPIN------SGKANKEFKVI----LNA-LID---------EKNPATFNQAMMEFGSLQCTPKSPNCE----SCTFSEICFAYKNNVIKDLPVKLKA-----  229
c_000001713769_5      137  VYRVIARFYGIDVPVN------NHLGKKFYMDY----AQK-LAP---------KKSCGDYNQGIMDFGSLICKPKAPLCD----KCMLAKDCIASKMKNVNYFPVRLKK-----  221
c_000004369364_1      140  VFRVLSRYTANFSDIG------TPVGKKEFTLL----ANQ-FLD---------KKQPGRFNQALMDLGALVCKPVKPICP----DCPLHSDCLAFAQDIISSLPVKAKK-----  224
c_000006057486_30     141  VYRVLSRYFGVSTPIN------SAKGIKEFKQL----AQI-LIN---------PKNPGTHNQSIMEFGALMCKPQKPNCT----NCPLNDSCFALQNNQIKTLIPIKTNK-----  225
c_000005494072_10     137  VYRVLSRYFGIEYAIN------SSKGIKKFKTL----AQS-LLL---------NENFGLHNQAIMDFGATICTPKKAKCD----SCIFCESCSALQHNKVSELPVKTPK-----  221
c_000001742634_3      140  VYRVLARYFGIKTAIN------STNGIKEFKLL----AQQ-LLP---------NNKIGAYNQAIMDFGATICTPKKPKCD----ICIFNDSCIALQKNKITELPTKTNK-----  224
```

```
c_000002826998_2        140  VFRVLSRYFGIETPMD------TSEGKKQFTHL----ANE-LLL---------KKKPAIYNQAIMDFGAIQCTPKSPDCK----HCFLKKKCQALQENKISKLPVKSKK-----  224
c_000003159439_6        137  VYRVLARYFGIRTSTN------STKGIKEFKQL----AQE-LID---------TKDPATFNQAIMEFGAIQCKPKNPNCN----NCPLNTSCIALQKKLITILPIKDKK-----  221
c_000004887214_1        142  VYRVLARYFGIRTSIN------TSKAKKEFKEV----AFE-LID---------SKNPGLFNQAIMEFGSLQCTPKNPTCS----ICPLNDTCVALQTNQVEVLPKKQPK-----  226
c_000005037037_2        140  VIRVLSRFFGIENTFQ------TVKEKKEFQEI----ADV-LIS---------KKSAADYNQAIMDFGAQICTPKNYQCE----KCPISANCFAYKNKQQNNLPLKKKK-----  224
c_000001059964_1        132  MIRVISRMMNNQEEPR------LAKTRKRLWQL----ARE-WIP---------EGEARRFNQAMMEIGALICLPSNPKCL----LCPMNNHCRALQKGTVSLIPLKPPK-----  216
c_000002498472_1        141  VERVFARIFNIDLIPG------SPEARQLNWEK----AEE-LLP---------KGHARNFNQALMELGALICKPKKPLCT----SCPLVTHCLALKYDLVPERPVPKKS-----  225
c_000004615912_3        130  VYRVLSRFFGIKAAID------ISKNKKIFEVL----AQE-VLD---------ASQAGKYNQAIMDFGATMCKPKSPDCP----LCPLSKKCIALKEELVTTLPIKSKK-----  214
c_000004766858_2        143  IERVLCRIEDIGEPPK------STPVQKQLWSL----AAE-WIP---------HGHAREFNQGLMELGATLCTPRNPSCL----LCPANAWCVAFHTGRTDTIPAPRVR-----  227
c_000002747260_18       149  VERVMARLFQIETPLP--------GAKPELTEL----AGR-LTP---------DIRPGDHAQAVMDLGATICTPKSPACG----ICPVMTGCTARAAGAQADLPRKTPK-----  231
c_000005371561_38       149  VERVMARLFDIHEPLP--------AAKPVLKAQ----AAA-LTP---------ATRPGDHAQAVMDLGATICTPRNPACG----ICPLRAPCAARAAGTQADLPRKTPK-----  231
c_000000169465_2        145  VKRVLARYYSIGGWPG------QKKVENQLWEV----AEK-NTPTN-----SEGGRCANYTQVMMDLGAMICTRSKPKCD----ECPLQADCIAYAQGAQADYPGKKPK-----  233
c_000003254110_11       145  VKRVLARYYAISGWPG------QKKVENQLWEV----AEK-NTPTN-----PEGGRCANYTQVMMDLGAMICTRSKPKCD----ECPLQPDCIAYAQGAQTDYPGKKPK-----  233
c_000004750284_20       145  VKRVLARYYAIEGWPG------QKAVENALWDV----AEK-NTP---------QNRCANYTQVMMDLGAMVCTRSKPKCD----ECPLQRDCVAFAQGKQTDYPGKKPK-----  229
c_000000141782_15       150  VKRVLARYHAVAGWPG------ETAVSRRLWAL----AER-YTP---------DNRTADYNQAMMDLGATVCVRRRPRCG----VCPLADGCRARREGNPEAYPGSRPR-----  234
c_000002717847_8        144  VKRVLARLFLQQEF----------KEGLFWNL----SAQ-CLD---------EDDPYSYQQGIMDIGATLCTPKAPLCP----ECPLNKNCLAFKENDFQLPSSKKPI-----  223
c_000004054799_1        131  VKRVLGRYKKISFKT-------ENEKQKKLWLI----SEE-LTP---------NKESFEYTQGIMDLGATHCSKSRPRCD----LCPVSNDCDSAFKEVSNKVSLRKAI-----  214
c_000006126673_1        156  IERVIARYHAINTPLP------KAKKEIKERAT----IYT-DNL---------PERPGDYAQALMDLGATICRPKSPLCY----SCPLKEGCTAYSLGQAEEYPKRQKK-----  240
c_000000339186_3        113  VIRVLTRIFGIRKNPK------DKATREYLWSL----AET-LVA---------KAENSSALNQSLMELGATVCTQRRPDCA----ACPVARSCTARRERATEIIPAKGEP-----  198
c_000003029168_2        145  VARVVSRLFALEGDLK------DTALQKKLWEI----TED-LVP---------NKRPGDFNQAMMELGAIICSTDSPRCT----LCPLAKPCHARVTGEPWRFPSESVT-----  229
c_000001863436_1        161  VARVLARLFAVDGDPR------SGDNQRRFWRL----AGA-LID---------RGRPGDFNQSLMELGATVCTPQGPSCL----LCPVREYCQAFATGDPTRYPAVAKR-----  245
c_000002830137_4        154  TTRVLSRLLAFHGDPS------QRVGQQQLLWAF----AES-LLP---------RKRVGEFNQSLMELGSEICIPRKPNCE----RCPVRMLCAAHRQGEQLSIPISRRR-----  238
c_000004612302_3        148  TARLFSRLLALSVDPS------SSEGRGRLWSL----AER-VVP---------RTRAGEFNQALIDLGATVCGV-EPRCQ----ECPVRSSCRSLEQGVVDRVPVRSPR-----  231
c_000001933926_1        134  IERVMSRLHRLETPLP--------AAKPEIAEV----AQR-LTP---------RQRAGDFAQALMDLGATVCRPRRPNCG----ACPWNASCLAHAAGIQETLPRKLPK-----  216
c_000003136334_1        146  VERVTARLHRISKPLP--------GAKKYLKEK----AAE-LTP---------SERSGDYAQAVMDLGATVCTPRKPNCP----LCPWSYICKAYLAGDAEKFPKRVPK-----  228
c_000005849454_9        143  VKRVAARWAGLELAADDLALEKAARALGEAWMT----DLP-LGK---------LSAPGELNEGLMELGATVCSPRNADCG----ACPLASGCVALADGMVDVLPLPKRA-----  233
c_000001169194_1        154  VRRVLARLFDKREP----------RARWLNRT----AAE-LVD---------KDRPGDWNQALMELGAMICSPKNPRCE----DCPVQLWCTAHEQGTQALSPAGVRR-----  233
c_000002786947_2        154  VRRVLARLFDKHEP----------GVRWLNHT----AAS-LVD---------KDRPGDWNQALMELGAMICAPKNPRCE----DCPVQRWCTAHEQGTQALSPAGIRR-----  233
c_000001279808_52       142  ISKIITVLWPN----------KGTVKV----AES-LVA--------AANSGYIWNSAMMDLASSLRAGNEIE-------AELGSTYFADVEVRDQFKPKRKKPSRTAK  219
c_000000358065_2        144  IERVLKRILNLKTEKE--------TKKENLHKL----KKV-FGI---------TNRSCDYVQALMEIGALLCKPKNPYCE----NCPITKNCLSFKKKDFEI---KKKN-----  223
c_000003745941_1        133  VIRILCRLTADNTVYT----DNTQAMKVFTPL----AEV-LLD---------TKTPGQHNEAMMEFGAMVCSKVNPTCD----ICPFNSVCLGRNHQPER-LPIKAAK-----  217
c_000004187032_4        144  IERVLKRYLHLKKDNE--------IQKNNLIKK----KSI-FGI---------SSRASDYAQALMELGAMICKPKNPECS----ACPILKSCKSYKKKDFDL---AKIR-----  223
c_000006067315_3        144  VERVLKRVFYLKNEKE--------ISKENLINK----KFF-FGK---------SNRSSDYAQAIMEIGALICKPANPLCF----ECPISKNCRSFKKNDFEI---TSKN-----  223
c_000005543774_1        114  VKRVLLRLFNLETKD----------------------PAN-LFK---------TKRNGDLAEALMEFGALICKPKDPLCY----KCLLKKNCLYYKSESKI-RFNKKIK-----  185
c_000006211484_1        136  VKRVLFRLDNININD----------------------AVN-LFK---------TQRNGDLAEALMEFGALVCRPKDPRCH----ECKIKKMCAYFKSKSKI-KFNRKIK-----  207
c_000002718976_3        145  VRRVFSRNLNIEENK--------INFDKFIKKN----KKK-FFI---------TKRNDDFVEALMEFGALICKPKDPNCL----TCCLNKTCKYFKSSKKI-KNIKNKM-----  226
c_000002992548_1        146  IKRVFSRLF--NANL--------EKNEKRIKRI----TKE-QLY---------TDRCGDLAEALMEFGAVICKPSNPMCN----ICNLKANCYFFKNKTSV-LVNKKRR-----  225
c_000000456751_1        110  VKRVFSRLLLINTSN--------QNIDNEIKKI----TYK-LPK---------IDRNADFAEALMEFGALVCKPQNPLCK----ICHLQKYCKFFNSENYI-SLNKKFS-----  191
c_000002655634_1        141  VKRFFSRLFLIDVSS--------QNFENEIKKI----TNK-LPK---------VNRNADLAEALMEFGAVVCKPKNPLCK----ICNLKKYCKFFNGENYT-LPNKKFF-----  222
c_000004232403_2        140  HRRVISRILGIKNLSR--------HNKNRINNY----LKQ-LVA---------IDRPGEINQAIMDIGSIICKPKEVFCD----FCPVQFSCKACVSSKPLEYPNKIYN-----  222
6u7t.after.N146.pdb     141  VMRVLSRLFLVTDDIA------KPSTRKRFEQI----VRE-IMA---------YENPGAFNEALIELGALVCTPRRPSCL----LCPVQAYCQAFAEGVAEELPVKMKK-----  225
Consensus_aa:                l.RVhsRhh.lp.............ppph..h....s...hhs..........ppssp@sQAhM-lGthlC..ppP.C......CPl...C.t........hP.........

Conservation:                                                        5   5
sp_P17802_MUTY_ECOL     226  -----QTLPE-------RTGYF----LLLQ-----HEDEVLLAQRPPS-GLWGGLYCFP---QFADEESLRQ-------------------------WLAQRQIA------A  281
sp_P83847_MUTY_GEOS     232  -----TAVKQ-------VPLAV----AVLAD----DEGRVLIRKRDST-GLLANLWEFP---SCETDGADG-----------------KEKLEQ----MVGEQYGLQ------V  294
c_000003652391_2        233  -----QPKPT-------RYGIV----YLARD----AHGAWLLERRPDK-GLLGGMLGWP---TSDWNDAPG-----------------------------DQPPFA------A  286
c_000000031207_6        224  -----RTTEQ-------QWWAA----AVVV-------CGDQVLVWIGDG-ELLAGHRGTP---LARLSGSGL---------------DADKMIG----KEFKRLGLD------GA  286
c_000000430975_4        234  -----RKKQR-------RYGVV----YWIER----RDGAISLRQRPEN-GIFGGMTEIP---GSLWKASPL-------------------SRH----SIHAQAPIP------A  293
c_000001007148_4        232  -----RDVTD-------VFELL----VVMR-----RGKRLLLIQHPDG-ERWGGMWGFP---ALSRDRALARS-------------RFDERLIE----SEVADWGCD------V  297
c_000001092848_3        244  -----AAKPS-------RRGAA----FVAVR----ADDAVLLRRRPES-GLLGGMSEVP---TSGWSAGSD--------------GEDGP---------SAASFP------A  302
c_000001151029_1        191  -----TKKRK-------RYGVV----YWIER----RDGAIVLRQRPEN-GIFGGMTEIP---GSLWQAAPL-------------------SRH----SIHAQAPIP------A  250
c_000001797282_1        253  -----AKGAL-------RRGAA----FVAIR----SGDEAVLLRTRPPE-GLLGNMAEPP---GSAWEPDYD---------------VAAAL---------LDAPLD------A  312
c_000001803648_25       236  -----PVKPT-------RYGIA----YIARR----GDGAWLLERRPDK-GLLGGMLGWP---GAQWGDQAV-----------------------------DKPPVS------A  289
c_000002078955_4        236  -----PVKPI-------RYGIA----YIARR----TDGAWLLERRPDK-GLLGGMLGWP---GAEWGAKAV-----------------------------ENPPVS------A  289
c_000002106160_4        224  -----QSLPW-------QNVLT----GVIW-----KRDSFLIIKRAGY-KHLNNLWLEP---GGRTLTKNN---------------PHPQLEE----LLKNKFKLE------V  286
c_000003607531_2        237  -----KVKAV-------RKGVV----FLALM----SDGSLFLRRRPEN-GLLGGMMEFP---STVWLERSL-------------------SPK----EICEAAPFK------T  296
c_000003872363_4        232  -----PAKPI-------RYGTA----YIARR----HDGAWLLERRPDN-GLLGGMLGWP---GSEWGEEAV-----------------------------ENPPFS------A  285
c_000004820107_1        241  -----KPKPQ-------KRGHV----FWISD----DSGNILFIRRPEN-EMLGGMLGLP---TTEWDIRP-------------------------------IKLP------Q  291
c_000005774797_1        223  -----RSKKK-------WNAFA----YVIYN----ADGNIGFVRRDKD-ALLGGMWALP---TSDWSTSPV-------------------FQP-----------PI------L  275
c_000005867021_3        236  -----PVKPT-------RYGIA----YVARR----PDGAWLLERRPDK-GLLGGMLGWP---GAEWGAQAV-----------------------------EKPPIV------A  289
c_000004546210_2        229  -----KKKRT-------ESMYF----LMIVF----DGKKLFLEKRKPK-GIWGGLWSFP---QIQLADDPQI-----------------------------WCEQQFTNN------I  286
c_000004551008_5        229  -----KKKRT-------ESTYF----LMIIF----DGKKLFLEKRKPK-GIWGGLWSFP---QIQLADDPKT-----------------------------WCEQQFTNK------I  286
c_000002511148_4        224  -----KSKPT-------KKVVW----LLPQG----PSGEVLLEKRKPK-GIWGGLWAFI---EAEKKNELEL-----------------------------ELSRRFKTD------C  281
c_000000577378_2        212  -----KSKPT-------KKVVW----LLPQG----PSGEVLLEKRKPK-GIWGGLWTFI---EAEKKNELEL-----------------------------ELSRRFKTD------C  269
c_000000598175_2        235  -----KVMPE-------KQAIM----VILKN----SKQEVFMQKRPPV-GIWGSLWCFP---QFEKRAFAEE-----------------------------WLKSNYGKL------I  292
c_000000754627_3        232  -----KAVPK-------KSCHQ----LIIK-----HNDKVLMEKRPNS-GIWGGLFGFP---EFNEYSELET-----------------------------FLAQQGLK-------  286
c_000000811118_1        223  -----KEKIE-------KKIDW----ILIK-----TKNKVLLKRNKEK-GIWQNLWVLP---EKDFFSSKKI-------------------------------SKI------L  273
c_000000899687_2        227  -----ISSKK-------IEVSA----GIII-----KNKKVYIQQRVKN-GLMGGLWEFP---GGKREQGES---------------PEECLKR----EIKEELRVN------V  289
c_000001176522_24       248  -----KVMPV-------RKVCW----QVSLS-----QGCVWLVQNPAR-GLWGGLWVLP---EGNR-----------------------------------------  288
c_000001345122_1        202  -----KIIPT-------KEKFF----LILVT----EAFQVALVKRPEK-GIWGGLWSLP---EFDSVQEIDL-----------------------------FLDGNEMVAT----LK  261
c_000001515736_6        220  -----IKRSS-------ESWWL----LIAVD----ARRRVWLQKRPQA-GIWAGLYSPP---VFEDYAALQS-----------------------------YAQAAWSQD-----GT  278
c_000001682161_6        223  -----KTIPT-------KERWF----LVLLT----KEGNIGLIKRPAK-GIWGGLWSLP---EFDEVKLLEN-----------------------------FITDNNVKCL----LK  282
c_000001923643_29       230  -----LKRSS-------EAWWL----LIAVD----GQGRVWLERRPQQ-GIWAGLYAPP---VFNEREALEQ-----------------------------AAARHWPAA-----RS  288
c_000002030994_5        224  -----KEMPE-------KHTVM----LILQN----DQQEVLMQKRPPV-GIWGGLWCFP---QFDKTALAEE-----------------------------WLDDSFGMS------L  281
c_000002038721_4        226  -----SQKKE-------KNINW----LLIA-----SKDKVLLKRNVNK-GIWQNLWLFP---EANSLNYEK-----------------------------FNL------L  275
c_000002523527_15       251  -----AAIPE-------RSTVM----LIVRH-----GRDVLVQLRPSS-GIWGGLWSLP---EMPVDTVPFD-------------SEAAEESA----LELARAYGS------P  315
c_000002608528_2        224  -----KVMPE-------KQTVM----LILKN----GKQEVFMQKRPPA-GIWGGLWCFP---QFDNHENVLE-----------------------------WLDNNYAVS------F  281
c_000002961510_2        227  -----VPSKK-------IEVSA----GIII-----KNKKIYIQQRSKN-ALMGGLWEFP---GGKREKGES---------------PEECLKR----EIKEELRVN------V  289
c_000002979152_8        232  -----KDKRI-------KQSVM----LYVSN----ENADVYLERRSST-GIWGGLWSLP---EFETNVKAND-----------------------------WLVKTFALE------S  289
c_000003248034_1        247  -----RKKKR-------VMTSV----IACE-----QGGQILMKQRSVNARLMPGFWELP---EVEGPFLNQ---------------ADL--------------ARWS------L  301
c_000003294679_3        229  -----KTNPL-------KTTNM----LFLRF-----EDRIYLERRPSS-GVWGGLWSFP---EISKDCSVSD-----------------------------WYQSRFGAY------P  285
c_000003333364_1        225  -----KPAVP-------VDVAS----GVLV-----HDGYIFIQKRPED-GVWPGLWEFP---GGTVEKGET---------------PAQCVVR----EFQEELEWD------V  287
c_000003402697_5        220  TEKPKKRKHR-------IEVGI----ACIW-----REGKYLVQARPKG-KSFEGSWEFP---GGKREKGEN---------------FRECVKR----EIQEEVGID------V  287
c_000003492925_4        212  ------VVPT-------REENI----MLYV-----YNDKLSLTQRKG--KFLHGLWGFE---NTEMPL-----------------------------------------  252
c_000003766664_5        239  -----RPPKK------IALVAC----VIRD-----EQGAVLLTQRPEK-GLFGGLWEVP---LKEVDETPS----------------KAGVGR----LVKETVGLT------A  301
c_000003948186_72       226  -----NNIPT-------RHTQM----LLIEN----TAGEVWLEKRPAQ-GIWGGLWSLP---ELATGQNAIT-----------------------------YCEQELQIA------V  283
c_000005057120_4        225  -----KVLAV-------KQLIF----LMLQD----QQTQFLLEKRPAS-GIWGGLWSFP---EFQSFAEIKS-----------------------------WCLEN-DIA------M  281
c_000005590109_5        225  -----KSLPV-------KKITF----LMCQD----QQGRILLEQRPKT-GIWGGLWSFP---EYESLANIQS-----------------------------CFLEK-GIP------I  281
c_000006063368_2        224  -----KTKTT-------KHVHW----LIPYS----HKGRFLLQKRLNK-GLWEGLWVFI---EKEKEEMLKK-----------------------------ECKSIFQIK------D  281
c_000000467631_2        204  -----KTLPV-------KNIMF----LMLKD----ELGLVLLEKRPAT-GIWGGLWSLP---EFETLSEIQS-----------------------------FCQQN-NFA------I  260
c_000000788235_1        225  -----KTLPE-------KKAVF----LLLIN----INHEIFIIKRPPS-GIWGGLWSLP---QFDTYLQAQQ-----------------------------WYEQYFSTH------L  282
c_000001187176_11       226  -----KRIPV-------RQCLM----PLLVN----PQGDIWLQRRPDS-GLWGGLWCPP---QLDSREQLDA-----------------------------LVIAQ-GWQ------A  282
c_000001660697_64       231  -----KAIPE-------RAATM----VIALH-----GETVLLQRRPQR-GIWGGLWSLPLVGEMDDALDAHP-----------------------------LDADTVRQAALAYGTV  296
c_000003378864_2        238  -----KPVPH-------HHIGV----GVIWK-----DDQVLIQLRPPE-GLLGGLWEFP---GGKQEPAET---------------IEETVRREIREELGVD----------V  300
c_000003910742_3        226  -----KVIPE-------KNAVM----LMLIN----NKDEVFIMKRPSP-GIWGGLWSFP---QFDDYDQAQA-----------------------------WHENYFGFP------C  283
c_000004116181_2        192  -----KKQAT-------RTCVF----ILIVN----AKQEILLQKRSPV-GIWGGLWSLP---EVEDTIEAIH-----------------------------WCSSR-KIT------V  248
c_000006223903_1        226  -----RPRPR-------REVVM----IMIKD----SRDRVLLERRASS-GIWGGLWSFP---ECPVEEALEP-----------------------------WCRMRFGAG------I  283
c_000001029068_2        199  -----KPKPV-------KTTAM----LIFDN----GKGGVYLQKRPAK-GIWGGLWSFV---ECSATDKAIT-----------------------------KTAHNFHPQ------A  256
c_000001128125_11       234  -----KDKPV-------KRAFF----LLLLN----SERQVLLRKRPQK-GVWAGLWSLP---EFDSLQGFKE-----------------------------SIESQTTV------I  290
c_000001595844_4        225  -----IKRKT-------RYFYY----FVFR-----KGSDYYIQKRTKK-DIWQSLYEFP---NIESKTPLN----------------DNEVKIV----FQEKYPRIK------F  287
c_000001719155_2        229  -----APSTK-------IAVSA----GVLF-----RRNRIYIQKRKAE-GLMGGLWEFP---GGKFESGES----------------PEQCLHR----EIKEELGVT------I  291
c_000002363038_6        234  -----KNKPV-------KQKWL----LLQKN----SAEQVFLYKRPQK-GIWAGLWSLP---EFDSYDRAVD-----------------------------FLNGEIHRH-----DA  292
c_000002561400_1        190  -----KEKPV-------KQAWF----VVQLN----AKNQIRLIKRPQK-GIWAGLWSFP---EFCSVTDAKT-----------------------------YLQTQTPG------L  246
c_000002566035_2        226  -----AVIPK-------RDTVF----AIMEN----NNGEILLEQRPSS-GIWGGLWCFP---EFSSPLRIEE-----------------------------MIKRKYGFN------I  283
c_000002598045_3        224  -----KSKLT-------KKVLW----LLPQG----PSGEILLEKRKEE-GIWGGLWSFI---ETEKKAELEL-----------------------------ALSRNFDSN------I  281
c_000004188575_41       226  -----KTIPT-------RQSHW----LILLT----DDRRVLLSKRPEK-GIWGGLWCFP---SFDDTHTL-----------------------------EQHIDNT------K  279
c_000000240049_1        224  -----KILPV-------REKRL----LIIRN----KQGHYLMEKRPPS-GIWGGMWSLP---ELTLDKSIAE-----------------------------SIERNWQLS------V  281
c_000002139456_2        224  -----KILPV-------REKRL----LIIRN----KQGYYLMEKRPSS-GIWGGMWSLP---ELTLDESIAE-----------------------------SVERNWQLN------V  281
c_000003535614_2        226  -----KILPI-------QQKRL----LIIRN----EQGAYLMEKRPPT-GIWGGLWSLP---ELALEAPIIT-----------------------------EVKNTWQII------V  283
c_000002657784_4        223  -----KQKPF-------KKVLW----LVVMN----KNGKVLLKKRNNL-GVWKGLWSFL---ESENIQDLGK-----------------------------ECLTMFEKR------K  280
c_000003800129_15       221  -----TPKKK-------VFLNY----EIYK-----RDGCIYMIKNMDL-GFWNELWMPP---YKIVE-------------------------------------------  262
c_000003716781_1        216  -----PTPID-------VELEM----FLVH-----DHGRVLLRRRDEG-GRMAGLWELP---TREVAAAELRLF----------PAELELEL-----------------  272
```

```
c_000001556689_3        216  -----LVKA--------VTEYA----AVIE-----DNGRLLLFRGQRP-SLVSDMWEFP---TLDSRLADTSSRSLTEKPASSHSRAQEAKLSR----YIKEQLGWS------V  293
c_000001834452_1        196  -----INWKE-------VHLLY----GVAS-----CPSGVLLEERKS--GWNQGLWEPP---SVPYDQKEE---------------------PDL----AWREFQPQR------G  253
c_000004255004_2        226  -------TKA-------VQWPL----TLAK-----WRSRILLHRRPDK-GILASLWELP---TPENLPAEL----------------IDETL----------------------  274
c_000000605438_3        219  -----IKKPI-------RKRAL----IIYA-----KNDKYALAQNEE--RLLSGLWGFK---QEEDFEF-----------------------------------------I  262
c_000000581237_15       219  -----KKKPI-------RKRTL----IIYT-----KDDKYALSQNSE--RLLSGLWGFK---QEEEFEF-----------------------------------------S  262
c_000002529579_2        219  -----IKKPI-------RKRTL----VIYH-----KANKYALAQNEE--RLLSGLWGFK---QEEDFEL-----------------------------------------L  262
c_000006097838_1        208  -----APVSK-------FYIA-----WHGNVFVQRRPSG-EVNSGFWEFP---NWLVDDTQ-------------------VQPAD----LASTKLALG------Q  267
c_000000583727_6        234  -----AATTN-------RRFIG----FLIA-----WRGKVFVQRRPSG-EVNAGFWEFP---NWQVNDAE-------------------LKPAA----LAKRELKLG------Q  293
c_000005807640_2        236  -----AATTK-------RRFIG----FLIT-----WHGKVFVQRRPSG-EVNAGFWEFP---NWEV-DAA-----------------ATADQ----LAKRELRLG------Q  294
c_000000134878_1        207  -----ETKPH-------YDVAV----GIIW-----DKKKLLITKRKEE-GLLGGLWEFP---GGKIKKNEK----------------IESAIKR----EIKEELSIN------I  269
c_000005136725_3        222  -----HTKPH-------YDVAV----GIIW-----DKQKLLITKRKEE-GLLGGLWEFP---GGKMKKNET----------------IKSAIKR----EIKEELSIN------I  284
c_000001684786_4        222  -----RKTIA-------VELEA----GIIR-----RGKRCLLERSESF-DFLEGLWLFP---LSLPGD----------------------GGIAA----RLGQALETD------I  279
c_000002971826_23       232  -----RKTVA-------VELEA----GIIR-----RGKRYLLERNEDL-DYLEGLWSFP---LARPDS----------------------GGIAA----RLRKRLATS------I  289
c_000002040695_2        233  -----TKKPC-------RWGHF----FFLVHIPSSGSPVSVLLEKEHG-PLLEGLWRPL---TQSWVHKKT----------------EAH--------------FPY------K  290
c_000003263657_32       239  -----KKERP------HHDHVA----GVIK-----LDDKFLIIQRPLD-GLLGGLWEFP---GGRRRESES----------------LTEGVTR----TILESTGLT------V  302
c_000005989041_2        237  -----KVKPV-------RKGIV----FLALM----SDGSLFLRRRPET-GLLGGMMEFP---STNWLERRI----------------SPE----EICEAAPFK------S  296
c_000002391082_2        234  -----KKQKP------HYQVAC----GVIC-----KGDRLLLAQRPSA-GMLGGLWEFP---GGKQEEGET----------------LQQCLQR----EIAEELAIE------I  297
c_000003839553_2        241  -----KKKKP------HYQVAA----GVIS-----KGRQLLIAQRPAE-VMLGGLWEFP---GGKQEEGET----------------LEECLVR----EIREELGIE------I  304
c_000003996707_2        231  -----KPTRP------HYEVTA----GVIW-----KGSKLLVAQRPSD-GMLGGLWEFP---GGKREPGES----------------LQECLRR----EIREELTID------I  294
c_000002044706_1        214  -----TKVTF-------VQRVA----LFLEK-----QDGSFLLVRRPAD-GLLAGMWELP---GRELGDGEAPE--------------RLASTLCG---EFGLTIRPD------  270
c_000006007075_1        224  -----KKKID-------KYFLI----KIYK-----KNKKILLIRNTKF-NFLKNFKIFP---MEEIPKP-----------------------------------------  267
c_000002094036_4        200  -----KKIIN-------KFYLA----TLYK-----NQNQILLIKNDKF-KFLKNLLIFP---MREITYP-----------------------------------------D  244
c_000002762689_1        224  -----KKNNN-------KYYLL----KVYK-----DKNKYLLIKNTKF-NFLKNFSIFP---MQELSKP-----------------------------------------  267
c_000005516980_2        224  -----KKNIN-------KYFIL----KVYK-----KDQRYLLIKNTKF-NFLKNLRIFP---MEELFQP-----------------------------------------  267
c_000003920004_1        233  ----KKSRMK------KYTRAY----IIIN-----GSEEILVRRRASK-GMLPSMLEVP---NDKWVTEKK----------------LLVRD------SSVSLFK---------  292
c_000004008511_2        227  ----KSTKRK------KYSRAY----IFYN-----EKNEILIRKRPSK-GMLASMLEIP---NDNWVINKK----------------SLVTD------TIAIKLK---------  286
c_000004481347_1        234  ----LEQKKI------KFTRAY----IIMN-----KKNEVLVNRRKSK-GMLASMLEVP---NDEWVDVKK----------------KLKRH------DVYKDLS---------  293
c_000003787733_3        226  -----KIIPH-------KEIVA----GIIW-----QRDKFLITKRPEN-ALLGELWEFP---SAEIQFNET----------------PIGALRR----QISKDFNIY------I  288
c_000000754657_2        226  -----KVIPH-------KEIVA----GIIW-----QREKFLITKREEN-ALLGGLWEFP---GGEIASSET----------------PVGALRR----QIKEKWDID------I  288
c_000000990943_3        226  -----KVIPH-------KDIVT----GIIW-----RGKKFLITKRSEN-ALLRGLWELP---GGEFESNET----------------PIEALRR----KIKEECDID------I  288
c_000004474996_2        225  -----KVIPH-------KEFVA----GMIC-----QGNKFLITKRNEN-ALLGGLWEFP---GWEIVSSET----------------PIGALRR----KIKEIVSSET------I  287
c_000001463500_11       213  -----SIKNI-------KHFIA----AIIW-----YNKEFLIMKRPVN-SMLGGLWELP---NTQIDLKIS----------------DDIQLFN----YIKNSLNFT------I  275
c_000001286181_5        218  -----VTKTQ-------ESFHW----FIYQ-----KNNKVMLCKNPDE-GIWPNLWVFP---KRELFQT-----------------------------------------K  262
c_000001293628_3        223  -----PRVRE-------ASYAV----LVTLN----PVGELLMVRRPPD-GLLGGLWEFP---AVEMEAEERNPWGGELR-VGDSLHGRCSARLE----DLGVQLSAR------R  301
c_000001535696_8        222  -----TKVRN-------RYFNY----IIPIAEADDGSRKTLLNQRKGK-DIWQNLWQFP---LIETKREID----------------LEEVKK-----RLCEVMHLAD---DPQ  291
c_000001614067_2        233  -----KEKEE-------RYGLF----FYLQN----KDGAVLFETNKSS-GLLANMDVLP---SIGWYEESN----------------RFKKSPK----FNKKKPNFL-----GM  297
c_000001765289_1        198  -----KKRKP------HYTIVA----GIIW-----RDNTFFIQRRPEK-AMLGGLWEFP---GGKVEEGES----------------LEAALKR----EIKEECGVV------P  261
c_000001961666_1        213  -----KTIRK--------GKCY----IIKRL----NDSKFLFIRNPAK-GLLGSMLLFP---TYGWLHSDH----------------DDYVK----SMIEKFIKN------I  273
c_000002018097_4        221  -----KPIPT-------KEYVA----VFIS-----YNDHIFINQRSED-GLLGGLWELP---MIEVFENSN----------------KIKALKT----YAKNKLGLS------I  283
c_000002843512_36       213  ------VVPT-------REQNI----MLYF-----YNDKLSLTQRTG--KFLHGLWGFE---AIEVPL-------------------------------------------  253
c_000002930199_1        220  -----RKPKI-------RYFNY----LFIK-----KNQQFLIQQRGTN-DIWKKLYELP---LIESKEKLN----------------REKIMA------HQYFKPFK------I  280
c_000003033795_2        214  -----KKIPI-------YDIAV----SVVE-----YQNKILITKRQNK-NFLPGLWEFP---GGKIEKKET----------------ATQAIIR----EVKEETNLT-------V  276
c_000003888107_2        215  -----KTIPT-------IKVST----CIIR-----KQNKLLILQRPID-TMLGGLWEFP---GGKIQNNES----------------RKKAVIR----EVMEETNLA------I  277
c_000004013286_6        242  -----PKIQA-------VREVA----VIVR-----KGKRVLVVQRPSD-GRWAGMYDFP---RFECSFRSKE--------------KIRQLMVR----GVGQQTGVP------I  306
c_000004852258_1        202  -----KTIPT-------KTIAA----ALVN-----HGDNIFITKRPLK-GLLGGLWELP---NIELVNGEI----------------PEDLLKI----KFADQFGLT------I  264
c_000005254087_1        230  -----KRTPHYDVVAGIIWQNG----VSPG-----EGGRFLIAQRPLN-GLLGGLWEFP---GGKQEPDET----------------LPQALER----EIREEMDMM------I  299
c_000005603677_1        227  -----LKIKK-------RYFYY----FVWRG----TNGGYYLRKRQEK-DIWQGLYEFP---MIELDKPLK----------------TSELKT-----LWAHWELEE------L  289
c_000003283462_3        213  -----KNIPH-------YDVVV----GMIW-----KEGRFLIQKRENQ-KHLSGLWELP---GGKIENNER----------------SELALIR----EIKEECDFE------V  275
c_000002687221_1        227  -----VSRPH-------HNVAV----GLIW-----KDNRILISKRNAS-GLLGGLWEFP---GGKIRSGES----------------GSSCVVR----KTQEILNVL------V  289
c_000003347358_19       244  ------VVPT-------REENI----LVRV-----YDDKLSLTQREG--KFLHGLWGFE---SVEVPH-------------------------------------------  284
c_000002701031_2        222  -----LKIRK-------RYFEF----LIIN-----ENEKILIQKREK--GIWKGLYQFP---LIENSVKKSE--------------RDIENSKS----WKKLFKHNS------I  285
c_000000582753_3        230  -----RKSTN-------RYFTY----YYIS-----DSNHLYIKKRVGK-GIWESLYEFP---LSENLIAFE----------------NEVLDK-----QIFKNSIND------F  291
c_000001713769_5        222  -----NAPKI-------IHFNY----LVLL-----DSDHMICMDRIKN-GIWKNLFQFP---MIESKKELN----------------KTQVLSN----EIFKSIAPIS---NSD  287
c_000004369364_1        225  -----IKKRT-------RYFYY----LVFK-----EKDLFLIQKRTQK-DIWQSLFEFP---NIEGMEELN----------------ESEIREA----IQKKYPNLK------Y  287
c_000006057486_30       226  -----IKIKK-------RHFNF----IVI-N----NNQHTYIHQRKDN-DIWKNLYQFP---LIESKHEIDTN------------KVIKSSIF-----QDFFKNNSY------T  291
c_000005494072_10       222  -----IKIKK-------RYFNY----LVVLS----KKNKITLNKRINK-GIWQNLFEFP---LIETVETITIK------------TLVEHEYL-----EKIINKKQL-----VD  289
c_000001742634_3        225  -----IKIKK-------RFFNY----LILIT----KDNKTIINKRINK-DIWLNLFEFP---LLETKTEINEN------------KLVKNDFF-----QKLILNDTT-----IN  292
c_000002826998_2        225  -----IKKKT-------RYFNY----LVLNF-----GKNTWVQKRTKK-DIWQNLYEFP---LLETDKIIS----------------QKELMKS----ELWKESVSDF----DF  289
c_000003159439_6        222  -----IKIKK-------RYFNY----LVIIN----SQNKIALQRRKHK-DIWKGLYEFP---LIETLTAIDIE------------QLLETEEF-----HQIINEDQA------I  288
c_000004887214_1        227  -----LKIKK-------RYFNY----LILKT----KNEETLVKQRKGK-GIWQNLYEFP---LLETEKEINYL------------ELIDHSIF-----KKMIEENNF-----D  293
c_000005037037_2        225  -----VKVKE-------RFFYF----LVLN-----IDNTTYIEQRQNN-DIWKGLFQFP---MIETQQLIN----------------KPKELEK----LITNKINVR------F  287
c_000001059964_1        217  -----KKVKI-------LYRGC----AVIL-----FQDGLILCKRNQP-GLLEGLWEFP---SVELGKGDN----------------LRKALQE----HLLEELGKP------F  279
c_000002498472_1        226  -----KKIIP-------IDMAT----GILI-----HNGMLFIQQRLPD-DVWGSLWEFP---GGRMELKET----------------PEQTVVR----EFLEETEFK------V  288
c_000004615912_3        215  -----IRKKK-------RYFHY----FISHF-----EGKILLRKRIQK-DIWQNLYEFP---MLETKDEELK--------------ISRVQKT----QLWQQLFHGNEKVKVL  284
c_000004766858_2        228  -----PEITR-------IEKLT----ALVQ-----NGDCWLVRKRPEG-GLMGGLWEFP---TWEELPSENS--------------AAAWLTQ----QLHEDYGVA------A  291
c_000002747260_18       232  -----KPKPT-------RHGIA----YLGRR----ADGAWLLERRPDK-GLLGGMLGWP---GGDWAETAT----------------------------EAPPID------A  285
c_000005371561_38       232  -----KPKPT-------RRGIA----YLGRR----ADGAWLLERRPEK-GLLGGMLGWP---GSDWAESPP----------------------------EASPPCP------G  286
c_000000169465_2        234  -----KALPE-------KSTYM----MVAQ-----FNSQVYLEQRPST-GLWGGLYGFI---EVSSIEEGIE--------------------------QLAKRGIS------V  289
c_000003254110_11       234  -----KALPE-------KATFM----MVAQ-----FNSQVYLEQRPST-GLWGGLYGFI---EVSSIEEGIE--------------------------QFKKRGVN------V  289
c_000004750284_20       230  -----KVTPV-------RQTYM----LIPM-----FGQQVFLEQRPQS-GIWGGLYGFL---EAQSAEEGIS--------------------------LLAQRGIE------V  285
c_000000141782_15       235  -----REKPL-------RRTRM----LLIVD-----GGRVLLERRPPS-GIWGGLWCPP---ECPPEVDVES--------------------------YCRHELGLE------V  291
c_000002717847_8        224  -------KKD-------VYLEF----SLYK-----DGPKYFLAQTEAL-GFWKKLWMPP---VKIVN-----------------------------------------  263
c_000004054799_1        215  ------QKPT--------VKLN----FILPH----TKNEILMHKKQAS-EYWESLWIPI---DGDEVKIK-----------------------------------------G  259
c_000006126673_1        241  -----KARPQ-------RYGYV----YWISD----DDGNILVHKRPEK-GLLGGLYGLP---TSSWETDIE-------------------TVEH----PIFISEQIK------P  301
c_000000339186_3        199  -----IRYED-------KQIDV----YLIT-----RGPRYLVQQRPEG-GVNAGFWEFP---NSDSGNKYS----------------APK-------------------  247
c_000003029168_2        230  -----VPLQSREE-VGIVIWAG----KPYN-----ADSLIFIVQRPPT-GLWAGLWEFP---HDEQTTDEP----------------HNAAALR----IAEATVGPT------M  298
c_000001863436_1        246  -----ARQKP-------VAARTAVVTSGGA-----EGRRFLMVKRPAD-GLLGGLWEFP---TVEAAGDVA----------------PTAQAVS----RLLARLTGGP---VPA  315
c_000002830137_4        239  -----TQTEH-------VVEAA----VVIR-----RGSRVLLRRHEDR-ERWAGLWDFP---RFPFDPDATN--------------NGHQQLVD----KSRELTGFE------I  303
c_000004612302_3        232  -----VRTTD-------LYQLL----LVIR-----RDDRLLLIQHPER-ERWGGMWGFP---ALSRTGPLLKS--------------RIDELQIA----SEAAEWGCD------V  297
c_000001933926_1        217  -----VARPT-------RHGVV----FWATR----GDGCVLLRRRPEH-GLLGGMMEFP---STEWRECPW----------------RIRSAV--------KAAPFA------A  276
c_000003136334_1        229  -----KRKPT-------RYATV----FWLLD----GRGNVLLRRREEK-GLLGGMMEVP---STDWLEENT----------------PTKENS--------MFAAAA------V  288
c_000005849454_9        234  -----KKWVD-------LEMVF----LVHR-----VGVRVGLQKRTS--GWSPGLYEPP---SAICESFATANE----------ASSAHESTAH----SLAREHR--------A  299
c_000001169194_1        234  -----VPARE-------ISFLL---VILQT-----EEGEVMLVQRPPE-GLLAGMWAFP---EQELAKPLD----------------CAATSRD---RAIELAMSLGA---EVV  301
c_000002786947_2        234  -----APARE-------IFFLL---VILQT-----EEGEVMLVRRPDK-GLLAGMWAFP---EQELHKPLD----------------CSTTSRD---RAIGLAMTLGA---EVV  301
c_000001279808_52       220  KKTTKRRKHR-------IEVGI----ACIW-----REGKYLVQARPKG-KSFEGSWEFP---GGKREKGES----------------FRGCVKR----EIEEEVGLN------V  287
c_000000358065_2        224  -----KKIIN-------KFYLA----TLYK-----HDDQVLLIKNDKF-KFLKNLLIFP---MKEISQS-----------------------------------------C  268
c_000003745941_1        218  -----KIKQE-------TIDRA----WVEQ-----GGKLLLHRANDKS-RRLSGLLELP---TLDVMGY----------------------SAT----ALMK------------  268
c_000004187032_4        224  -----KKNKE-------KYFIL----KVYK-----KNQKYLLVKNTKF-NFLKNLAIFP---MEELSKP-----------------------------------------  267
c_000006067315_3        224  -----KFNKI-------KYFEA----NIYQ-----NQNKYLLIKNSKF-KFLNNLLIFP---MKEVEKN-----------------------------------------K  268
c_000005543774_1        186  -----IESKT-------YDIFC----YLRK-----NKKQIALTKNNDL-GFLKKFNLPE---IKATSKK-----------------------------------------N  230
c_000006211484_1        208  -----IQSKN-------YDIFC----YLQK-----NKKQIALTKNNDL-GFLKKFNLPK---IKETSEK-----------------------------------------N  252
c_000002718976_3        227  -----IKSKN-------YDIFC----YI-N-----TKKQIALTKNNQI-SFLKDFNIPE---IKE-TNNVSN----------------------------------N  272
c_000002992548_1        226  -----KKEQK-------YNVYC----YLKK-----KKKEIALTKNKHL-SFLSNFNIPE---TKIETSNNKK----------------------------------N  273
c_000004456751_1        192  -----IKEKK-------INIYC----YLNR-----YKKEIALTKNKNL-SFLNNFKMPE---IQMVVANNNLK----------------------------------K  240
c_000002655634_1        223  -----TKEKK-------FNVYC----YLNK-----YKREIALAKNRNL-GFLYNFKMPE---IQMEVTNNNLK----------------------------------R  271
c_000004232403_2        223  -----KKIPT-------RNLAA----ALIE-----YGDYILISKRPIN-GQLGGLWEFP---NIELKNGKS----------------PEKSLKE----SINNHYGFT------I  285
6u7t.after.N146.pdb     226  -----TAVKQ-------VPLAV----AVLAD----DEGRVLIRKRDST-GLLANLWEFP---SCETDGADG----------------KEKLEQ----MVGEQ----------  283
Consensus_aa:                ....................h.....hh.........phhl.pp....shh.shh.hP.....p.....................................p...........

Conservation:                                  6
sp_P17802_MUTY_ECOL     282  DNLTQLTAFRTTFSHFH---LDIVPMWLPV----SSFTG-------------------CMDEGNA---------LWYNLAQPPSVG--LAAPVERLLQQLR--TGAPV------  350
sp_P83847_MUTY_GEOS     295  ELTEPIVSFEHAFSHLV---WQLTVFPGRL----VHGG----------------------PVEEPY---------RLAPEDELKAYA--FPVSHQRVWREYK--EWASGVRRPD-  366
c_000003652391_2        287  DWRTLRGEVRTTFTHFH---LILRIRVAEL----PDDI----------------------TPAVG---------FLLSKHAFRPSD--LPTV-----------------  338
c_000000031207_6        287  ELIGHADRFQHAITYRK---LFIHPLVYRY----QGE----------------------IPAGV---------ETIPIDGGDR----LPALHRKSIAAAR--ELLLEVFR---  352
c_000000430975_4        294  DWKPHPGRVYHSFTHFD---LELKIMTINS----EVT----------------------PEGT---------FWCRKEEVKAQA--LPSLMHKIMTHMY--KA---------  354
c_000001007148_4        298  EFQEWMDEFQHGVTRFR---LHLQPIVAMW----RSGSW----------------------EGPVTA---------EWVGIDELATRP--LSVAGRRVATRVE--RILRHNGAQA-  370
c_000001092848_3        303  PWEPC-GTVTRTTTHFT---LALSVYRARV----ADMP----------------------APAGS---------WWCAPGLLAGEA--LPTVMKKVVEAAI--SGATKARRPIG  373
c_000001151029_1        251  DWKPQPGLVYHSFTHFD---LELKI-----------------------------------------------------------------------------------  272
c_000001797282_1        313  RWKRLPGLVRHGFTHFP---LELTVFVARV----ALATP----------------------APAGT---------RFTPRSALDDEP--LPGLMRKVLAHAF--DPKPEPEKKPR  385
c_000001803648_25       290  DWQDPGAEVRHSFTHFD---LRLSLRISVV-----GNA----------------------KPTAG---------HFIADADFDPET--LPTVMRKAYRIAR--TKFDGDANG--  358
c_000002078955_4        290  DWQVLNAEVRTTFTHFH---LRLSLRVATI-----GNK----------------------PPATG---------DFISKNAFDPET--LPTVMRKAYRLAC--TQFDDN-----  355
c_000002106160_4        287  TIKNKIGEIRHRYSHFG---ITLSGFHCRL----INGSK-------------------PEVNRPF---------RWIRLNEINQYA--FPKANHKLFNLMQ--E----------  351
```

```
c_000003607531_2        297  QWVKIAGTVNHAFTHFQ---LQLQVRVAVV----NDQT---------------------QISGG---------IWVDRDRLDEFA--LPNLMKKVWAFAS--KRIQPLPRLDC  368
c_000003872363_4        286  DWQRLDAQVRTTFTHFH---LRLSLRIVVA-----KNV----------------------IPTVG---------MFIPSDKFDPDD--LPTVMRKAHALAD--AAFDHG-----  351
c_000004820107_1        292  ESEQTNLSVKHSFTHFD---LIL-------------------------------------------------------------------------------------------  311
c_000005774797_1        276  SEWKGIGQIRHIFTHID---L-------------------------------------------------------------------------------------------  293
c_000005867021_3        290  DWHVLGAEVRTTFTHFH---LRLMVHVATV-----NEN----------------------SPDVG---------KFIADDCFDPAD--LPTVMRKAHKLAL--GSFVGD-----  355
c_000004546210_2        287  LWVDSWEPFKHSFSHFD---LYIHPIQIRM---DGQFM-----------------NTAYKSKT---------QSFSLNQLDSLG--LSAPVKGLVNKLY----------  351
c_000004551008_5        287  LWVDPWEPFKHSFSHYD---LYIHPIQIRM----DGQFM-----------------NTAHTPKT---------QSFSLNQLGSLG--LSAPVKGLVNQLY----------  351
c_000002511148_4        282  LNIKKLSKVKHSFSHYN---LEAIPYLAKL----ESNK---------------------KYKNS---------VWVDYKNVESLG--LPAPIKKTINQLT--KP---------  344
c_000000577378_2        270  LNIKKLGKVKHSFSHYN---LEAIPYLATL----GSNK---------------------KYKNS---------VWVDYKNVESLG--LPAPVKKTINQLT--KP---------  332
c_000000598175_2        293  QESIALAEIKTTFSHFH---LYIQPLIINV----ETPIE-----------------LGVMEDNES---------LWYNITTEFNGG--LAAPVQTLINITK--NIEKEIENGTN  369
c_000000754627_3        287  SDLNEVAPFTVVFSHFE---LTINPHVLNI----EKAPD-----------------VVNDKQL---------VWYPLDQSIEVG--LAAPTKKLVKQIT--AIV--------  353
c_000000811118_1        274  PVEEDFSSFEHHLSHKK---LLISVKVLTI----KNKDR-----------------LSIDRRIF---------NWVNISDSVNMG--IPKPVKDILKKYD----------  338
c_000000899687_2        290  VSLNKVMTIKTTYTQFR---VTLTVFNCKL----QKKQI-----------------RPDGCEQW---------KWVSLSKLKKYP--FPAANVKIVKYLT--EK---------  356
c_000001176522_24       289  --GEALSAFRTTFTHFH---LDITPYLWHT----HSDD---------------------MPQQAG---------QWLPLAQALKMN--IPAAVKRVLQEVI--EYEQNR-----  354
c_000001345122_1        262  EQAVI----KHTFSHYH---LRMKPMM-------------------------------------------------------------------------------------  281
c_000001515736_6        279  RSWQDLPGFLVLTHKD---LHLHPVLVPV----DQAYA----------------ALVSEADES---------CWADAAAWAELG--LPAPIRKLLDAEL--A----------  345
c_000001682161_6        283  DSVEILKQLKTTFSHYH---LMMNPVMLQI----ERE----------------------AIEEV---------SWFELENLNGVG--LPTPIKKILRESF--A----------  343
c_000001923643_29       289  ADWTELPAFLVLTHRD---LHLHPVRVPV----TKGSGP---------------DAAMADGQG---------CWADARAWADMG--LPAPVRKLLDAQF--AG---------  356
c_000002030994_5        282  SEADKLSKIKTTFSHFH---LHIQPLIINT----ETPLK-----------------MGVMEHDDS---------LWYNVTTEFNGG--LAAPVQTLINITK--DIVKGTQNDTN  358
c_000002038721_4        276  EEKESLPLIVHNLSHQK---LSIFTKKYFL----KNKDR-----------------LPIDRRGF---------VWVNKKESVKMG--IPKPVKTILENHL----------  340
c_000002523527_15       316  ARAYLAGELTHVFTHFR---LLIRAIRVDL----DTVAL----------------EGDAAS---------RWLSLDDLDMLG--TPAPVRKLLEDQA--RFGLF------  383
c_000002608528_2        282  DEAIKLSEIKTTFSHFH---LHIHPLIINM----ETPLK-----------------IGVMEHDD-----------------------------------------------  321
c_000002961510_2        290  ISLNKVMTIKTTYTQFR---VTLNVFNCEL----QKKQI-----------------RPDGCEQW---------KWVSLSSLKKYP--FPAANVKIVKYLT--EK---------  356
c_000002979152_8        290  PSLQAYAEIEEEFSHFH---LSAKVLRIEL----DSDKR-----------------LGVMEETNA---------LWYNMHKDNPGG--FPAPIKKLLSRNL--L----------  356
c_000003248034_1        302  KHPKKLCKFRHSITFRL---YDVSIYTAIL----EGE----------------------LPEPF---------CWVPQRKLATIP--LTTVTRKALAVLS--PNSSDRNDEQN  372
c_000003294679_3        286  LKISRWKTLQHKFSHFD---LKIHLFKAEP----DIAIK----------------NIANNDSA---------LWYKSDSQTRIG--LAAPVKKLLETQF--NIQ--------  353
c_000003333364_1        288  AASETIAVIRHGYTSFR---VTLTAFLLRF----GEPPEGW-------------PEPVLHDATEG---------VWVSFAELDRYT--FPAGHRKLIDQMS--RDMRYAWLV--  366
c_000003402697_5        288  SVRPHFYEELHEEHTE---LLLRFHRCQI----QAGEP----------------SPQEEQEI---------KWVSPDDFGDID--FLKTNHKALDKLK--EMRV-------  356
c_000003492925_4        253  CASEYIGEVTHAYTHFK---LICKVYVYDE-------------------------LDAEQN---------SYFTSSKIQKLA--ISKVDDKILKLYL--DTIQTY-----  316
c_000003766664_5        302  RIEGQCSSVRHVLTHRD---MDIAVFALSV----KGREPS---------------PALSTYTGG--------MWIHSPEALEDLG--VGRVTRKILEAAA--THEISGQER--  378
c_000003948186_72       284  DSCTELAPRRTTFSHFH---LLITPLHIKS----SQQTT----------------HVMEDQQQ---------LWYNMRKPASVG--LAAPVKRLLRQTS--NQ---------  350
c_000005057120_4        282  QSVQHIEKQRTTFSHFH---LDYTAVLVKT----ENRIN----------------NVMESNQS---------VWYKSDKINALG--LPAPIKKLLQNQN--RNEV-------  350
c_000005590109_5        282  KLIKQLDEQRTTFSHYH---LDYTGIMVKI----ENPIN----------------IVMEANQS---------VWYKSTQLKFLG--LPTPIKKLLQNQY----------  346
c_000006063368_2        282  PSLIQIKKVKHSFSHFN---LEAIPYLVKI----KEKK----------------------SIKNT---------VWVDSKNIESLG--IPSPVKKTIKEIR--VP---------  344
c_000000467631_2        261  TSKKQLDGQRTTFSHYH---LDYTPIEVKI----ENPIN----------------NVMEANQS---------VWYKSEQINSLG--LATPIKRLLQNQ----------------  324
c_000000788235_1        283  PSATTESHFSHTFSHFR---LQAQTLLITQ----STPIK----------------RVMEGTPS---------LWYNSNTTFMGG--FPAPIKTIINKLN--EDF--------  350
c_000001187176_11       283  AEPEALSPLRTTFSHFH---LDIQPLLVRV----DAPDR----------------VTESGQ---------VWYNLREPSRLG--LAAPVKKLLQRAE--LNTQADERTSP  356
c_000001660697_64       297  SDVEAAGALTTTFTHFR---LHMHLLHADI----TKP----------------ATLDDDW---------RWVPLAQLNSVG--LPAPVKLALETLV--QPSLL------  363
c_000003378864_2        301  EVGPLVTQVRHAYSHFKITLHAYHCRLLRG------TPH----------------PKVAVES---------RWVPLNGLDQFA--FPKANKRVLEVLEQPPAMPEVTP---  375
c_000003910742_3        284  ENTEKETKLSTTFSHFR---LHIHALLIRQ----TPPIK----------------RVMEGTPS---------LWYNRSTKFTGG--FPAPIEKLFTRIL--T----------  349
c_000004116181_2        249  KTSEVLPTLRTTFSHFH---LQYTTLVVRT----DNPLN----------------FVMEVDQA---------VWYKAEQIPGLG--LPAPVKRLLDNLI--LNEDNND----  320
c_000006223903_1        284  LFESAWSVLRHSFTHFE---LAITPQPARA----PKRLT----------------RRMQSTGL---------LWYKPGEPF------------------------------  332
c_000001029068_2        257  QVRKKLPLINHSFTHYY---LTI-----------------------------------------------------------------------------------------  276
c_000001128125_11       291  ESLVEWNYFRTTFTHYH---LDIYPLVYRLD--EGEVSPVFTIG--------KSQNNDEKASAEM---------NWIELDRLNKGEVGVPAPVNKMLERLS--ETLDK------  374
c_000001595844_4        288  NIKNKLGPVTQLLTHQK---VIGQFFDITL----VGE----------------------KDLDI---------APISKVDLQVFP--FPKMMMMYINRIM--A----------  348
c_000001719155_2        292  HIDEKLMTVKHSYTRFR---VILHVFLCRV----LSGRV----------------SPTHCEEW---------DWVKTEELDRYP--FPAANVKIIKSLR--KNQGRS-----  362
c_000002363038_6        293  HSLSEGEPFRHSFSHYH---LDIFPLIF-------QASTDYEIGQP---SNRVQETPAEYSVAHH---------QWFDINRLKSGEIGVPVPIRKMISSL---------------  370
c_000002561400_1        247  EGLIEWDDFRTTFSHYH---LDIYPLIFDE----RGAPKNTAKI--------ISEPGGVYDAEQG---------SWFDIQLLLDGGLAVPAPVNKLIGRLD--ELVE-------  327
c_000002566035_2        284  KQQKEYKPIRHAFTHFH---LTIRPIHIKI----RIYKD----------------KDYDTKVL---------KWIHPKTNPNLG--LPAPVLSMLIDLN--NTKRAVLNES-  358
c_000002598045_3        282  SNIKKLKKVKHNFSHFN---LEAIPYLVKV----KKGK----------------------KIKNT---------VWVNFKNVESLG--LPAPVKKTINQLT--MP---------  344
c_000004188575_41       280  TVSKQPAVFRHTFTHFK---LDIHPHLYHA----KQSAA----------------ENITGQA---------IWYKIEDALTLG--LPAPVKKTLTLTS--KG---------  345
c_000000240049_1        282  NSHNDLAAFRTTFSHFH---LDITPCEVDI----TQSAY----------------AIAENDRF---------QWH--ADITQLA--LAAPISAIFRENN--SLK--------  347
c_000002139456_2        282  NSHNDLAAFRTTFSHFH---LEITPCEVDV----AQSAH----------------AIAENDRF---------QWH--ADITQLA--LAAPISAIFRENN--SLK--------  347
c_000003535614_2        284  KKHHDLPVFRTTFSHFH---LDITPCEIEI----KSISN----------------AIGDDSCY---------QWQ--VDISKLA--LAAPVSAIFNAK--------------  345
c_000002657784_4        281  KDLKTYPDINHSFSHLN---LNAKLFLLKL----NEDIN----------------YNINKGKY---------VWVDKKGLKALG--IPTPVKNIFEVLK--LNDQTS-----  351
c_000003800129_15       263  ---NNNREIKQLSHRE---LHLEFLVTSG-----------------------------TPKGG---------KWINKNEFNNLP--TPKPISDKLEQYA--KSFL------  320
c_000003716781_1        273  LQREEVGELRHSITKHR---IRARVLRAVP----GRDE----------------------LPEGW---------AWFERSEIEALG--VTGMTSKALTLGR-----ARSR----  337
c_000001556689_3        294  RLGPKLGEIRHGMTDRR---ISCHVYEASL----SGSQTTMAGA------PTGLEKSDHNIPDS---------GWFTIEEVMTLA--LAASARKTLTTLL--DIDLTTAHNPG  380
c_000001834452_1        254  KLGEMMGSARTTTRHR---IQARVHQVEG-------------------------------W---------NGKGAVDPSSVP--LSSLGRKVLSIAG--VLGGLLLLSPD  317
c_000004255004_2        275  QQGDPVALIHHAITDMR---ITAPVYLLQC----KIPV----------------KGPEW---------RWVSLANLHAYP--LSSLSTKALSAAR--QHLPTTRKLGG  346
c_000000605438_3        263  ENMTGLGEFKQHYTHFT---LDARVVLLKD----------------------------KKQE---------NYFTVEEIHDLA--LSGADRKALALLE--KYLQ-------  322
c_000000581237_15       263  KDMIKLGEFKQHYTHFT---LEASVVVLVQN----------------------------KEQE---------RYFRNEIHALA--LSGADRKGLALIE--KHLA-------  322
c_000002529579_2        263  NNMTELGEFKQHYTHFT---LDGSVVLVTN----------------------------KKQK---------NYFTIDEIHDLA--LSGADRKALALLE--KYLQ-------  322
c_000006097838_1        268  VEPESLCIINHSITCYR---NRLEMFVIFR----TRKPL----------------LAKLKG---------KWVPLAKLDEIP--LTTAHRKAANQFA--G----------  331
c_000000583727_6        294  AKFQPLCNINHSITRYR---NRLEMIVAEP----KRKPR----------------LAKLKG---------EWLPLAKLAELP--LTAAHRKASDYFL--KTANSR-----  362
c_000005807640_2        295  AELEPLCTINHSITRYR---NRLEMFVIEP----ARKPR----------------LAKLNG---------EWIPLAKLDELP--FTTSHHKAANQFC----------------  357
c_000000134878_1        270  KPETIIHQVDHHYSHFS---VTINAINCKY----NGGDI----------------KLSGPTDF---------KWIHTSKLHSYA--F----------------------  322
c_000005136725_3        285  KPEKLIHQVDHHYSHFS---VTINAINCYY----NGGKI----------------KLSGPTDF---------RWVKTSRLHSYA--FPKASIKLFNAIK--SIE-------  352
c_000001684786_4        280  ETGEELKAIKHSITFRR---LTLRPRLLET----GSFTI----------------RGKPSF---------RWALPEELGSVL-PVSSLCLKIADRLS----------------  343
c_000002971826_23       290  EEGAELEAIRHSITFRR---LTLRPRILGA----GSFST----------------RAKPSF---------RWGRADELGNVL-PVSSLCLKIADRLS----------------  353
c_000002040695_2        291  ASWTLLGQIQHVTTHFT---LKLDLWYTIR----------------------------TVKTQG---------QWVLNQMNDYP--LSSLTKKSLAFLV--KNLELEKEQAR  359
c_000003263657_32       303  KTDLVLPTVLHVTTHFK---ITLYGFGCTF----IKGAA----------------KALTCHDA---------RWVSQTQLADFA--FSRAHNRLIDIMK--KEAAKGQVEMF  378
c_000005989041_2        297  QWVKITGTVTQPFTHFE---LQLQVRVAVV----ENKP----------------KSSEG---------IWVDRNRLDDFA--LPNIMKKVLAFAL--RHNQHFSREFF  368
c_000002391082_2        298  EVGARLTSVDHAYSHFS---ITLHALEARY----LRGEP----------------QAIGCADW---------KWIVPDELDDYA--LPRADRLIVEFLR--QDGRQLGLFDS  373
c_000003839553_2        305  EAVARLTSVDHAFSHFS---ITLHAIAARY------------------------------------------------------------------------------------  331
c_000003996707_2        295  EVGGLLAKVDHGFSHFS---ITLNLFEAMY----RGGEP----------------EAVECDDW---------RWIDPAELDDLP--MPRADRRALEELS-EEEGQRVPVLMA  371
c_000002044706_1        271  ----LVVTSEHRFSHRH---WTTRVYRA------RVEDE---------------CSGPADASHWSEGL-----RWTCAGDLGKLA--VPEASRKTLRPCLRASRGSLPG----  353
c_000006007075_1        268  --KNLQKTLNFKMSMNN---MNIVIKIMKI----SSN----------------------ISDS---------SWIKETKFNHYT--FPTFTKKIFNSLE--NYL--------  327
c_000002094036_4        245  SLLKSKNKLNLKMSMNN---MNICIDFVKI----KKK----------------------PKNG---------LWIESSKLENCM--IPTFTKKIFASVK--NNL--------  306
c_000002762689_1        268  --LSSTDIVNFKMSMNN---MNIKIEHKNK----VGS----------------------LPTS---------HWIDPKKLKNYT--LPTFTKKIVRYLE--SNK--------  327
c_000005516980_2        268  --KNFNENLNFKMSMNN---MNIKIQYSKN----NIK----------------------LPSS---------YWVDQKKLNNYM--LPTFTKKVVKYLE--NNK--------  327
c_000003920004_1        293  KKFTKCKPILVSFSHFD-------------------------------------------------------------------------------------------  309
c_000004008511_2        287  NKLKLKGFVDVSFSHFD---LETEVFFTKV----KKDY----------------------FANQ---------KWIKKNNIKISG--LPTVMKKIIEVAL----------  346
c_000004481347_1        294  VKYKRIGKITVSFSHFD---LDIGIYKSKV----VKKK----------------------FKNF---------NWIKSNKIDTSG--MPTVMKEIVKKSL--KI---------  355
c_000003787733_3        289  NIGKKIGHVKHAYSHFR---ITLTLYQCKT----NRSVK----------------NIYKEY---------KWIIPTEVNNYP--FPKSNHKLFK----------------  348
c_000000754657_2        289  NIKKKVGNVKHAYSHFK---ITLTLYRCQS----QGSVK----------------SNNKEY---------RWITPAEIKYFP--FPKSNHKLFDVLN--SNGWNV-----  357
c_000000990943_3        289  NIEKEVGYVKHAYSHFK---ITQTLFQCQT----QESVK----------------SINKEY---------RWITPIEVNNYP--FPKSNHKLFTILN--SDGWNV-----  357
c_000004474996_2        288  NIKNKVGYVKHAYSHFK---ITLTLYQCQT----QESVK----------------IMNKEY---------RWITPSEIQYFP--FSKSNHKLFNLLN--SNGWNV-----  356
c_000001463500_11       276  QIQNKICTIKHSYSHFD---IRVSLYHCRI----NDINI----------------KNMKNV---------CWIRPKDINNYA--FSKSNHKLFKILN--TNGWTI-----  344
c_000001286181_5        263  QNINELPSFKHSLSHKD---FHISPRIINI----PDDME----------------TDDKKT---------IWIEKNKIAKLG--APKPVLDIVKKIL--NQNDKGLL---  333
c_000001293628_3        302  VRFSPLPEVRTTFTHLK---AVYRPMLVIG----TRCEG----------------SNPDSEH---------KWVLPDQVEDLP--LPVAQRRILDLAW--AALEGC-----  371
c_000001535696_8        292  VLRYNHKIIVHKLSHQQ---LHTHFWIVVI----SNELD----------------------GSILLSEIENYP--VPVLIANFINTFK--NSYFWPNNSIR  359
c_000001614067_2        298  KWKILDQNLNHIFTHFK---LNCSLAIATI----NDENAL----------------VNDLSKSSY---------RFVQKTNINELA--LPSLIKKILNALK--NNEILDF----  371
c_000001765289_1        262  SIKKRIGAVDHAYSHFS---ITFHGYHCIE----NGEK----------------------------------------------------------------------  292
c_000001961666_1        274  NRILSKNVVRHEFSHFK---LELEVYYIET----RFP----------------------QITHG---------EWLKTKQAGKQ---LPSVMKKVLDNI----------  331
c_000002018097_4        284  DILNCFDKVKHSYSHYK---LLVTPFECEV----LNA----------------------DSRIG---------RWIRVGEIDKYA--FSKVNHKLFNAIG--IKNY-----  347
c_000002843512_36       254  CASKYIGEVTHAYTHFK---LIIQVYVYDE----------------------------LDAEKN---------SYFSSQEIQKLA--ISKVDDKILNLFL--NSIECD-----  317
c_000002930199_1        281  SNIKNKYTIEHNLSHQK---LKISFWDVDA----KNIN----------------------TQKGF---------EKITLDNISNYP--FAKPLEKYFQTQH--LIT--------  344
c_000003033795_2        277  SNPIFLGNITHKYSHFG---VNISFFISFP----KSIKS----------------FNSSQEY---------RWIRIKNILNYP--LPKANHKMLDILK--KL---------  342
c_000003888107_2        278  KNPNYLGQIKHQYSHFK---VLISLFLADV----NDISS----------------LKPNQNY---------KWTTRKQLDKFA--LPKANYKMLEILD--KLN--------  344
c_000004013286_6        307  QIGNKIGQIRTVVTRFR---TTVDCYEAHF----SNIGKPNRSKNDISVNTLQKEQQEAGSLWHM---------QWATRKQIDSFP--MSVPARRVIRLLA--DFD--------  392
c_000004852258_1        265  KVGHKLGMINHTTFHFK---MNVLLYNCQA----KDQP----------------------LSKNTE---------KWVSFSDLDQYA--FSKANHKLFKLVE--LRND-------  330
c_000005254087_1        300  RTDRFLTSIKHAYTHFR---ITLTAFHATH----LRGEP----------------QHIGVADH---------AWVTLADLDKYA--------------------------  351
c_000005603677_1        290  VWALISGQEQQQLSHQK---IIARFLWPKQ----PLPQ----------------------ALSADW---------ELVDAQSWPNYP--LPRLIAQYWEKKQ--LAFF-------  355
c_000003283462_3        276  VPKKKIESIKHKYSHFS---ITLTAFHCKI----KNGAK----------------VPSKQST---------RWITPGEIGDYP--FPRANHKLFNQLE--NFYSNG-----  345
c_000002687221_1        290  APVAFVKQIKHAYSHFS---ITVDAYSCTF----RGGRP----------------RALGCADY---------RWIYPHETGKYA--FHSAGHKLFDKIL--ETAS-------  358
c_000003347358_19       285  CASEYIGKVTHAYTHFK---LICKVYVYYE----------------------------NHPEEA---------DYFSEEEIQKIA--ISKVDEKIVNLYG--NFLKYD-----  348
c_000002701031_2        286  QILKISATIKHQLTHQT---IFARFWHIDT----KIQ----------------------KVENS---------KYISKKELKNYP--ISRLLEKYLENIG--MI---------  347
c_000000582753_3        292  AIRKVHLKIKHVLSHQN---IYAKLIEVEV----LDVI----------------------RLENY---------IKVEKEDFESYP--IHRLMEKMMEQVG--K----------  353
c_000001713769_5        288  IILFNSSPIIHKLSHKT---IYAKFWILPV----NGSNS----------------------------------NSIKFSDVNKYP--VPRLIEKFLDKFN--YKHF-------  348
c_000004369364_1        288  KIKHKSAPSKQLLTHQK---IIGIF----------------------------------------------------------------------------------------  309
c_000006057486_30       292  ISKYNQKQIVHKFTHQH---LFTTFWIINT----NTCSK----------------------------------NTVNWQNLSNFA--LPTLIQNFVDDFK--K----------  349
c_000005494072_10       290  VSLFNPKPIIHKLSHQH---IYTSFWIVNY----NGSVS----------------------------------NTISIDELKKYP--VPVLIDNFIKDFE--M----------  347
c_000001742634_3        293  PYLFNEKTVIHKLSHQH---IYTKFWIVKL----DVNLE----------------------------------QTIPLNSIEKYP--VSTLIQNFIKEFN--F----------  350
```

```
c_000002826998_2        290  QIKKISLPFRQNLTHRK---IIATFFEIDF----MDKFS--------------------NKNNNY---------IKIERKNLSKFA--FPKIIDLYLKDNS--LNLAMK-----  358
c_000003159439_6        289  TSLYSDNIKAHKLSHQH---IFSKFWIVET----KQNIL----------------------------------NSILIEDLEKYP--VPKLIQNFIEKFD--NIKLNPATK--  354
c_000004887214_1        294  LQLLTPKVVVHKLSHQH---LYIKFWEIRV----PTFRS----------------------------------ATISWQELLKLP--FPIVIFKFIREFL--ETSMPNFDTPL  361
c_000005037037_2        288  KVKKRSKNFTQRLTHQK---LNAIFIEIDL----KTK----------------------INKQF---------INTDINNLSKFA--FP------------------------  336
c_000001059964_1        280  QLGKEIHTLEHRYTSFR---TRVSFFTATV----HVLPE--------------------SSQKRV---------RFVKTSEVE-----------------------------  326
c_000002498472_1        289  EIVEKITNTVHHYTRYK---VTLHCFLLRL----KKGSDT---------------DPVLHAAQDF---------NWVPFKALQEYA--FPAGHRKLITYL--------------  355
c_000004615912_3        285  QTKK----YQQELTHQK---ISAYFWEIN-----LNVAP--------------------VNIYDF---------ILIDSKNLAKFA--FPKIIALYLENNS--LNL--------  345
c_000004766858_2        292  RIGPELFKMEHDYTRFH---ATLHCRLCTV----AEIPS--------------------VLPDNA---------QWLTLGEISALA--LPRVFQRLRKRLL--DEVSA------  359
c_000002747260_18       286  PWRDPGAEVRHTFTHFH---LRLALRVADL----PASA---------------------TPRTG---------QFIPHHAFRPGQ--LPTVMRKAYDLAS--ATFLGN-----  352
c_000005371561_38       287  NWRTVPGEVRHTFTHFH---LVLELRRADL----PEDC---------------------TTMRG---------QFLAPGGFRPSD--LPTAMRKAFDLAR--DG---------  349
c_000000169465_2        290  DETRTLEGFRHTFSHFH---LDITPVVAVV----NSAPS------------------KRVAETAF---------RWFSLNEPIEVG--LAAPTTKIIQQLM--R----------  355
c_000003254110_11       290  EETKTLETFRHTFSHFH---LDITPVVAVV----NSPPT------------------KRVADTAF---------RWFSLGEPIEVG--LAAPTTKIIKQLI--G----------  355
c_000004750284_20       286  EEIQQLDEFRHTFSHFH---LDITPLIAVV----NSTPQ------------------KRVAENES---------RWFLLDEPIEVG--LAAPTTKIIKTLA--K----------  351
c_000000141782_15       292  TRPQAWPGLRHTFTHFQ---LDIQPVRLRL----TAAPA------------------RLMDGDRH---------VWYNVESPDARG--LAAPVARLLQGLA--LEDTRA-----  362
c_000002717847_8        264  ---KFNAANKHALSHRH---IHFRFKQIAN----LPT-----------------------GMKG---------EWFSKKDLANIA--LPKPISDKLLQDG-------------  319
c_000004054799_1        260  NQKYSKIKVNHPLSHLN---LDMKIKTFKI----DKKCD------------------------------------------LDSN--L------------------------  296
c_000006126673_1        302  KPYKQLPFIMHVFTHFS---LKLYLCKTSE----FICDN--------------------VPEEH---------KWVFISTLEDIG--FPSVFEKVKKII--------------  362
c_000000339186_3        248  -NGQPFASARHSITRYR---IQLNAHLIHA------------------------------KRGKG---------MWCTSDEIRSLP--FSSAHLKLLGKLP--HRLDK------  308
c_000003029168_2        299  QKLTKLVTITTTVMHTR---TTLHVFEGML----DLETS------------------TMQEASSR---------LWVRLDGLHKHP--FSSPQHEIIKTLD--KEAALQLAGLV  374
c_000001863436_1        316  TPRVELGEVVHHFSHIR---MTITAEHRVA----SSEQVRL--------------RSDDEVGQEA---------RWVSAVELVDLP--MSAAMRKIEGLYR--ELGSGAI----  391
c_000002830137_4        304  EPVRQLTTIQHGVTRFQ---ITLQCHLAQC----VSGRK--------------------RGPHL---------SWRMPHELEHLP--LSVTARKIGRLVA--KLAD-------  369
c_000004612302_3        298  EVEEWLQEFRHGVTRFR---LHLQPVVASW----RAGHW--------------------IGPVVA---------EWVEIDELTARP--LSVAARRVADGVQ--AGQRCGGASG-  370
c_000001933926_1        277  AWRELPGAVRHSFTHFH---LELQIMTVET----NEEVEG------------------------------------------------------------------------  309
c_000003136334_1        289  DWQMLSGKVRHTFTHFH---LEL-----------------------------------------------------------------------------------------  308
c_000005849454_9        300  KLGAYQGQVRHTITHHR---IRAHVYAAEG------------------------------Q---------APTHLEDPTQVP--LTGLARKVLDRIT--SA---------  354
c_000001169194_1        302  GVASALAEVQHRFTHLQ---ARYRPWVVPV----AKL----------------------LNGEGN---------VWMTPGEPVDFP--IPVAQHKVLDALA--ECGATLGSSES  373
c_000002786947_2        302  GHASALAEIQHRFTHLQ---ARYRPWVVPV----AKL----------------------LTGEGS---------MWMTPGEPVDFP--IPAAQHKVLDALA--ECRATPDSLES  373
c_000001279808_52       288  SVRSHFYEELHEEERTE---LLLRFHRCQI----QAGEP------------------QALENQEL---------KWVSPDEFDSVG--FLKTNHKALEKLK--SMRV-------  356
c_000000358065_2        269  SLSKPDHKLNLKMSNMN---MNISIDFSNN----KKK----------------------LKNG---------LWVEKTKLENHM--IPTFTKKIFASVK--HNL--------  330
c_000003745941_1        269  RGNTPIVTRKRGISNSQ---ITEPIWRLSD----NSEA---------------------DHAEM---------EWIDLEQIKKVT--LSGPHKRWINEIL--TDFNKQK----  336
c_000004187032_4        268  --KNFNENFNFKMSNMN---MNIKIQYLKN----LKK----------------------FPSS---------YWFDKRRLDNYM--LPTFTKKVVKYLE--KN---------  326
c_000006067315_3        269  FNTSLNKKINIKMSNMD---MKIILNKKNK----LIK----------------------IKNS---------YLVDKKNIKDLI--LPSFTKKIFNSVS--NYL--------  330
c_000005543774_1        231  KNWKFLCNYKNSISNKK---LNLNLYYKFS----SRI----------------------PSKY---------NWYSL-KKNKEF--IPTFTKKIFKQIT--HLY--------  291
c_000006211484_1        253  KNWRFLCNYKNSISNKK---LNINLYYKFT----SKI----------------------PSEY---------NWYSL-NKNNEF--IPSFTK-------------------  303
c_000002718976_3        273  QNWIFLKNYKNSISNLK---LNINLYYKFS----KKI----------------------PRSF---------NWYSL-KENKEF--IPSFTKRIFRQVS--TLF--------  333
c_000002992548_1        274  NGWIYLCNYKNNISNIK---MNINLFYKFI----KNK----------------------PKKY---------MWYPIDRRSNEF--IPSFTK-------------------  325
c_000000456751_1        241  RGWYYLCSYKNNISNIK---MNIDLFYKFT----KKI----------------------PKKF---------NWYSIDRSNQEF--IPSFTKKIFKKIA--KVYS-------  303
c_000002655634_1        272  REWNYLCNYKNNISNIK---MNINLFYKFT----KNK----------------------PKKF---------DWYSVDKSSQEF--MPSFTKKII----------------  326
c_000004232403_2        286  EVGKNIGLIQHTFTHFK---MNITLFCCLL----KYDQ--------------------KIDSSA---------KWVRLFELDQYA--FSRGNHKLFQLYQ--KENV-------  351
6u7t.after.N146.pdb     284  ELTEPIVSFEHAFSHLV---WQLTVFPGRL----VHGG--------------------PVEEPY---------RLAPEDELKAYA--FPVSHQRVWREYK--EWAS-------  349
Consensus_aa:                ........hp+.ho+bp...l.hp.h.h.............................................@h..pph.phs..hs...p+hhp.h.............
```

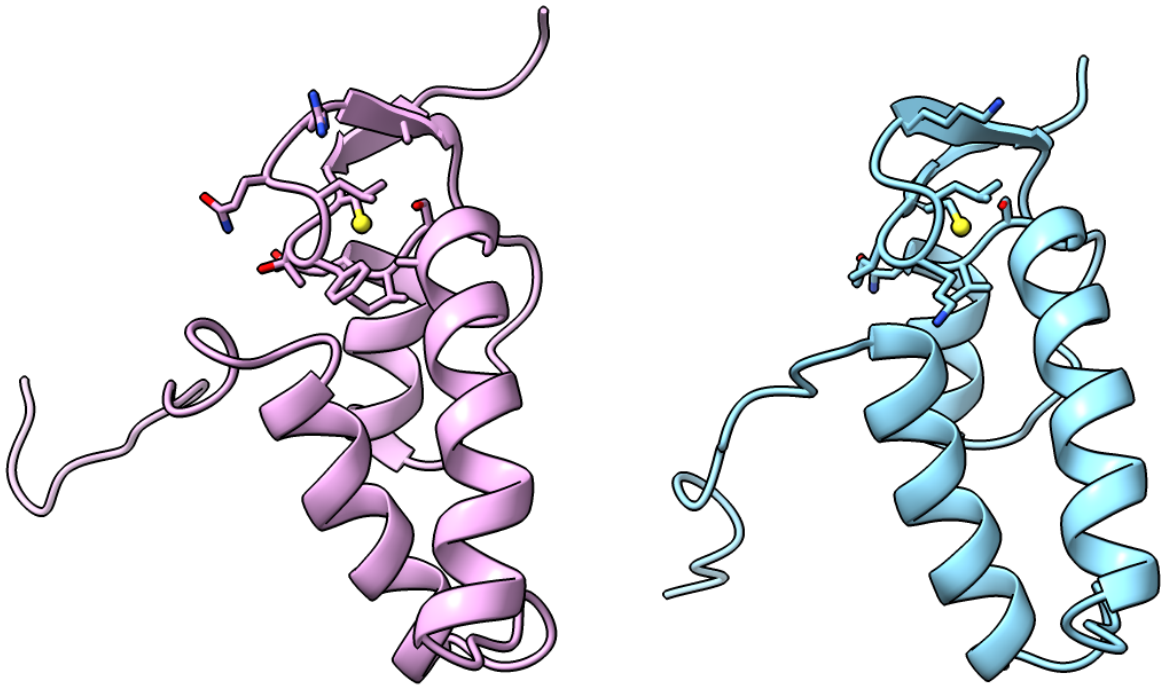
1139
1140

1141
1142 **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

1143
1144 We were interested in determining how similar the amino sequences of LCHF MutY
1145 representatives were to existing MutY enzymes. We visualized this in the form of a percent
1146 identity matrix that was generated by *Clustal Omega*.
1147

1148



1149
1150 **S3 Fig. MutY gene neighbors.** Structural homology for *Ec* YggX (left) and nearest neighbor to
1151 *Thiotrichaceae* MutY (right). The structure solution NMR structure for *Ec* YggX (PDB ID 1yhd)
1152 (103) is superimposable to the structure predicted by *Colabfold* for the nearest neighbor to
1153 *Thiotrichaceae* MutY, with RMSD of 0.99 Å for 69 pruned pairs selected from 88 possible pairs.
1154 The Cys residue critical for function is highlighted with a yellow sphere for the sulfhydryl group.
1155

1156

1157 **S4 Table. Metabolic gene identification**

			Marinosulfonomonas MAG 1	*Marinosulfonomonas* MAG 2 [a]	*Rhodobacteraceae*	*Thiotrichaceae*	*Flavobacteriaceae*
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

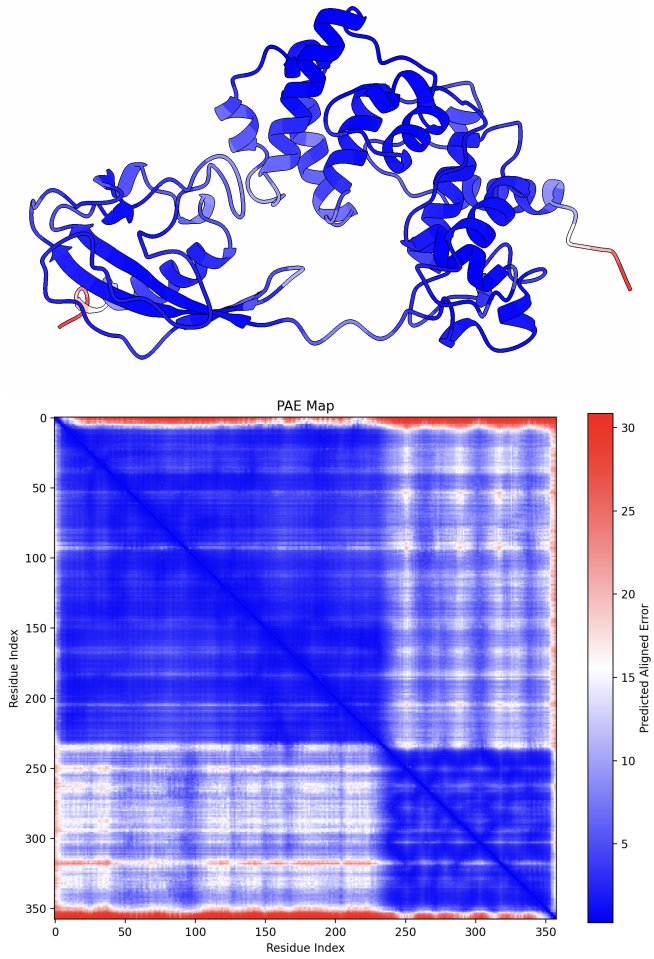
1158

1159 [a] *Marinosulfonomonas* MutY contig belongs to two separate MAGs and each are reported separately as MAG 1 and
1160 MAG 2, respectively.
1161 [b] KEGG ID gene not identified in any MAG and not reported in Table 2.
1162 [c] Completeness and contamination scores generated by CheckM v1.0.5 as described in Brazelton et al 2022 (35).
1163 A KEGG ID analysis was used to identify the potential metabolic strategies of the MutY
1164 encoding organisms at the LCHF. The full metabolic KEGG ID search is shown above.

S5 Fig. *Colabfold* structure prediction pLDDT scores

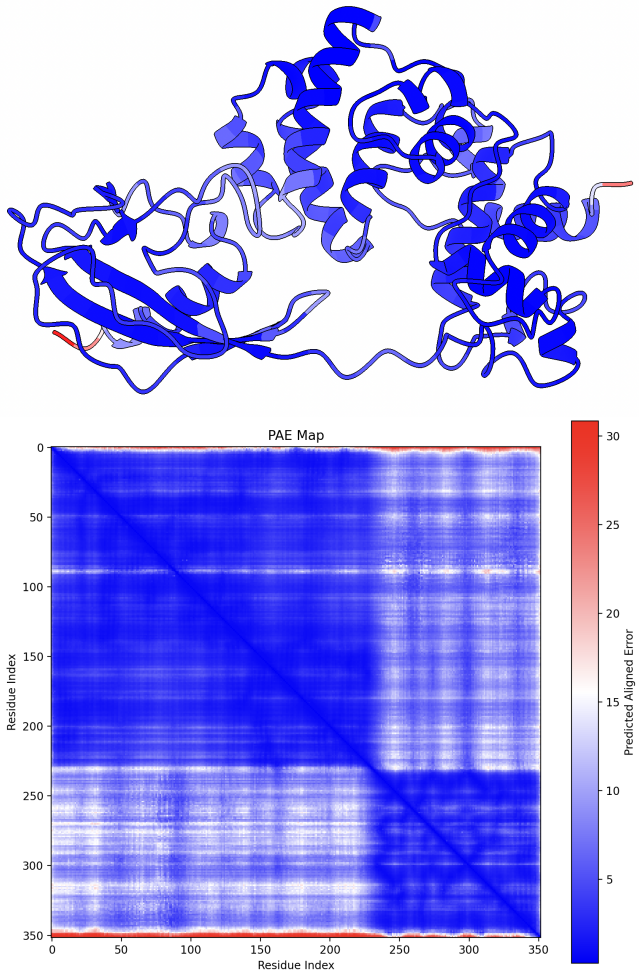
pLDDT scores represent the confidence in the prediction calculated by *Colabfold*.

Marinosulfonomonas MutY Structure Prediction



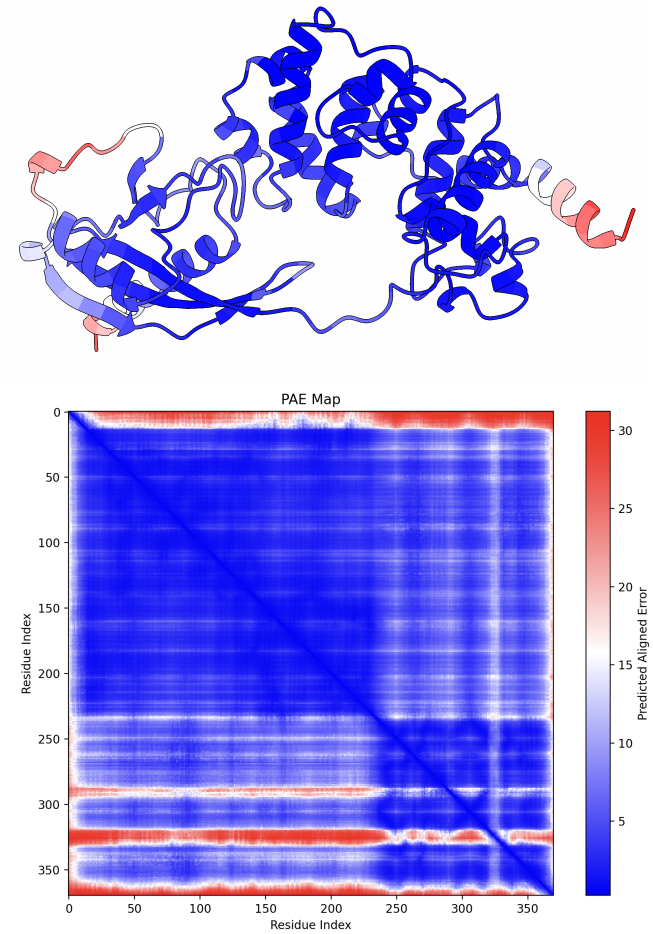
pLDDT score: 93 (±9)
ptm score: 0.75

Rhodobacteraceae MutY Structure Prediction



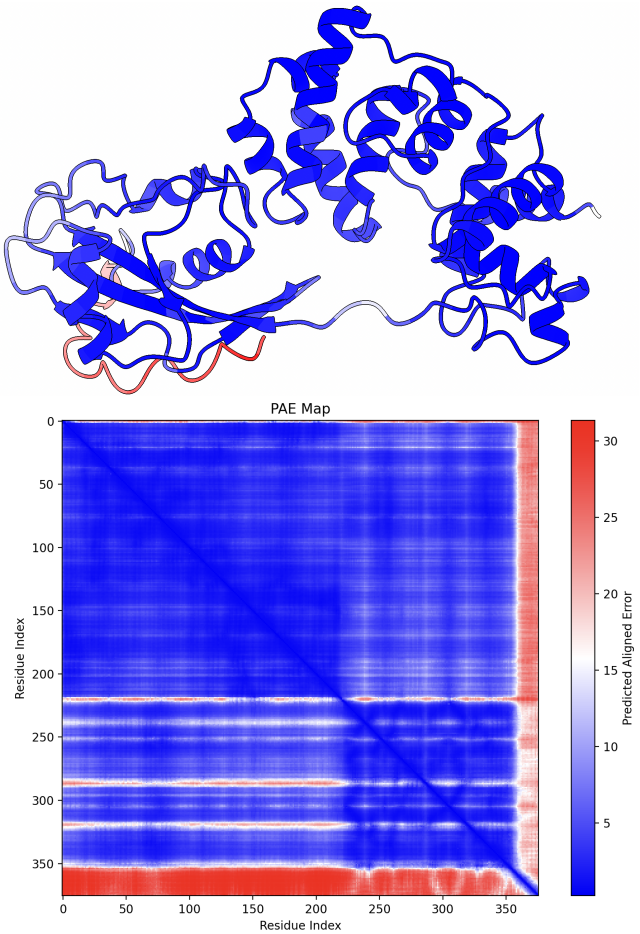
pLDDT score: 94 (±6)
ptm score: 0.79

Thiotrichaceae MutY Structure Prediction



pLDDT score: 92 (±12)
ptm score: 0.83

Flavobacteriaceae MutY Structure Prediction



pLDDT score: 92 (±14)
ptm score: 0.84

1177
1178 **S6 Table. Ligand binding affinity (kcal / mol)**

Adenosine	*Gs* MutY (3g0q)	*Marinosulfono-monas* MutY	*Rhodobac-teraceae* MutY	*Thiotrich-aceae* MutY	*Flavobac-teriaceae* 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)	*Marinosulfono-monas* MutY	*Rhodobac-teraceae* MutY	*Thiotrich-aceae* MutY	*Flavobac-teriaceae* 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

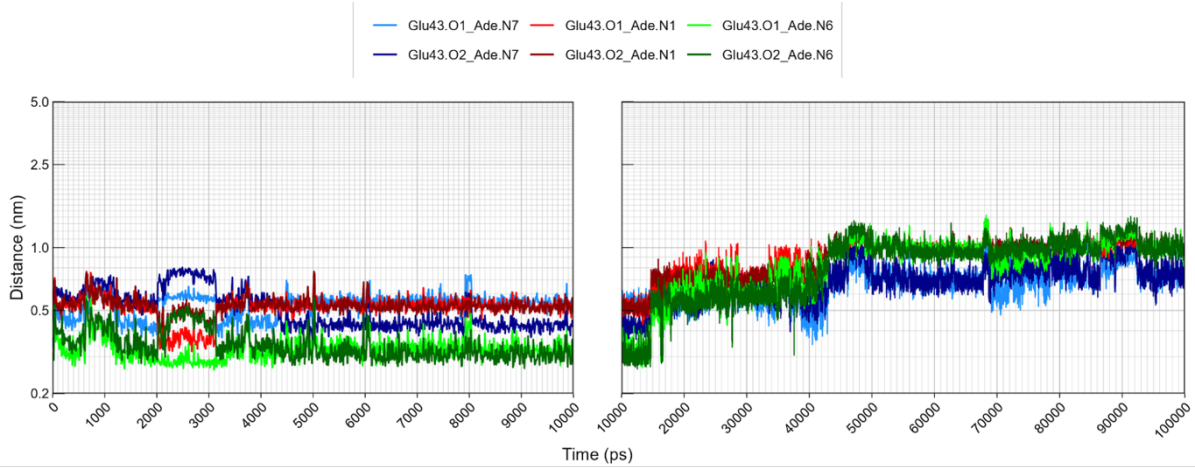
1179
1180 *Binding affinities are reported for the binding modes generated by *AutoDock VINA*. Each mode represents a
1181 predicted ligand pose, which differs by a combination of position, orientation, and rotamer conformation. The
1182 receptor structure was obtained from PDB ID 3g0q for *Gs* MutY and through structure prediction for the LCHF
1183 *Marinosulfonomonas* MutY, *Rhodobacteraceae* MutY, *Thiotrichaceae* MutY, and *Flavobacteriaceae* MutY. The
1184 binding mode representing the starting complex for molecular dynamics analysis is highlighted.
1185

1186 **S7 Fig. Molecular dynamics.** Molecular dynamics simulations were calculated by *GROMACS*
1187 with the Amber99SB and GAFF force fields for MutY complexed to adenosine and OG. For
1188 each MD simulation, short range interaction energies, distances between the ligand and
1189 functionally relevant residues, and representative structures are shown. Note that the Y-axis is
1190 logarithmic for distance. The adenosine and OG ligands are shown with all atoms wrapped in
1191 transparent surfaces. For the adenosine complexes, the protein structure was truncated so as to
1192 focus on the NTD (residues 8 - 220 in *Gs* MutY, and corresponding residues for the LCHF
1193 MutYs). Catalytic residues are shown: Glu43 and Asp144 in the *Gs* MutY protein and
1194 corresponding residues in the LCHF MutYs. The distance versus time plot for the adenosine
1195 complex, tracks potential contacts between the catalytic Glu (atoms OE1 and OE2) and the
1196 hydrogen bond donors and acceptors on adenosine (atoms N1, N6 and N7). For the OG complex,
1197 the iron-sulfur cluster domain and inter-domain linker were omitted so as to focus on the OG-
1198 recognition site found at the interface between NTD (residues 29-137 in *Gs* MutY) and CTD
1199 (residues 234-360 in *Gs* MutY). Residues that interact with OG are shown: Thr49, and Ser308 in
1200 the *Gs* MutY protein and corresponding residues in the LCHF MutYs. The distance versus time
1201 plot for the OG complex tracks potential contacts between the critical Ser/Thr residues and the
1202 hydrogen bond donors and acceptors on OG (atoms N1, N2, O6, N7 and O8). The total short
1203 range interaction energy (black trace) is the sum of short range Leanord-Jones (salmon trace) and
1204 Coulombic (sky blue) interaction energies. (A) Molecular dynamic simulation for *Gs* MutY NTD
1205 complexed with adenosine. The ligand complex persisted for the entire 100,000 ps, with changes
1206 in location and orientation evident at 16,000 ps and 42,000 ps in the distance plot. Hydrogen
1207 bonds between catalytic Glu43 and the Hoogsten face of the adenine base were observed during
1208 the first 16,000 ps. These consistently involved direct contact with N6, as evidenced by close
1209 distance (green traces) and inspection of structures. N7 was also engaged (blue traces), with
1210 relevance for catalysis, via bridging water molecules (O red and H white). (B) Molecular
1211 dynamic simulation for *Gs* MutY complexed with OG. The ligand complex was stable for 92,000
1212 ps, with the OG ligand bound to a cleft between the NTD (white) and CTD (gray). The
1213 functionally relevant hydrogen bond between the amide N of Ser308 and atom O8 of OG was
1214 frequently observed (not shown), sometimes accompanied by a second OG-specific hydrogen
1215 bond between the hydroxyl oxygen of Ser308 and atom N7 of OG (sky blue trace in the distance
1216 plot). (C) Molecular dynamic simulation for *Marinosulfonomonas* MutY NTD complexed with
1217 adenosine. In the first 3,000 ps, the adenine base approached closely catalytic Glu49 (green
1218 traces), often directly hydrogen bonded and occasionally bridged by a solvent molecule.
1219 However, the complex was relatively unstable, and the ligand departed the active site and found
1220 a new binding site by 8,000 ps. Favorable VDW interactions characterize both binding sites, but
1221 favorable Coulombic interactions are diminished substantially at the second site. (D) Molecular
1222 dynamic simulation for *Marinosulfonomonas* MutY complexed with OG. The initial ligand
1223 complex was unstable with a hinge-like motion creating new contacts between the NTD (white)
1224 and CTD (gray). After nearly escaping at ~4,000 ps, the OG ligand found several alternate sites
1225 on the NTD and CTD. (E) Molecular dynamic simulation for *Rhodobacteraceae* MutY NTD
1226 complexed with adenosine. The complex was relatively unstable. The adenine base initially
1227 hydrogen bonded with catalytic Glu45 during the first 3,800 ps but then changed orientation and
1228 drifted to a new site distinct and different from its original docking site. Note, catalytic E43 is
1229 not visible in the 10,000-ps representative structure as the new position of adenosine blocks its
1230 view. (F) Molecular dynamic simulation for *Rhodobacteraceae* MutY complexed with OG. The
1231 ligand complex was unstable and dissociated completely within 48 ns. Functionally relevant

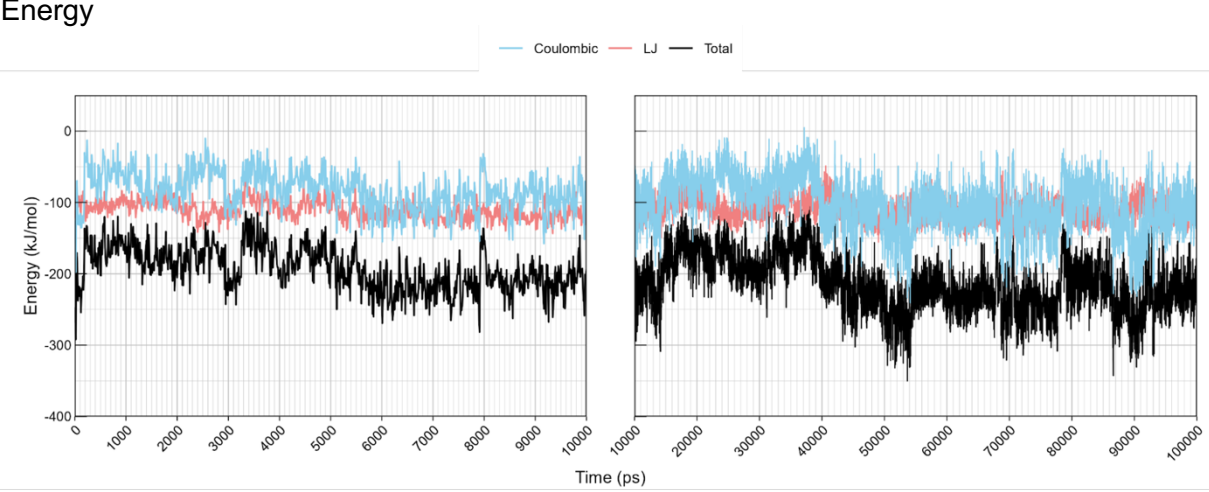
1232 hydrogen bonds between Thr299 and OG observed for the initial starting structure were lost as
1233 the ligand moved to new positions on the NTD and CTD before dissociation. (G) Molecular
1234 dynamic simulation for *Thiotrichaceae* MutY NTD complexed with adenosine. Note that Ser
1235 replaces active site Tyr for this LCHF MutY, as is also the case for *Ec* MutY. The complex was
1236 relatively stable with the ligand persisting in the active site throughout the simulation. Hydrogen
1237 bonds between catalytic Glu46 and the Hoogsteen face of adenine were evident by close distance
1238 to N7 (blue traces) and N6 (green traces) and by inspection of structures. Water frequently
1239 bridged N7 to Glu46 as seen in the representative structure at 10,000 ps. (H) Molecular dynamic
1240 simulation for *Thiotrichaceae* MutY complexed with OG. The ligand complex was stable for the
1241 entire 100,000-ps simulation with the OG ligand bound to a cleft between the NTD (white) and
1242 CTD (gray). Hydrogen bonds between Ser306 and OG were frequently observed. (I) Molecular
1243 dynamic simulation for *Flavobacteriaceae* MutY NTD complexed with adenosine. The ligand
1244 persisted in the active site throughout the simulation, with the ligand periodically finding new
1245 orientations as evident in different distance traces vying for close approach to catalytic Glu43.
1246 For example, N7 of the adenine base was very close to Glu43 (blue trace) during the first 2,700
1247 ps, suggesting catalytic engagement, but slipped out of reach at later time points. Water
1248 frequently bridged contacts between Glu43 and the adenine base. (J) Molecular dynamic
1249 simulation for *Flavobacteriaceae* MutY complexed with OG. The ligand complex was relatively
1250 stable. Functionally relevant hydrogen bonds between Ser305 and the Hoogsten face of OG can
1251 be inferred from recurring close distances up until 13,000 ps when the ligand adopts a new pose
1252 at the NTD-CTD interface.
1253

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

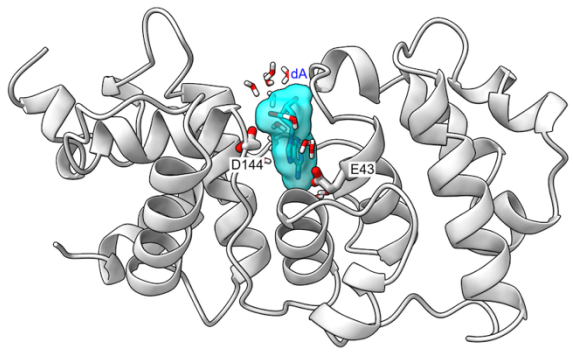
Distance Adenosine - Glu43



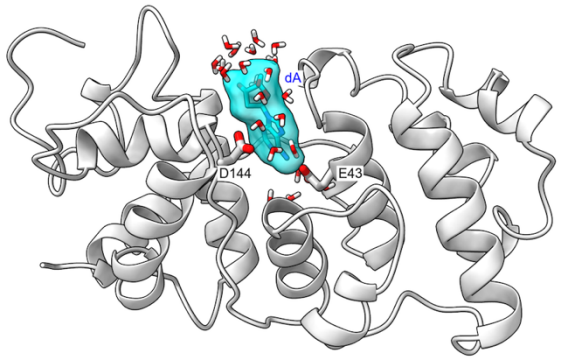
Energy



Adenosine and G. stearothermophilus MutY NTD 0,000 ps

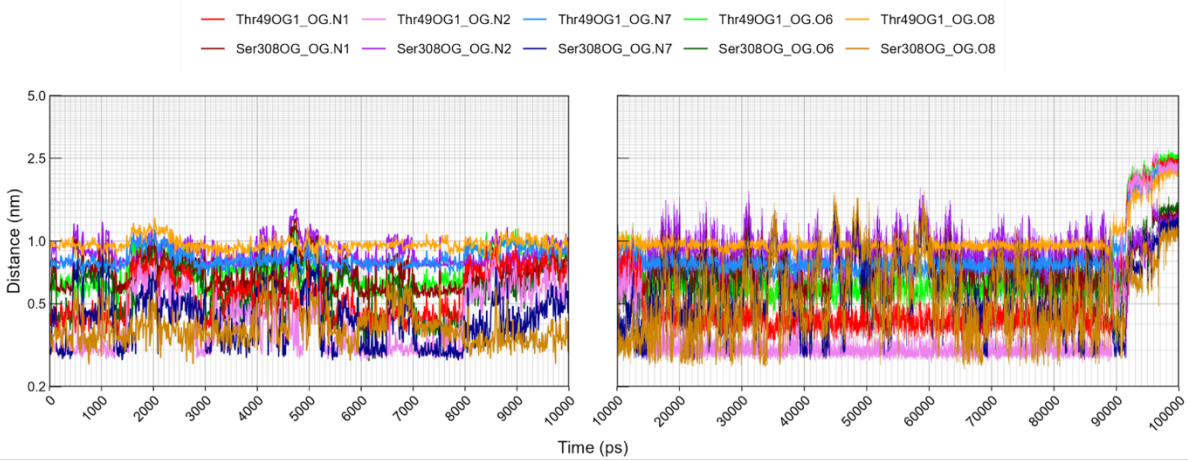


Adenosine and G. stearothermophilus MutY NTD 10,000 ps

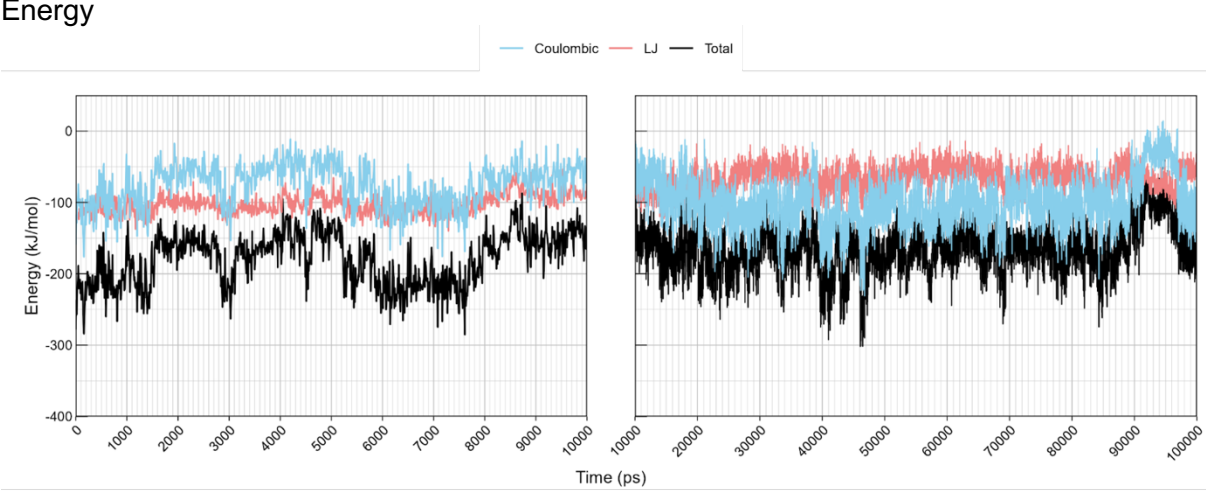


(B) *Geobacillus stearothermophilus* MutY complexed with OG

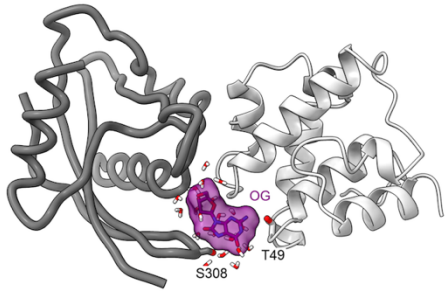
Distance OG - Thr49 and Ser308



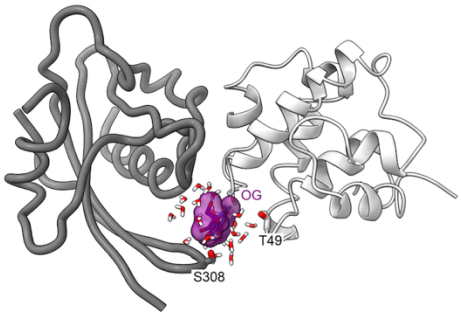
Energy



OG and Geobacillus stearothermophilus MutY 0,000 ps

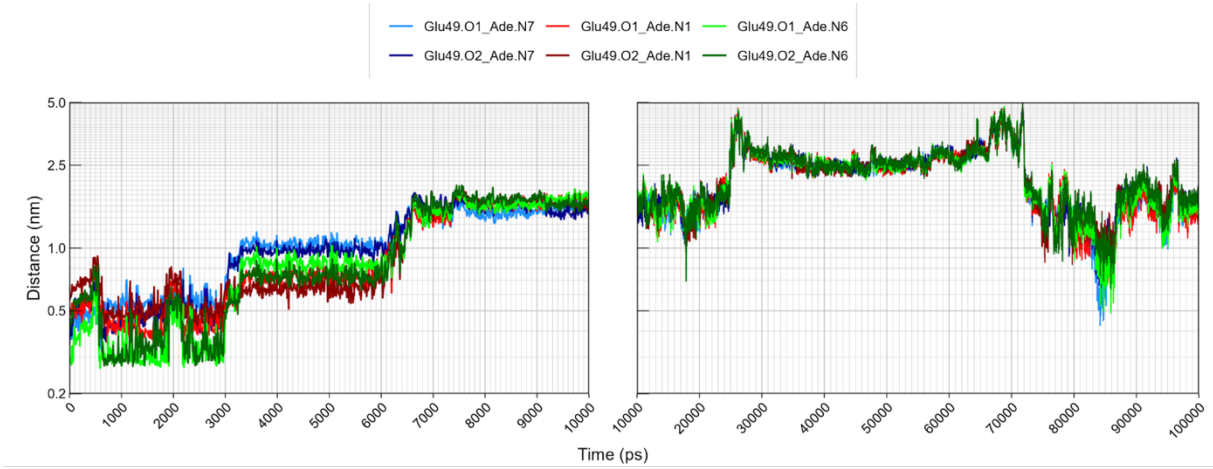


OG and Geobacillus stearothermophilus MutY 10,000 ps

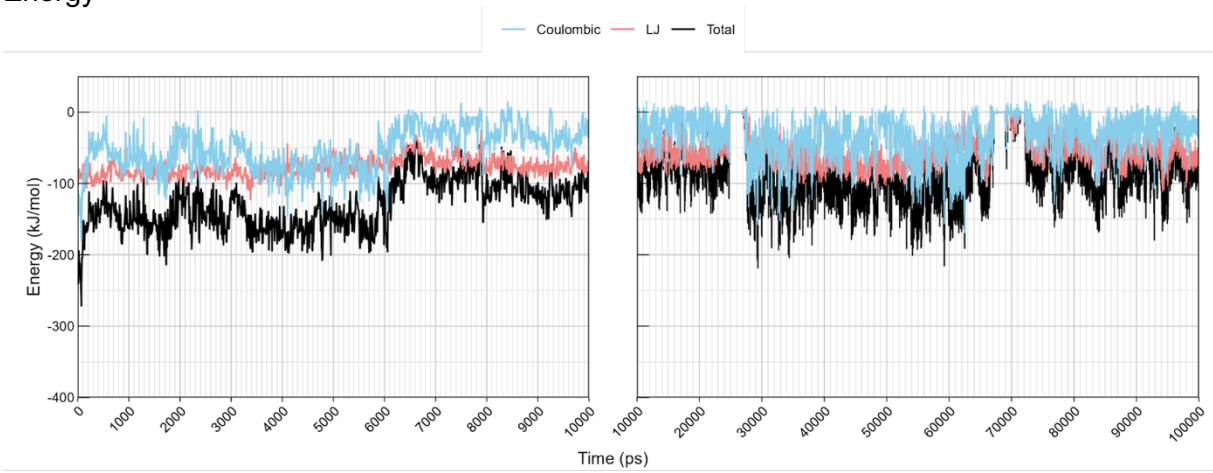


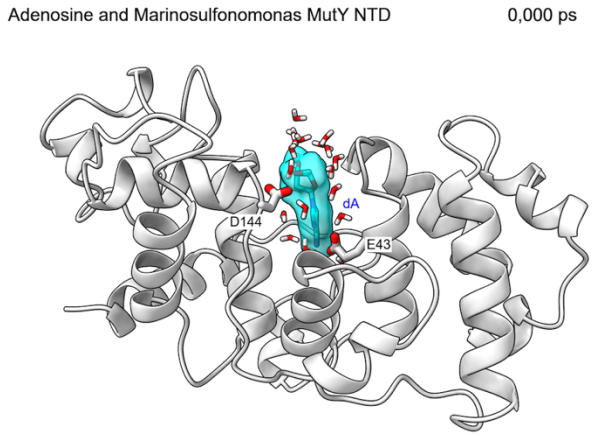
(C) *Marinosulfonomonas* MutY NTD complexed with adenosine

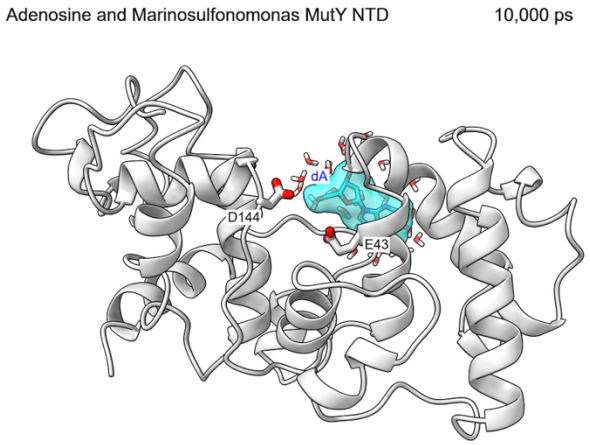
Distance Adenosine - Glu49



Energy

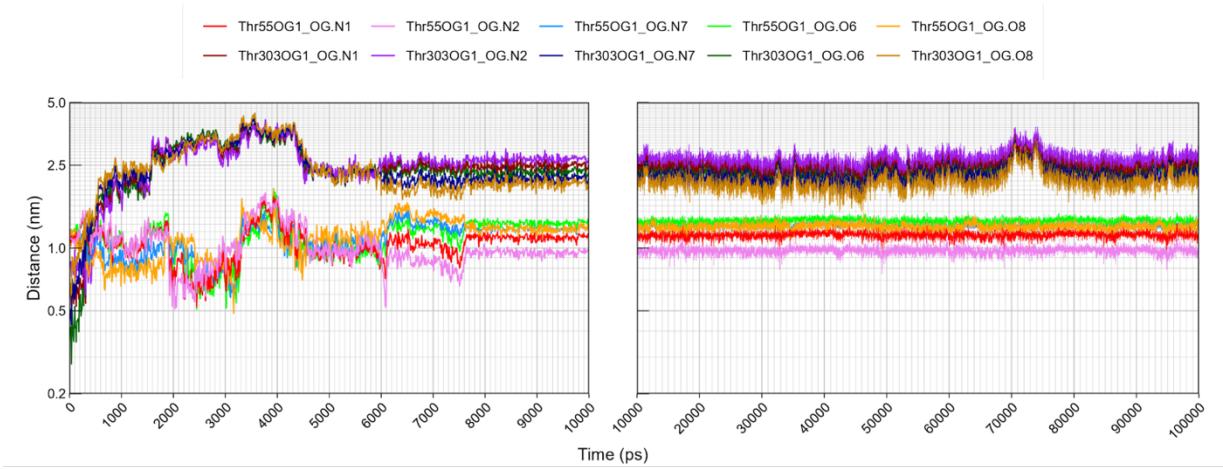




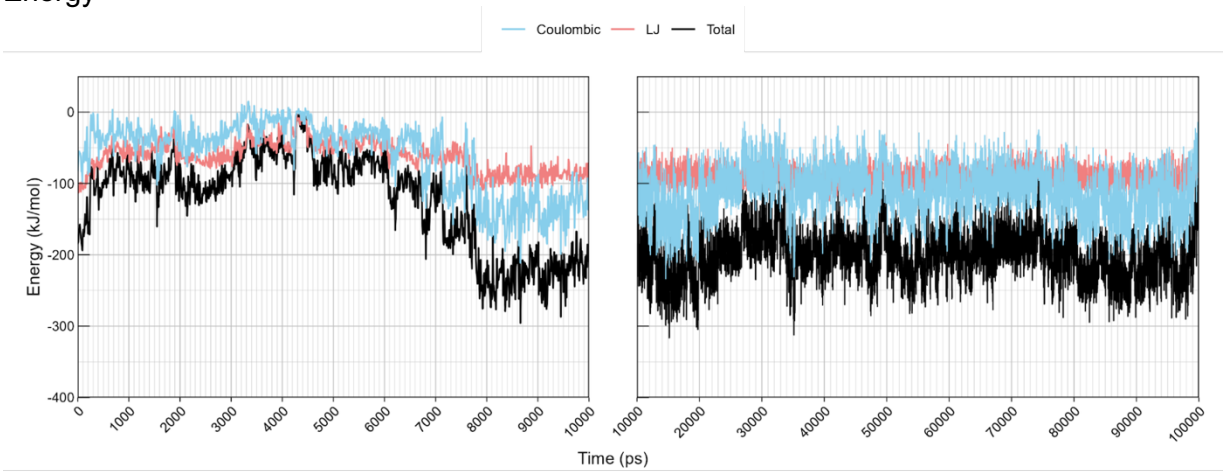


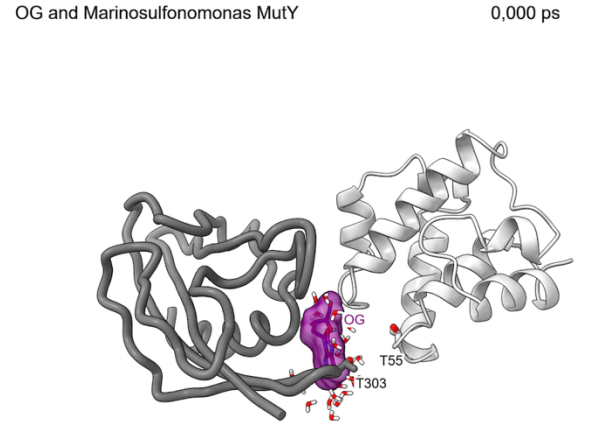
(D) *Marinosulfonomonas* MutY complexed with OG

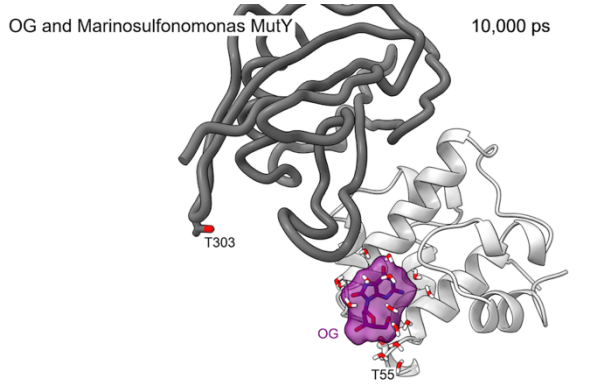
Distance OG - Thr55 and Thr303



Energy

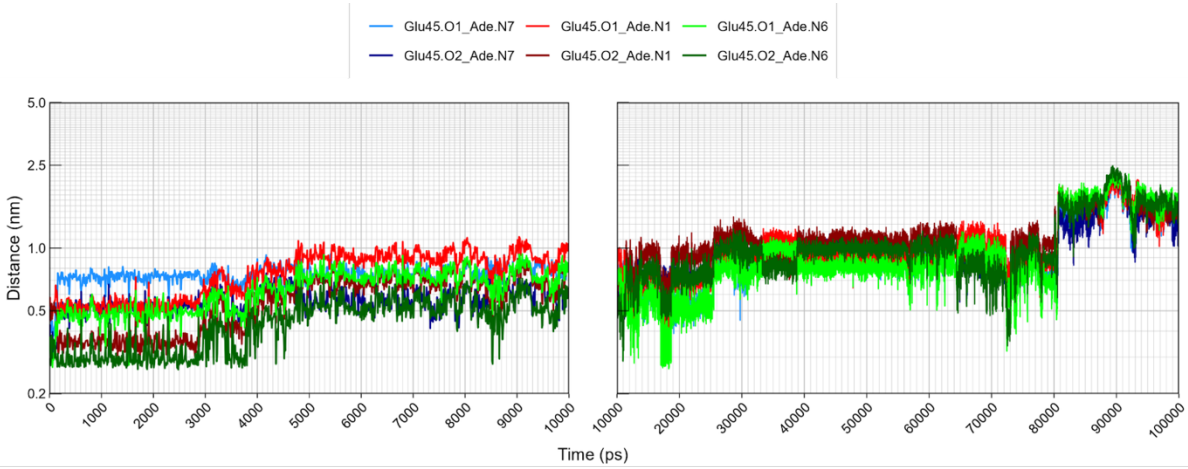




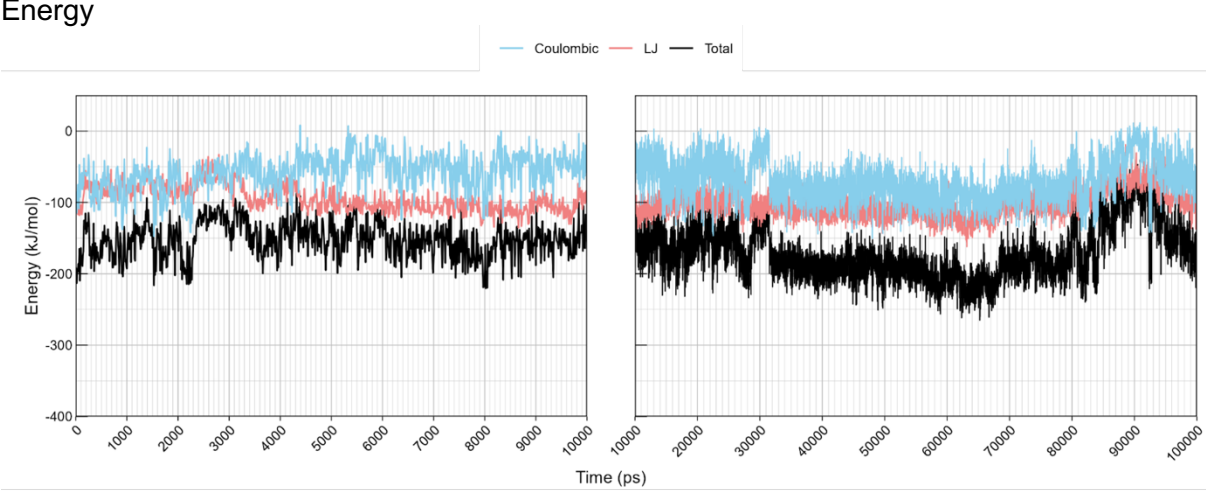


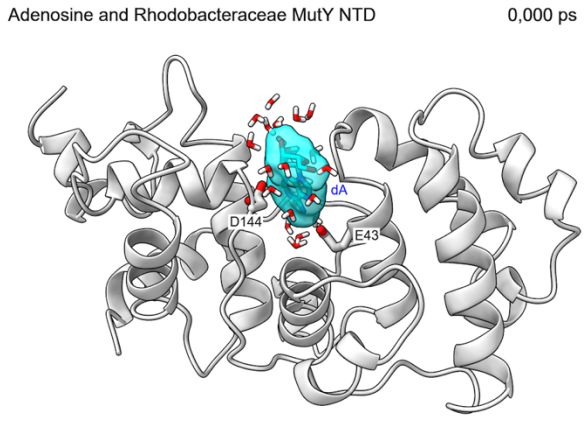
(E) *Rhodobacteraceae* MutY NTD complexed with adenosine

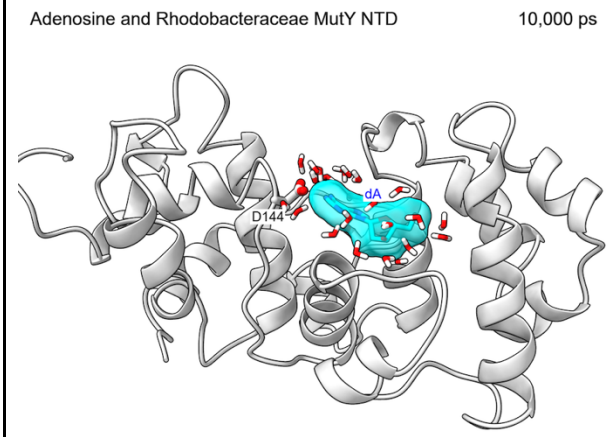
Distance Adenosine - Glu45



Energy

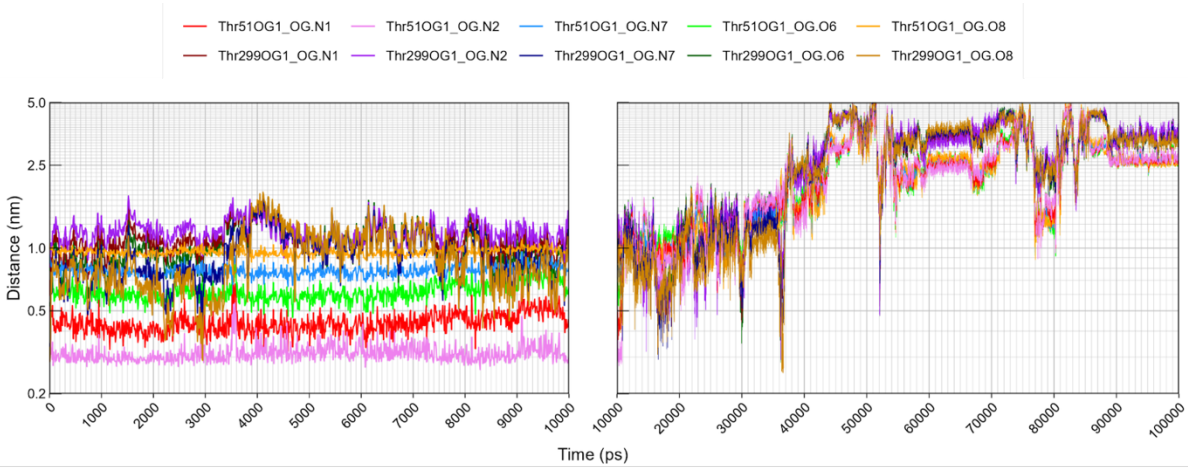




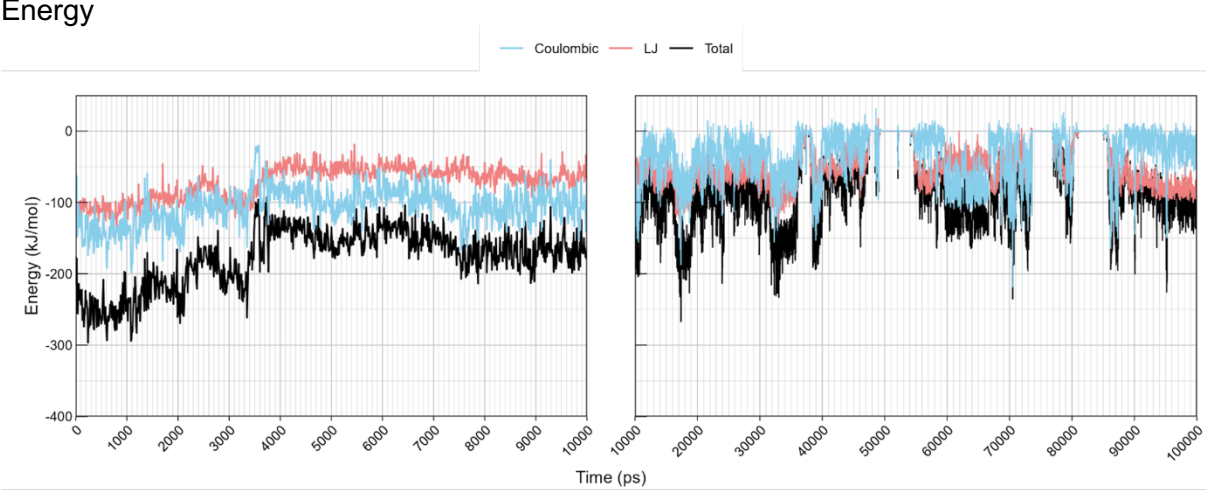


(F) *Rhodobacteraceae* MutY complexed with OG

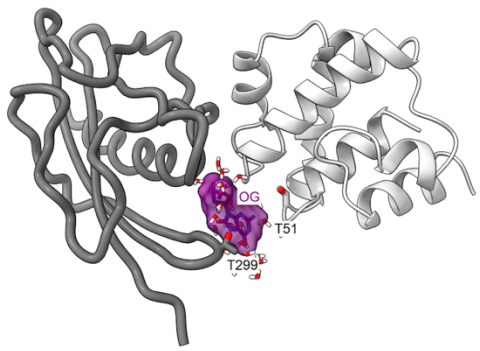
Distance OG - Thr51 and Thr299



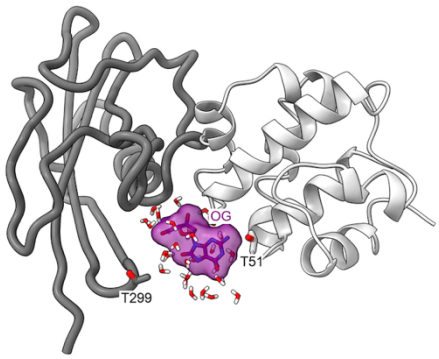
Energy



OG and Rhodobacteraceae MutY 0,000 ps

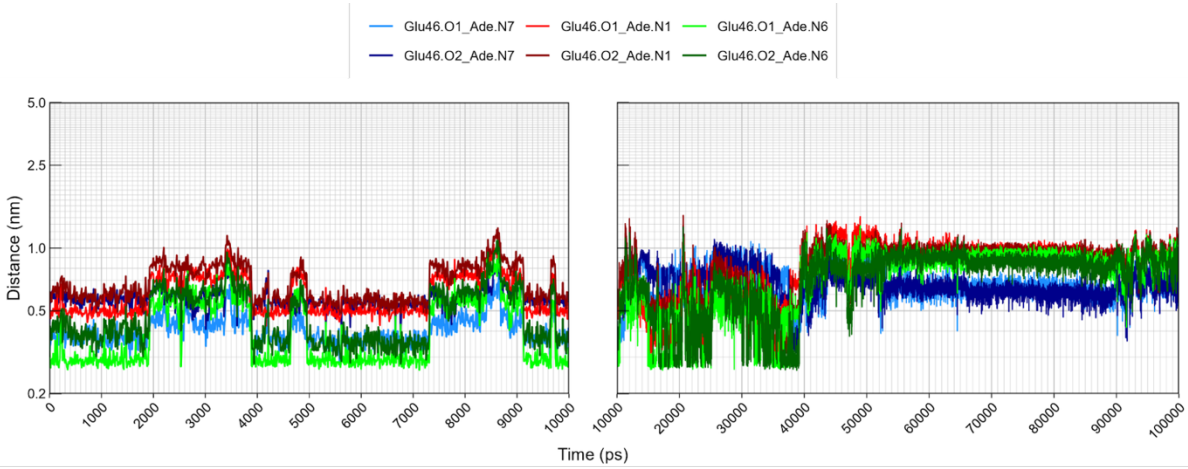


OG and Rhodobacteraceae MutY 10,000 ps

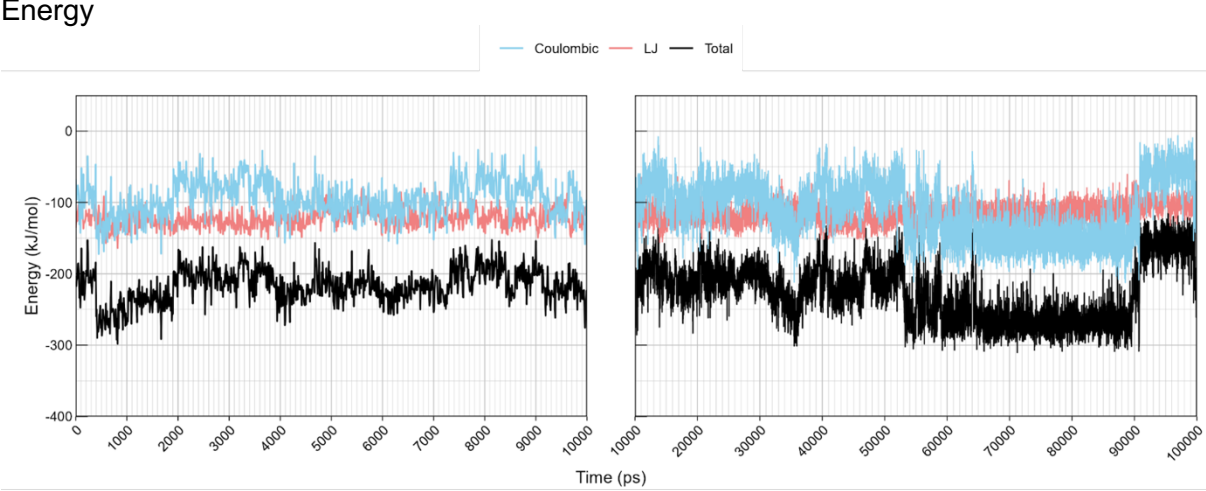


(G) *Thiotrichaceae* MutY NTD complexed with adenosine

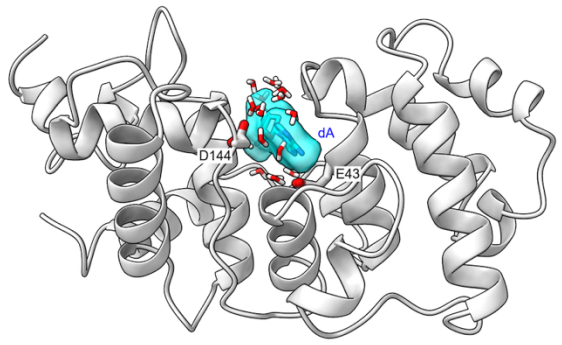
Distance Adenosine - Glu46



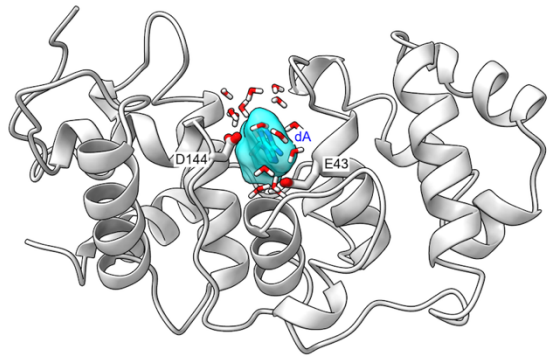
Energy



Adenosine and Thiotrichaceae MutY NTD 0,000 ps

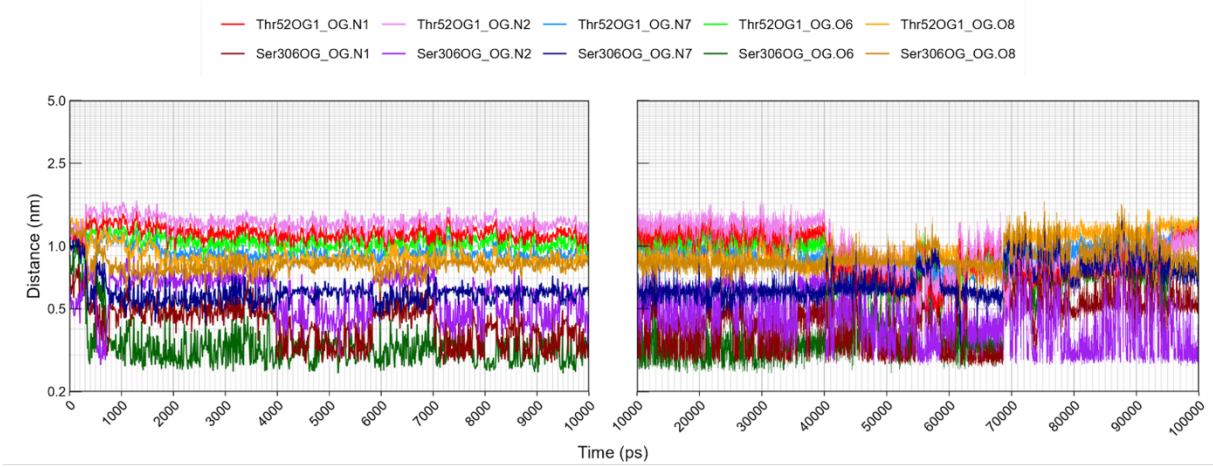


Adenosine and Thiotrichaceae MutY NTD 10,000 ps

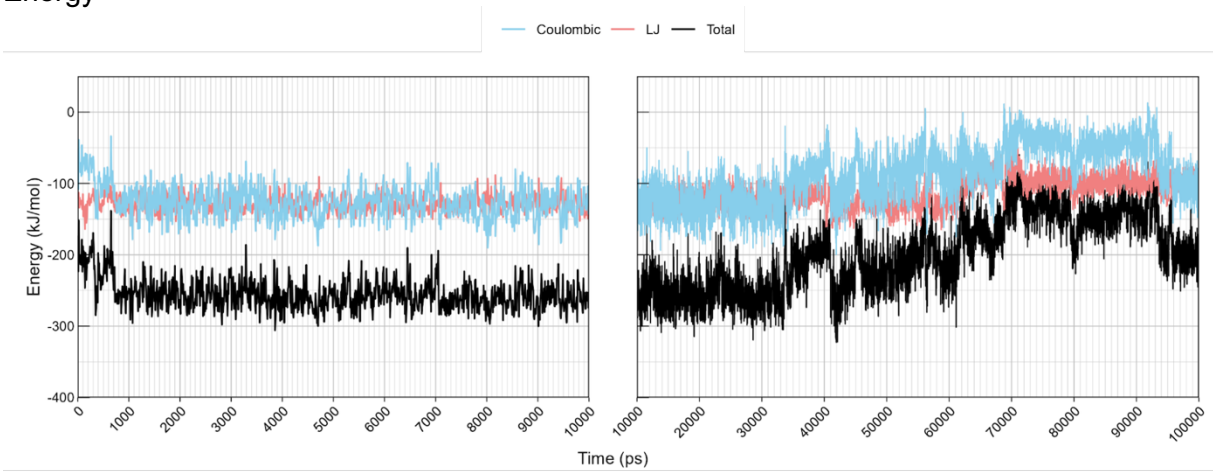


(H) *Thiotrichaceae* MutY complexed with OG

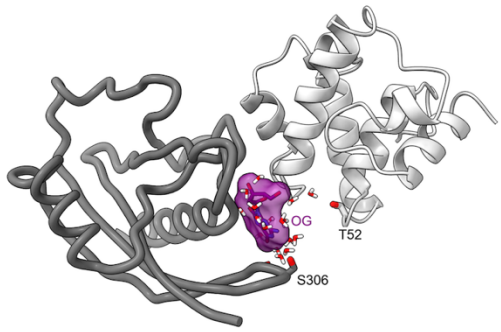
Distance OG - Thr52 and Ser306



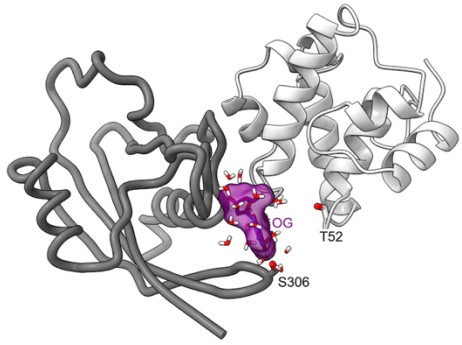
Energy



OG and Thiotrichaceae MutY 0,000 ps

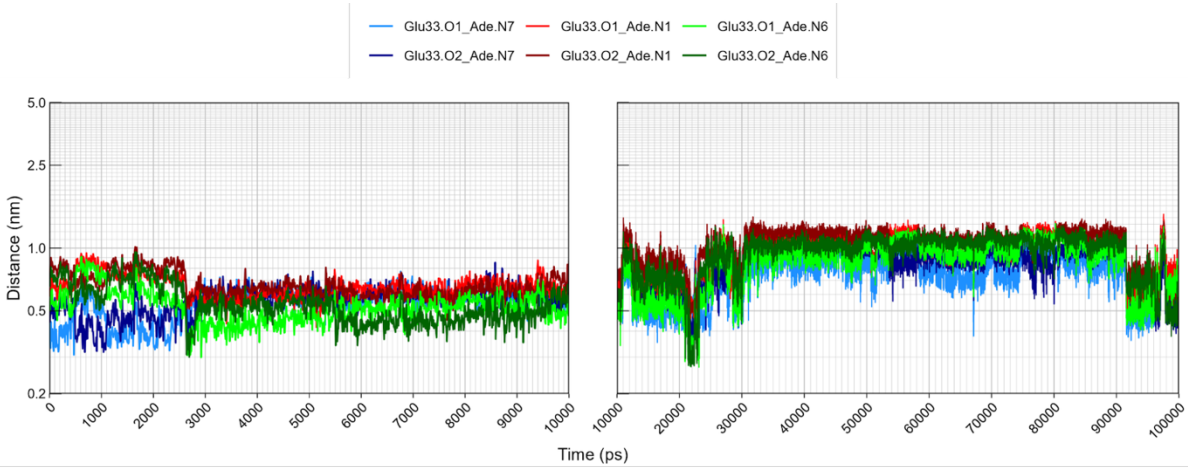


OG and Thiotrichaceae MutY 10,000 ps

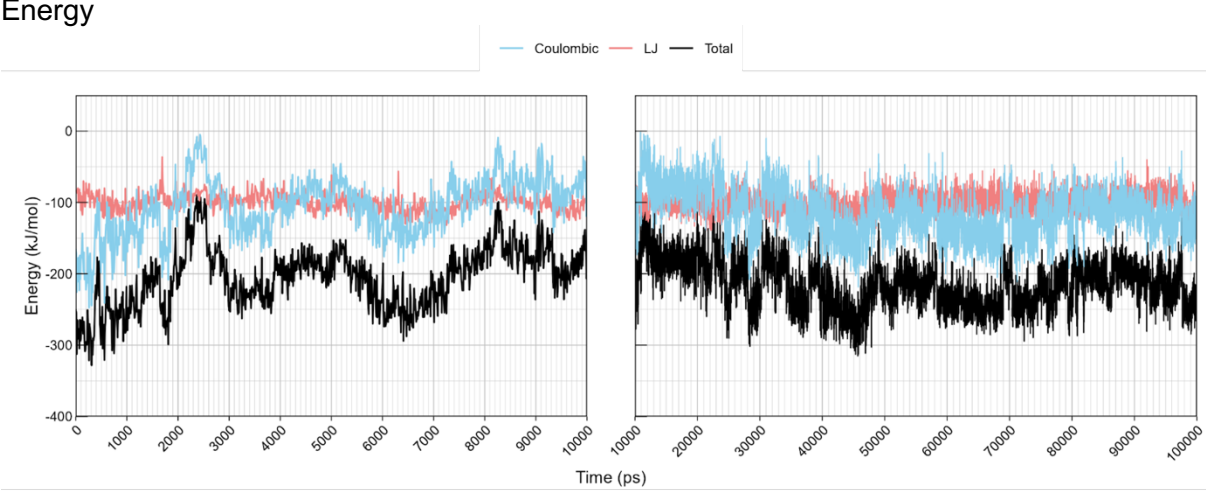


(I) *Flavobacteriaceae* MutY NTD complexed with adenosine

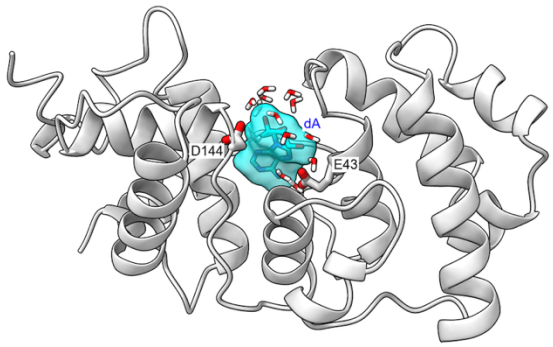
Distance Adenosine - Glu33



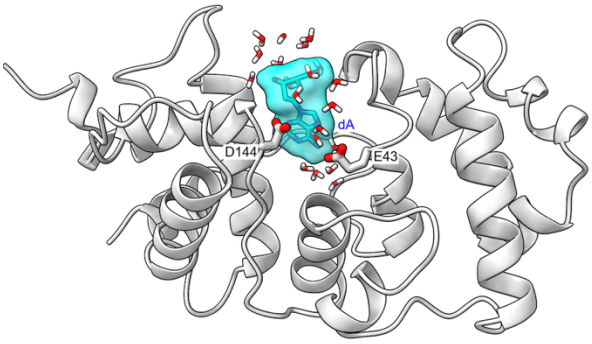
Energy



Adenosine and Flavobacteriaceae MutY NTD 0,000 ps

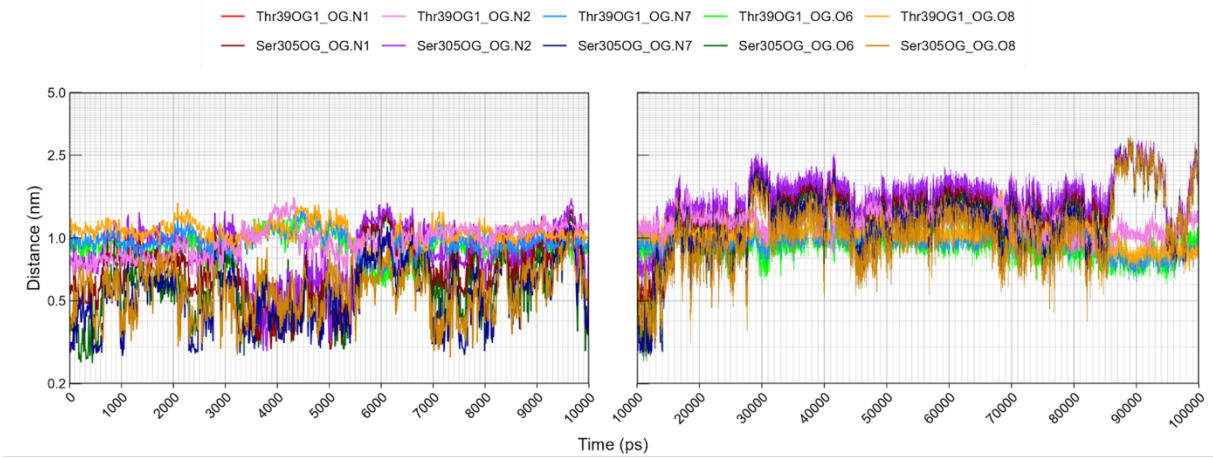


Adenosine and Flavobacteriaceae MutY NTD 10,000 ps

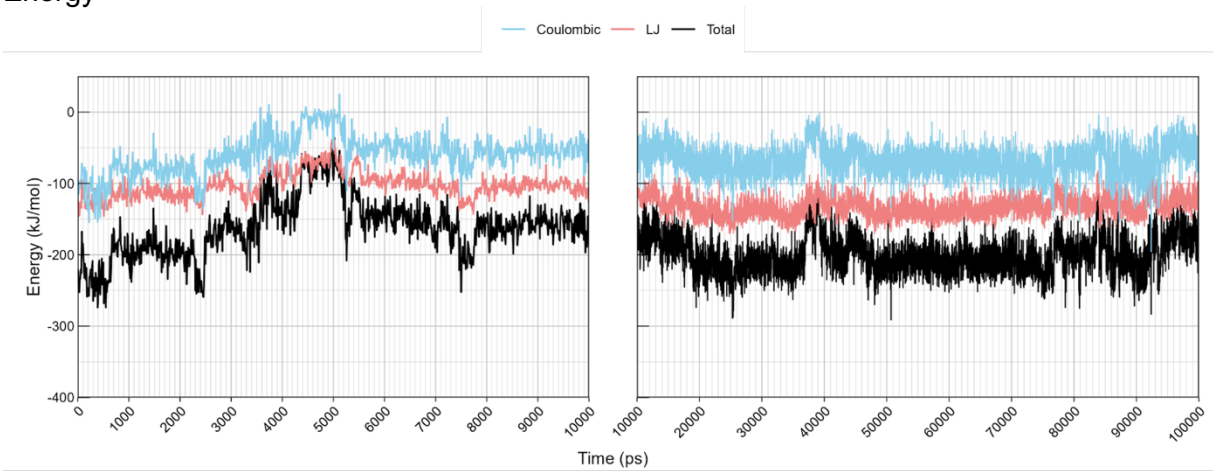


(J) *Flavobacteriaceae* MutY complexed with OG

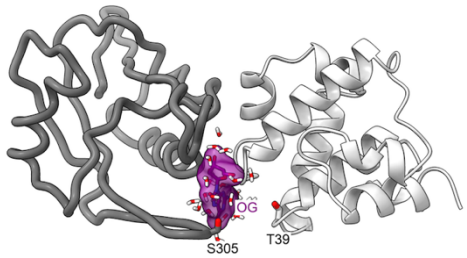
Distance OG - Thr39 and Ser305



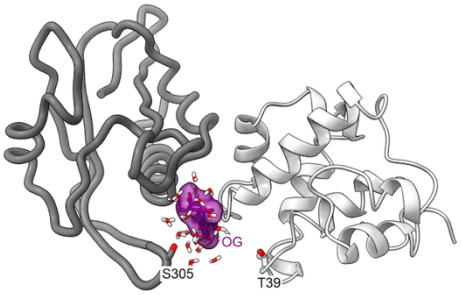
Energy



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 helixes 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

1348
1349 **S9 Table. Rifampicin resistance assay.**

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

1350

1351 [a] Confidence intervals (95%) determined by a bootstrap method, see **Materials and methods** for details.
1352 [b] Mutation frequency reported as median number of resistant colonies per 10^8 viable colonies. Fold change was
1353 calculated by dividing Rif^R frequency by the frequency measured for cultures expressing *Ec* MutY.
1354