

1 **Metagenome mining and functional analysis reveal oxidized guanine**

2 **DNA repair at the Lost City Hydrothermal Field**

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

4 Brazelton, Martin P. Horvath^{*}

5

6

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

8 USA

9

10 *Corresponding author: martin.horvath@utah.edu

11

12 [¶]These authors contributed equally to this work.

13

14

15

16

17

18

19 Abstract

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

33

34 Introduction

35 The Lost City Hydrothermal Field (LCHF) resembles conditions of a younger planet and
36 thus provides a window to study the origin of life on Earth and other planets (1–3). Located 15
37 kilometers from the Mid-Atlantic Ridge, the LCHF comprises a series of carbonate chimneys at
38 depths ranging from 700–800 meters below the ocean surface (3). The temperature and pH of the
39 LCHF are both elevated, with temperature ranging from 40°C to 90°C and pH ranging from 9 to
40 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.

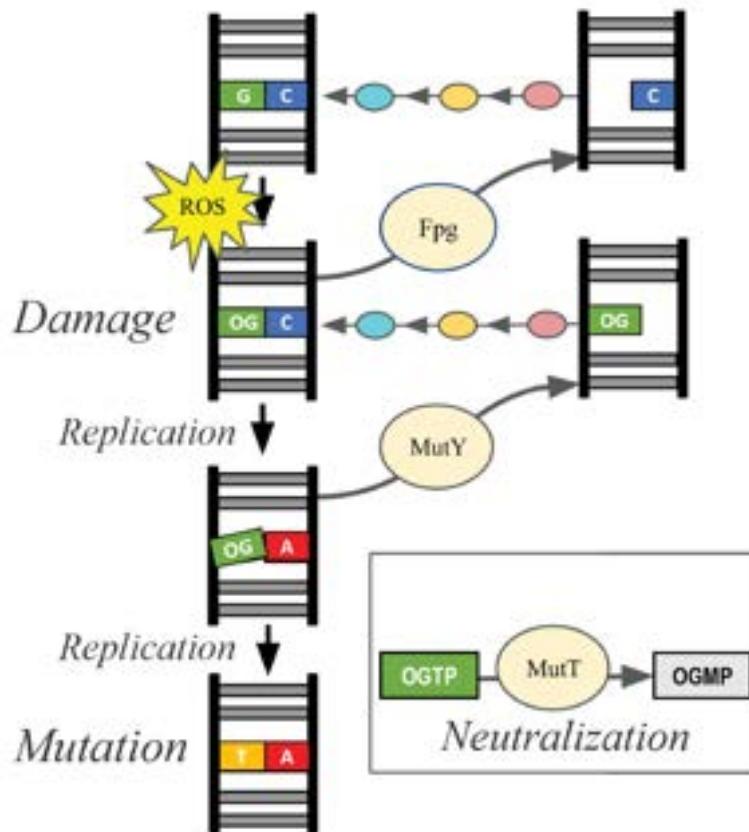


Fig 1. Overview of the GO System.

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).

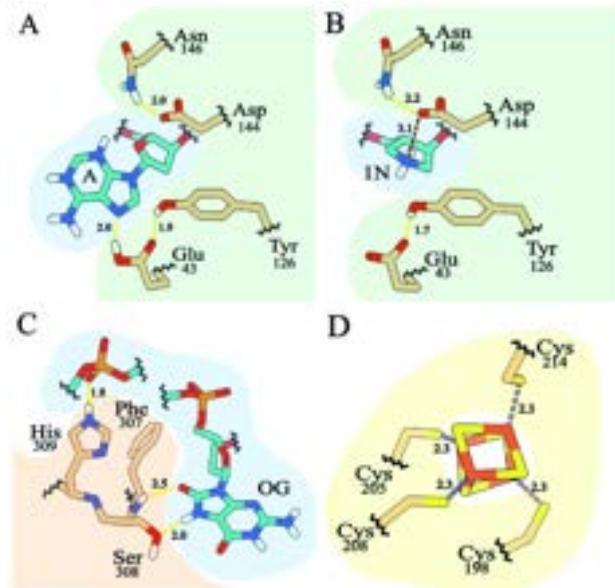


Fig 2. MutY chemical motifs. Panels A-C show interactions between MutY residues and DNA, with DNA highlighted in blue. Panel A was made from PDB ID 3g0q, a structure which describes interaction with the competitive inhibitor fluoro-A (31); panels B-D were made from PDB ID 6u7t, a structure which describes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

143
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

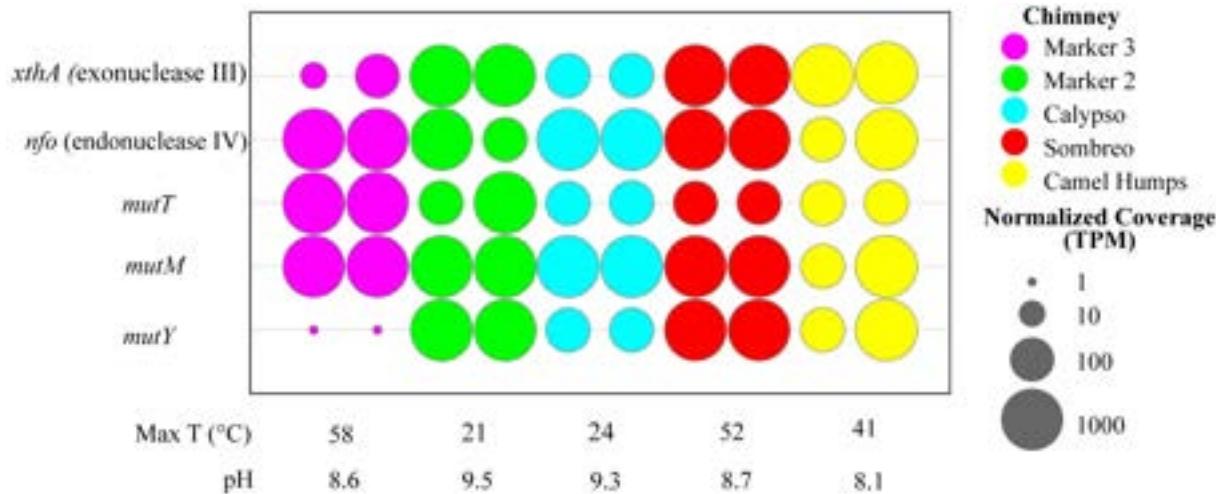
157 confidently to metagenome assembled genomes (MAG)s, allowing for additional gene inventory
158 analyses that revealed metabolic strategies involving sulfur oxidation and nitrogen reduction.
159 These results have important implications for understanding the repair of oxidative guanine
160 damage 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).



175
176 **Fig 3. Abundance of GO system gene homologs.** Listed on the vertical axis are genes encoding
177 DNA repair enzymes. Genes *xthA* and *nfo* are generally necessary for DNA repair involving base
178 excision repair in bacteria, including the particular GO system investigated here. Together, *mutT*,
179 *mutM* and *mutY* constitute the GO system that deals specifically with oxidized guanine. Across
180 the horizontal axis are the various LCHF sites, coded by color, from which samples were
181 collected in duplicate, along with the reported temperature and pH. The normalized coverage of
182 each gene is reported as a proportional unit suitable for cross-sample comparisons, the
183 transcripts/fragments per million (TPM).
184

185 Metagenomic mining for MutY genes

186 Having determined that GO system gene homologs are abundant at the LCHF, we
187 focused our efforts on the final safeguard of the pathway, MutY. A BLASTP search against the
188 LCHF metagenome with query MutY sequences from *Geobacillus stearothermophilus* (*Gs*
189 MutY) and *E. coli* (*Ec* MutY) preliminarily identified 649 putative MutY candidates on the basis
190 of sequence identity, excluding hits with less than a 30% sequence identity cut-off or E-values
191 exceeding 1E-5 (Fig 4A). Structure-guided alignments of these preliminary hits were examined
192 for presence and absence of MutY-defining chemical motifs. We paid particular attention to the
193 chemical motif associated with OG recognition as these residues in the C-terminal domain
194 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**).

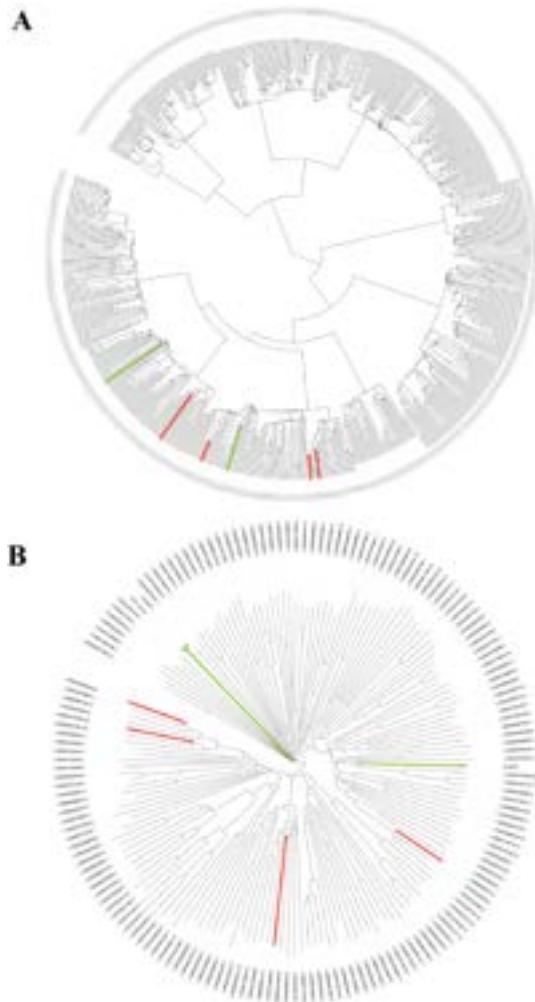


Fig 4. Phylogeny for LCHF MutYs. (A) 649 sequences were identified as LCHF MutY candidates due to sequence similarity to existing MutYs (green branches) and aligned to reconstruct evolutionary relationships. (B) A subset of 160 members contained all necessary MutY-defining chemical motifs. Alignment 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),
216 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.

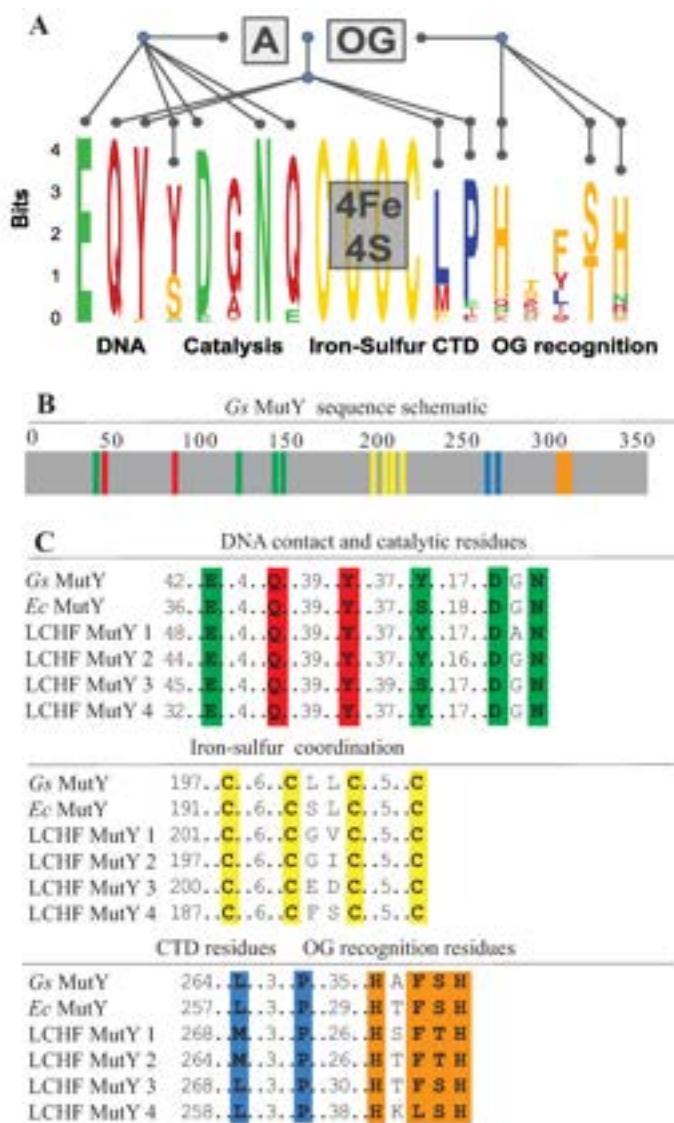


Fig 5. LCHF MutY chemical motifs. (A) Conservation and diversity of MutY-defining chemical motifs are depicted with a sequence logo for the 160 LCHF MutYs. These motifs are associated with biochemical functions including DNA binding, enzyme catalysis, attachment of the iron-sulfur cofactor, and recognition 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).

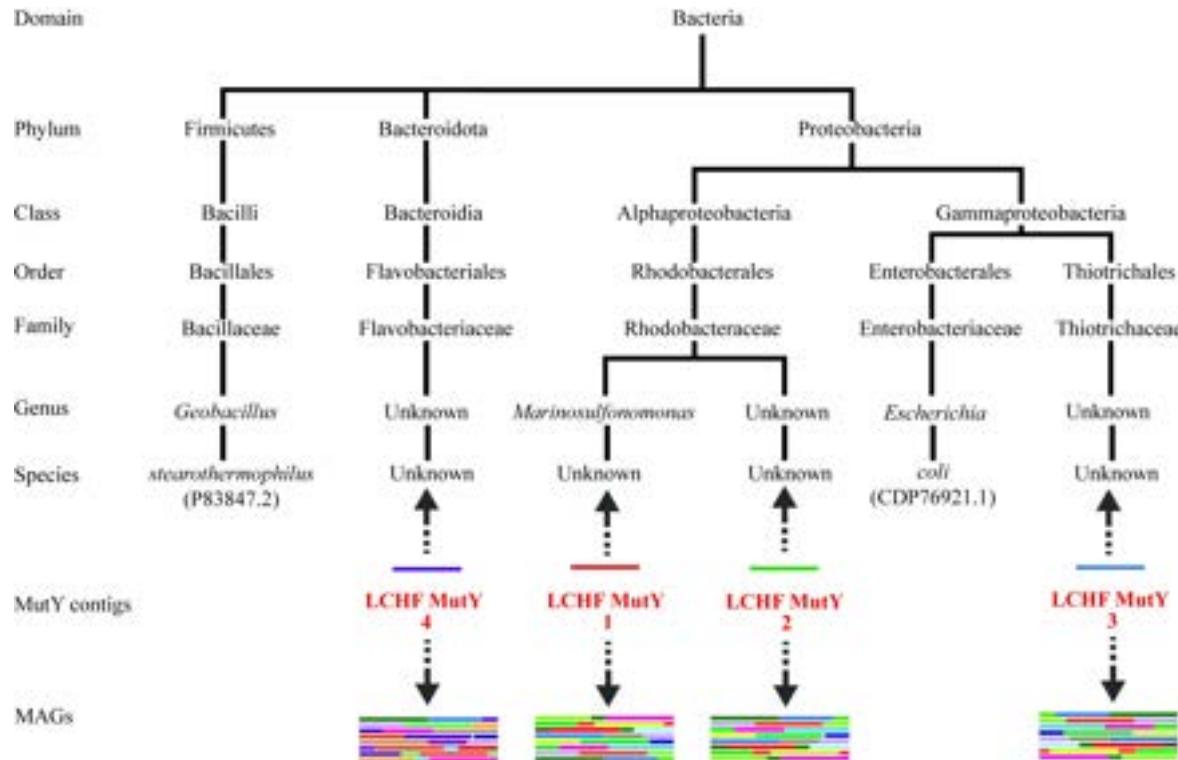


Fig 6. Taxonomic classification. LCHF MutY-encoding contigs were found in several branches of bacteria. The classification places these in relation to MutY of *G. stearothermophilus* and *E. coli* (accession IDs included). LCHF MutYs were mapped to their respective microbes by two methods indicated by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

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

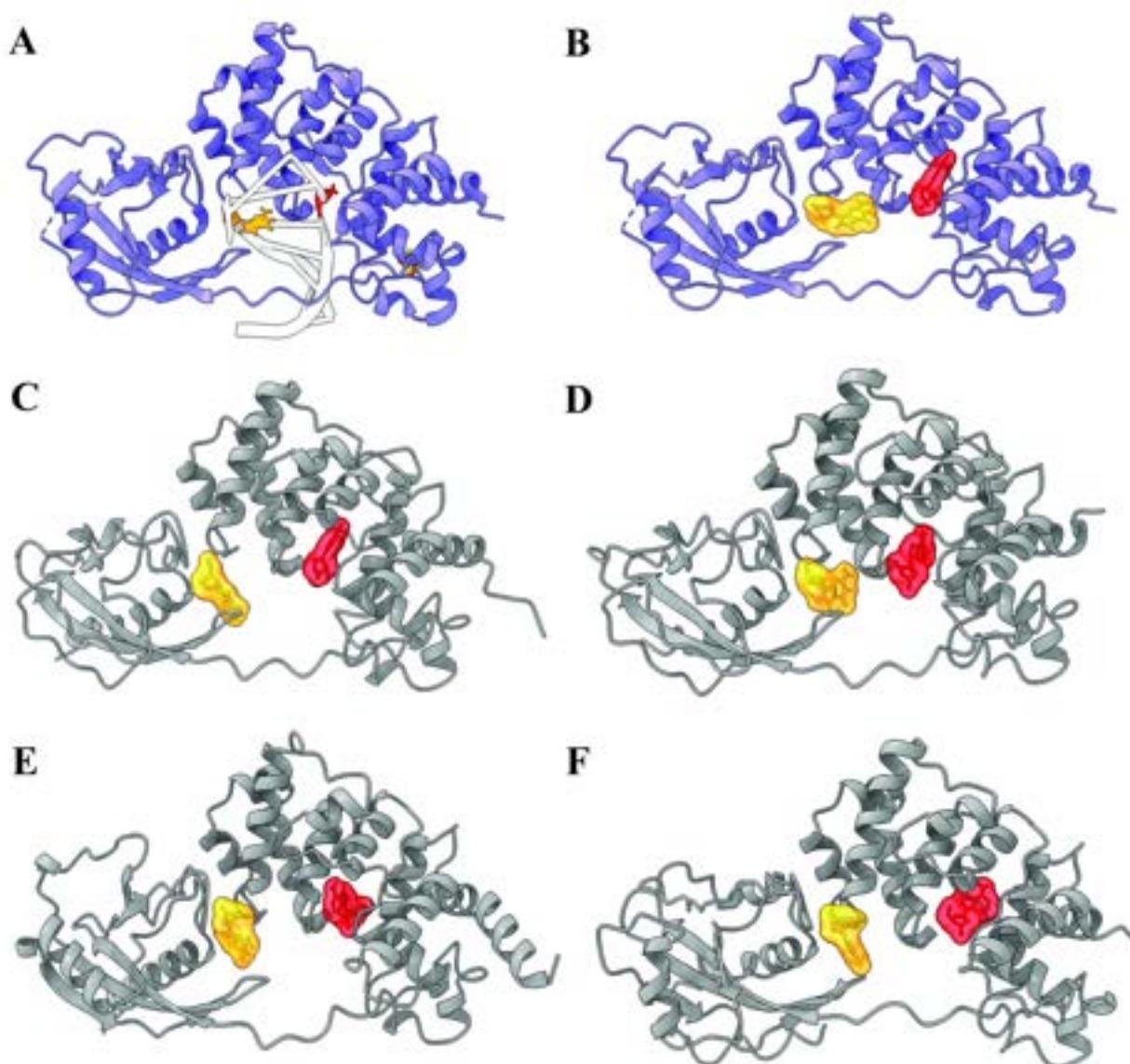
355

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.



373
374 **Fig 7. Structure predictions and virtual docking of MutY ligands.** (A) The x-ray crystal
375 structure of *Gs* MutY (blue ribbon; PDB ID 3g0q) in complex with DNA (white helix) highlights
376 the positions of the adenosine nucleoside (red) and OG (yellow). (B-F) Virtual docking of
377 ligands. Adenosine and OG were separately docked to identify binding surfaces for these ligands
378 in the structures of *Gs* MutY (B), which served as a positive control, and the four representative
379 LCHF MutYs: *Marinosulfonomonas* MutY (C); *Rhodobacteraceae* MutY (D); *Thiotrichaceae*
380 MutY (E); and *Flavobacteriaceae* MutY (F).
381

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

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

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 *Marinosulfolonomonas* 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 *Marinosulfolonomonas* 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).

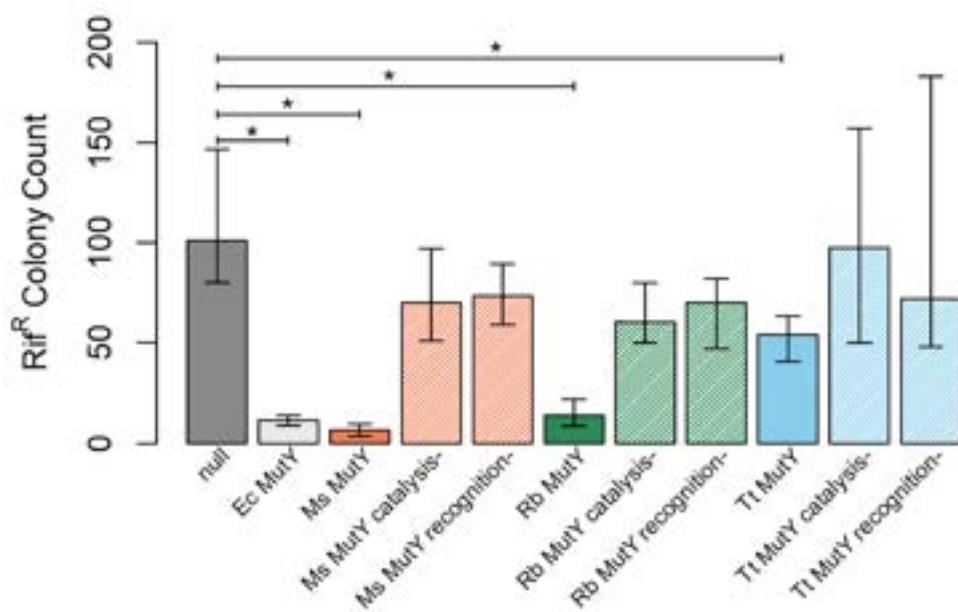


Fig 8. Functional analysis. Bars represent median Rif^R colony counts for *E. coli* cultures expressing MutY, MutY variants, or no MutY (*null*) from a plasmid DNA. Error bars represent 95% confidence intervals as determined by bootstrap sampling (see Supporting Information Table

545 S9 for tabulated values). *Marinosulfolonomonas* (*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 O₂ 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 O₂ 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 O₂.

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

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

677

678 **Materials and Methods**

679 **Metagenomic sequencing and analysis of LCHF fluid samples**

680 Generation, assembly, and annotation of metagenomes from the Lost City Hydrothermal
681 Field (LCHF) have been described previously (35), and are briefly summarized here. In 2018,
682 the remotely operated vehicle (ROV) Jason collected samples of fluids venting from chimneys at
683 the LCHF, which is located near the Mid-Atlantic Ridge at 30 °N latitude and a depth of ~800 m.
684 Whole-genome community sequences ("metagenomes") were generated from the fluid samples,
685 and assembled metagenomic contigs were binned into metagenome-assembled genomes
686 (MAGs). Potential gene homologs encoding enzymes involved in the GO DNA repair system
687 were identified by conducting KEGG (82,83) orthology assignment using the *BlastKOALA* v2.2
688 program (84). The selected genes that were identified include: *mutT* (KEGG ID: K03574), *mutM*
689 (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 Å³
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 *Marinosulfolonomonas* 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 DH5 α
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).

813 Acknowledgements

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

831 References

- 832 1. Kelley DS, Karson JA, Früh-Green GL, Yoerger DR, Shank TM, Butterfield DA, et al. A
833 Serpentinite-Hosted Ecosystem: The Lost City Hydrothermal Field. *Science*. 2005 Mar
834 4;307(5714):1428–34.
- 835 2. Amador ES, Bandfield JL, Brazelton WJ, Kelley D. The Lost City Hydrothermal Field: A
836 Spectroscopic and Astrobiological Analogue for Nili Fossae, Mars. *Astrobiology*. 2017

837 Nov;17(11):1138–60.

838 3. Kelley DS, Karson JA, Blackman DK, Früh-Green GL, Butterfield DA, Lilley MD, et al.

839 An off-axis hydrothermal vent field near the Mid-Atlantic Ridge at 30° N. *Nature*. 2001

840 Jul;412(6843):145–9.

841 4. Brazelton WJ, Schrenk MO, Kelley DS, Baross JA. Methane- and Sulfur-Metabolizing

842 Microbial Communities Dominate the Lost City Hydrothermal Field Ecosystem. *Appl*

843 *Environ Microbiol*. 2006 Sep;72(9):6257–70.

844 5. Russell MJ, Hall AJ, Martin W. Serpentinization as a source of energy at the origin of life.

845 *Geobiology*. 2010 Dec;8(5):355–71.

846 6. Proskurowski G, Lilley MD, Seewald JS, Früh-Green GL, Olson EJ, Lupton JE, et al.

847 Abiogenic hydrocarbon production at lost city hydrothermal field. *Science*. 2008 Feb

848 1;319(5863):604–7.

849 7. Lang SQ, Butterfield DA, Lilley MD, Paul Johnson H, Hedges JI. Dissolved organic

850 carbon in ridge-axis and ridge-flank hydrothermal systems. *Geochim Cosmochim Acta*

851 [Internet]. 2006 Aug [cited 2023 Nov 28];70(15). Available from:

852 <https://par.nsf.gov/biblio/10088640-dissolved-organic-carbon-ridge-axis-ridge-flank-hydrothermal-systems>

853 8. Lang SQ, Brazelton WJ. Habitability of the marine serpentinite subsurface: a case study of

854 the Lost City hydrothermal field. *Philos Trans R Soc Math Phys Eng Sci*. 2020 Feb

855 21;378(2165):20180429.

856 9. Bruskov VI, Malakhova LV, Masalimov ZK, Chernikov AV. Heat-induced formation of

857 reactive oxygen species and 8-oxoguanine, a biomarker of damage to DNA. *Nucleic Acids*

858 *Res*. 2002 Mar 15;30(6):1354–63.

859 10. Kino K, Hirao-Suzuki M, Morikawa M, Sakaga A, Miyazawa H. Generation, repair and

860 replication of guanine oxidation products. *Genes Environ*. 2017 Aug 1;39:21.

861 11. Nghiem Y, Cabrera M, Cupples CG, Miller JH. The mutY gene: a mutator locus in

862 *Escherichia coli* that generates G.C----T.A transversions. *Proc Natl Acad Sci*. 1988

863 Apr;85(8):2709–13.

864 12. Cabrera M, Nghiem Y, Miller JH. mutM, a second mutator locus in *Escherichia coli* that

865 generates G.C----T.A transversions. *J Bacteriol*. 1988 Nov;170(11):5405–7.

866 13. Maki H, Sekiguchi M. MutT protein specifically hydrolyses a potent mutagenic substrate

867 for DNA synthesis. *Nature*. 1992 Jan;355(6357):273–5.

868 14. Michaels ML, Cruz C, Grollman AP, Miller JH. Evidence that MutY and MutM combine

869 to prevent mutations by an oxidatively damaged form of guanine in DNA. *Proc Natl Acad*

870 *Sci U S A*. 1992 Aug 1;89(15):7022–5.

871 15. Prakash A, Doublie S, Wallace SS. Chapter 4 - The Fpg/Nei Family of DNA Glycosylases:

872 Substrates, Structures, and Search for Damage. In: Doetsch PW, editor. *Progress in*

873 *Molecular Biology and Translational Science* [Internet]. Academic Press; 2012 [cited 2023

874 Mar 20]. p. 71–91. (Mechanisms of DNA Repair; vol. 110). Available from:

875 <https://www.sciencedirect.com/science/article/pii/B9780123876652000043>

876 16. Denver DR, Swenson SL, Lynch M. An evolutionary analysis of the helix-hairpin-helix

877 superfamily of DNA repair glycosylases. *Mol Biol Evol*. 2003 Oct;20(10):1603–11.

878 17. Trasviña-Arenas CH, Demir M, Lin WJ, David SS. Structure, function and evolution of the

879 Helix-hairpin-Helix DNA glycosylase superfamily: Piecing together the evolutionary

880 puzzle of DNA base damage repair mechanisms. *DNA Repair*. 2021 Dec 1;108:103231.

881 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

multiple Amber force fields and development of improved protein backbone parameters. *Proteins*. 2006 Nov 15;65(3):712–25.

51. Manlove AH, McKibbin PL, Doyle EL, Majumdar C, Hamm ML, David SS. Structure–Activity Relationships Reveal Key Features of 8-Oxoguanine: A Mismatch Detection by the MutY Glycosylase. *ACS Chem Biol*. 2017 Sep 15;12(9):2335–44.

52. Lee AJ, Majumdar C, Kathe SD, Van Ostrand RP, Vickery HR, Averill AM, et al. Detection of OG:A Lesion Mispairs by MutY Relies on a Single His Residue and the 2-Amino Group of 8-Oxoguanine. *J Am Chem Soc*. 2020 Aug 5;142(31):13283–7.

53. Liu X, Shi D, Zhou S, Liu H, Liu H, Yao X. Molecular dynamics simulations and novel drug discovery. *Expert Opin Drug Discov*. 2018 Jan;13(1):23–37.

54. Guterres H, Im W. Improving Protein-Ligand Docking Results with High-Throughput Molecular Dynamics Simulations. *J Chem Inf Model*. 2020 Apr 27;60(4):2189–98.

55. Nikkel DJ, Wetmore SD. Distinctive Formation of a DNA-Protein Cross-Link during the Repair of DNA Oxidative Damage: Insights into Human Disease from MD Simulations and QM/MM Calculations. *J Am Chem Soc*. 2023 Jun 21;145(24):13114–25.

56. Wang L, Lee SJ, Verdine GL. Structural Basis for Avoidance of Promutagenic DNA Repair by MutY Adenine DNA Glycosylase. *J Biol Chem*. 2015 Jul 10;290(28):17096–105.

57. Majumdar C, Nuñez NN, Raetz AG, Khuu C, David SS. Chapter Three - Cellular Assays for Studying the Fe–S Cluster Containing Base Excision Repair Glycosylase MUTYH and Homologs. In: David SS, editor. *Methods in Enzymology [Internet]*. Academic Press; 2018 [cited 2021 Jun 22]. p. 69–99. (Fe-S Cluster Enzymes Part B; vol. 599). Available from: <https://www.sciencedirect.com/science/article/pii/S0076687917303816>

58. Garibyan L, Huang T, Kim M, Wolff E, Nguyen A, Nguyen T, et al. Use of the *rpoB* gene to determine the specificity of base substitution mutations on the *Escherichia coli* chromosome. *DNA Repair*. 2003 May 13;2(5):593–608.

59. Wu EY, Hilliker AK. Identification of Rifampicin Resistance Mutations in *Escherichia coli*, Including an Unusual Deletion Mutation. *J Mol Microbiol Biotechnol*. 2017;27(6):356–62.

60. Loeb LA, Preston BD. Mutagenesis by apurinic/apyrimidinic sites. *Annu Rev Genet*. 1986;20:201–30.

61. Michaels ML, Tchou J, Grollman AP, Miller JH. A repair system for 8-oxo-7,8-dihydrodeoxyguanine. *Biochemistry*. 1992 Nov 17;31(45):10964–8.

62. Pope MA, Porello SL, David SS. *Escherichia coli* apurinic-apyrimidinic endonucleases enhance the turnover of the adenine glycosylase MutY with G:A substrates. *J Biol Chem*. 2002 Jun 21;277(25):22605–15.

63. Lu AL, Lee CY, Li L, Li X. Physical and functional interactions between *Escherichia coli* MutY and endonuclease VIII. *Biochem J*. 2006 Jan 1;393(Pt 1):381–7.

64. Komine K, Shimodaira H, Takao M, Soeda H, Zhang X, Takahashi M, et al. Functional Complementation Assay for 47 MUTYH Variants in a MutY-Disrupted *Escherichia coli* Strain. *Hum Mutat*. 2015 Jul;36(7):704–11.

65. Jaenicke R. Protein stability and molecular adaptation to extreme conditions. *Eur J Biochem*. 1991 Dec 18;202(3):715–28.

66. Závodszky P, Kardos J, Svingor Á, Petsko GA. Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proc Natl Acad Sci*. 1998 Jun 23;95(13):7406–11.

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 LACTOBACILLI. *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. Šlesak I, Šlesak 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. Carty A, support) BR (author of parallel. boot: Bootstrap Functions (Originally by Angelo
1110 Carty for S) [Internet]. 2022 [cited 2023 Nov 27]. Available from: <https://cran.r-project.org/web/packages/boot/>

1111 102. Davison AC, Hinkley DV. *Bootstrap Methods and their Application* [Internet]. Cambridge:
1112

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

1122 Supporting Information

1123

1124 Data and code availability

1125 Primary data and code for analysis of data have been deposited with GitHub:

1126 <https://github.com/paytonutzman/Lost-City-MutY-Discovery>. These include FASTA files for the
1127 discovered MutY enzymes, predicted structures, virtual docking outcomes, MD trajectories and
1128 scripts to analyze these trajectories, rifampicin resistance frequency data, and the *R* code to
1129 report statistics, median and 95% confidence intervals.

1130

1131 S1 Dataset. Alignment of Lost City MutY homologs

1132 Chemical motifs are highlighted in columns. Alignment was generated by *Promals3D* guided by
1133 the structure of *Gs* MutY. It was necessary to align sequences in the first block including up to
1134 N146 separately from the second block and third block because otherwise the C-terminal domain
1135 residues were aligned inconsistently. The homologs flagged with dark red highlighting were
1136 eliminated because of missing chemical motifs. The homolog flagged with light pink
1137 highlighting required manual adjustment so as to align the H-x-FSH motif.
1138

Conservation:	9	5	579	67	7999	5	5	65	7	55	6	9	98989	9	5	5	8	65969	67	66	5	859												
sp_P17802 MUTY ECOL	28	KTPYKVWLS	IVMLQ	YQVATV	VIP	-YFERM	MARF	PTT	DLAN	ALP	DEL	VH	LG	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PET	FEV	VAL	PVG	GR	TAGA	ILS	SLG	KH	FP	PI	LG	
sp_P83847 MUTY GEOS	34	RDPYKVWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_000003652391_2	37	RDPYKVWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_00000031207_6	26	EDPYPAI	WLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG
c_00000430975_4	38	PNVYHVLWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_000001007148_4	35	RDPYVWRWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_000001092848_3	48	PDPYRVWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_000001151029_2	1	---	WLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG
c_000001797282_1	58	PDPYRVWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_000001803648_25	40	PDPYRVWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_000002078955_4	40	AEPYRWRWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_00000201616_4	28	KEPYKIVWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_000003607531_2	38	KPVYKIVWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_000003877363_4	38	PDPYHIVWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_000004820107_1	43	PNPYVWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_000005774797_1	29	PNPYVHVLWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_000005867021_3	40	PDPYRVWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_000004546210_2	31	PSKYKTVWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_000004515008_5	31	PSKYKTVWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_000002511148_4	26	KTPYRWRWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_00000577378_2	14	KTPYRWRWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_00000598175_2	37	INPYRWRWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_000007554673_3	34	KTPYKVWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_00000811118_1	25	KDPYKVWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_00000896687_2	29	NDPYKVWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_00001176522_24	31	RTPYRWRWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_00001445221_1	4	ISPYRWRWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_0000151573_6	18	RDPYRWRWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_00001682161_6	25	PTPYRWRWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_00001923643_29	28	RTPYRWRWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_00002030994_5	28	KSPYHIVWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_00002038721_4	28	KSPYHIVWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_000020253257_15	45	RTPYRWRWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_00002068528_2	26	INPYRWRWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_00002961510_2	29	KDPYSIWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_00002971528_2	34	RTPYRWRWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_00003248043_1	46	KNPYHIVWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_00003294679_3	31	KTPYIWIWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_00003333364_1	28	YDPYEWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_00003402697_5	28	REPYKIVWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_00003494815_4	25	DDPYHIVWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_0000376664_5	41	PTPYRWRWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_00003948187_72	28	ITPYRWRWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE	V	DL	K	W	E	A	V	R	N	R	Y	YAR	NRN	HLKAA	Q	VAT	HLG	KPF	PI	LG	
c_0000505120_4	28	ITPYRWRWLS	IVMLQ	YQVATV	VIP	-YFERM	R	FPT	T	LEA	LA	DA	DE</																					

c_000003607531_2 151 VERVITRVGIFKPLP---EARPIIKDAKA---LAP-KTT---RGRPGDYAQAVMDLGATVCRPKPLCD---LCPWEKYCIANQNFQDVPRKAPK--- 236
c_000003872363_4 149 VERVMARLFNVITPLP---AAKPELTAAH---AAA-LTP---VDRAGDYAQAVMDLGATCTPKSPCG---ICPCTEYCAALPGEADANLPLRPPK--- 231
c_000004820107_1 156 IERMARILYATEPLP---DSKANLNLAA---GLS-EER---KDRPGDYAQALMDLGATCTPKSPCG---LCPVNDSTAKRQGIAESLPLIKKKK--- 240
c_000005774797_1 140 INVLVSLVNVITDANL---RPSARLARIA---AET-LAH---PKRSQDVAQALMDLGATCTPKSPCG---TCPICDCKAHSLRGTHELPVKAPQ--- 222
c_000005867023_3 153 VERVMARLYIHSPLP---AVKPEMLTH---ATA-LTP---EFRAGDYAQALMDLGATCTPKSPCG---ICP1C1DCKAHSLRGTHELPVKAPQ--- 222
c_000004546210_2 144 VRKVLSPFVGKGSWGP---ESVKSKELWDL---SAK-SLP---VDFNFTVTCQIMDLGATLCSRNPNCs---ICP1QRCYTAHQHNLVDEIPSKKPK--- 228
c_000004510085_5 144 VRKVLSPFVGKGSWGP---ESVKSKELWDL---SAK-SLP---VDFNFTVTCQIMDLGATLCSRNPNCs---ICP1QRCYTAHQHNLVDEIPSKKPK--- 228
c_000002511148_4 139 VRKVLARHAFIEGKEL---TSSKRKEWLQI---SED-LLP---DNKIEVTTDQIMDLGATLCSRNPNCs---LCPLNQDCLALKEDLTVLPSKTPK--- 223
c_000000577378_2 127 VRKVLARYFIEGKEL---TSSKRKEWLQI---SEG-LLP---DNKIEVTTDQIMDLGATLCSRNPNCs---LCPLNQDCLALKEDLTVLPSKTPK--- 211
c_00000598175_2 150 VRKVLTRWFGWP---KADVLKLNLL---AEQ-TTP---ASRNADYTDQIMDLGATLCSRNPCE---DCLPKLADLAFOOGNCNTVYPTSKPK--- 234
c_000000754627_3 147 VRKVLARFFMVCGWYG---VKVUNQWLH---SEQ-LTP---KNNVTEFNAQALMDLGASLCSRNPCE---ACPLNSRCGAFNAGKVKEFPHSKPK--- 231
c_000000811118_1 138 VRKVLRLGSSKSKP---KYKTEKIS---ANT-LTP---NSKHEISQALMDLGATCRRNPNCs---ECPVDEDCANQKEIODEIPLIKKK--- 222
c_00000899687_2 142 VRKVLRLFCINKNGT---TSAFENCLWGE---AED-LLP---VRKPGDFNQALMDLGATVCPASPCQ---QCPPLRT1CAFLKLNKNEFPSSKN--- 226
c_000001765224 161 VRKVLRLCPFVGEGWP---ERQVEAHWTL---AEL-LPP---QSEVMPDYTDQIMDLGATLCSRNPCE---ACPFQAECAVYRQGRQVAAELPAPRPR--- 247
c_000001345122_1 117 VRKVLRCFADTWSWGP---NKKTEKVLWDR---ADE-FTP---QERFDYTDQIMDLGATLCSRNPCE---TCPVCEADCAKRDMDKV1QFYSNPK--- 201
c_000001515736_6 131 VRKVLTRVLAFLDGLA---QSNRERQLWEY---AQQ-CLPTE---NLHEAMPAQ---CMMGLGASVTSRKPCL---TCPCLSECDRAAARAGNEFNYFVTRR--- 219
c_000001682161_6 138 VRKVLRCFDAINISWGP---NQKTESQWLWQ---AEQ-LTP---SKRVDYTDQIMDLGATVCRSPKSC---ICPVAENQARQLNQVQAKYFPSKPK--- 222
c_000001923643_29 141 VRKVLTRVLAFLDGLA---VARNERLELWL---AQQ-CLPTE---DLQZQMRPTYQALMDLGATCRRNPCE---DCLPKLQDCAARAFGPNENYFVTRR--- 229
c_000002030994_5 139 VRKVLRSKFTWGP---KADVLKLNLL---AEQ-TTP---KTRNADYTDQIMDLGATLCSRNPCE---DCLPLQDCLALQFOGNCNTVYPTSKPK--- 223
c_000002038721_4 141 VRKVLRLVGLGNKSP---HSKTYKELKF---AEE-LTP---KNDFTTYTDQIMDLGATCRRNPCE---NCP1QRCYTSQNLQNLVDEIPSKKPK--- 225
c_000002523527_15 156 VRKVLRFARVGIFHGP---ERAEINRMLW---AEQALPAG---RHQADHMYAVTDQIMDLGATCRRNPCE---DCLPLRACLAFOQNGNTDYPPTKPK--- 250
c_000002608528_2 139 VRKVLTRVLFVGEWP---KSGVLKLLWLL---AEQ-TTP---QTRNADYTDQIMDLGATLCSRNPCE---DCLPLRACLAFOQNGNTDYPPTKPK--- 223
c_000002961510_2 142 VRKVLRLSLFCINENG---TSASENRLWQ---ADE-LLP---VRRPGDFNQALMDLGATVCPASPCQ---QCPPLT1CAFLKLNKNEFPSSKK--- 226
c_000002979152_8 147 VRKVLRLARYHIEGKEL---KSAVLKLWEEK---AEL-LHP---AERNADYTDQIMDLGATLCSRNPCE---TCPVWQCAARKSWSALYFPGKKPK--- 231
c_000003248024_1 159 VRKVLTRVLDKCNVA---STAQKQJTA---ARL-LVES---VCPHSACTNQALMEVGATVCPASPCQ---DCLPWNQCSAKACTQKTRPVKARP--- 246
c_000003294679_3 144 VRVVFTHRAKIGWP---TTMVQNLWQ---AED-NTP---SNRVADETDQIMDLGATCRRNPCE---TCPVNDQCLALHSRGQEAEPFHKNK--- 228
c_00000333364_1 141 VEVRLARLFDVDAPIK---SSQAQSFELWET---TQGQD---FPLTTECZAHYLGIVFHRVPAER--- 224
c_000003402697_5 142 ISKIIWTWLNPG---QDTEV---ANE-LVS---LSQSGYAANNQDQMLLASSRAGNPNVE---GALGETYFSDSEAIICRLKFKPKRKEPKK--- 219
c_000003492925_4 134 VRKILCRLKHLKTP---SDEKLWRI---AYN-LVD---KRNPFYNDQIMDLGATVCPASPCQ---TCPPLSD1CQGONGPDTIYPTKKR--- 211
c_000003766664_5 154 VIRLSSLRLCDLEEEVG---ETAKAKERLWQD---AER-LVD---PQNAFASNQALMELGAMLTPTSPGCD---QCPWREDRCAAAGTVELRPIKKAK--- 238
c_000003948184_72 141 VRKVLTRHVRVAGWP---RASVNRNLWQ---AER-YTP---KRRVERYTDQIMDLGATVCRVGQPRC---TCPVAKHKECALRHDVCQSFQPKAR--- 225
c_000005057124_4 140 VRKVLTRYFAGWITG---NSKISNWLWEI---SSR-YTP---KIRADYTDQIMDLGATLCSRNPCE---ICP1QADCLAKKONKTAELPTPKPA--- 224
c_000005509101_5 140 VRKVLTRVLAFLDGLA---GTTKVNKLWQI---SSD-YTP---LLRILADYTDQIMDLGATCRRNPCE---COPINSACLAKIEKISSELEPTPKPK--- 224
c_000006063368_2 139 VRKVLTRHRIEATDLS---KSSTSLKLWKL---SEL-LLP---DVRDITYTDQIMDLGAIKVCKSPCE---ICPVNNDCLALAKNVLNDCLPVKPK--- 223
c_00000467631_2 119 VRKVLRLARYHISGWTG---SAKISSELWKL---SAR-YTP---AERCADYTDQIMDLGATLCSRNPCE---COPISSDCVAR1GDKVKLFLPTPKPA--- 203
c_00000788235_1 140 VRKVLTRHAYIEGKEL---KGHLKLWLSI---AES-HLP---DTRYNADYTDQIMDLGATVCPASPCQ---GCPH1HADCSAQAGTPTAYPTPKPK--- 224
c_000001187176_11 141 VRKVLTRHAYIEGKEL---EKPVHDLRWA---AER-HTP---HSRVAEYTDQIMDLGATVCRVGQPRC---TCPVAKHKECALRHDVCQSFQPKAR--- 225
c_000001660697_64 141 VRKVFARVFCIGEWP---EADLWLP---ADCIQPYTDQIMDLGATVCPASPCQ---ICP1QACQARMLEGEPTAYQOARPR--- 225
c_000003378864_2 152 VRKVLRLSLYDVSDB---VGTAAATRTWLWAYSTELLN---QADDATHTDQIMDLGATLCSRNPCE---TCPVSVRCDAALARDTVPLRPIKPR--- 237
c_000003910742_3 141 VRKVLTHYHAVEGWP---KSCVLQKJLWQ---AES-HLP---ATRYADYTDQIMDLGATVCPASPCQ---TCPVNSNCALSKGNFTAYPTPKPK--- 225
c_000004116181_2 107 VRKVLRFANQWGP---NAVSNQWLWAI---SAR-YTP---VERVADYTDQIMDLGATVCRTRPQ---ICPVNENCLARLKKS1TEIIFASPKV--- 191
c_000006223903_1 142 VRKVLRFANQWGP---RAEVARLWSL---AER-HTP---TRRVAYAATDQIMDLGATVCPASPCQ---TCPLAGGCKAHQHQNQHEDYFASPKP--- 225
c_000001029068_2 114 VRKVLRFANQWGP---QGKTIQWLWKL---SEK-HTP---NTRNNDYTDQIMDLGATVCRTRPQ---TCPVPLARCLAKKHTHQHLYFASPK--- 198
c_000001128125_11 149 VRKVLRFANQWGP---EKKIEADLWHP---ADE-LTP---PKRDRADYTDQIMDLGATVCRTRSAKCG---ICP1QADCLAKKONKTAELPTPKPA--- 233
c_000001595844_4 140 VRKVLRLTRYFAGWITG---NSKISNWLWEI---SSR-YTP---KIRADYTDQIMDLGATLCSRNPCE---ICP1QADCLAKKONKTAELPTPKPA--- 224
c_000001719155_2 144 VRKVLRLSLFLLKENGP---TRKSENLIWT---MQL-LLP---ETGAGFNQDQIMELGATVCPKPNLCL---TCPPLKRNQAOAKSGKQNLYPPRKR--- 228
c_000002363036_6 149 VRKVLRSKYAIEGWP---EKRVENQMLGNGW---ADN-LTP---ETFDADYTDQIMDLGATLCSRNPCE---TCPVKTDCRALKTQDQASFNSPKPK--- 233
c_0000025614001_1 105 VRKVLTRFQFVGHP---ECKVENRLWHR---ADE-LTP---SFRVADYTDQIMDLGATLCSRNPCE---TCPVHSGQCALKGKEVHLLPSFKPK--- 189
c_000002566035_2 141 VRKVLTRHRYKIKGP---EKKIENKLWLS---AEL-YTP---RKLTRQYTDQIMDLGATVCPASPCQ---ICP1LTLT1CFAFMKGQHDFGPKKPK--- 225
c_000002598045_3 139 VRKVLRLARFLKIEGNLQ---ASSTTKRLWQI---SES-LLP---GDRDITYTDQIMDLGATVCPASPCQ---NCPVQNDQCSAFKEGLV1DPLPKVPK--- 223
c_000004188575_41 141 VRKVLRFARLYAIEASWP---NKAQEWMQ---ADN-LLP---DQRIAYAIQDQIMDLGATLCSRNPCE---COPLQSNQCAAYRGTETDPFLAKPK--- 224
c_000002400409_4 139 VRKVLRFARLYAIEGWP---QSPVSKLWTI---AEQ-FTP---ENQLADYTDQIMDLGATVCRTRSAKCG---ICP1QACQARMLEGEPTAYQOARPR--- 225
c_000002139456_2 139 VRKVLRFARLYAIEGWP---QSOVSKLWTI---AEQ-LTP---KQQLADYTDQIMDLGATVCRTRSAKCG---ICP1QACQARMLEGEPTAYQOARPR--- 225
c_000003535614_2 141 VRKVLRFARLYAIEGWP---KPALEKQWLW---AER-YTP---TEELADYTDQIMDLGATLCSRNPCE---TCPPLNGKCALANNCVAAFLPTPKPK--- 225
c_000002657784_4 138 VRKVLRLARYKIEKGP---KALNIRLWRI---SEL-LTP---EERKVDLQDQIMDLGATVCRSSPKC---TCPVPSRCUCLANRKKL1QVPLRKL--- 222
c_000003800129_15 143 VRKVLVGRYLNQYSE---QEKWFWL---SKK-LLD---KSNPFLFOCQIMDLGATCRRNPCE---TCPPLGRCWLSKRNQSYTYYKKKK--- 220
c_00000371671_1 130 VRVRLARLWQ---TKAPEERIEG---AEE-LDV---GVNDPSGSNQIMELGATVCPKPNLCL---TCPVMTECRARARGLERAVPFLPV--- 215
c_000001556689_3 131 VARIMCRLAGIADTR---RSKVRQRLAE---AAD-AIB---CHPGTGTNQALMELGARVTCRTRPSPRC---QCPCCSYECAEHLGIEEIPPRRSK--- 215
c_000001834452_1 108 VRVCAARFLRQLKSPGL---SALSARLWQ---GET-LHG---VSAGLACLSNQALMELGATVCPKPNLCL---TCPVPLC1QAYQQRVADFFGPKKPK--- 195
c_000004255002_2 148 ARVYVARYLSSLTN---LRAINDET---AEQ-MVS---HSPRDGFNQDVMELGATVCPKPNLCL---TCPPLAYCQARSGVAFQLQLRPTR--- 225
c_000006005438_3 141 VRVILYRFFVAFKEA---NDKLWEM---AYD-LYD---KONAYIYNQDQIMMDIGAISCTHNPCL---TCPVPLC1QAYQQRVADFFGPKKPK--- 218
c_00000581237_15 141 VRVILYRFFFAMTSC---NDKLWEM---SYE-LYD---RENSYIYQDQIMMDIGAISCTHNPAD---TCPVPEPLCQGKESPL---LYPEKKKK--- 218
c_000002529579_2 141 VRVILYRFFVAFKEA---NEKLWEM---AYD-LYD---KONAYIYNQDQIMMDIGAISCTHNPAD---TCPVPEPLCQGKESPL---LYPEKKKK--- 218
c_000006079838_1 114 VARVLTRLLGIERPSK---QKQVEQLW---TAD-LVXKHAVIDLQTSRNRNATPAFNQAMMELGALVCPTRPNLCH---DCLPLKGRCTARKQNKVETIPALAKR--- 207
c_00000583727_6 140 VARVLTRLLGIEQSPK---DKPVTRLRLWV---ADE-LVXKHTVTLQKQRNQASAFNQAMMELGALVCPTRPNLCH---DCLPLKGRCTARKQNKVETIPALAKR--- 233
c_000005807640_2 142 VARVLTRLLGIEESPK---ARPVADRLWV---ADE-LVXNHAVTLLRKPQPSQNSQASAFNQAMMELGALVCPTRPNLCH---DCLPLKGRCTARKQNKVETIPALAKR--- 235
c_000001348781_2 124 VRVRSVSLRQYTAQDPD---KNKKKMEKF---MSL-IIN---NRPQDGINQDVMGDRYQKCPFLP---TCPVHNGKCALKLGQISDLP1KIKK--- 206
c_000005136725_3 139 VRVRSVSLRQYTAQDPD---KNKKKMEKF---MNA-IIN---DKPGRDINGQDVMGDRYQKCPFLP---TCPV1QYCKALKLGQIAQDLPV1KIKK--- 221
c_000001684786_4 137 VERVMTLRLAENP---QVSTKRLKEI---AGG-WMP---ADKASSDNQAMMELGALVCPSPDPC---TCPVPEREVCAAAEGDPEFPLRPLPK--- 221
c_000002971823 147 VERVILTRLRLADNPR---TAAKYLK1KEI---AEG-WMP---RGRKAGYTDQIMMDIGAISCTHNPCL---TCPVHNGKCALKLGQISDLP1KIKK--- 231
c_000002040694_2 150 IARVFSRSLQYQGKSE---ALTTSLRQ---ISL-FLP---FKRIGDYTDQIMDLGALICRPNPCE---TCPVHNGKCALKLGQISDLP1KIKK--- 232
c_000003263657_32 154 IARVFSRSLQYQGKSE---KQKWEI---LTTSLRQ---KQKWEI---TCPVHNGKCALKLGQISDLP1KIKK--- 232
c_000005989041_2 151 IERVTRIFGICFTFP---EAKTIVKDCAAK---LAP-KTT---RGRPGDYAQIMDLGAIKVCPKPNLCH---TCPVHNGKCALKLGQISDLP1KIKK--- 236
c_000002391082_2 149 IERVTRIFGICFTFP---RTIAEELIARIA---GEE-LLA---RGEADYTDQIMMDIGLARVTCRTRPILP---TCPVPGVCRQAELDDPSALPYKVP--- 233
c_000003839553_2 156 IERVTRILRLIEGDP---RAQIKAELIARIA---VEA-LLA---RQGADYTDQIMMDIGLARVTCRTRPILP---TCPVPGVCRQAELDDPSALPYKVP--- 240
c_000003996707_2 146 IERVTRILRLIEGDP---RAKAVITLRL---GER-LLA---RGRGADYTDQIMMDIGLARVTCRTRPILP---TCPVPGVCRQAELDDPSALPYKVP--- 230
c_0000020440706_1 129 IERVTRILRLIEGDP---TAAVSRLWQ---ATD-VLP---ONRPQDGINQDVMMDIGLATVCPKPNLCH---TCPVPLC1QAYQQRVADFFGPKKPK--- 213
c_0000060007057_1 144 IERVIIKRLNLTKEKE---ISKENI1KK---KKI-LGM---SDRSRQDIAQIMELGALVCPKPNPCK---TCP1PTKNC1SKKKDQEFL---ISKN--- 223
c_000002094036_4 120 IERVIIKRLNLTKEKE---TKENKLHQQ---KFG---TNRNDYVTDQIMDLGAIKVCPKPNLCH---TCP1PTKNC1SKKKDQEFL---KPKN--- 199
c_000002762689_1 144 IERVIIKRLNLTKEKE---INKENL1KK---KKI-FGY---SNRASDQDIAQIMELGALICRPNPCE---TCP1PTKNC1SKKKDQEFL---TKIK--- 223
c_000005516980_2 144 IERVIIKRLNLTKEKE---IQKENL1KK---KSV-FGL---SQRSSDQDIAQIMELGALICRPNPCE---TCP1SKKKCSFRKKDQFLN1---IKNT--- 223
c_00000392004_1 150 IERKMLARLYQDQSIN---LINKK1TSL---SKF-YES---KQKSSNLLDQIMDGY1SIC1QPNRKNQ---IC113RECI1ANQRK1S1N1P1PKK1S--- 232
c_00000400480511_2 144 IERKMLARLYQDQSIN---LINKK1TSL---SDN-FIS---KNSSD1QDQIMDGY1SIC1QPNRKNQ---IC113QRNQNAFYNQNLQD1FVFLKLS--- 226
c_000001286181_5 144 IERKMLARLYQDQSIN---LINKK1TSL---TDAISLRL---SDN-FIS---TCPVQDFNQDQIMDGY1SIC1QPNRKNQ---TCPV1QYCKALKLGQIAQDLPV1KIKK--- 221
c_000001293628_3 137 VRVLSVSLRQYTAQDPD---TTEGKVVYKEL---LXQ-NSP---SONIRDYVQIMEMGVSQVCPNPKCD---TCP1QK1LCSFLKKQDF1IPTKKN--- 217
c_000001535696_2 143 VRVLSVSLRQYTAQDPD---TTEGKVVYKEL---AFK-VMD---TCPVPLC1QAYQQRVADFFGPKKPK---TCPVPLC1QAYQQRVADFFGPKKPK--- 217
c_000001614067_2 150 VRVLSVSLRQYTAQDPD---ESK1F1KNI---AEE-HLP---DNRHGDYTDQIMDLGSLIC1QPKSPRK---TCP1PL1A1CDVCGTTSQAKQY1PKLPK--- 232
c_000001765289_1 116 ACRVMSRILG1KLN1---TSWNLNRLKNT---LSN-1IP---EHTPGNQDIAQIMDLGSLIC1QPKSPRK---TCP1PL1A1CDVCGTTSQAKQY1PKLPK--- 197
c_000001961666_1 131 VERVTRILRLIDKNNK---KVNKKK1K---TQS-LIP---QNKPGDFNQDVMMDLGSLIC1QPKSPRK---TCP1PL1A1CDVCGTTSQAKQY1PKLPK--- 212
c_000002018097_4 138 LKRVISYRFLKQ1KLN1---RNTLTHRNQ---LNM-LLP---NGRPGDQDIAQIMDLGSLIC1QPKSPRK---TCP1PL1A1CDVCGTTSQAKQY1PKLPK--- 220
c_000002843512_36 135 VRKRLCRLQQLKLT---TDEKLWKL---AYD-LVN---KRNPDYTDQIMMDIGLATVCPKPNFCE---ACPLSD1CQGGSOPALYPTPKKPK--- 212
c_000002930199_1 135 VRVLSRSLRQYTAQDPD---TNKGREEFQSI---ANN-LLP---NKNTGLYNAQIMDGFQSIQCKY1NPKCN---TCP1PL1A1CDVCGTTSQAKQY1PKLPK--- 219
c_000003033795_2 131 LERIGYRILGLKTTK---RNKKRVRV---LEK-NQQ---ANNPQDGVNAQIMDGFQSIQCKY1NPKCN---TCP1PL1A1CDVCGTTSQAKQY1PKLPK--- 213
c_000003888107_2 132 LERIGYRILGLKLSK---FQNQKVKF---LEK-IQD---KNNPGDQDIAQIMDLGSLIC1QPKSPRK---TCP1PL1A1CDVCGTTSQAKQY1PKLPK--- 214
c_000001765289_1 157 SRVLSLRSFAIKGDL---RQMPKHLWDL---AEF-VLP---EKNVGFNQDQIMELGQSCQCRQYQPRC---TCP1PL1A1CDVCGTTSQAKQY1PKLPK--- 241
c_000004852258_1 119 HSR1IARVLGKVNQTS---RNLNRINNY---LKK-LVR---EGNPGEINQDQIMDGMNS1QKSNK1V---TCP1PL1A1CDVCGTTSQAKQY1PKLPK--- 201
c_0000050630677_1 145 VRVVLRSR1LTDFAEDV---TTPAKKQWLQW---AAK-LVP---ADRPQDYNQDQIMELGQTLQDLP1KTDQ---TCP1PL1A1CDVCGTTSQAKQY1PKLPK--- 229
c_000005254087_1 142 VRVVLRSR1LTDFAEDV---SSAGKKWFK---AQA-LLP---EDDPANONYDQIMMDGFQ1HCRPKPNSCG---TCP1PL1A1CDVCGTTSQAKQY1PKLPK--- 226
c_000003283462_3 130 IERKIMARMLGK1KTTN---FNYKR1KKT---LLN-LIS---SSRPQDGNQDVMMDLGQSVCPNPKNCE---ACPLSD1CQGGSOPALYPTPKKPK--- 212
c_000002687221_1 144 AIRVSRMMSIDSYP---KSRK1KKT---LSE-HID---TIRPGYV1CQIMDGFQ1VCKYSPSCY---TCP1PL1A1CDVCGTTSQAKQY1PKLPK--- 226
c_000003347358_19 166 VRKILCRLKELRP---SDTELWDL---AYT-LVD---KINPDYTDQIMMDIGLATVCPKPNFCE---TCP1PL1A1CDVCGTTSQAKQY1PKLPK--- 243
c_000002701031_2 137 VRVLSVSLRQYTAQDPD---SSDGKKEFQOLL---AEE-LLI---KEKAGENNAQIMELGALQCTPKSPNCN---TCP1PL1A1CDVCGTTSQAKQY1PKLPK--- 221
c_00000582753_3 145 VRF1F1SLRQYGVSTP1N---SGRANKERFV1---LNA-1ID---EKNPATD1QDMEIGFSQLOQ1PKSPNCN---TCP1PL1A1CDVCGTTSQAKQY1PKLPK--- 229
c_000001713769_5 137 VRVLSVSLRQYTAQDPD---NLHGLKKFYM---AQK-LAP---KRSKCGDYNQDVMMDLGSLIC1QPKSPRK---TCP1PL1A1CDVCGTTSQAKQY1PKLPK--- 221
c_000004369364_1 140 VRVLSVSLRQYTAQDPD---TFIGVKK1---ANQ-FLP---KQKPGFNRQDVMMDLGQ1HCRPKPNSCG---TCP1PL1A1CDVCGTTSQAKQY1PKLPK--- 224
c_000006057486_30 141 VRVLSVSLRQYTAQDPD---SARG1KEFQKL---AQI-LIN---PKNPGDHNQDVMMDLGQ1HCRPKPNSCG---TCP1PL1A1CDVCGTTSQAKQY1PKLPK--- 225
c_000005494072_10 137 VRVLSVSLRQYTAQDPD---SSKG1KEFQKL---AQS-LLI---NENFGLNQDQIMDGFATC1QPKSPRK---TCP1PL1A1CDVCGTTSQAKQY1PKLPK--- 221
c_000001742634_3 140 VRVLSVSLRQYTAQDPD---STNG1KEFQKL---AQQ-LLP---NNK1GHNQDQIMDGFATC1QPKSPRK---TCP1PL1A1CDVCGTTSQAKQY1PKLPK--- 224

c_000002826998_2 140 FFRVRLSRYFGIETPMHD---TSEGGKKQFTHL---ANE-LLL---KKPAINYNAIMDFGAIQCTPKSPDCK---HFLKKQCALQENKISKLKVSKK--- 224
 c_000003159439_6 137 VYRLARYFGIRTSTN---STKGKIKEFKQL---AQE-LID---TKDPATFNQAIMEGFAIQCKPKPNPCN---NCPNLTSCIALQKKLTILPIKDKK--- 221
 c_000004887214_1 142 VYRLARYFGIERT---TSKAKKEKEFKEV---AFL-LID---SKNPKLNQAIMEGFSLQCTPKPNPCN---KCPNLTDCVALQTNQEVLPKQPK--- 226
 c_000005037037_2 140 VYRLARYFGIERT---TVEKKEKEFQE---AFL-LID---KKSADADYNAIMMEFGLSLOCCTPKPNPCN---KCPNLTSCAFYKNNQCNLPLKKKK--- 224
 c_000001059961_1 132 MTRVISRMNNSKIF---LAKTRKRWLQ---AWE-WIP---EGEARMEFQAIMMEGALICLPSNPKL---LCPNNTSCALQKQGTVSLIPLPKPK--- 216
 c_000002498472_1 141 VERVFAFARINFLDILPG---SPEARLWNEK---AEE-LLL---KGHANRNFQAIMELGALICLPSNPKL---SCPPLVTHCALKDYLVPERPVPKS--- 225
 c_000004619512_3 130 VYRLSRLSKFQKIAID---ISNKKKIIEFL---AQE-VLD---ASQAKYNAIMDFGATMKPKSPDCK---LCPNNTSCALQKQGTVSLIPLPKPK--- 214
 c_000004766858_2 143 IERVLCRIEDIGEPEK---STPVLQQLWLS---AWE-WIP---HGHAREFNAIMELGALICLPSNPKL---LCPANACWAFHTRGRDIDIPAPVR--- 227
 c_000002747260_18 149 VERVMARFLQIETPLP---GAKPELTD---AGR-LTP---DIRPGDHAQAIMDLGATCTPKSPAG---ICPVMTCATRAAGAQADLPRKTPK--- 231
 c_000005371561_38 149 VERVMARLFDIHEPLP---AAKPVLKQ---AAA-LTP---ATPGDHAQAIMDLGATCTPKSPAG---ICPRLPACAAARAACTQADLPRKTPK--- 231
 c_00000169465_2 145 VYRLARYSISGGWP---QKVKENQLEWV---AEK-NTPTN---SEGGRCANYTQAIMDLGAMICLTSKPKCD---ECPLQADQIAAYQGQADYGPKKPK--- 233
 c_000003254110_11 145 VYRLARYSISGGWP---QKVKENQLEWV---AEK-NTPTN---PEGGRCANITQAIMDLGAMICLTSKPKCD---ECPLQDQDIAAYQGQADYGPKKPK--- 233
 c_000004750284_20 145 VYRLARYSIAIEGWP---QKVENALWDV---AEK-NTPTN---QNRCANITQAIMDLGAMICLTSKPKCD---ECPLQDCVAFQGQKTDYGPKKPK--- 229
 c_000001417892_15 150 VYRLARYHAVAGWP---ETAVSRSLWL---AER-YTP---DNRKTDYNTQAIMDLGATVQVRRRRCG---VCPADGCRARREGNEFAYPGSRRP--- 234
 c_000002717847_8 144 VYRLARLFLQEEF---KEGLFWLN---SAQ-CLD---EDDPYSSQGIMDQGATLCPKAPLC---ECPLNKNCLAFKENDFQLPSSKKP--- 223
 c_000004057999_1 131 VYRLVGRYKQKISFKT---ENEKQVQLWL---SEE-LTP---NKESEPYTQAIMDLGATHQSKSPRCD---LCPVSMCDSAFKEVSNNKVSLLRKA--- 214
 c_000006126673_1 156 IERVLARYHANTPLP---KAKKEKEKERT---IYT-DNL---PERPGDYAQAIMDLGATCQPKSPLY---SCPPLKEGTAYSLQAEYYPKQKK--- 240
 c_000003393183_3 156 VYRLVGRYKQKISFKT---ENEKQVQLWL---SEE-LTP---NKESEPYTQAIMDLGATHQSKSPRCD---LCPVSMCDSAFKEVSNNKVSLLRKA--- 214
 c_000003029168_2 156 VYRLVGRYKQKISFKT---ENEKQVQLWL---SEE-LTP---NKESEPYTQAIMDLGATHQSKSPRCD---LCPVSMCDSAFKEVSNNKVSLLRKA--- 214
 c_000001863436_1 145 VARVSSRLFALEGDLK---DTALQKWLWEI---TLD-LPV---NKRPGDFNQAIMELGATVQAIMDLGAMICLTSKPKCD---LCPPLAKPCHARVYGEFWFSESVT--- 229
 c_000002830137_4 161 VARVSLARLFAVGDPR---SGQNDRFLRGLA---AGA-LID---RGRGPDFNQAIMELGATVQAIMDLGAMICLTSKPKCD---LCPVREYQOAFATGDTYRPAVAKR--- 245
 c_000004612302_3 154 TTRVLSRLAFQGDPS---ORVGQOLLWAF---AEE-LLL---RKGVEFNLQAIMELGSEICLTSKPKCD---RCPVRLMCAAHROQEGLSIPIPSRR--- 238
 c_000001933926_1 148 TARLSRLAFQGDPS---SSGEGRGLWL---AER-VVP---RTRAGEFNLQAIMELGATVQVGP-EPRCQ---CPRVRSRCSLESQGVWDRVPVRSRPR--- 231
 c_000003136334_1 134 IERVMSRLHRLLETPLP---AAKPEAEV---AQR-LTP---RQAGDFQALDILGATVQVGP-EPRCQ---CPRVRSRCSLESQGVWDRVPVRSRPR--- 216
 c_000002717847_8 146 IERVTLARLHRLIKEK---GAKKYLKEK---AEE-LTP---SERSDGQAQAIMDLGATVQVGP-EPRCQ---LCPWPSYQKAYLAGDAEKFPKRVRK--- 228
 c_000005849454_9 142 IERVTLARLHRLIKEK---GAKKYLKEK---AEE-LTP---LAPGECNLQAIMELGATVQVGP-EPRCQ---LCPWPSYQKAYLAGDAEKFPKRVRK--- 228
 c_000001169194_1 154 IERVTLARLHRLIKEK---GAKKYLKEK---AEE-LTP---LAPGECNLQAIMELGATVQVGP-EPRCQ---LCPWPSYQKAYLAGDAEKFPKRVRK--- 228
 c_000002786947_2 154 IERVTLARLHRLIKEK---GAKKYLKEK---AEE-LTP---LAPGECNLQAIMELGATVQVGP-EPRCQ---LCPWPSYQKAYLAGDAEKFPKRVRK--- 228
 c_000001279803_52 142 IERVTLARLHRLIKEK---GAKKYLKEK---AEE-LTP---LAPGECNLQAIMELGATVQVGP-EPRCQ---LCPWPSYQKAYLAGDAEKFPKRVRK--- 228
 c_000003745941_4 144 IERVTLARLHRLIKEK---GAKKYLKEK---AEE-LTP---LAPGECNLQAIMELGATVQVGP-EPRCQ---LCPWPSYQKAYLAGDAEKFPKRVRK--- 228
 c_000004187032_4 133 IERVTLARLHRLIKEK---GAKKYLKEK---AEE-LTP---LAPGECNLQAIMELGATVQVGP-EPRCQ---LCPWPSYQKAYLAGDAEKFPKRVRK--- 228
 c_000006067315_3 144 IERVTLARLHRLIKEK---GAKKYLKEK---AEE-LTP---LAPGECNLQAIMELGATVQVGP-EPRCQ---LCPWPSYQKAYLAGDAEKFPKRVRK--- 228
 c_000005543774_1 144 IERVTLARLHRLIKEK---GAKKYLKEK---AEE-LTP---LAPGECNLQAIMELGATVQVGP-EPRCQ---LCPWPSYQKAYLAGDAEKFPKRVRK--- 228
 c_000006211484_1 144 IERVTLARLHRLIKEK---GAKKYLKEK---AEE-LTP---LAPGECNLQAIMELGATVQVGP-EPRCQ---LCPWPSYQKAYLAGDAEKFPKRVRK--- 228
 c_000002718976_3 145 IERVTLARLHRLIKEK---GAKKYLKEK---AEE-LTP---LAPGECNLQAIMELGATVQVGP-EPRCQ---LCPWPSYQKAYLAGDAEKFPKRVRK--- 228
 c_000002992548_1 146 IERVTLARLHRLIKEK---GAKKYLKEK---AEE-LTP---LAPGECNLQAIMELGATVQVGP-EPRCQ---LCPWPSYQKAYLAGDAEKFPKRVRK--- 228
 c_00000456751_1 110 IERVTLARLHRLIKEK---GAKKYLKEK---AEE-LTP---LAPGECNLQAIMELGATVQVGP-EPRCQ---LCPWPSYQKAYLAGDAEKFPKRVRK--- 191
 c_000002655634_1 144 IERVTLARLHRLIKEK---GAKKYLKEK---AEE-LTP---LAPGECNLQAIMELGATVQVGP-EPRCQ---LCPWPSYQKAYLAGDAEKFPKRVRK--- 222
 c_000004232403_2 140 IERVTLARLHRLIKEK---GAKKYLKEK---AEE-LTP---LAPGECNLQAIMELGATVQVGP-EPRCQ---LCPWPSYQKAYLAGDAEKFPKRVRK--- 222
 64t.after.N146.pdb 141 VYRLSRLFVTDIA---KPTSTRKRFQI---VRE-IMA---YENPQAFNQYLIELEGALVTPRPSCL---LCPVQAIQAFQEVAAELPVKMKR--- 225
 Consensus_aa: 1. RVsRhh.l.pph.h.s.hhs...ppspisQAhm-lGth1c..ppf.C...CPI...C.t...hP....

Conservation:
 sp_P17802 MUTY ECOL 226 ----QTLPE---RTGYF---LLLQ---HEDEVLLAQRQPS-GLWGGI^{TC}QF---QFADEEISLRQ---WLAQRQIA---A 281
 sp_P83847 MUTY GEOS 232 ----TAVKQ---VPLAV---AVLAD---DEGRVLLRKRQST-GLLANN^{WEP}---SCETDQGADG---KEKLEQ---MVGEQYQIQL---V 294
 c_000003652391_2 233 ----QPKPT---RYGVY---YLARD---AHGAWLLEPRPDK-GLLGGMLWGP---TSDWDNADPG---DQPPFA---A 286
 c_0000031207_6 224 ----RTTEQ---QWQAA---AVVV---CQDVLVWVWIDGGD-ELLAGHRGPT---LARLSSGGL---DADKMIG---KEFQRGLGD---GA 286
 c_00000430975_4 234 ----RKKQR---RYGVY---YMIER---RDGAISLQRQPN-GLFFGQTEP---GSLWKSAPL---SRH-SIHAQAPIP---A 293
 c_000001007148_4 232 ----RDVTD---VFEEL---VVMR---RGKRLLIIQHDPD-ALSRDRLARALS---RDFERLIE---SEVADNGCD---V 297
 c_000001092848_3 244 ----AAKPS---RRGAA---FVAVR---ADDAVLLRERPES-GLLGGM^{WEP}---TSGWSAGSD---GEDGP---SAASFP---A 302
 c_000001151029_1 191 ----TKKAK---RGGAA---FVAVR---RDGAIVLQRQPN-GLFFGQTEP---GSLWQAAPL---SRH-SIHAQAPIP---A 250
 c_000001797282_1 253 ----AKGAL---RGGAA---FVAVR---SGDEANVLLRTRPPE-GLLGGM^{WEP}---GSAWEFHD---VAAAL---LDAFLD---A 312
 c_000001803646_25 236 ----PVKPT---RYGIA---YIAR---GGDGAWLLERPRDK-GLLGGMLWGP---GAQWGDQAV---DQPPVS---A 289
 c_000002078955_4 236 ----PVKPT---RYGIA---YIAR---TDGAWLLERPRDK-GLLGGMLWGP---GGAEGAKAV---ENPPVS---A 289
 c_000002016160_4 224 ----QSLPQ---QNVLT---GIVMT---KRDVLLFIRKRAQ-^YHLNNL^{WEP}---GGRITLWQD---LCPQLEEE---LKKNPKFKE---V 286
 c_000003607531_2 237 ----KVKAV---RKGVY---FLAIM---SDGSLFLRERPES-GLLGGM^{WEP}---STWLRSL---SPK-ECEAAAPFK---T 296
 c_000003872363_4 232 ----PAKPI---RYGTA---YIAR---HDGAWLLERPRPD-GLLGGM^{WEP}---GSEWGMV---ENPPFS---A 285
 c_000004820107_1 241 ----KPKPQ---RKGHV---FWISD---DSGNHLLIRPEN-EMLGGM^{WEP}---TTEWIRP---IKLP---Q 291
 c_000005774797_1 223 ----RSKKK---WNAFA---YVYIN---ADGNIGFVVRDKD-ALLLGWMLW---TSDWSTLSPV---FQP---PI---L 275
 c_000005867021_3 236 ----PVKPT---RYGIA---YIAR---PDGAWLLERPRDK-GLLGGMLWGP---GAEGWQAQAV---EKPPIV---A 289
 c_000004546210_2 229 ----KKKRT---ESMFY---LMIVF---GGKLLFLEKKRKEK-GLWGGI^{WEP}---QIOLADDPKT---WCEQQFTNN---I 286
 c_000004515008_5 229 ----KKKRT---ESTYF---LMIF---DGKGLLFLERKKEK-GLWGGI^{WEP}---QIOLADDPKT---WCEQQFTNN---I 286
 c_000002511148_4 224 ----KSKPT---KKVWV---LLPGQ---PSGEVLLERKKEK-GLWGGI^{WEP}---EAEKKNELEI---ELSRRFKTD---C 281
 c_00000577378_2 212 ----KSKPT---KKVWV---LLPGQ---PSGEVLLERKKEK-GLWGGI^{WEP}---EAEKKNELEI---ELSRRFKTD---C 269
 c_00000598175_2 235 ----KVMPE---QKAIM---VILRN---SQQEVMQFMRPPV-^YWIGSNC^{WEP}---QFEKRAFAEE---WIKSNYQKGL---I 292
 c_00000754627_3 232 ----KAVPK---RKSCHQ---LILQ---HNDKVLIMERPRNS-^YWIGGQIFGP---EFNREXESELET---FLAQQGLK---286
 c_000008111118_1 223 ----KEKIE---KKIDW---ILIK---SKDVKVLLRNVRN-^YWIGWQNLWLP---EDKFFSSSKL---SKI---L 273
 c_00000899687_2 227 ----ISSKK---IEVSA---GIIH---KNNKVVYIQRVRK-^YQWVKN-GLMGGM^{WEP}---PEECLR---EIKEELRVN---V 289
 c_000001176522_24 248 ----KVMFV---RKVCW---QVSLIS---QGCVWLWQNPAR-^YWIGWQNLWLP---EGNR---FLDGNEWVAT---L 288
 c_00000145122_1 202 ----KLIPT---KFK---LILMF---EAFQVQLVRKPEK-^YWIGWQNLWLP---EDFVSQEIDL---WCEQQFTNN---I 286
 c_00000151573_6 220 ----KTRSS---ESWIL---LLAVD---ARRWLWQLRPOA-^YWIGWQNLWLP---FVDEYALQOS---FIDTNNVNCCL---L 282
 c_000001682161_6 223 ----KTIPT---KERWF---LVLIT---KEGNILKIRRAK-^YWIGWQNLWLP---EFDEVKLLLEN---AAHAWPAA---RS 288
 c_000001923643_29 230 ----LKRSS---EAWIL---LLAVD---QGGRVLWLERPQO-^YWIGWQNLWLP---VFNEREALEQ---WLDSSFQGMS---L 281
 c_000002030994_5 224 ----KEMPE---KHTVM---LILQN---DQEDEVLMQRPV-^YWIGGQIFGP---QDFDITALAEK---EANSLNYPE---L 275
 c_000002038721_4 226 ----SOKKE---KINW---LLIA---SKDVKVLLRNVRN-^YWIGWQNLWLP---EFETLSEILOS---ENL---L 275
 c_000002523527_15 251 ----AAIPE---RSTVM---LILVRN---RDVLVQLRQPPS-^YWIGGQIFGP---SSEAAESA---LELARAYGS---P 315
 c_000002608528_2 224 ----KVMPE---KOTVM---LLRN---GQKEVFLMQRPPA-^YWIGGQIFGP---FDDVTDVTFD---WLDNNYAVS---F 281
 c_000002961510_2 227 ----VPSKK---IEVSA---GIIH---KNKKVYIQRQSKR-^YWIGGQIFGP---PEECLKR---EIKEELRVN---V 289
 c_000002979152_8 232 ----DDKRR---KQSVN---LYVSN---ENADVLERLRSRST-^YWIGGQIFGP---WLVNTFALE---S 289
 c_000003248034_1 247 ----RKKR---VMTSV---IACE---QGGQIMKMRQSVNARLMPGFWLP---EVEGFPLNQ---ADL---ARWS---L 301
 c_000003294679_3 229 ----KTNPL---LFLRV---EDRIVLERRPFS-^YWIGGQIFGP---EISKDVNPSLSD---WYQSRFGAY---P 285
 c_000003333364_1 225 ----KPAVP---DVVAS---GVLV---HDGYQIQRKPED-^YWIGGQIFGP---GTTVEKGET---PAQCVVR---EFQEELWED---V 287
 c_000003402697_5 220 ----TEPKPKKHKR---IEVGI---ACIV---REGKYLIVQARPKG-^YWIGGQIFGP---GKGRKEGEN---FRECVKR---EIQEEVGIID---V 287
 c_000003492925_4 212 ----VVP---REENI---MLYV---YNDLKLSTQTRQK-^YWIGGQIFGP---KAGVRL---LVKETVGLT---A 252
 c_000003766668_5 239 ----RPKLM---ALVAM---VIRD---EGQAVLITRQPEK-^YWIGGQIFGP---LCPVQAI---FIDTNNVNCCL---L 251
 c_000003948186_72 226 ----KTRSS---ESWIL---LLAVD---ARRWLWQLRPOA-^YWIGGQIFGP---KAGVRL---LVKETVGLT---A 251
 c_0000050505712_4 225 ----NNIPT---RHTOM---LLIEN---TAGEWLERKRPQ-^YWIGGQIFGP---YCECEQELQIA---V 283
 c_000005590109_5 220 ----KVLAV---KQLF---IMLDQ---QTOFLERKRPQ-^YWIGGQIFGP---WCLEN-DIA---M 281
 c_000006063368_2 225 ----KSLPV---KKITF---IMCD---QGRILRERKRPQ-^YWIGGQIFGP---CFLK-GIP---I 281
 c_00000788235_1 224 ----KTKT---KHWB---LILPVS---HKGRLFLERKRNLT-^YWIGGQIFGP---ECKSIFQIJK---D 281
 c_000001187176_11 226 ----KTKLP---KHWB---LILPVS---HKGRLFLERKRNLT-^YWIGGQIFGP---FQONQ-NFA---I 260
 c_000001660697_64 226 ----KTRSS---ESWIL---LLAVD---ARRWLWQLRPOA-^YWIGGQIFGP---WYEOYFSTH---L 282
 c_000003378864_2 231 ----KAIPE---RAATM---VIAH---GETVLLQRLRQPR-^YWIGGQIFGP---LIAQ-GWQ---A 282
 c_000003910742_3 238 ----KTPKK---HIIHG---GVVIR---DDQVILQRLRPE-^YWIGGQIFGP---LDAADTVRQALAYGT---V 296
 c_00000416181_2 226 ----KVIPE---KNAMV---LMLN---NKEDEVLMRERPSP-^YWIGGQIFGP---IEETVRREIREELGV---V 300
 c_00000441181_2 192 ----KQQT---RITCFV---LLIVN---AKQEIIQLQRSPV-^YWIGGQIFGP---QFDDYDQQA---WHENYQFF---C 283
 c_000006223903_1 226 ----RPRPR---REVVM---IMIKD---SRDRVLLERRASS-^YWIGGQIFGP---WCSSR-KIT---V 248
 c_000001029068_2 199 ----KPKPV---KTTAM---LIFDN---GKGGVYLQQRPAK-^YWIGGQIFGP---WCRMRFGAG---I 283
 c_000001128125_11 234 ----KPKPV---KTTAM---LIFDN---GKGGVYLQQRPAK-^YWIGGQIFGP---ECSATDAK1T---KTAHNHFPQ---A 256
 c_000001595844_4 234 ----KPKPV---KTTAM---LIFDN---GKGGVYLQQRPAK-^YWIGGQIFGP---SIESQTTV---I 290
 c_000002363038_6 225 ----IKRKT---RYFY---FVFR---KGSDDYIYQKRTKE-^YWIGGQIFGP---PEOQYPRIK---F 287
 c_000002566035_2 229 ----APSTK---IAVSA---GVLF---RRNRIYIQRKAE-^YWIGGQIFGP---PEOQYPRIK---EIKEELGV---I 291
 c_000002561400_1 234 ----KPKPV---KOKWL---LLQKN---SAEQVFLYIQRPK-^YWIGGQIFGP---LTDLDESIAE---FLNGEIHRR---D 292
 c_000002598045_3 204 ----KPKPV---KOKWL---LLQKN---SAEQVFLYIQRPK-^YWIGGQIFGP---EFCVSDTDAK---YLTQTPQG---L 246
 c_00000467631_2 225 ----KPKPV---KOKWL---LLQKN---SAEQVFLYIQRPK-^YWIGGQIFGP---EFCVSDTDAK---MIKKRYGFN---I 283
 c_00000788235_1 226 ----KPKPV---KOKWL---LLQKN---SAEQVFLYIQRPK-^YWIGGQIFGP---EFCVSDTDAK---AISRNFDSDN---I 281
 c_000001187176_11 234 ----KPKPV---KOKWL---LLQKN---SAEQVFLYIQRPK-^YWIGGQIFGP---EFCVSDTDAK---EIQEIDNT---K 279
 c_00000240409_1 224 ----KLPF---REKRL---LILRN---KQGHYLMERPPS-^YWIGGQIFGP---SIERNQQLS---V 281
 c_000002139456_2 224 ----KLPF---REKRL---LILRN---KQGHYLMERPPS-^YWIGGQIFGP---SVERNNQQLN---V 281
 c_000003535614_2 226 ----KLPF---QOKRL---LILRN---KQGHYLMERPPS-^YWIGGQIFGP---EVKNNTQOII---V 283
 c_000002657784_4 223 ----KQKPF---KTVW---LVVM---KNGKVLKLLRNRL---ESENIQDQLGK---ECLTMFKEKR---K 280
 c_000003800129_15 221 ----TPKKK---FVFLN---EYXK---RDGICYLMERPPD-^YWIGGQIFGP---TREVAEAEELRF---P 262
 c_000003716781_1 216 ----PPID---VELEM---FLVH---DHRGVLLERRDEG-^YWIGGQIFGP---PAAELEL---P 272

c_000001556689_3 216 --LVKA --VTEYKA --AVIE --DNGRLLRLGPGR --SLPSDLMWEFP --TLDSRLADTSSRSLTKEPKASSHSRAQEAKLRSR --YIKEQLGWS --V 293
 c_000001834452_1 196 --INWKE --VHLLY --GVAS --CPSVLLKEERKS --GWNGLWFP --SEVYDQIWEFP --PDL --ANREFQPOR --G 25
 c_000004255004_2 226 --TKA --VQWPL --TLAK --WRSRILLHRRPDK --GILASLWEFP --TPENLPAEL --IDETL -- 274
 c_00000605438_3 219 --IKKPI --RKRAL --IIYI --KNDKYLALQNEE --RLLSGLWFP --QEEDEF -- 262
 c_00000581237_15 219 --IKKPI --RKRTL --IIYT --KDDKYLALQSNE --RLLSGLWFP --QEEDEF --S 262
 c_000005259579_2 219 --IKKPI --VYIH --KANXYALALQNEE --RLLSGLWFP --QEEDEF --L 262
 c_00000697838_1 208 --APVSK --RRFIG --FIIA --WHGNWVVFQQRPSG --EVNGSLWEFP --NLWLDDTQ --VQPAD --LASTKLALG --Q 267
 c_00000583727_6 234 --AATTK --RRFIG --FLIA --WRGKVFVQQRPSG --EVNGSLWEFP --NWQVNDAE --LKPAA --LAKRELRLLQ --Q 293
 c_000005807640_2 236 --AATTK --RRFIG --FLIT --WHGKVFVQQRPSG --EVNGSLWEFP --NWEV-DAA --ATADQ --LAKRELRLLG --Q 294
 c_00000134878_1 207 --ETTKPH --YDVAV --GIW --DCKKLLITTRKKEE --GLLGGLWFP --GGKMKKNEK --IESAIKRR --EIKEELINS-I 265
 c_000005136725_3 222 --HTKPH --YDVAV --GIW --DKQKLLITTRKKEE --GLLGGLWFP --GGKMKKNET --IKSAIKR --EIKEELINS-I 284
 c_00001684786_4 222 --RKTKIA --VELEA --GIIR --RGCRCLLERSSEF --DFLGLWLWFP --LSPLPGD --GGIAA --RLQGALDTE-I 275
 c_00002971826_23 232 --RKTKVA --VELEA --GIIR --RGCRKYLLENEREDL --DYLEGLWLWFP --LARPDS --GGIAA --RLRKLATS --I 283
 c_00002040695_2 233 --TKKPC --RWGHT --FLVH1PSSGSPVSVLLEKEHG --PLLEGLWLWFP --TQSHWVHKKT --EAH --FPY --K 290
 c_00003263657_32 239 --RKTKP --HIDHVA --GVK --LDDFLIITQRLD --GLLGGLWFP --GGRRESES --LTGVT --TILESTGLT --V 302
 c_000005989041_2 237 --KVKPV --RKGIV --FLALM --SDGSFLRRLRPFET --GLLGGLWFP --STNWLERRI --SPE --EICEAAFP-L 296
 c_000002391082_2 234 --KVKPV --HYQVAC --GVIC --KGDLHLLAQARPSA --GMLGGLWFP --GGKQEEGET --LQCLQR --EIAEELAIE --I 297
 c_000003839553_2 241 --KVKPV --HYQVAA --GVIS --KGQRLLIAQARPAE --VMLGGLWFP --GGKQEEGET --LEELCVR --EIRELGIE --I 304
 c_000003936707_2 231 --KPTPR --HYEVT --GVIV --KGSKLVVAQARPSD --GMLGGLWFP --GGKREGEPS --LQECLR --EIREELTD --I 294
 c_0000044047601 214 --TKVTF --VQRVA --FLFLEK --QDGSSFLVVRPAD --GLLGGLWFP --GRELDGEAPE --RLASTICG --EFIGLITRPD --270
 c_00000607075_1 224 --KKKID --KYFLI --TLYK --NQNLKLLIKNDKF --KFLNKLWFP --MREITYP --D 244
 c_000002094036_4 200 --KKLIN --KFYLA --TLYK --NQNLKLLIKNDKF --KFLNKLWFP --MREITYP --D 244
 c_000002762689_1 224 --KKNNN --KYLLI --KVKY --DCKNKLILLIKNDKF --NFKLNSLWEFP --MQLSKP -- 267
 c_000005516980_2 224 --KKNNN --KYFLI --KVKY --DCKNKLILLIKNDKF --NFKLNSLWEFP --MEEFLQP -- 267
 c_000003920004_1 233 --KKSRMK --KVTRAY --IIIN --GSEELVVRASK --GMLGSLSLWEFP --NDKWRVTEKK --LLVRD --SSVSLFK --292
 c_000004008511_2 227 --KSTKRK --KYSRAY --IFYN --EKEELIIRKRPSK --GMLGSLSLWEFP --NDNWVINKK --SLVTD --TIAIKLK --280
 c_000004481347_1 234 --LEQPKKI --KFTRAY --IIIM --KKNVLEVNRRRSK --GMLGSLSLWEFP --NDEWDVVKK --KLKRH --DVYKDLS --293
 c_000003787733_3 226 --KIIPH --KEIVIA --GIW --QRDFKLITTRKPFN --ALLGGLWFP --SAEIQLQNET --PIGALAR --QISKDFNIX --I 288
 c_00000754657_2 226 --KIVPH --KEIVIA --GIW --QRDFKLITTRKPFN --ALLGGLWFP --GGEIASSET --FVGALAR --QIKEKWDID --I 288
 c_00000090943_3 226 --KIVPH --KDIIVT --GIW --RGKFLITTRKSEN --ALLGGLWFP --GGEFESNET --PIEALR --KIKEKCDID --I 288
 c_000004474996_2 225 --KIVPH --KEEVA --GMIC --QGKFLITTRKRNEN --ALLGGLWFP --GWEIVSSET --PIGALAR --KIKEEWAID --I 287
 c_000001463500_11 213 --SIKIN --KHTPIA --AIW --YNEFELIMKRPVN --GMLGSLSLWEFP --NTQDILKIS --DDIQLFN --YIKNSLNFT --I 275
 c_000001286181_5 218 --VTKTO --ESPHW --FIYQ --KNNVKEMLCKNDE --GIWPNLWFP --KRELQF -- 262
 c_000001293628_3 223 --PRVRE --ASYAV --LVTLN --PFGELLMVRVPD --GMLGSLSLWEFP --AVEMAEERPNWPGELR --VGDLSHGRSARLE --DLGVOLSAR --R 301
 c_000001535696_8 222 --TKVRN --RYENY --IPIIAEADDSGKTRSLNQRKG --DIWONLWQFP --LIERTRIED --LEEVKK --RLCEVMILAD --DPQ 290
 c_000001614067_2 233 --KEKEE --RYGLF --FQLQN --KDGAVLFETNKGSS --MLNGLWFP --SIGWEYESN --RFKKSPK --FNKKPPNFL --G 297
 c_000001765289_1 198 --KKKPK --HYTIVA --GIW --RDNTF1IQRKEE --AMLGGLWFP --GGKVEEGES --LEAALR --EIKEECCV --P 261
 c_000001961666_1 213 --KTIRK --GKCY --IIKRL --NDNSKFLIRNPBK --GLLGGLWFP --TIGWHLSDH --DDYVK --SMIEKFKN --I 273
 c_000002018097_4 221 --KPTPR --KEYVA --VFIS --YNDHFIQINRSQED --GLLGGLWFP --MIEVFENS --KIKALKT --YAKNKLGLS --I 283
 c_000002843512_36 213 --VVPT --REQNI --MLYF --YNDKSLSLQRTG --FHLGLWFP --AIEVPL -- 253
 c_000002930199_1 220 --RKPKP --RYFNY --LFIK --KNQFCFLIQQRGTN --DIWONLWQFP --LIESKEKLIN --REKIMA --HQYKPFK --I 280
 c_000003033795_2 214 --KKPPI --YDIAV --SVVE --YQNLKLLITQRKQ --NLFGLWFP --GGKIKKEET --ATQAIR --EVKEETNL --V 276
 c_00000388107_2 215 --KTPTR --IKVST --CIIR --QNKQLLILQORPID --MLNGLWFP --GGKIQNNES --RKKAVIR --EVMEETNLA --I 277
 c_000004013286_6 242 --PKQIA --VREVA --VIVR --KGKRVLVLVQRPSD --GRWAGLWFP --RFECSFRSK --KIRQLMVR --GVGGQZGPV --306
 c_000004852258_1 202 --KTPTR --KTTIAA --ALVN --HGDNIFITTRKPLR --GLLGGLWFP --NIELVNGEI --PEDLILK --KFAFDQGLT --I 264
 c_000005254087_1 230 --KTPRPHDVAGIIWONG --VSPG --EGPGRFLIAQRPK --GLLGGLWFP --GGKQREPDET --LQPALE --EIREEMDDM --I 299
 c_000005603677_1 227 --LKKIKK --RYFYY --FVNRG --TNGGYLLRKRQEK --DIWQGLWFP --MIELDKPLK --TSELKT --LWAHWELE --I 283
 c_000003283462_3 213 --KNIPH --YDVFV --GMIN --KEGRFLIQRKRENO --KHLGSLWFP --GGKINNER --SELALIR --EIKEEDCFE --V 275
 c_000002687221_1 227 --VSRPH --HNVAV --GLIW --KDNRLISKRNAS --GLLGGLWFP --GGKIRSGES --GSSCVR --KTQEEILNVL --V 281
 c_000003347358_19 244 --VVPT --REENI --LVRV --YDDKLSLTQREG --KFLHGLWFP --SVEVPH -- 284
 c_000002701031_2 222 --LKKRK --RYFEEF --LILN --ENEKILQKREK --GIWGLWFP --LIEINSVSKSE --RDIENSKS --WKKLFKINS --I 285
 c_000000582753_5 230 --RKSTN --RYTYT --YYIS --DSNHYL1KRRKGK --GIWNSLWEFP --LSENLIJAEF --NEVLDRK --QIFKNSND --F 293
 c_000001713769_5 222 --NAPKI --IHPNY --LVLL --DSDHMCMDRKN --GIWNSLWEFP --MIESKKEIN --KTQVLSN --EIKFSIAPIS --NSD 287
 c_000006057486_30 225 --IKKKR --RYFYY --LFLK --EKLDFLIIQRKQK --DIWNSLWEFP --NIEGMEELN --ESEIREA --IQCQYKPNL --Y 287
 c_0000015494072_10 222 --IKKKR --RYFYY --LFLK --EKLDFLIIQRKQK --DIWNSLWEFP --NIEGMEELN --ESEIREA --IQCQYKPNL --Y 287
 c_000001742634_3 225 --IKKKR --RFFNY --LILIT --DKDNTT1IQRKRN --DIWNLWFP --LIEETTEINEN --KVIKSSII --QDFNNNSY --T 297
 c_000002826998_2 225 --IKKKR --RYFNY --LILIT --DKDNTT1IQRKRN --DIWNLWFP --LIEETTEINEN --KLVNDNFF --QKLILNDTT --I 292
 c_000003159439_6 222 --IKKKR --RYFNY --LILVNI --NQQNITYI1QRKRN --DIWNLWFP --LIEKHEIDTN --KVIKSSII --QDFNNNSY --T 297
 c_000004887214_1 227 --LKKIKK --RYFNY --LILKT --KNEET1LVQQRKG --GIWQNLWFP --LLETEKEINYL --ELIDHSIF --KRMIEENN --D 293
 c_000005037037_2 225 --VKVK --RFFYF --LVLN --IDNTT1IQRQNN --DIWGLWFP --MIEQQLIN --KPKEKELE --LITKINVR --F 287
 c_00000105964_1 217 --KVKVI --LYRGC --AVIL --FQDGLL1IQRKRN --GLLGGLWFP --SVELGKGDN --LRLAKQJ --HLLELGKP --F 275
 c_000002498472_1 226 --KPKIP --IDMAT --GILI --HNGMLF1IQRPLD --DWGWSLWEFP --GGRMLKET --PEQTQVR --EFLTEEFK --F 287
 c_000004615912_3 215 --IRKKK --RYFHY --FISH --EGKRL1IQRKQ --DIWONLWFP --MLETDKEELK --ISRVQKT --QLWQFLFGHNGEVKVL --284
 c_000004766858_2 228 --PETR --IEKLT --ALVQ --NGCDWLVRKRP --GLMGLWFP --TWELPLSNS --AAWLTQ --QLHEDYVGA --297
 c_000002747260_18 232 --KPKPT --RHHIA --YLGR --ADGAWLLELLRDPK --GLLGGLWFP --GHDWAEAT --EAPPID --A 285
 c_000003571561_38 232 --KPKPT --RHHIA --YLGR --ADGAWLLELLRDPK --GLLGGLWFP --GHDWAEAT --EAPPID --A 285
 c_00000169465_2 234 --KALPE --KSTYM --MVAQ --FNSVQVLECRPST --GLWGGLWFP --EVSSIEEGIE --EASPPPC --G 286
 c_000003254110_11 234 --KALPE --KATFM --MVAQ --FNSVQVLECRPST --GLWGGLWFP --EVSSIEEGIE --OLAKRIGIS --V 285
 c_000004750284_20 230 --KTPPV --RQTYM --LIPM --FQCVQVLECRPQ --GIWGGLWFP --EAQSAEREGIS --OFKKRGNV --V 285
 c_000001041782_15 235 --REKPL --RRTRM --LLIVD --GGRVLLLERRPS --GIWGGLWFP --ECPPEVDOVS --LLAQRIE --V 285
 c_000002717847_8 224 --KRD --VYLEF --SLYK --DGPKYFLQTAEL --GFWRKWMF --VKIVN --YCRHEGLE --V 263
 c_000004054799_1 215 --QRPT --VRLN --FILPH --TKNEILMHKQAS --EYESWLWFP --DGEVWIKK --G 255
 c_000001626673_1 241 --KARPO --RYGIV --YWLSD --DDGNL1VHHRPEK --GLLGGLWFP --TSSWEDEL --TVEH --PIFISEQIK --P 302
 c_00000339186_3 199 --IRYED --KQD1V --YLIT --RGPY1VWQRPEG --GVNGLWFP --NSDNGKNYS --APK 247
 c_000001863436_1 230 --VPLQSLREE --GCV1IWIAG --KPYN --ADSL1IIVQRCP --GLWGGLWFP --HDEQTTDEP --HNAALR --IAEATVGP --F 285
 c_000001808634_1 246 --ARQPK --VAARTAVTTSGGA --EGRFLMVKRDPK --GLLGGLWFP --TVEAAGDVA --PTQAVS --RLLARLTGPP --V 315
 c_000002830137_4 239 --TQTEH --VVEAA --VVIR --RGSVLLRHRHEDR --EWRGWLWFP --RFPFDPPDTN --NQHQLVOL --KSRELTGFE --I 303
 c_000004612302_3 232 --VRTTD --LYOLL --LIVR --DRDRL1IQRHPER --ERWGWLWFP --ALSR7GPLLS --RIDELOTA --SEAAEWCCD --V 297
 c_000001933926_1 217 --VARPT --RHHGV --FWATR --GDDGCVL1IQRPEH --GLLGGLWFP --STEWRECPW --RIRSVA --KAAPFA --A 276
 c_000003136334_1 229 --KRKPT --RYATV --FLWLL --GRGNV1L1RREK --GLLGGLWFP --STDWLENT --PTKENS --MFAAAA --V 288
 c_000005849454_9 234 --KKWVD --LEMVT --LVRH --WCVGRVGLQRTS --GWSGLWFP --SAICSEFATANE --ASSAESTH --SLAREHR --A 299
 c_000001169194_1 234 --VPARE --ISFLL --VILQT --EEGEV1MLVQRPEK --GLLGGLWFP --EQELAKPLD --CAATSDR --RAIELMLSTG --EVV 301
 c_000002786947_2 234 --APARE --IFPLL --VILQT --EEGEV1MLVQRPEK --GLLGGLWFP --EQELAKPLD --CSTTSRD --RAIGLAMLTG --EVV 301
 c_000002712988_52 220 --KTTKTRKRHR --IEVGI --ACIW --REGKYL1VQARPKG --GFSKGEWSFP --GGKREGEKS --FRGCVR --EIEEEVGLN --V 287
 c_000003029168_2 234 --KKKIN --KFYLA --TLYK --HDDQV1L1KNDKF --KFLNKLWFP --MKEIQSOS --SAT --ALMK --266
 c_000003085605_2 224 --KKKIN --KFYLA --TLYK --HDDQV1L1KNDKF --KFLNKLWFP --MKEIQSOS --SAT --ALMK --266
 c_000003745491_1 218 --KIKQK --TIDRA --WVEQ --GGKLL1LHRANDKS --RLLSL1LWEFP --TLDVWGMY -- 267
 c_000004187032_4 224 --KIKKNE --KFYIL --KVKY --KNQYKL1L1KNDKF --KFLNKLWFP --MEELSKP -- 267
 c_00000607315_3 224 --FKNPK --KFYEA --NIQY --NQNLKLL1KNDKF --KFLNKLWFP --MKEVEKN -- 267
 c_000005543774_1 186 --IESKTI --YDIFC --YLRK --NKKQ1I1KNDKF --GFLKK1LWEFP --IKATSKK -- 230
 c_000006211484_1 208 --IQSKN --YDIFC --YLQK --NKKQ1I1KNDKF --GFLKK1LWEFP --IKETSEK -- 252
 c_000002718976_3 227 --IKSKN --YDIFC --YI-N --TKQK1I1KNDKF --SFLRD1LWEFP --IKE-TNNVSN -- 272
 c_000002992548_1 226 --KKEQK --YNYC --YLKK --KKE1I1KNDKF --SFLSN1LWEFP --TK1ETSNKKK -- 273
 c_00000465751_1 192 --IKEKK --YLNR --YKKE1I1KNDKF --SFLNN1LWEFP --IQMVVANNNLK --K 240
 c_000002655634_1 223 --TEKKK --FNVY --YLKN --YKKE1I1KNDKF --GFLYNN1LWEFP --IQMVEWTNNNLK --R 272
 c_000004232403_2 223 --KKKIP --RNLLA --ALIE --YGDYL1L1KNDKF --GFLYNN1LWEFP --NIELNNGRS --PEKSILKE --SINHHYFT --I 285
 67ut. after.N146.pdb 226 --TAVQK --VPLAV --AVLAD --DEGRVLLRKRDPK --GLLAAW1LWEFP --SCETDGDG --KEKLEQ --MVEQ -- 285
 Consensus_aa:h.....hh.....ph1..pp.....shh..h.p.....p.....p.....p

c_000003607531_2	297	QWVKVIACTVNLIAFTHFQ--LQLQVRVAVV--NDQT-----QISGG-----IWVDRDRLEFA--LPNLMKVWAFAS--KRIQPLRLOC	368
c_000003872363_4	286	DWQRQLDAQVRNFTHFH--LRLSLRIVVA--KNV-----IPTVG-----MFIPSDFKFDPD--LPTVMRKAHALAD--AAFDHG-----	351
c_000004820107_1	292	EESEQTNLNSVSKNSFTFH--LII-----	311
c_000004820107_1	276	SEWKGICQIQRNFTHFH--LII-----	293
c_000005867023_3	290	DWVFLVGAEVTRFTHFH--LRLMWHVATV--NEN-----SPDVG-----KFIADDFCPDAD--LPTVMRKAHLAL--GSFVGD-----	355
c_000004546210_2	287	LWVDSWEPFKMSFTHFH--LYIHP1Q1RMA--DQGFM-----NTAYAKS5T-----QFSLSNLQDLSLG--LAPAVKGLVNKLKY-----	351
c_000004546210_2	287	LWVDSWEPFKMSFTHFH--LYIHP1Q1RMA--DQGFM-----NTAHTPKT-----QFSLSNLQDLSLG--LAPAVKGLVNQLY-----	351
c_000005810085_5	287	LWVDSWEPFKMSFTHFH--LYIHP1Q1RMA--DQGFM-----NTAHTPKT-----QFSLSNLQDLSLG--LAPAVKGLVNQLY-----	351
c_000002511148_4	282	LNIKKLSKVKSFTHFH--LEAT1PYLAKL--ESNK-----KYKNS-----WVWDYKVNVESLG--LAPAPIKKTINOLT--KP-----	344
c_000000577378_2	270	LNIKKLSKVKSFTHFH--LEAT1PYLAKL--ESNK-----KYKNS-----WVWDYKVNVESLG--LAPAPIKKTINOLT--KP-----	332
c_000000598175_2	293	QESTIALAEIKMFTHFH--LYIQLP1IINV--ETPIE-----LGVMEDNES-----LWNNITTERFNCG--LAPAVQTLINITK--NIEKEIENGTN	369
c_000000598175_2	287	SDLNEVAPEPTVTFH--LTINPHVFLN1--EKAPD-----VNDKQL-----WVYPLDQSIEVG--LAAPTKKLVQJLT--ATV-----	353
c_000000811118_1	274	LNWDFWEPPFKMSFTHFH--LII-----LSIDDRKIF-----NNWNSLDSVNMG--IPIKPKVDILKLYD-----	338
c_000000899687_2	290	VSLNKKVMTKMTYQFR--VTLTFLVNCNLK--QKRCQI-----RPGCEW-----KVWSLSLKLKKY--FPAANVKVVKYL--EK-----	356
c_000001176522_24	289	GEALSAFRFTTFH--LDITTPFLWHT--HSDD-----MPQAG-----QWPLPLAQALMNN--IPIAAVKRVLQEV--EYEQRN-----	354
c_000001345122_1	262	EQAVI-----KTFHFH--LRLMKPMM-----	281
c_000001515736_6	279	RSWQDLPGLFLVLFH--LHLHPLVFLPV--DQAYA-----ALVSEADES-----CWADAAAWEALG--LAPAPIRKLLDAEL--A-----	345
c_000001682161_6	283	D5VLENLQKLHTFHFH--LMMNPFMLQ1--ERE-----AIEEV-----SWELENLNGV--LAPTPKKILRESF--A-----	343
c_000001923643_29	289	ADWTDELPAFLVLFH--LHLHPLVFLPV--HGSGP-----DAMADGCG-----CWADARAWADMG--LAPAPVRKLLDAEF--AG-----	356
c_000002030994_5	282	SEADKLSKL1KHTFHFH--LHQ1QPL1IINT--ETPLK-----MGVMHEHDS-----LWNVNTTEFNCG--LAPAVQTLINITK--DVKGTQNDTN	358
c_000002038721_4	276	EEKESELVPLVNLNLSHOK--LS1SF1TCKYF1--KRNDR-----LP1DRRGF-----WVWNKVESLG--IPIKPKVTLLENH-----	340
c_000002523527_15	316	ARAYLAGELTFTHFH--LRLRAIRVDI--DTVAL-----EGDAAAS-----RLWLSLDDLDMLG--LAPAPVRKLLLEDQA--RFLGF-----	383
c_000002608528_2	282	DEA1KLSEIKMFTHFH--LHLHPL1IINM--ETPLK-----IGVMHEHD-----KWSLSSLKLY--FPAANVKVVKYL--EK-----	321
c_000002961510_2	290	ISLNKQVMTKMTYQFR--VTLTFLVNCNLK--QKRCQI-----RPGCEW-----KWSLSSLKLY--FPAANVKVVKYL--EK-----	356
c_000002979152_8	302	KP1QAYAEIHEHFHFH--LSAKVLRLEI--DSDKR-----LGWMEETNA-----LWNNHHNDKPNHG--FPAPIKLLLSRN--L-----	356
c_000003248024_1	286	KP1QAYAEIHEHFHFH--LSAKVLRLEI--DSDKR-----LGWMEETNA-----LWNNHHNDKPNHG--FPAPIKLLLSRN--L-----	356
c_000003294679_3	288	ASSETIAVIRGYSYFR--VTLTFLAFLR--GEPEW-----PEPVADLHEDEG-----CWPVRKLLATIP--LTTVRKRALAVL--PNSSDRNDEQN	372
c_00000333364_1	253	SVRPHFEEELHEFTE--LRLRFAHRCQ1--QAGEP-----SPQEEQE1-----KWSLSSLLKLY--FPAANVKVVKYL--EK-----	316
c_000003402697_5	284	CASEYEFEEVHAYTFH--LICKVYVYDE-----LDAEON-----NIANSDA-----LWIKSDSQTTRG--LAAPVKKLLETFP--NIQ-----	353
c_000003492925_4	288	SVRPHFEEELHEFTE--LRLRFAHRCQ1--QAGEP-----SPQEEQE1-----MHW1HSPEAEDLG--FGRVTRKILEAAA--THEISQER-----	378
c_000003766664_5	302	RIEQCCS9RVBLFHRD--MIAVAFALSV--KGREPS-----PALSTYTTGG-----WVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	350
c_000003948184_72	284	PSLQI1QKVKVMSFHFH--LEAT1PYL1V1--ERE-----HVMEDQO-----LWNNHDKPNHG--FPAPIKLLLSRN--L-----	350
c_000005057124_4	282	QSVQHIEKQMSFTHFH--LDTAVLWVKT--ENRIN-----NVMSNQS-----WVYKSDKINALQ--LAPAPIKLLQCNQ--RNEV-----	350
c_0000055090105_5	282	RIEKT1QKVKHMSFHFH--LEAT1PYL1V1--ERIN-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	346
c_0000060633682_5	282	PSLQI1QKVKVMSFHFH--LEAT1PYL1V1--ERIN-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	344
c_000004676312_2	261	TSKQD1QKVKHMSFHFH--LEAT1PYL1V1--ERIN-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	324
c_000007882235_1	283	PSATTEFSHFSFTHFH--LQATQ1L1LTO--STPIK-----RVMEGTPS-----LWNVNTTEFNCG--LAPAPIKTT1INLN--EDF-----	350
c_000001187176_11	283	AEEAELPSHFSFTHFH--LDTQ1P1RVR--DADPR-----VTESGQ-----LWNVNTTEFNCG--LAPAPIKTT1INLN--EDF-----	356
c_000001660697_64	297	SDVEAGALHTTFHFH--LHMHLLHAD1--TKP-----ACLDDD-----RWWPLAQLNSVG--LAPAVKLAETLV--QPSL-----	363
c_000003378864_2	301	EVGPLTQVHAYSHK1T1HAYHCRLLRG--TPH-----PKVAVES-----RWWPLAQLNSVG--LAPAVKLAETLV--QPSL-----	375
c_000003910742_3	264	ENTEKETKLMSFTHFH--L1H1HALL10--TPPIK-----RVMEGTPS-----LWNVNTTEFNCG--FPAPIKLLFTRL--T-----	349
c_000004116181_2	249	KTSEVPLT1RHTFHFH--LQYTT1LVR7--DNFLN-----FVMEDQO-----LWVYKAEQ1PGLG--LAPAVKRLLDNL--IINEDNN-----	320
c_000006223903_1	284	LSLMPAVS1RHTFHFH--LA1TQ1QPARA--PKRLT-----RPMQSTGL-----LWYKPFGEF-----	332
c_000001029068_2	257	QVRKKLPL1MSFTHFH--LTT-----	276
c_000001128125_11	291	WNL1DRLNKGEGVGVPAVNPMLERL-----ETLDK-----	374
c_000001595844_4	288	NIK1K1LGPV1HTL1HOK--V1Q1FFD1T1--VGE-----KOLD1-----APISKVDLQVFP--FPKMMMMYINRIM--A-----	348
c_000001719155_2	292	HIDEK1LMTVKSFTHFH--V1LHVF1CRV--LSGRV-----SPTHECEW-----DWWVKEELDRDYP--FPAANVKI1KSLR--KNQGRS-----	362
c_000002363036_6	293	HSLSSEGEPRFMSFTHFH--LDT1PL1P-----QASTDYIEQCP--SNRQVETPAEYVAH-----QWF1DQLSK1GEG1VGPV1PKM1SSL-----	370
c_000002561400_1	247	EGLIE1QPLDHTTFHFH--R1D1PL1FDE--RGPAPKNTAK1-----ISEPQGVYDAEQQ-----SWF1DQLL1DGG1LAPAVPNKL1GRLD--ELVE-----	327
c_0000025660035_2	284	KQK0YKEPK1P1K1HTFHFH--L1T1R1P1H1K1--R1YKD-----RDYDFTKVL-----K1KNT-----VWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	358
c_000002598045_3	282	SN1KKL1KVKHMSFHFH--LEAT1PYL1V1--KKGK-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	344
c_000004188575_41	282	TD1K1L1KVKHMSFHFH--LEAT1PYL1V1--KKGK-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	344
c_000002400409_4	282	TSKQD1QKVKHMSFHFH--LEAT1PYL1V1--KKGK-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	345
c_000002139456_2	282	NSHNDLAAFR1T1HAYHFH--LDT1PCEV1D1--TQSAY-----A1AENDRF-----QW1H1EDAL1TG--LAPAVKLAETLV--QPSL-----	347
c_000003535614_2	284	NSHNDLAAFR1T1HAYHFH--LDT1PCEV1D1--TQSAY-----A1AENDRF-----QW1H1EDAL1TG--LAPAVKLAETLV--QPSL-----	347
c_000002651784_4	284	KKHD1L1KVKHMSFHFH--LEAT1PYL1V1--K1S1N-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	345
c_000003800129_15	263	EN1K1L1KVKHMSFHFH--LEAT1PYL1V1--K1S1N-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	351
c_000003716781_1	273	LNQEEV1K1L1KVKHMSFHFH--LEAT1PYL1V1--K1S1N-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	320
c_000001556689_3	294	R1LGP1K1L1KVKHMSFHFH--LEAT1PYL1V1--K1S1N-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	331
c_000001834452_1	254	LGKEMGP1K1L1KVKHMSFHFH--LEAT1PYL1V1--K1S1N-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	331
c_000004255002_2	275	QGQDPVAL1HAY1D1M--IT1VAP1L1V1Q1C--K1F1V-----K1F1V-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	346
c_000006050438_3	263	EMNTGL1GEFKHAYHFH--LDARV1L1K1D-----K1Q1E-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	322
c_000005136725_3	263	KM1K1L1KVKHMSFHFH--LEAT1PYL1V1--K1S1N-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	322
c_0000058123715	263	KM1K1L1KVKHMSFHFH--LEAT1PYL1V1--K1S1N-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	322
c_000002529579_2	263	NDN1K1L1KVKHMSFHFH--LEAT1PYL1V1--K1S1N-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	322
c_000006079783_1	268	VEPEL1C11L1KVKHMSFHFH--LEAT1PYL1V1--K1S1N-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	322
c_00000583727_6	294	AEKQFPLCN1S1T1HAYHFH--LDT1PCEV1D1--K1S1N-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	322
c_000005807640_2	295	AELEPLCT1N1S1T1HAYHFH--LDT1PCEV1D1--K1S1N-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	357
c_00000134878_1	270	KP1T1Q1VQD1HAYHFH--V1T1N1NCKY--NGGDI-----KLSGPTDF-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	322
c_000005136725_3	265	KP1T1Q1VQD1HAYHFH--V1T1N1NCKY--NGGKI-----KLSGPTDF-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	352
c_000001684786_4	280	ETGEAL1K1A1HAYHFH--V1T1N1NCKY--NGGKI-----KLSGPTDF-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	343
c_000002971822_23	290	EGEAL1K1A1HAYHFH--V1T1N1NCKY--NGGKI-----KLSGPTDF-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	343
c_000002040694_2	291	ASW1T1L1Q1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	353
c_000003263657_32	303	KTD1L1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	353
c_000000598904_1	297	QWV1N1Q1T1G1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	358
c_000002391082_2	298	EVG1L1T1V1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	358
c_00000392004_4	305	EVAV1L1T1V1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	331
c_00000400408511_2	295	EVG1L1T1V1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	346
c_00000060007075_1	271	EVG1L1T1V1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	353
c_0000020440706_1	268	--NNNE1K1E1HAYHFH--L1H1H1L1H1-----	331
c_000003223903_1	294	WV1N1Q1N1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	331
c_000005807640_2	295	WV1N1Q1N1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	331
c_000000134878_1	270	WV1N1Q1N1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	331
c_000005516980_2	268	WV1N1Q1N1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	327
c_00000392004_4	293	WV1N1Q1N1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	309
c_00000400408511_2	287	WV1N1Q1N1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	343
c_00000060007075_1	290	WV1N1Q1N1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	343
c_00000204418347_1	294	WV1N1Q1N1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	355
c_000003787733_2	289	WV1N1Q1N1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	348
c_00000754657_2	289	WV1N1Q1N1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	357
c_000000999043_3	289	WV1N1Q1N1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	357
c_000004474996_2	288	WV1N1Q1N1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	356
c_000001463505_1	276	QZ1NQK1V1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	344
c_000001286181_5	263	QZ1NQK1V1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	333
c_000001293628_3	262	QZ1NQK1V1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	327
c_000001535696_8	292	QZ1NQK1V1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	331
c_000001614067_2	298	QZ1NQK1V1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	344
c_000001765289_1	262	QZ1NQK1V1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	292
c_000001961666_1	274	QZ1NQK1V1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	331
c_000002018079_4	284	QZ1NQK1V1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	347
c_000002843512_36	254	QZ1NQK1V1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	317
c_000002930199_1	281	QZ1NQK1V1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	344
c_000003033795_2	277	QZ1NQK1V1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	342
c_00000388107_2	278	QZ1NQK1V1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	342
c_0000040413288_6	278	QZ1NQK1V1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	344
c_000004852581_1	307	QZ1NQK1V1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	392
c_000005254087_1	265	QZ1NQK1V1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	330
c_000005603677_1	300	QZ1NQK1V1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	351
c_0000052382462_3	290	QZ1NQK1V1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	355
c_000002687221_1	290	QZ1NQK1V1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF1-----LWVYKTSK1QPLK--LAPAPIKLLQCNQ--RNEV-----	355
c_000003347358_19	285	QZ1NQK1V1Q1V1T1HAYHFH--V1T1N1NCKY--NGGKI-----RPGF	

c_000002826998_2	290 QIKKISLPRFRQNLTHRK---IIATFFEIDF---MDKFS-----NKNNNY-----IKIERKNLSKFA--FPKIIDLYLKDNS--LNLLAMK-----	358
c_000003159439_6	289 TSLYSNDNIKKLQLSHQH---IFSKFWIVET---KQNIL-----NSILIEDLEKYP--VPKLIQNFIEKED--NIKLNPNATK-----	354
c_000004887214_1	294 LQLLTPKVVVHQLSHQH---LYIKFWERV---PTFRS-----ATISWQELLKLP--FPVIVIFKFIREFL--ETSMPNFDTPL-----	361
c_000005370372_2	288 KVKKRKSKNTRLTHQK---LNAIFIEIDL---KTK-----INTDINNLNSKFA--FP-----	336
c_000001059961_1	280 QLGKEIHTTLEIHYTFR---TRVSFFTATVY---HVLPE-----INKQF-----RFPVKTSEVE-----	326
c_000002498472_1	289 EIVEKIKNTVHHTYRK---VTLHCFLLRL---KKGSDT-----DPVLHQAQDF-----NWVPPFKALQEYA--FPAGHRKLITYL-----	355
c_000004615912_3	285 QTKK---YQELTHQK---ISAYFWEIN---LNVAP-----VNIYDF-----ILIDSKNLAKFA--FPKIIALYLENNS--INL-----	345
c_000004766858_2	292 RIGPELFKMEHDYTRFH---ATLHCRCLTVY---AEIPS-----VLPDNA-----QWNLGEISALAL--LPRVFQLRKRL--DEVSA-----	359
c_000002747260_18	286 PWRDPGAEVRHTFTFH---LRLALRVAIDL---PASA-----TPRTG-----QFIPHAFRPGQ--LPPTVMRKAYDLAS--ATFLGN-----	352
c_000005371561_38	287 NWRTVTPGEVRHTFTFH---LVLELRRADL---PEDC-----TTMRC-----QFLAFCGERPSD--LPTAMRKAFDLR--DG-----	349
c_00000169465_2	290 DETRTLEGFRHTFTFH---LDITPVAVV---NSAPS-----KRAVETAF-----RWFSLNEPIEVG--LAAPITKIIQQIM--R-----	355
c_000003254110_11	290 EETKLTETFRHTFTFH---LDITPVAVV---NSPPT-----KRVADTAF-----RWFSLGEPIEVG--LAAPITKIIKQD--G-----	355
c_000004750284_20	286 EEIQOLDEFTRHTFTFH---LDITPLIAVY---NSTPQ-----KRAVANES-----RWFLLDDEPIEVG--LAAPITKIIKTLA--K-----	351
c_000004171782_15	292 TRPQAWPGLRHTFTFH---LDIQFVRLRL---TAAPA-----RIMDGDRIH-----VWNVESPDLARG--LAAPFVARLLQGLA--LEDTRA-----	362
c_000002717847_8	264 --KENANKALSHRH---IHFRFKQIAN---LPT-----GMKG-----EWSKSKDLANIA--LPKPISDKLLQDG-----	319
c_000004054799_1	260 NQKSYKLLKVNPLSHLN---LDMKIKTFKJ---DKKCD-----LDSN--L-----	296
c_000006126673_3	302 KPKYKQDLPFIMUVFTHFS---LKLYLCKTSE---FICDN-----VPEEH-----KWFISTLEDIG--FPFSVFEKVKVKKI-----	362
c_000000339188_3	248 -NGQPFASARISITRYR---IQLNAHLIHA-----KRGKG-----MNCTSDEIRSLR--FSSAHILKLLGKLP--HRLDK-----	308
c_000003029168_2	299 QKLTKLVTITTTTVMTR---TTLHVFEGML---DLETS-----TMEAASSR-----LWVRLDGLHKHP--FSSPQHEIIKTLID--KEAALJLAGIV-----	374
c_000001863436_1	316 TRPVELGEVHHFSHHR---MTTITAEHRVVA---SSEQVRL-----RSDDDEVGOEA-----RVWSADEVLDL--MSAAMRKIEGLYR--ELGSGAI-----	391
c_000002830137_4	304 EVFVQLTTIQIGVTFQ---ITLQCHLACQ---VSGRK-----RGPHL-----SWRMPHELEHL--LSVTARKIGRIVLA--KLAD-----	369
c_000004612302_3	298 EVEEWLQEFRIGVTFR---LHLQPVVASW---RAGHW-----IGPVVA-----EWEVDELTARE--LSVAARRVADGVQ--AGQRCCGGASG-----	370
c_0000001933926_1	277 ANRELPGAVRHSFTFH---LELQIMTVET---NEEVEG-----DQ-----	309
c_000003136334_1	289 DQMLSGKVRHTFTFH---LEL-----	308
c_000005849454_9	300 XJGAYQQCVRHTFTFH---IRAHVYAAEG-----Q-----APTHLEDPQTVP--JTGALARVLDRT--SA-----	354
c_000001169194_1	302 GVASALAEVQIRFTHLQ---ARYRPWVVV---AKL-----LNGEQN-----VWMTFGEPVDFP--IPVAGHVKVLDAIA--ECRATPDLES-----	373
c_000002786947_2	302 GHASALAEIQIRFTHLQ---ARYRPWVVV---AKL-----LTGEQS-----MMMTFGEPVDFP--IPAAQHVKVLDAIA--ECRATPDLES-----	373
c_000001279803_52	288 SVRSHFYEELEFETE---LLLRFHRCQI---QAGEP-----QALENQEL-----KWSPEDEFDSVG--FLKTNHKALEKLK--SMRV-----	356
c_000000358065_2	269 SLSKPDHKLNLKMSNNMN---MNIKIQYLNK---LKK-----LNKG-----LWVEKTKLENHM--IPTFTKKIFASVK--HNL-----	330
c_000003745941_1	269 RGNTPIVTRKGIGISQ---ITEPISIDFRLSD---NSEA-----DHAEM-----EWDLQEIKKVV--LSPGHPKRWINEIL--TDFNKQK-----	336
c_000004187032_4	268 --KNNENFNKMSNNMN---MNIKIQYLNK---LKK-----FSSS-----YWFDKRRLDNYM--LPTFTKKVVKYLE--KN-----	326
c_000006067313_5	269 FTTSNLNKKINIKMSNNMD---MNIKILNKKN---LIK-----IKNS-----YLVDKNNIKDLL--LPSFTKKIFNSVS--NYL-----	330
c_000005543774_1	231 KWNKFPLCNKYSISNKK---LNINLNYKF5S---SRI-----PSKY-----NWYSL_KKNKEE--IPTFTKKIFKQJT--HLY-----	291
c_000006211484_1	253 KWNRFPLCNKYSISNKK---LNINLNYKF7-----SKI-----PSEY-----NWYSL_NKNNEE--IPSFTK-----	303
c_000002718976_3	273 QNWFPLCNKYSISNKK---LNINLNYKF5S---KKI-----PRSF-----NWYSL_KENKEE--IPSFTKRIFRQVS--TLF-----	333
c_000002992548_1	274 NGWYIPLCNKYSISNKK---MNINLNYKF1-----KKR-----PKY-----MWYPIDRRSNEEF--IPSFTK-----	325
c_00000456751_1	241 RGWYIPLCNKYSISNKK---MNIDLNYKF7---KKI-----PKF-----NWYSIDRSNEEF--IPSFTKKIFFKIA--KVYS-----	303
c_000002655634_1	272 REWNYILCNKYSISNKK---MNINLNYKF7-----KKR-----PKF-----DWSVDFKSQEE--MPSFTKKII-----	326
c_000004232403_2	286 EVGKHNGLIQITFTFH---MNITLFCCLL---KYDQ-----KLDSSA-----KWRFLFELDQYA--FSGRNHKLQF1Q--KENV-----	351
6u7t.after.N146.pdb	284 ELTEPIVSFEMAFSHLV---WQLTVFPGR---VHGG-----PVEEPY-----RLAPEDELKAYA--FPVSHQRVWREYK--EWAS-----	349
Consensus_aa:hpt.ho+bp.....1.hp.h.....h...phh.h.....phh.phs...p+hhp.h.....	

1139

1140

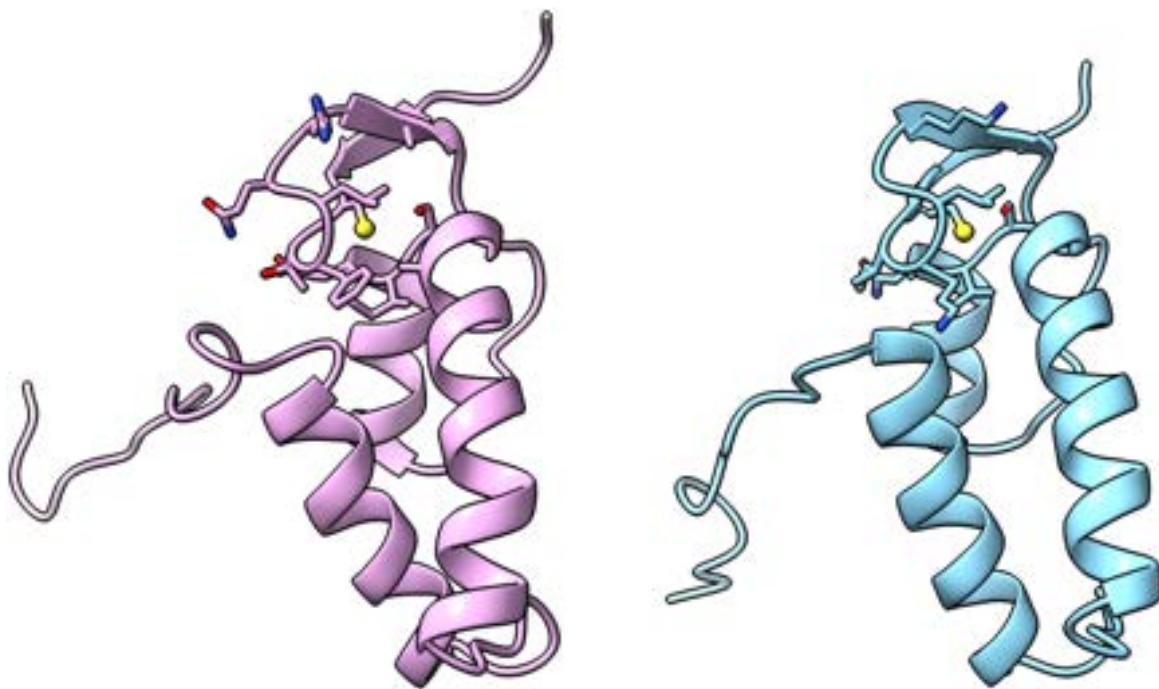
1141
1142

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)
1151 and nearest neighbor to
1152 *Thiotrichaceae* MutY (right). The structure solution NMR structure for *Ec* YggX (PDB ID 1yhd)
1153 (103) is superimposable to the structure predicted by *Colabfold* for the nearest neighbor to
1154 *Thiotrichaceae* MutY, with RMSD of 0.99 Å for 69 pruned pairs selected from 88 possible pairs.
1155

1156

1157 **S4 Table. Metabolic gene identification**

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

1158

1159 ^a*Marinosulonomonas* MutY contig belongs to two separate MAGs and each are reported separately as MAG 1 and
1160 MAG 2, respectively.

1161 ^bKEGG ID gene not identified in any MAG and not reported in Table 2.

1162 ^cCompleteness and contamination scores generated by CheckM v1.0.5 as described in Brazelton et al 2022 (35).

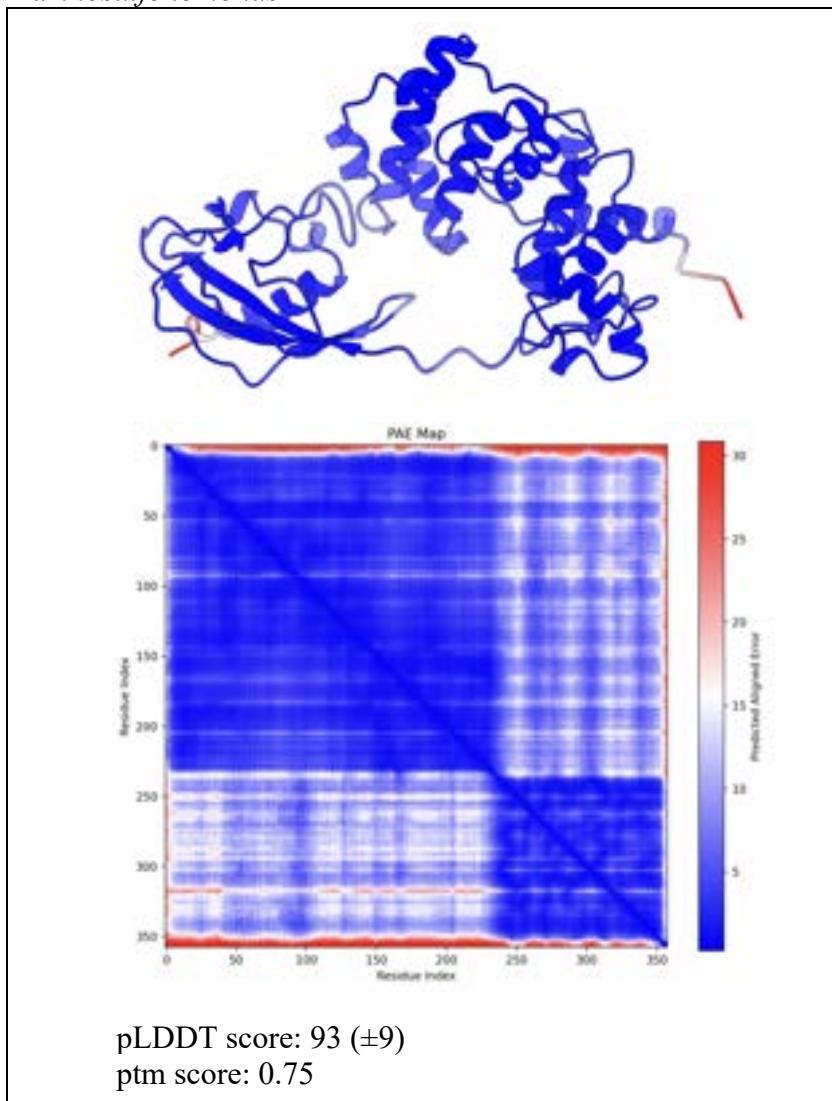
1163 A KEGG ID analysis was used to identify the potential metabolic strategies of the MutY
1164 encoding organisms at the LCHF. The full metabolic KEGG ID search is shown above.

1165 **S5 Fig. Colabfold structure prediction pLDDT scores**

1166 pLDDT scores represent the confidence in the prediction calculated by *Colabfold*.

1167

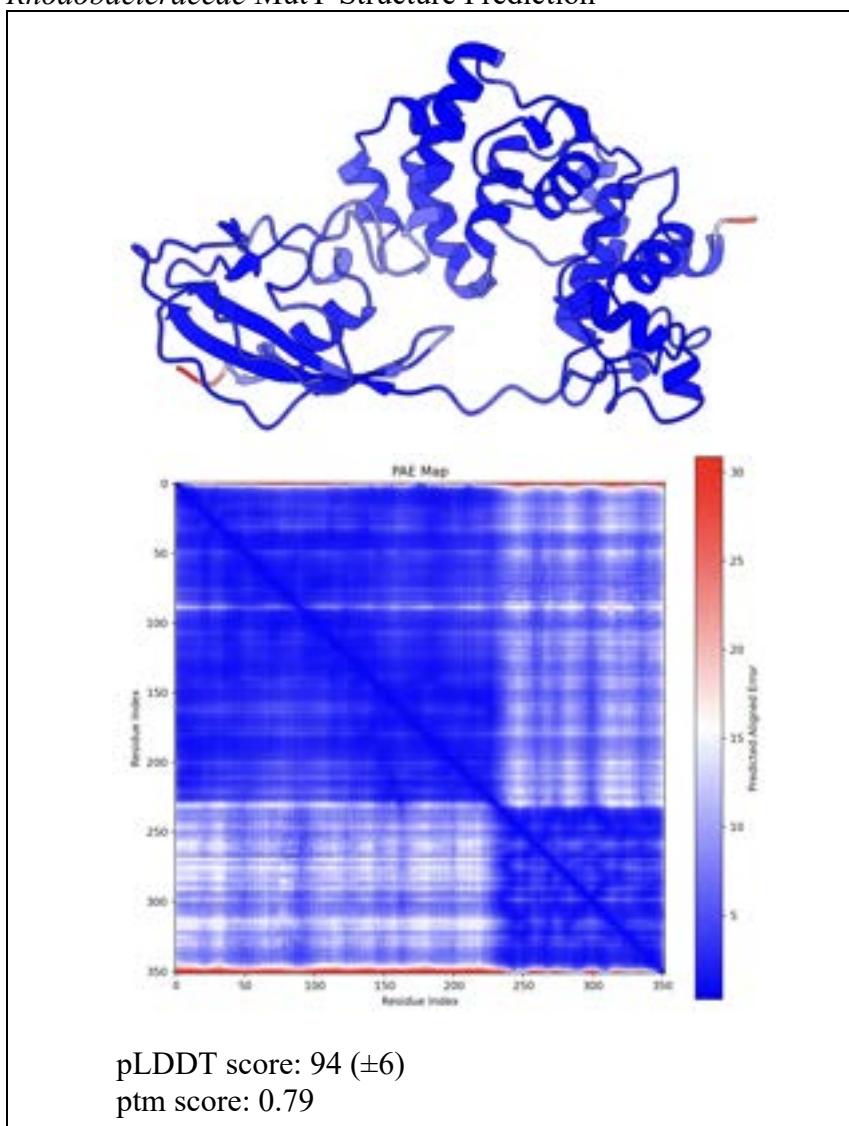
1168 *Marinosulphonomonas* MutY Structure Prediction



1169

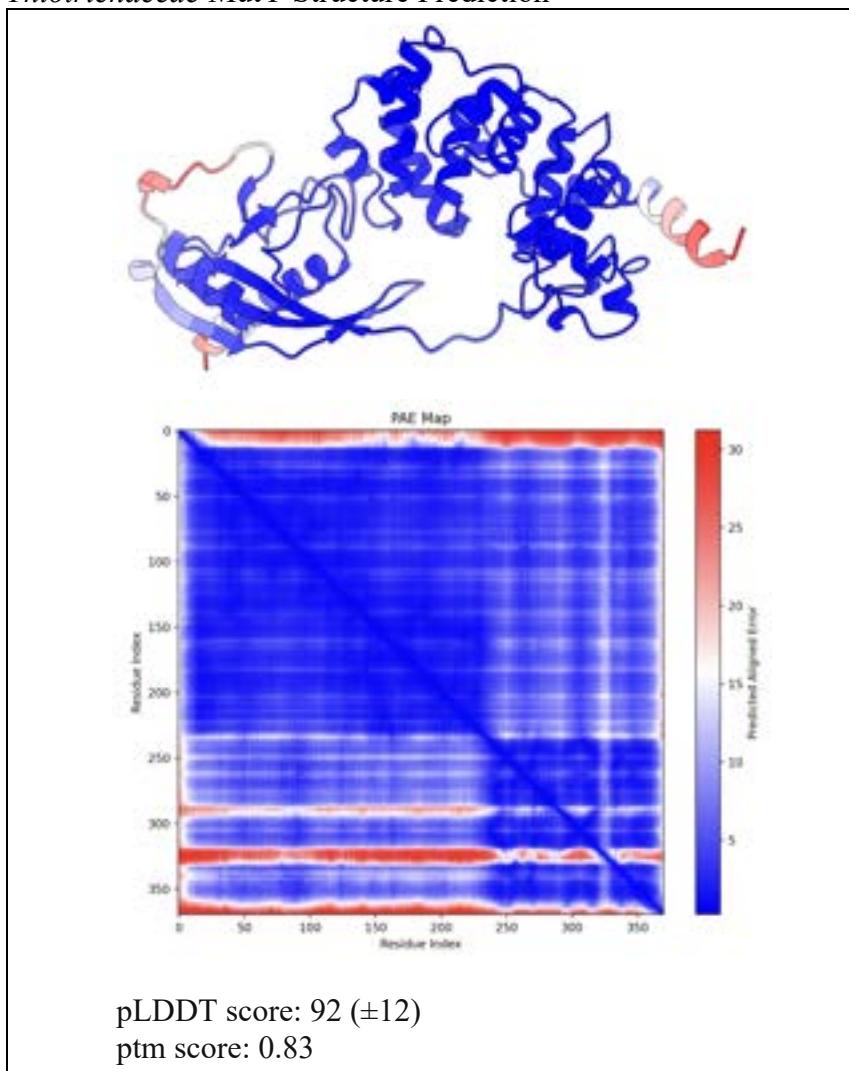
1170

1171 *Rhodobacteraceae* MutY Structure Prediction



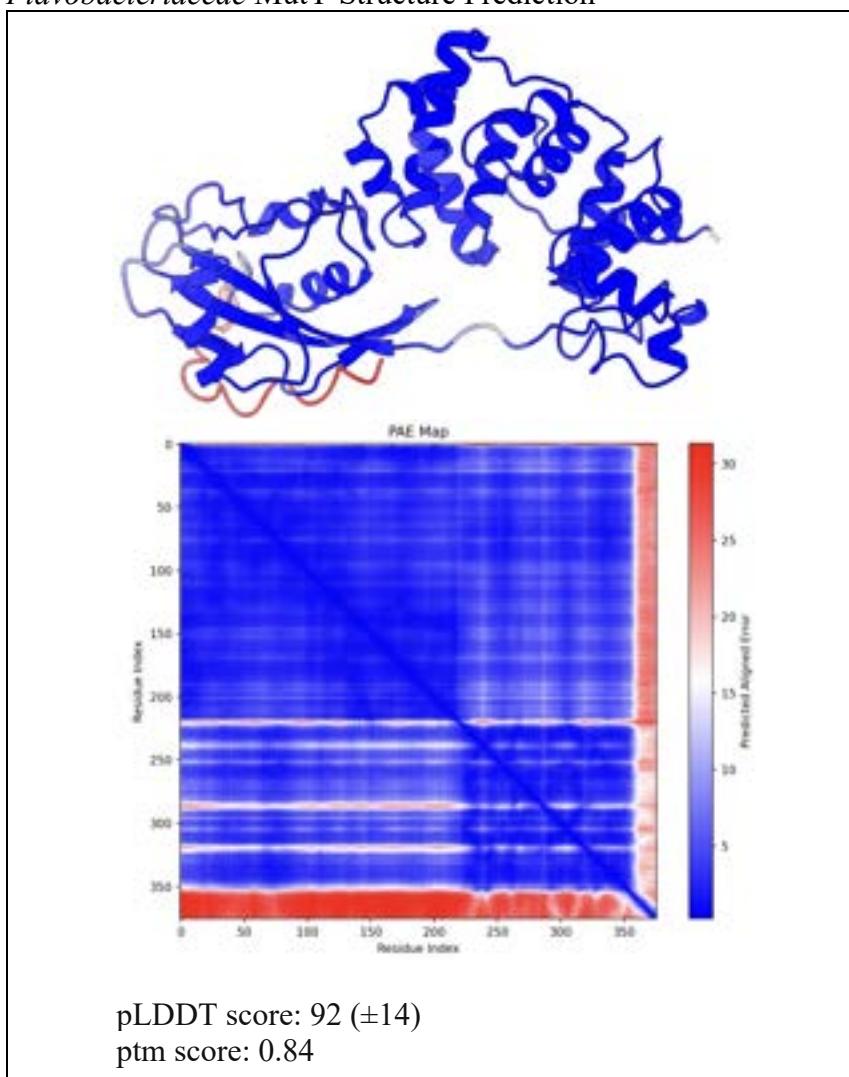
1172

1173 *Thiotrichaceae* MutY Structure Prediction



1174

1175 *Flavobacteriaceae* MutY Structure Prediction



1176

1177
1178

S6 Table. Ligand binding affinity (kcal / mol)

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

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

1179
1180
1181
1182
1183
1184
1185

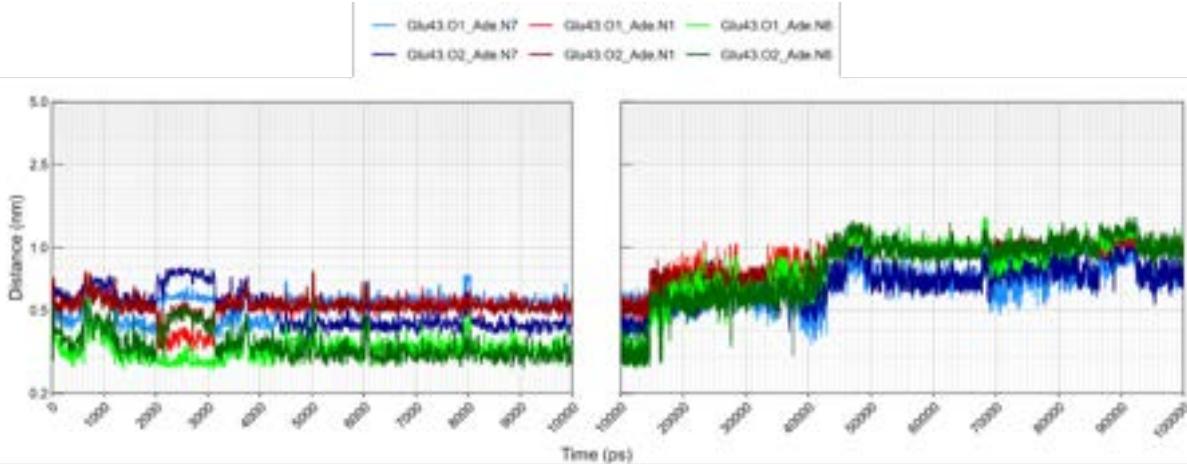
*Binding affinities are reported for the binding modes generated by *AutoDock VINA*. Each mode represents a predicted ligand pose, which differs by a combination of position, orientation, and rotamer conformation. The receptor structure was obtained from PDB ID 3g0q for Gs MutY and through structure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

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 Leonard-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

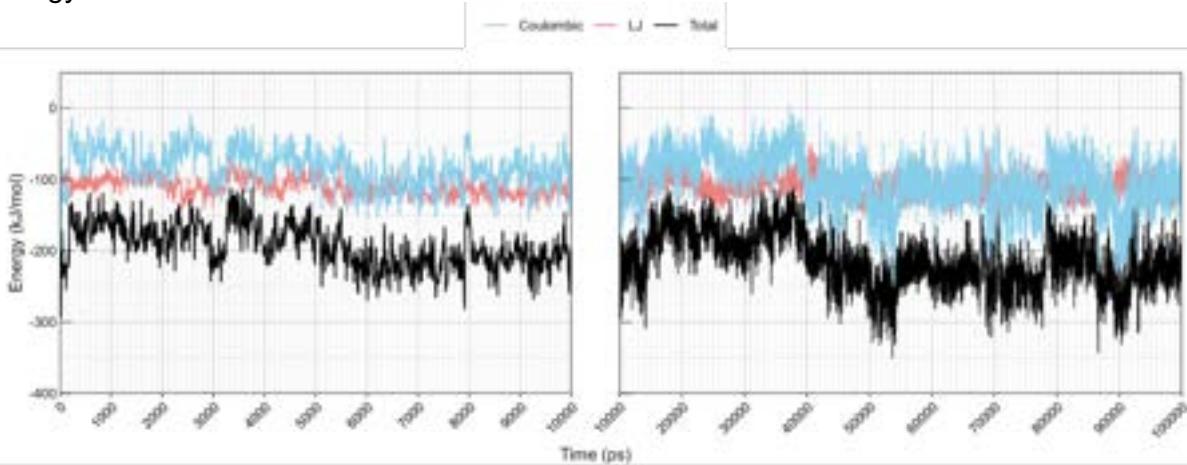
1232 hydrogen bonds between Thr299 and OG observed for the initial starting structure were lost as
1233 the ligand moved to new positions on the NTD and CTD before dissociation. (G) Molecular
1234 dynamic simulation for *Thiotrichaceae* MutY NTD complexed with adenosine. Note that Ser
1235 replaces active 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

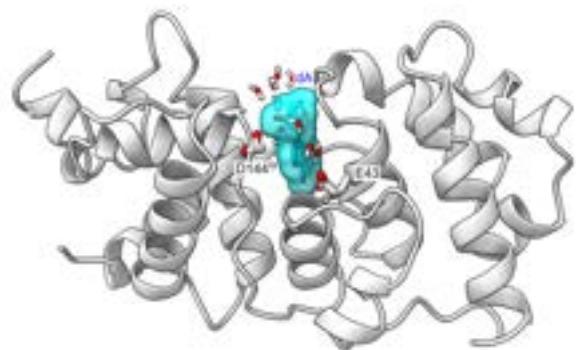


Energy



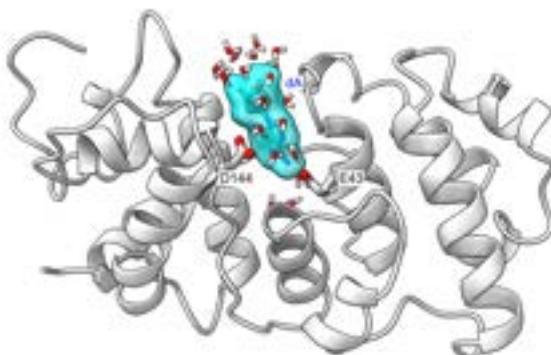
Adenosine and *G. stearothermophilus* MutY NTD

0,000 ps



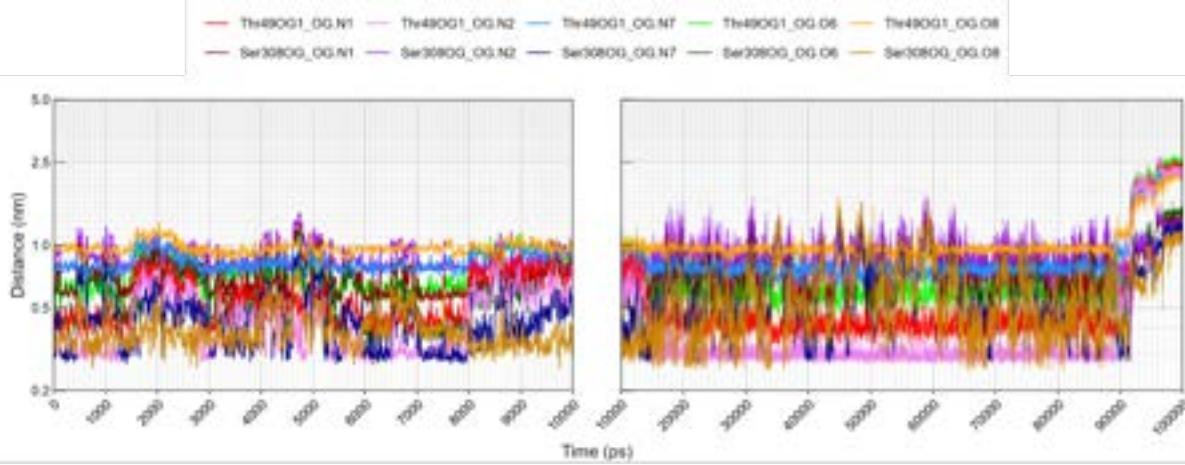
Adenosine and *G. stearothermophilus* MutY NTD

10,000 ps

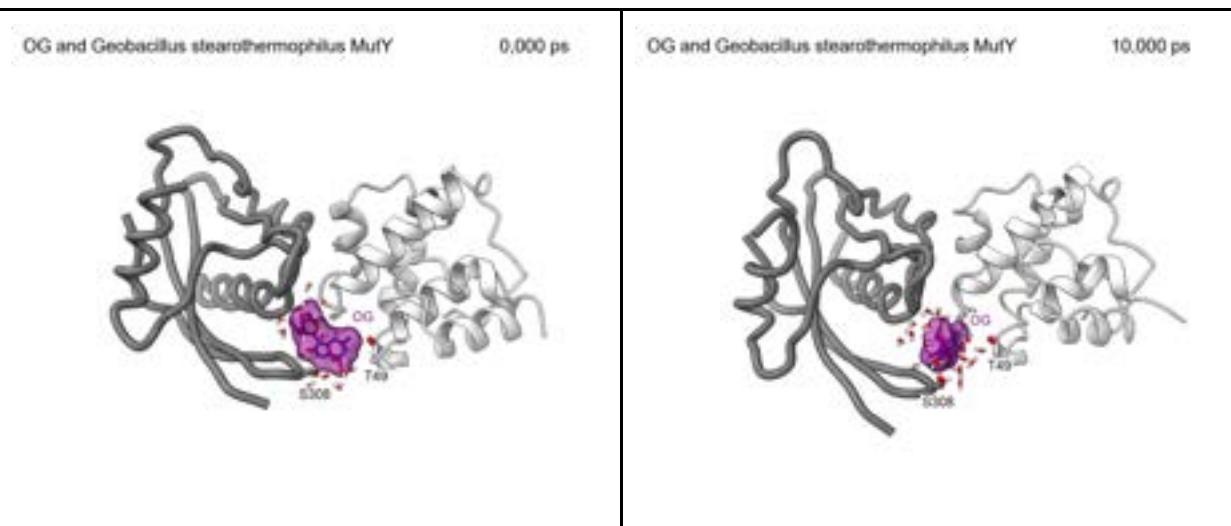
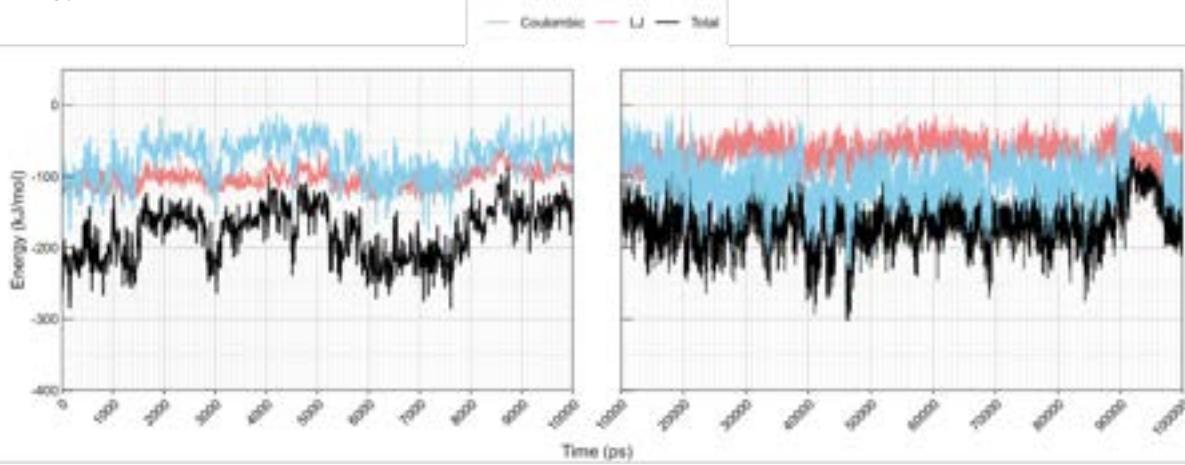


(B) *Geobacillus stearothermophilus* MutY complexed with OG

Distance OG - Thr49 and Ser308

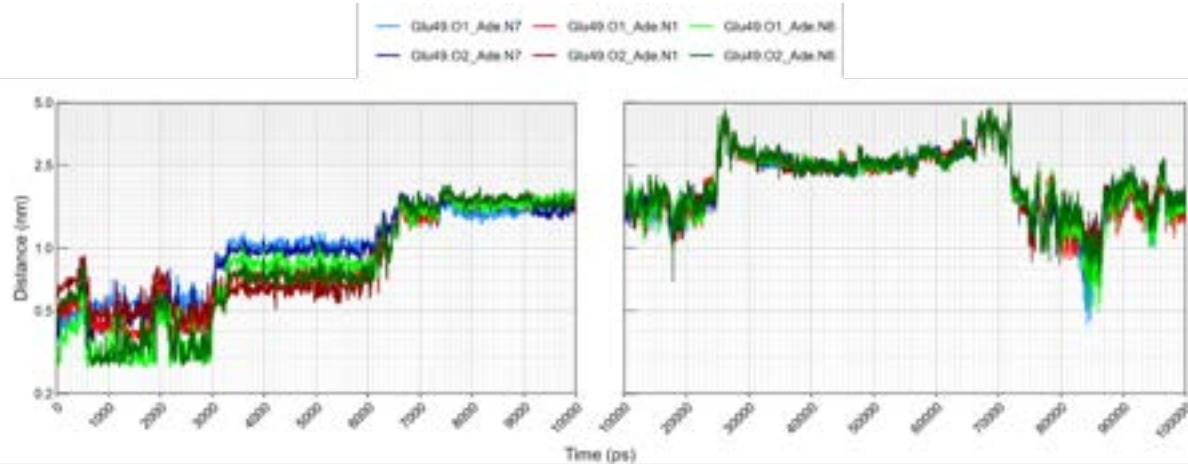


Energy

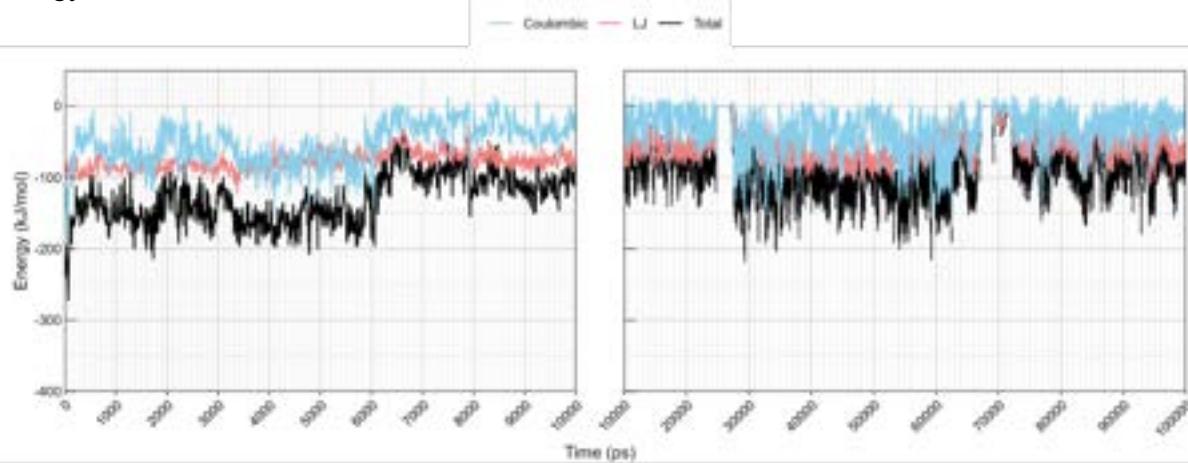


(C) *Marinosulfonomonas* MutY NTD complexed with adenosine

Distance Adenosine - Glu49

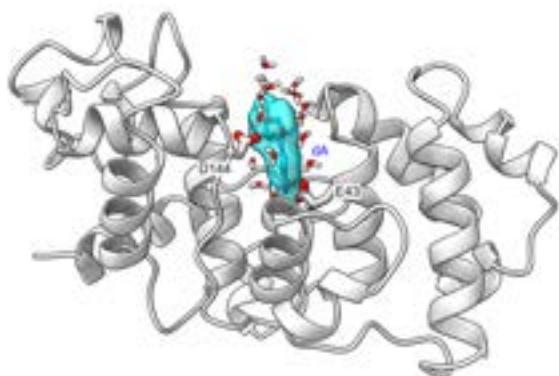


Energy



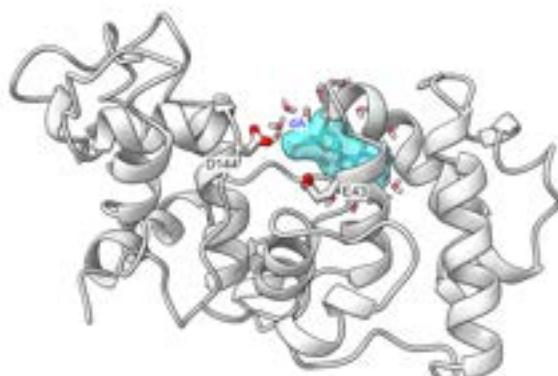
Adenosine and *Marinosulfonomonas* MutY NTD

0,000 ps



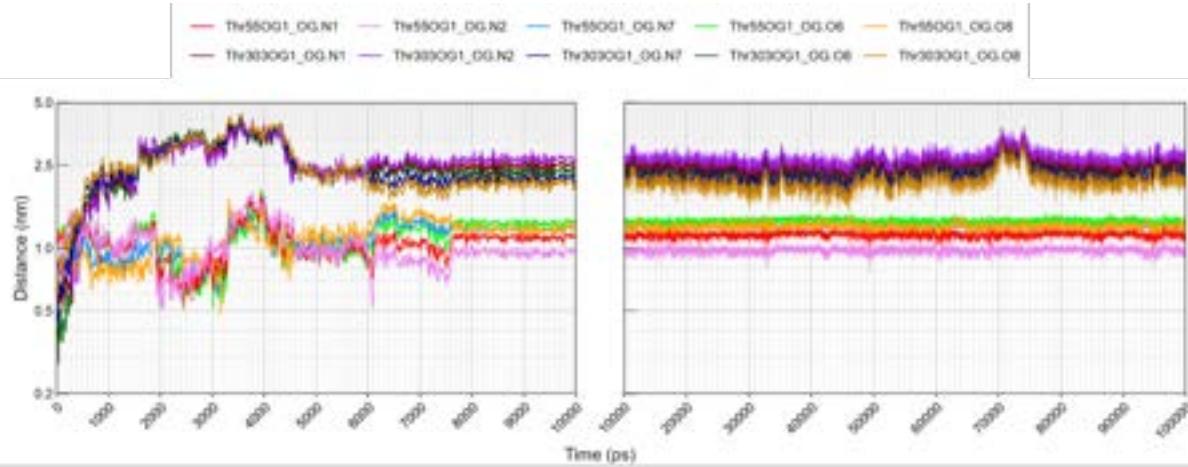
Adenosine and *Marinosulfonomonas* MutY NTD

10,000 ps

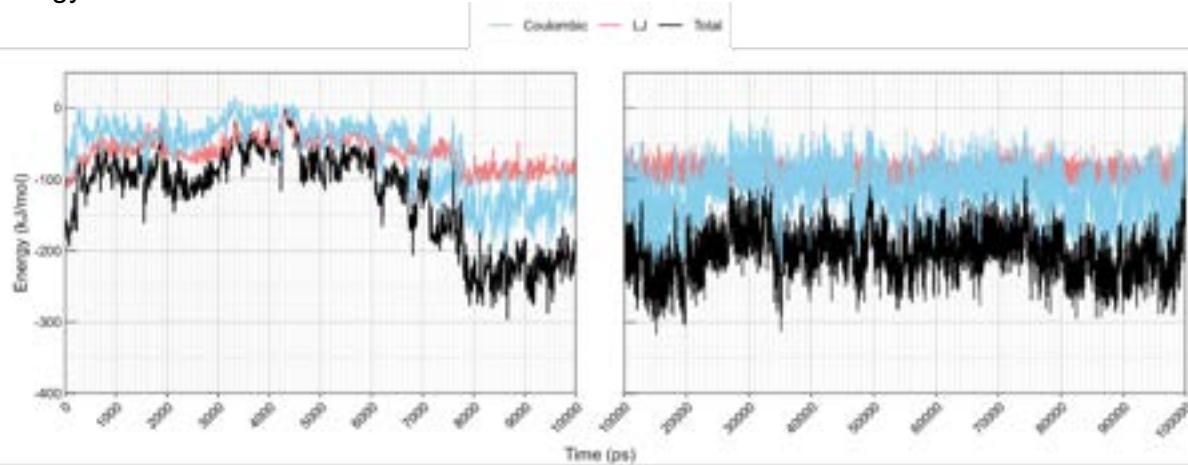


(D) *Marinosulfonomonas* MutY complexed with OG

Distance OG - Thr55 and Thr303

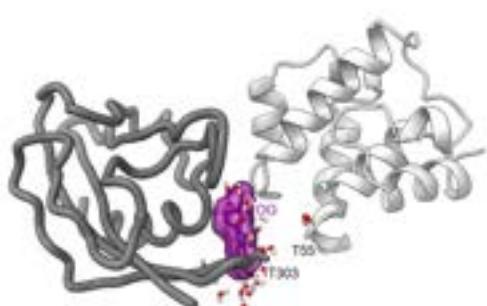


Energy



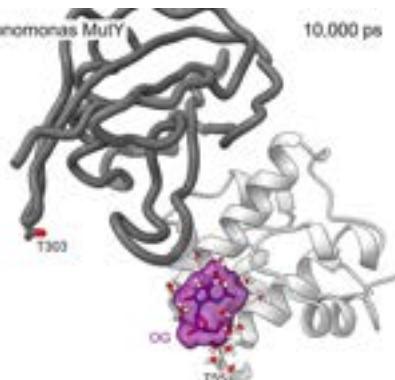
OG and *Marinosulfonomonas* MutY

0,000 ps



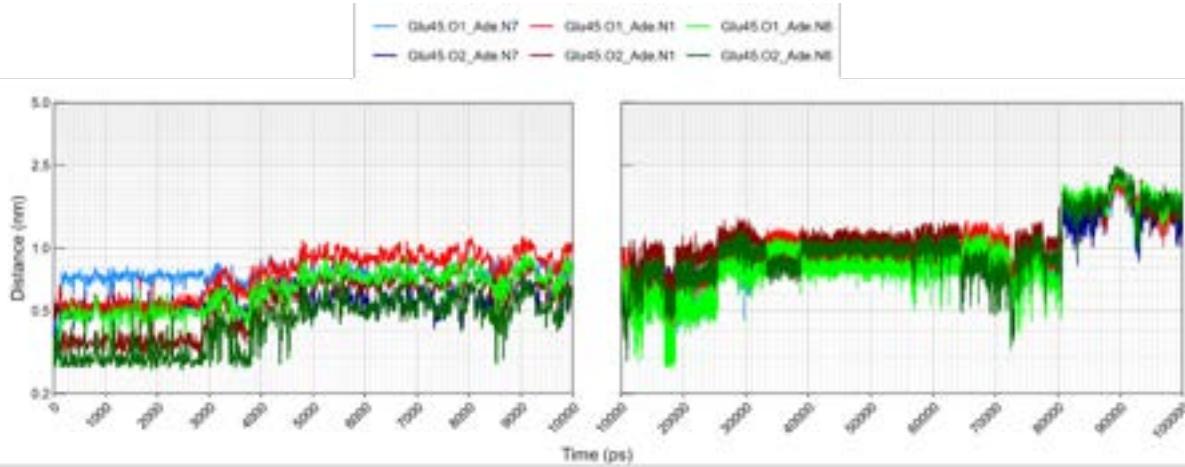
OG and *Marinosulfonomonas* MutY

10,000 ps

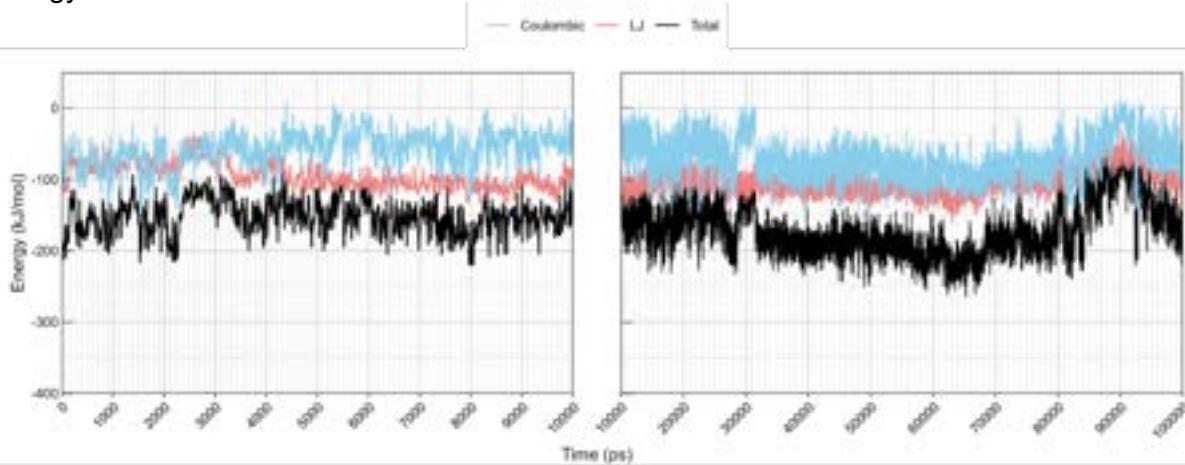


(E) *Rhodobacteraceae* MutY NTD complexed with adenosine

Distance Adenosine - Glu45



Energy



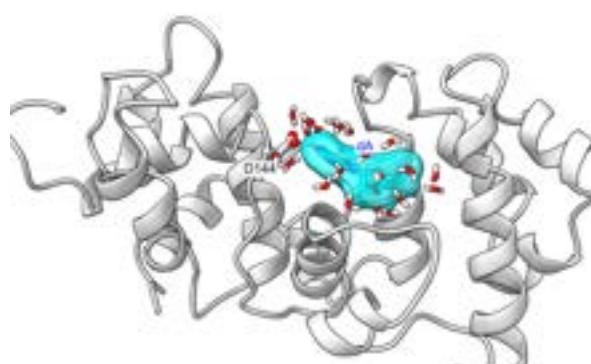
Adenosine and *Rhodobacteraceae* MutY NTD

0,000 ps



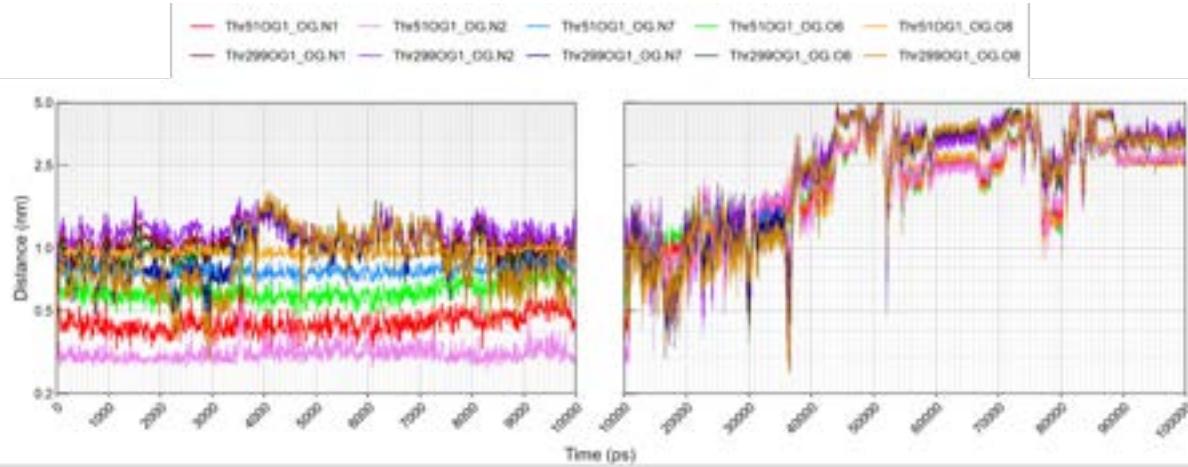
Adenosine and *Rhodobacteraceae* MutY NTD

10,000 ps

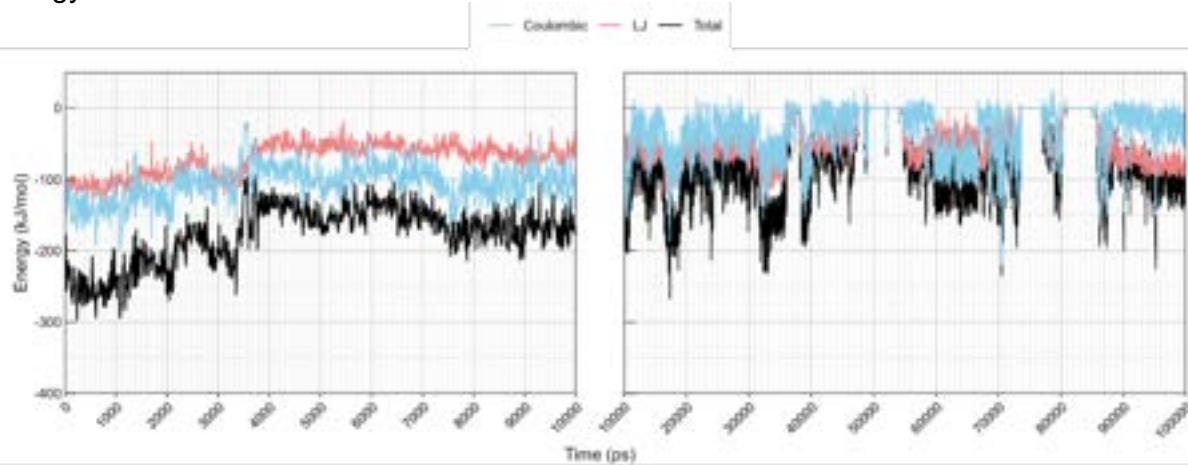


(F) *Rhodobacteraceae* MutY complexed with OG

Distance OG - Thr51 and Thr299

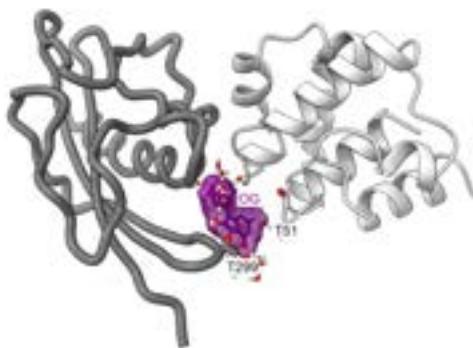


Energy



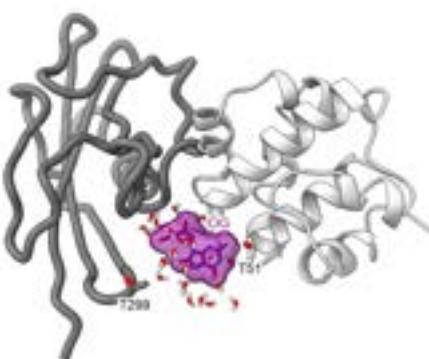
OG and *Rhodobacteraceae* MutY

0,000 ps



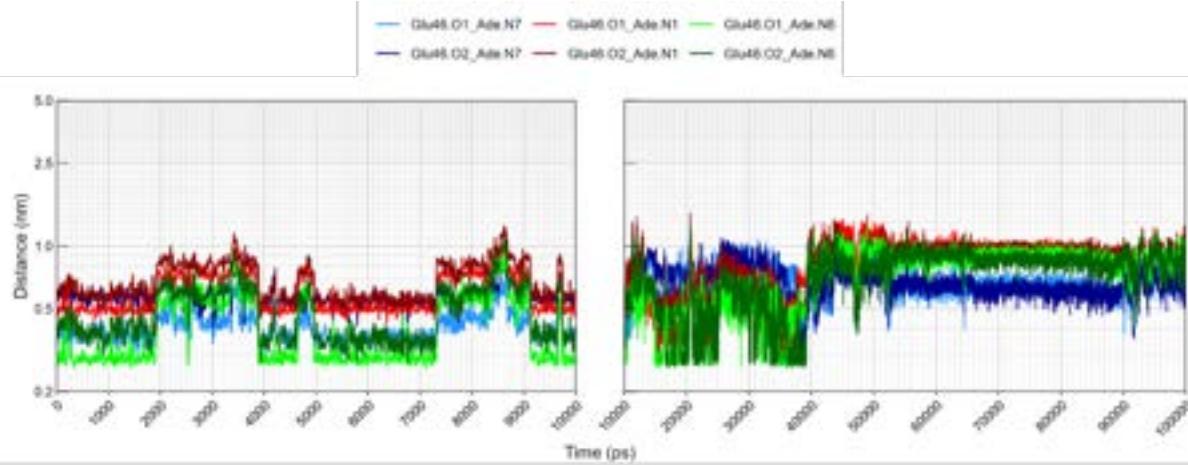
OG and *Rhodobacteraceae* MutY

10,000 ps

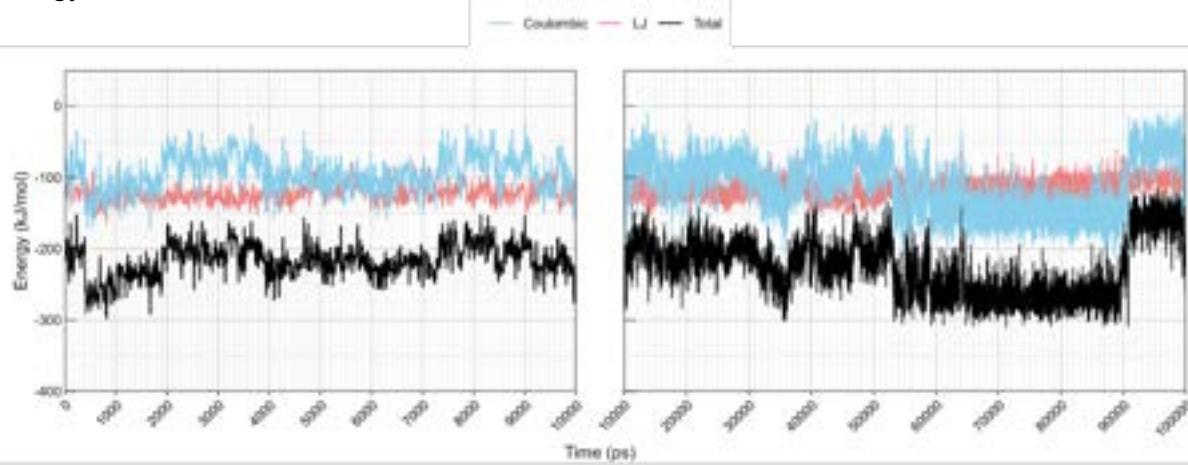


(G) *Thiotrichaceae* MutY NTD complexed with adenosine

Distance Adenosine - Glu46

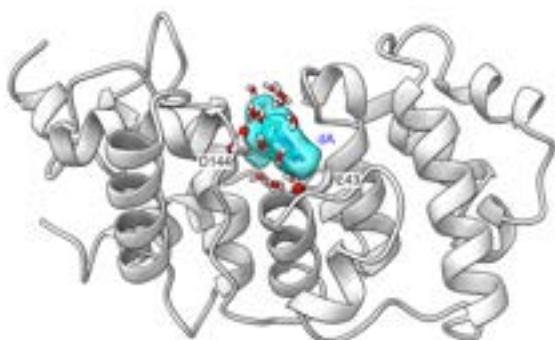


Energy



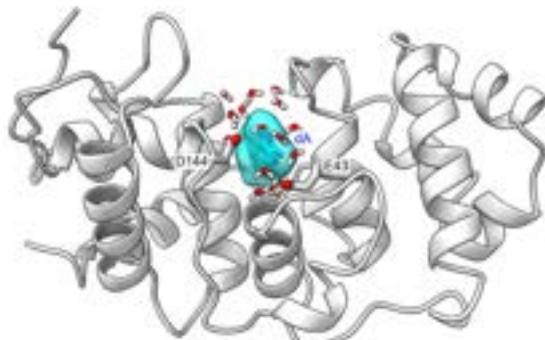
Adenosine and *Thiotrichaceae* MutY NTD

0,000 ps



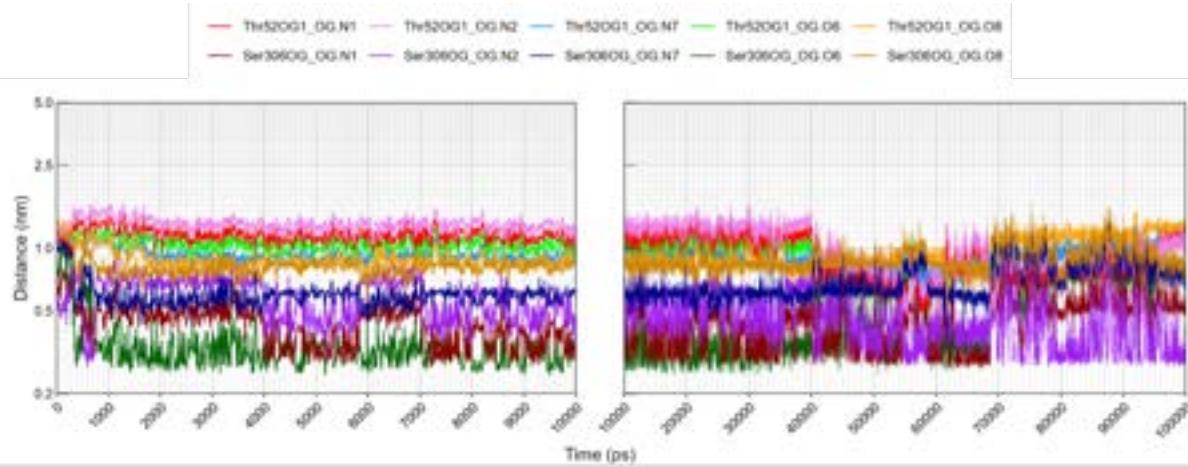
Adenosine and *Thiotrichaceae* MutY NTD

10,000 ps

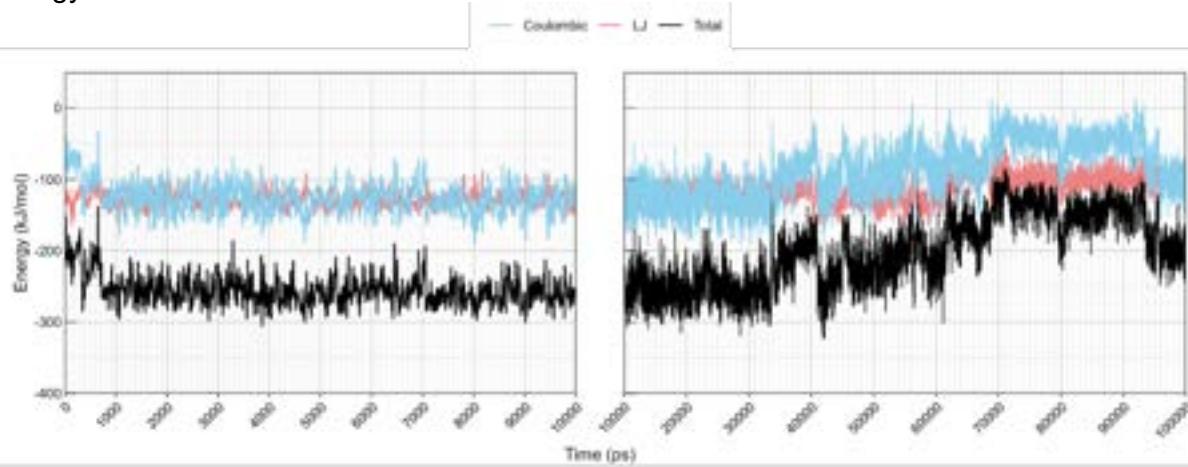


(H) *Thiotrichaceae* MutY complexed with OG

Distance OG - Thr52 and Ser306

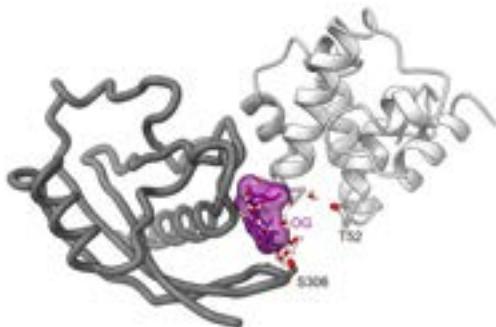


Energy



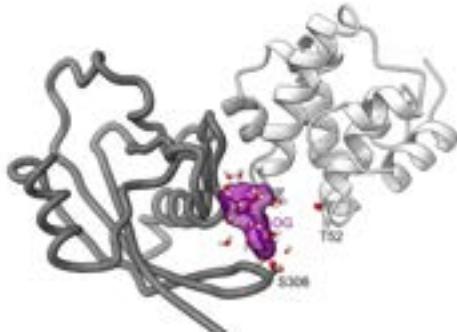
OG and *Thiotrichaceae* MutY

0,000 ps



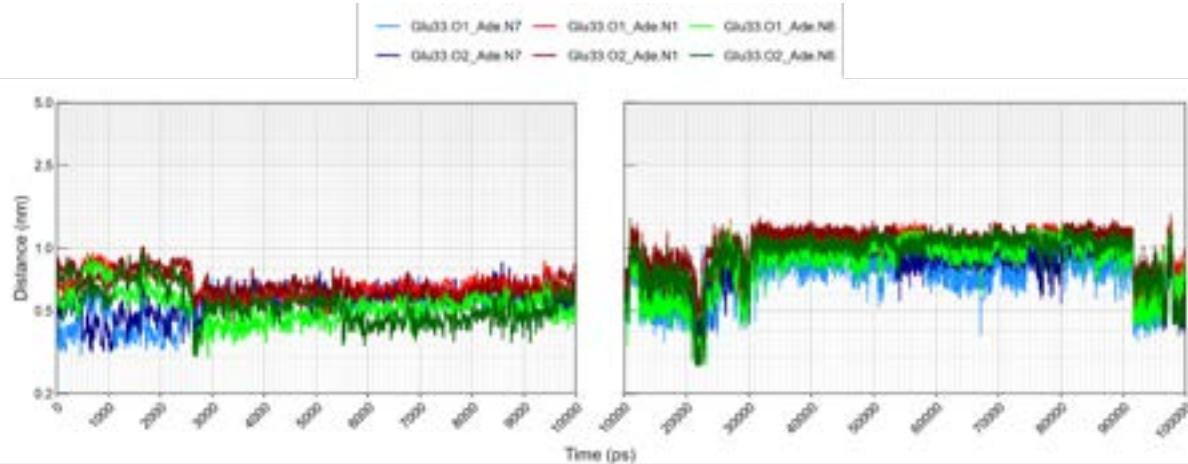
OG and *Thiotrichaceae* MutY

10,000 ps

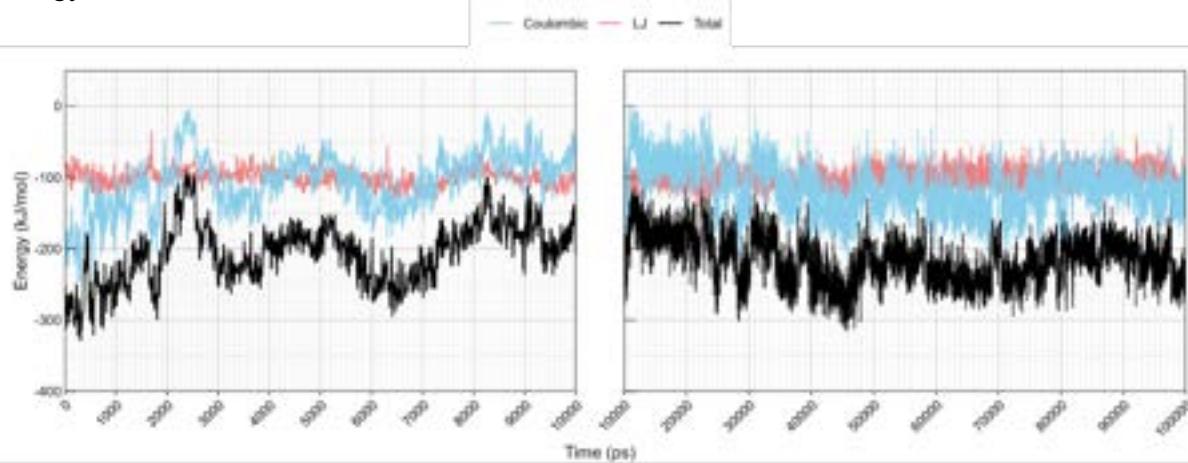


(I) *Flavobacteriaceae* MutY NTD complexed with adenosine

Distance Adenosine - Glu33

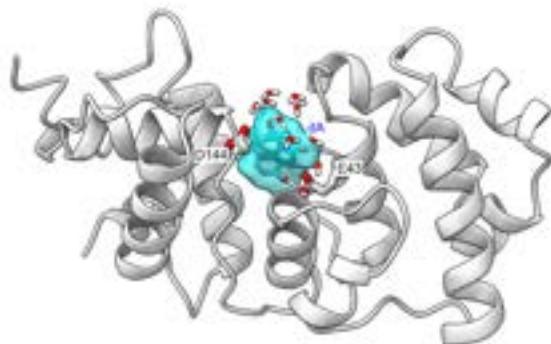


Energy



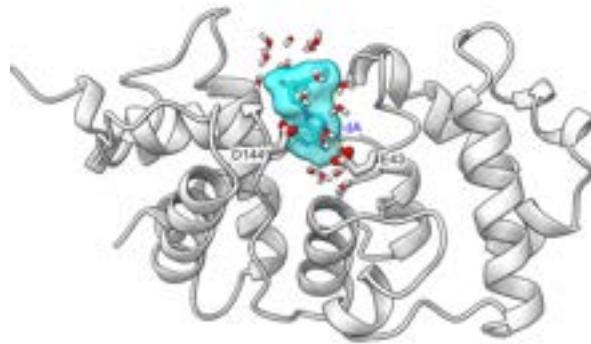
Adenosine and *Flavobacteriaceae* MutY NTD

0,000 ps



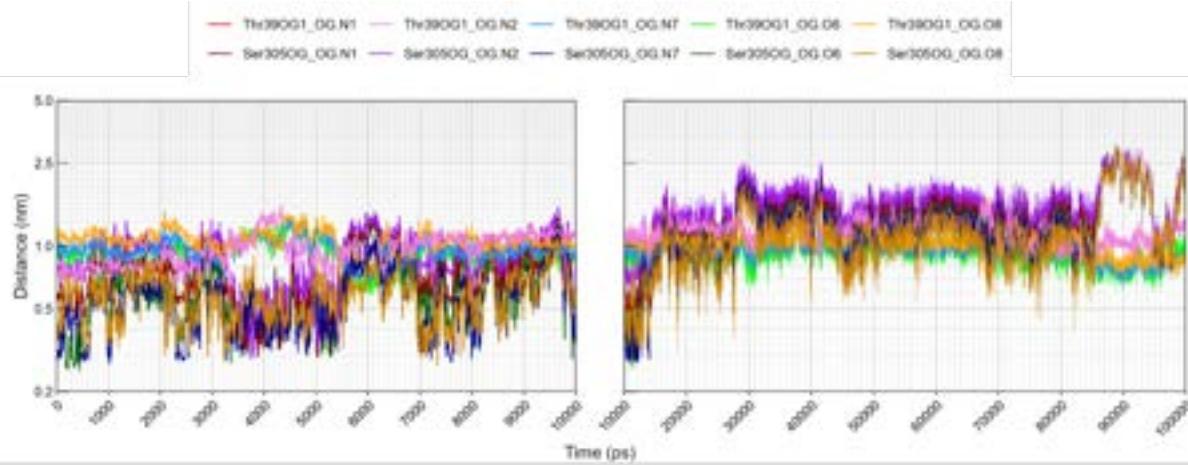
Adenosine and *Flavobacteriaceae* MutY NTD

10,000 ps

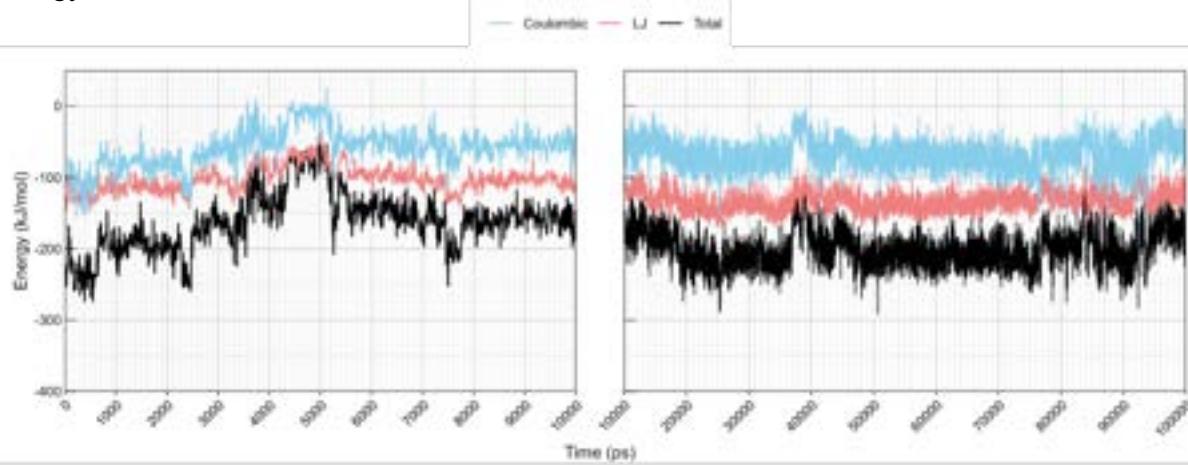


(J) *Flavobacteriaceae* MutY complexed with OG

Distance OG - Thr39 and Ser305

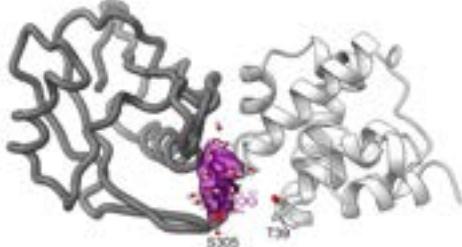


Energy



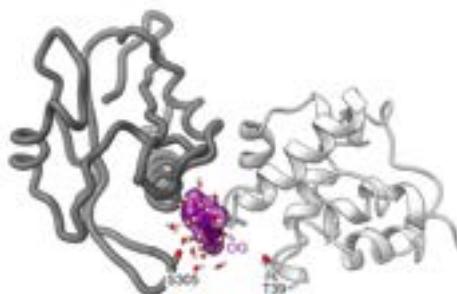
OG and *Flavobacteriaceae* MutY

0.000 ps



OG and *Flavobacteriaceae* MutY

10.000 ps



1264
1265 **S8 Movies. Molecular animations.** Structures for each MD trajectory were sampled at 200-ps
1266 intervals from 0 – 10,000 ps and at 1,000-ps intervals from 10,000 – 100,000 ps and movies were
1267 recorded with *ChimeraX*. Residues belonging to the NTD are depicted with a traditional ribbon
1268 style cartoon and colored light gray. Residues belonging to the CTD are depicted with a licorice
1269 cartoon style and colored dark gray. Solvent molecules that are within 4 Å of both the ligand and
1270 protein are shown (O, red; H, white). Each movie highlights particular features and events with
1271 time paused, the scene rotating about the y axis, and a brief caption. Links to view and download
1272 each movie are provided.
1273
1274 (A) Molecular animation for *Gs* MutY NTD complexed with adenosine. Adenosine remains
1275 within the active site pocket throughout the entire 100,000-ps simulation. At ~45,000 ps
1276 adenosine rotates within the active site to place its sugar in close proximity to the catalytic Glu43
1277 residue, demonstrating a limited degree of flexibility within the active site pocket. Link to view
1278 movie: <https://youtu.be/CO-ljoafn1Q>
1279
1280
1281 (B) Molecular animation for *Gs* MutY complexed with OG. OG remains wedged between NTD
1282 and CTD for most of the trajectory. Interactions with the Hoogsteen and Watson Crick face of
1283 OG relevant for OG recognition are highlighted at pauses. A new pose emerges at 90,000 ps just
1284 prior to departure of the ligand from the NTD-CTD interface. Link to view movie:
1285 <https://youtu.be/sVujlyoViRU>
1286
1287 (C) Molecular animation for *Marinosulfonomonas* MutY NTD complexed with adenosine.
1288 Adenosine slips back toward the entrance of the active site pocket within 1ns and exits
1289 completely by ~6,000-7,000 ps. It then settles in a pocket on the surface of the protein defined by
1290 the loop containing Ser24 and a helix from Ala58 to His65. It remains there until ~25,000 ps
1291 when it begins to move freely in the solvent, engaging, disengaging, then re-engaging with the
1292 surface of the protein for the remainder of the 100,000-ps simulation. Link to view movie:
1293 <https://youtu.be/8RAHtBEzjTA>
1294
1295 (D) Molecular animation for *Marinosulfonomonas* MutY complexed with OG. The two domains
1296 adopt a different disposition with a new inter-domain interface early in the simulation. The OG
1297 ligand finds two new binding sites on the NTD, each distinct from the original binding site, and
1298 persists complexed with the NTD until the end of the 100,000-ps simulation. Link to view
1299 movie: <https://youtu.be/IUCuN82XJ-U>
1300
1301 (E) Molecular animation for *Rhodobacteraceae* MutY NTD complexed with adenosine. Similar
1302 to the *Marinosulfonomonas* MutY NTD simulation, adenosine is completely outside the active
1303 site pocket relatively early in the simulation by ~5,000 ps. It then settles on the surface of the
1304 protein and wedges into a groove with residues Gly126 and Tyr128 on one side and Gln49 and
1305 Arg93 on the other side, and remains at this binding site for the rest of the 100,000-ps simulation.
1306 Link to view movie: https://youtu.be/qvgICZj6k_o
1307
1308 (F) Molecular animation for *Rhodobacteraceae* MutY complexed with OG. The animation
1309 features a highly dynamic OG-MutY complex that dissociates completely by 48,000 ps. The OG

1310 ligand disengages from functionally relevant interactions at the NTD-CTD interface to find a
1311 new site on the NTD by 4,400 ps, nearly escapes at 13,000 ps, and samples several alternate sites
1312 on the NTD or the CTD or at a new site at the NTD-CTD interface prior to exiting this region
1313 and exploring new sites on the surface of the NTD. The molecular animation is discontinued at
1314 48,000 ps with the complex dissociated. The NTD-CTD structure remains intact for the
1315 remainder of the 100,000-ps simulation but the OG ligand did not rebind (not shown). Link to
1316 view movie: <https://youtu.be/M9N8OXkGre4>
1317

1318 (G) Molecular animation for *Thiotrichaceae* MutY NTD complexed with adenosine. Adenosine
1319 remains in the active site pocket for the entire 100,000-ps simulation. Similarly to the *Gs* MutY-
1320 adenosine simulation, the ligand rotates within the active site pocket at ~43,000 ps to place its
1321 sugar within close proximity of the active site Glu46 residue, demonstrating limited flexibility
1322 within the active site pocket. Link to view movie: https://youtu.be/WyWPjDp_pqI
1323

1324 (H) Molecular animation for *Thiotrichaceae* MutY complexed with OG. The complex persists
1325 for the entire 100,000-ps simulation. The initial complex features interaction of Ser305 with the
1326 Watson-Crick-Franklin face of the OG base. This pose persists until transition to a new pose at
1327 ~69,000 ps with the deoxyribose sugar closer to Ser305 and the base wedged between two
1328 helices that converge at the NTD-CTD interface. Link to view movie:
1329 <https://youtu.be/B64q0FWSrBs>

1330 (I) Molecular animation for *Flavobacteriaceae* MutY NTD complexed with adenosine.
1331 Adenosine remains within the active site pocket for the entire 100,000 ps, It starts with its sugar
1332 facing the catalytic Glu33 residue. At ~3,000 ps it rotates to bring the base portion deeper within
1333 the active site. It remains in this general orientation for the remainder of the 100,000 ps with the
1334 sugar engaging Glu33 in the ~60,000 – 80,000-ps time window. Link to view movie:
1335 <https://youtu.be/X7az3V8-wMY>
1336

1337 (J) Molecular animation for *Flavobacteriaceae* MutY complexed with OG. During the first 9,800
1338 ps, Ser305 makes hydrogen bonds with the Hoogsteen face of OG in a manner relevant for
1339 recognition. At 10,000 ps, a new pose emerges with the base wedged between helices in the
1340 NTD and CTD and thus removed from the FSH recognition loop. The complex with this new
1341 pose persists for the remainder of the 100,000-ps simulation. Link to view movie:
1342 <https://youtu.be/FJuTrgaGxNQ>
1343

1344
1345
1346
1347

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

1350

1351 ^aConfidence intervals (95%) determined by a bootstrap method, see **Materials and methods** for details.

1352 ^bMutation frequency reported as median number of resistant colonies per 10⁸ viable colonies. Fold change was
1353 calculated by dividing Rif^R frequency by the frequency measured for cultures expressing *Ec* MutY.

1354