

**Supplementary online material**

**Uracil DNA glycosylase (UDG) activities in *Bradyrhizobium diazoefficiens*, a G+C rich nitrogen fixing symbiont and structural and functional characterization of a new class of UDG, *BdiUng* having broad substrate specificity**

Ullas Valiya Chembazhi<sup>1,#</sup>, Vinod Vikas Patil<sup>2,3,#</sup>, Shivjee Sah<sup>1</sup>, Wayne Reeve<sup>4</sup>, Ravi P. Tiwari<sup>4</sup>, Euijeon Woo<sup>2,3\*</sup>, and Umesh Varshney<sup>1,4,5,\*</sup>

<sup>1</sup>Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore, 560012, India, <sup>2</sup>Functional Genomic Research Center, Korea Research Institute of Bioscience and Biotechnology, 125 Gwahak-Ro, Yuseon-Gu, Daejeon, 34141, Republic of Korea, <sup>3</sup>Department of Bio-Analytical Science, University of Science and Technology, Daejeon 34113, Republic of Korea, <sup>4</sup>Centre for Rhizobium Studies, School of Veterinary and Life Sciences, Murdoch University, Murdoch WA 6150, and <sup>5</sup>Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, 560064, India

<sup>#</sup>Contributed equally to this work

<sup>\*</sup>Joint corresponding authors

### **Generation of *BdiUng* A59Y mutation its purification and activity assays**

PCR based quick-change mutagenesis was performed to generate A59Y mutation in *BdiUng*. Primer extension reactions were carried out taking 50 ng of template (pTrc*BdiUng*) and 10 pmol of the forward and reverse primers (Table S2) in a 20 µl reaction with Pfu DNA polymerase. The reaction was heated at 94 °C for 4 min followed by 18 cycles of heating for 1 min at 95 °C, 45 seconds at corresponding annealing temperature and 10 min at 70 °C. Final extension was also carried out at 70 °C for 10 min. An aliquot (8 µl) from the reaction was digested with 1 unit of DpnI (NEB) for 14 h in NEB cut smart buffer and transformed into *E. coli* TG1. Plasmids were isolated and restriction digested from colonies that were ampicillin resistant. The mutation was ascertained through sequencing. The mutant protein was purified from *E. coli* TG1Δ*Ung*::*cm* strain using Ni-NTA column chromatography as described for the wild type *BdiUng* and subsequently assayed for base excision activities as described.

### **Supplementary references:**

1. Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, **4**, 406-425.
2. Dopazo, J. (1994) Estimating errors and confidence intervals for branch lengths in phylogenetic trees by a bootstrap approach. *Journal of molecular evolution*, **38**, 300-304.
3. Rzhetsky, A. and Nei, M. (1992) A Simple Method for Estimating and Testing Minimum-Evolution Trees. *Molecular biology and evolution*, **9**, 945.
4. Nei, M. and Kumar, S. (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press.
5. Kumar, S., Stecher, G. and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution*, **33**, 1870-1874.

**Table S1:** List of strains and plasmids used.

Strain/plasmid	Details	Reference
<i>E. coli</i> CC102	F', <i>ara-600</i> , $\Delta(gpt-lac)5$ , $\lambda^+$ , <i>relA1</i> , <i>spoT1</i> , <i>thiE1</i> , F128-(used to screen for C to T or G to A mutation)	Cupples and Miller(1989)
<i>E. coli</i> TG1	K-12 <i>supE thi-1</i> $\Delta(lac-proAB)$ $\Delta(mcrB-hsdSM)5$ , ( $r_K^- m_K^-$ )	Sambrook, J. et. and Fritsch
<i>E. coli</i> MG1655 $\Delta ung::kan$	MG1655 containing $\Delta ung::kan$ (kan cassette inserted through BamHI site of <i>ung</i> )	Bharti, S.K. and Varshney U (2010)
<i>E. coli</i> CC102 $\Delta ung::kan$	<i>E. coli</i> CC102 strain containing $\Delta ung::kan$ from MG1655 $\Delta ung::kan$	This study
<i>E. coli</i> TG1 $\Delta ung::Cm$	<i>E. coli</i> TG1 containing $\Delta ung::Cm$	Lab stock
<i>Bradyrhizobium diazoefficiens</i> USDA 110	Strain available at Murdoch University, Murdoch, Australia.	<a href="http://www.ncbi.nlm.nih.gov/pubmed/25685260">http://www.ncbi.nlm.nih.gov/pubmed/25685260</a> and <a href="http://www.nature.com/articles/srep16825">http://www.nature.com/articles/srep16825</a>
pTrc99c	<i>E. coli</i> expression vector ( $Amp^r$ )	Amersham
pTrcNdeIHis	A pTrc99C derivative wherein the original NdeI site was mutated by digesting it with NdeI and filling in the ends with T4 DNA polymerase followed by self-ligation. Subsequently, NcoI-HindIII fragment is replaced by NcoI-HindIII fragment from pET14b so as to include a sequence for hexaHis tag	Lab stock
pTrcBdiUng	pTrc99c with <i>Bdiung</i> ORF cloned in its NdeI/HindIII sites	This study
pTrcEcoUng	pTrc99c harbouring <i>EcoUng</i> ORF under lac promoter	Kumar, N. V. and Varshney, U.
pTrcBdiUng A59Y	pTrcBdiUng with <i>BdiUngA59Y</i> mutation	This study
pTrcBll3023	pTrcNdeIHis with <i>Bll3023</i> ORF cloned in NdeI/HindIII sites	This study
pTrcBlr5068	pTrcNdeIHis with <i>Blr5068</i> ORF cloned in NdeI/HindIII sites	This study
pTrcBlr6661	pTrcNdeIHis with <i>Blr6661</i> ORF cloned in NdeI/HindIII sites	This study

**Table S2:** List of DNA oligomers used.

Oligomer	Size	Sequence	Remarks
ssU9	24	ctcaagtgUaggcatgcaagagct	ssDNA with U at position 9
GU9	37	ctcaagtgUaggcatgcttttgcgatgcctgcaactga	G:U pair in stem at position 9 of a stem loop structure
HmU	24	ctcaagtg(HmU)aggcatgcaagagct	ssDNA containing 5-Hydroxymethyl uracil at position 9
ssX9	24	ctcaagtg(X)aggcatgcaagagct	ssDNA containing xanthine at position 9
ssF9	24	ctcaagtgFaggcatgcaagagct	ssDNA with abasic site at position 9 (A stable AP DNA mimic of ssU9)
ssU9-Compl-G	24	cttgcctgcctgcaacttgagtga	Complementary oligonucleotides used in annealing reactions
Compl-T	26	ggcacttcgcctttagcagccagtc	
Compl-G	26	ggcacttcgcctgtagcagccagtc	
G-oligo	30	cgatcatggagccacagagctcccgttacag	
I	26	gactggctgctac(I)aggcgaagtgcc	Inosine at position 14
Ecoung-up-Fp	18	tggataacctgtaatcatg	Forward and reverse primers flanking <i>ung</i> gene used for knockout confirmation
Ecoung-dn-Rp	18	ctcgatgattccgctcag	
<i>Bdiung</i> -Fp	25	gcatttcatatgctcacagaattcg	Forward and reverse primers with NdeI, HindIII sites respectively to amplify <i>blr0248</i> ORF
<i>Bdiung</i> -Rp	25	gcccaaagcttcgggttgccggccg	
U1-hairpin	22	ctagaggatcc Uttt ggatcct	U at 1st position of tetraloop hairpin
U2-hairpin	22	ctagaggatcc tUtt ggatcct	U at 2nd position of tetraloop hairpin
U3-hairpin	22	ctagaggatcc ttUt ggatcct	U at 3rd position of tetraloop hairpin
U4-hairpin	22	ctagaggatcc ttU ggatcct	U at 4th position of tetraloop hairpin
<i>BdiUngA59Y</i> Fp	25	gccaggatccctaccagcacgagac	Primers used for site directed mutagenesis
<i>BdiUngA59Y</i> Rp	25	gtctcgtgctggtagggatcctggc	
Bll3023 Fp	25	tcgagtcatatgtacctcgtcacc	Forward and reverse primers with NdeI, HindIII sites respectively to amplify <i>Bll3023</i> ORF
Bll3023 Rp	25	gatgacaagcttgggtgaaaggcg	
Blr6661 Fp	26	tgacctatgtcgtgacgctcatg	Forward and reverse primers with NdeI, HindIII sites respectively to amplify <i>Blr6661</i> ORF
Blr6661 Rp	28	ctcctaagcttgggtctatgggtgttg	
Blr5068 Fp	26	gagggcatatgacactttgaggagt	Forward and reverse primers with NdeI, HindIII sites respectively to amplify <i>Blr5068</i> ORF
Blr5068 Rp	25	aaggccaagctgtcaagactagt	
TTUTT	5	ttutt	Used for determining minimum substrate required for <i>BdiUng</i>
TUTT	4	tutt	
TTUT	4	ttut	
UTT	3	utt	

**Table S3:** Percentage (G+C) content of organisms in which homologues of *BdiUng* has been identified.

<b>Organism</b>	<b>% (G+C) content</b>	<b>Characteristic</b>
<i>Bradyrhizobium diazoefficiens</i>	64.1	Nitrogen fixer
<i>Frankia</i> sp. Eu11c	72.3	Nitrogen fixer
<i>Streptacidiphilus rugosus</i>	71	Acidophile
<i>Kitasatospora azatica</i>	71.6	Acidophile
<i>Nitrosospira</i> sp. NpAV	53.14	Nitrite oxidizer
<i>Intrasporangium calvum</i>	70.7	A type strain
<i>Streptomyces xylophagus</i>	69.8	Xylanase producer
<i>Janibacter</i> sp. HTCC2649	68.44	Marine
<i>Chromobacterium haemolyticum</i>	62.84	Human pathogen

**Table S4:** Sequences considered for constructing evolutionary tree

No.	Protein	Secondary Accessions & Ids	Species
1	UNG_ECOLI	P12295	Escherichia coli (strain K12)
2	UNG_SHIDS	Q32CT9	Shigella dysenteriae serotype 1 (strain Sd197)
3	UNG_CITK8	A8AD07	Citrobacter koseri (strain ATCC BAA-895 / CDC 4225-83 / SGSC4696)
4	V5CNW9_ENTCL	V5CNW9	Enterobacter cloacae S611
5	D2TU36_CITRI	D2TU36	Citrobacter rodentium (strain ICC168)
6	E3G7U3_ENTLS	E3G7U3	Enterobacter lignolyticus (strain SCF1)
7	W1DYL0_KLEPN	W1DYL0	Klebsiella pneumoniae IS46
8	UNG_KLEP7	A6TCJ2	Klebsiella pneumoniae subsp. pneumoniae (strain ATCC 700721 / MGH 78578)
9	W1H494_KLEPN	W1H494	Klebsiella pneumoniae ISC21
10	W1DPM6_KLEPN	W1DPM6	Klebsiella pneumoniae IS43
11	W1HRH7_KLEPN	W1HRH7	Klebsiella pneumoniae IS39
12	UNG_SALTY	P67073	Salmonella typhimurium (strain LT2 / SGSC1412 / ATCC 700720)
13	UNG_SALAR	A9MGV9	Salmonella arizonae (strain ATCC BAA-731 / CDC346-86 / RSK2980)
14	A0A090VYD8_ESCVU	A0A090VYD8	Escherichia vulneris NBRC 102420
15	A0A089PQ15_PLUGE	A0A089PQ15	Pluralibacter gergoviae
16	UNG_ENT38	A4WDE9	Enterobacter sp. (strain 638)
17	A0A0D7EHL0_RAOOR	A0A0D7EHL0	Raoultella ornithinolytica
18	A0A085I9K6_9ENTR	A0A085I9K6	Kluyvera ascorbata ATCC 33433
19	G5LMQ3_SALET	G5LMQ3	Salmonella enterica subsp. enterica serovar Alachua str. R6-377
20	H5V6R8_ESCHE	H5V6R8	Escherichia hermannii NBRC 105704
21	A0A0A2W1E1_BEABA	A0A0A2W1E1	Beauveria bassiana D1-5
22	A0A089UDV1_9ENTR	A0A089UDV1	Cedecea neteri
23	I2B6E0_SHIBC	I2B6E0	Shimwellia blattae (strain ATCC 29907 / DSM 4481 / JCM 1650 / NBRC 105725 / CDC 9005-74)
24	A0A071LPK9_9ENTR	A0A071LPK9	Mangrovibacter sp. MFB070
25	E6WF59_PANSA	E6WF59	Pantoea sp. (strain At-9b)
26	U3TZT6_9ENTR	U3TZT6	Plautia stali symbiont
27	A0A0Q4MGN7_9ENTR	A0A0Q4MGN7	Erwinia sp. Leaf53
28	A0A0J9BFU5_9BACT	A0A0J9BFU5	bacteria symbiont BFo1 of Frankliniella occidentalis
29	D8MVU8_ERWBE	D8MVU8	Erwinia billingiae (strain Eb661)
30	A0A085JIK1_9ENTR	A0A085JIK1	Tatumella ptyseos ATCC 33301
31	A0A0A1B204_9ENTR	A0A0A1B204	Pantoea sp. PSNIH2
32	A0A0L7SZ48_9ENTR	A0A0L7SZ48	Erwinia iniecta
33	W0L5U5_9ENTR	W0L5U5	Chania multitudinisentens RB-25
34	UNG_ERWT9	B2VEC7	Erwinia tasmaniensis (strain DSM 17950 / CIP 109463 / Et1/99)
35	A0A0F7H788_SERFO	A0A0F7H788	Serratia fonticola

36	H3RG71_PANSE	H3RG71	<i>Pantoea stewartii</i> subsp. <i>stewartii</i> DC283
37	A0A014N9E9_9ENTR	A0A014N9E9	<i>Erwinia mallotivora</i>
38	A0A083ZVJ9_9ENTR	A0A083ZVJ9	<i>Serratia</i> sp. DD3
39	L0MLK8_SERMA	L0MLK8	<i>Serratia marcescens</i> FGI94
40	V3V4K5_SERS3	V3V4K5	<i>Serratia</i> sp. (strain ATCC 39006)
41	D4GKN2_PANAM	D4GKN2	<i>Pantoea ananatis</i> (strain LMG 20103)
42	G7LSC0_9ENTR	G7LSC0	<i>Brenneria</i> sp. EniD312
43	W1IS03_9ENTR	W1IS03	<i>Xenorhabdus szentirmaii</i> DSM 16338
44	A0A0M2KL85_9ENTR	A0A0M2KL85	<i>Erwinia tracheiphila</i>
45	I3AM20_SERPL	I3AM20	<i>Serratia plymuthica</i> PRI-2C
46	UNG_PECAS	Q6D208	<i>Pectobacterium atrosepticum</i> (strain SCRI 1043 / ATCC BAA-672)
47	A0A0U5L6G4_9ENTR	A0A0U5L6G4	<i>Erwinia gerundensis</i>
48	E0SBI5_DICD3	E0SBI5	<i>Dickeya dadantii</i> (strain 3937)
49	MUG_ECOLI	P0A9H1	<i>Escherichia coli</i> (strain K12)
50	MUG_SHIDS	Q32BP9	<i>Shigella dysenteriae</i> serotype 1 (strain Sd197)
51	D2TRS0_CITRI	D2TRS0	<i>Citrobacter rodentium</i> (strain ICC168)
52	W1G114_ECOLX	W1G114	<i>Escherichia coli</i> ISC11
53	MUG_CITK8	A8APW0	<i>Citrobacter koseri</i> (strain ATCC BAA-895 / CDC 4225-83 / SGSC4696)
54	MUG_SALAR	A9MPV1	<i>Salmonella arizonae</i> (strain ATCC BAA-731 / CDC346-86 / RSK2980)
55	MUG_SALTY	Q7CPR7	<i>Salmonella typhimurium</i> (strain LT2 / SGSC1412 / ATCC 700720)
56	MUG_ENT38	A4WEK3	<i>Enterobacter</i> sp. (strain 638)
57	MUG_CROSS8	A7MJU3	<i>Cronobacter sakazakii</i> (strain ATCC BAA-894)
58	A0A090V372_ESCVU	A0A090V372	<i>Escherichia vulneris</i> NBRC 102420
59	E3G2N1_ENTLS	E3G2N1	<i>Enterobacter lignolyticus</i> (strain SCF1)
60	A0A0J5MK11_PLUGE	A0A0J5MK11	<i>Pluralibacter gergoviae</i>
61	MUG_KLEP7	A6TE50	<i>Klebsiella pneumoniae</i> subsp. <i>pneumoniae</i> (strain ATCC 700721 / MGH 78578)
62	W1DXH6_KLEPN	W1DXH6	<i>Klebsiella pneumoniae</i> IS46
63	W1HTC0_KLEPN	W1HTC0	<i>Klebsiella pneumoniae</i> IS39
64	V5A8Z5_ENTCL	V5A8Z5	<i>Enterobacter cloacae</i> S611
65	H5V6S8_ESCHE	H5V6S8	<i>Escherichia hermannii</i> NBRC 105704
66	A0A0D7EJY6_RAOOR	A0A0D7EJY6	<i>Raoultella ornithinolytica</i>
67	A0A089UCX2_9ENTR	A0A089UCX2	<i>Cedecea neteri</i>
68	A0A085IA79_9ENTR	A0A085IA79	<i>Kluyvera ascorbata</i> ATCC 33433
69	A0A014M110_9ENTR	A0A014M110	<i>Erwinia mallotivora</i>
70	A0A0L7TFD6_9ENTR	A0A0L7TFD6	<i>Erwinia iniecta</i>
71	A0A0A1AVI2_9ENTR	A0A0A1AVI2	<i>Pantoea</i> sp. PSNIH2
72	D8MX55_ERWBE	D8MX55	<i>Erwinia billingiae</i> (strain Eb661)

73	A0A097QYU6_HAFAL	A0A097QYU6	Hafnia alvei FB1
74	I2B552_SHIBC	I2B552	Shimwellia blattae (strain ATCC 29907 / DSM 4481 / JCM 1650 / NBRC 105725 / CDC 9005-74)
75	A0A0J8YRT2_9BACT	A0A0J8YRT2	bacteria symbiont BFo1 of Frankliniella occidentalis
76	G7LRY8_9ENTR	G7LRY8	Brenneria sp. EniD312
77	A0A0U5GHM7_9ENTR	A0A0U5GHM7	Erwinia gerundensis
78	B2VL29_ERWT9	B2VL29	Erwinia tasmaniensis (strain DSM 17950 / CIP 109463 / Et1/99)
79	MUG_SERP5	A8GJV5	Serratia proteamaculans (strain 568)
80	MUG_PECAS	Q6D9D7	Pectobacterium atrosepticum (strain SCRI 1043 / ATCC BAA-672)
81	D4GN66_PANAM	D4GN66	Pantoea ananatis (strain LMG 20103)
82	H3RA23_PANSE	H3RA23	Pantoea stewartii subsp. stewartii DC283
83	E6WD71_PANSA	E6WD71	Pantoea sp. (strain At-9b)
84	A0A0F7H902_SERFO	A0A0F7H902	Serratia fonticola
85	I3AIH5_SERPL	I3AIH5	Serratia plymuthica PRI-2C
86	A0A0G3CMJ9_9ENTR	A0A0G3CMJ9	Pragia fontium
87	I0QNN3_9ENTR	I0QNN3	Serratia sp. M24T3
88	L0MJZ4_SERMA	L0MJZ4	Serratia marcescens FGI94
89	A0A0T9L0T9_9ENTR	A0A0T9L0T9	Yersinia nurmii
90	W0I2B2_9ENTR	W0I2B2	Sodalis praecaptivus
91	V3TWE1_SERS3	V3TWE1	Serratia sp. (strain ATCC 39006)
92	H2IUG4_RAHAAC	H2IUG4	Rahnella aquatilis (strain ATCC 33071 / DSM 4594 / JCM 1683 / NBRC 105701 / NCIMB 13365 / CIP 78.65)
93	G8NW51_GRAMM	G8NW51	Granulicella mallensis (strain ATCC BAA-1857 / DSM 23137 / MP5ACTX8)
94	A0A0Q8PLL6_9ACTN	A0A0Q8PLL6	Kitasatospora sp. Root187
95	A0A0F0HJP7_9ACTN	A0A0F0HJP7	Streptomyces sp. NRRL F-4428
96	D9VKL8_9ACTN	D9VKL8	Streptomyces sp. C
97	L8ELN0_STRRM	L8ELN0	Streptomyces rimosus subsp. rimosus ATCC 10970
98	A0A0M9ZRN1_9ACTN	A0A0M9ZRN1	Streptomyces sp. AS58
99	A4IGM6_XENTR	A4IGM6	Xenopus tropicalis
100	H3B891_LATCH	H3B891	Latimeria chalumnae
101	G1KKQ2_ANOCA	G1KKQ2	Anolis carolinensis
102	V8PA34_OPHHA	V8PA34	Ophiophagus hannah
103	H3CLQ6_TETNG	H3CLQ6	Tetraodon nigroviridis
104	I3KSN2_ORENI	I3KSN2	Oreochromis niloticus
105	H2VE25_TAKRU	H2VE25	Takifugu rubripes
106	A0A0P7UYP6_9TELE	A0A0P7UYP6	Scleropages formosus
107	A0A0G2KZ26_DANRE	A0A0G2KZ26	Danio rerio



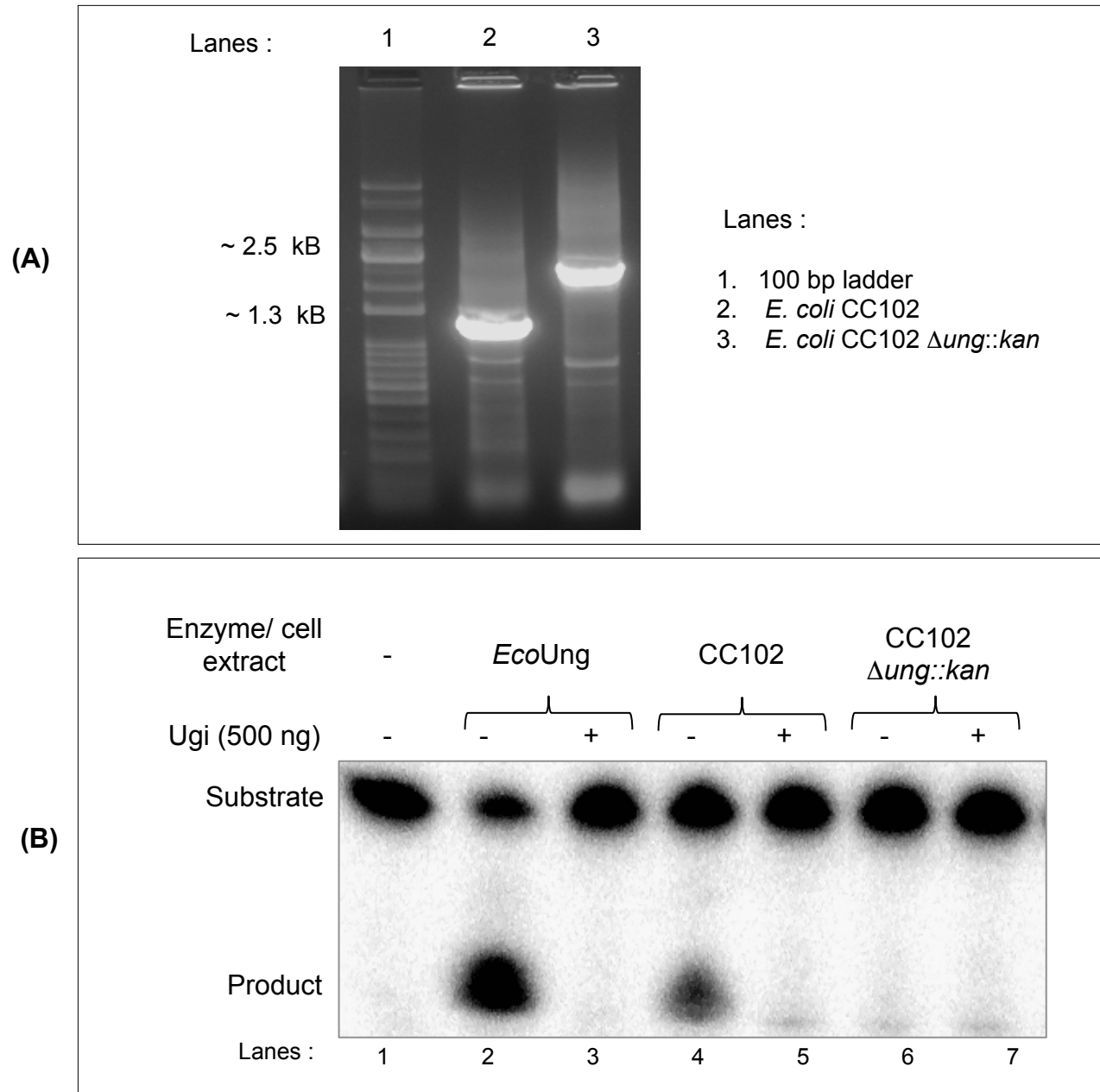
108	H0WN54_OTOGA	H0WN54	Otolemur garnettii
109	H0VAS3_CAVPO	H0VAS3	Cavia porcellus
110	G1U181_RABIT	G1U181	Oryctolagus cuniculus
111	G1S7F9_NOMLE	G1S7F9	Nomascus leucogenys
112	G1THF7_RABIT	G1THF7	Oryctolagus cuniculus
113	SMUG1_HUMAN	Q53HV7	Homo sapiens
114	L8Y2M5_TUPCH	L8Y2M5	Tupaia chinensis
115	F7DGX4_HORSE	F7DGX4	Equus caballus
116	H2Q638_PANTR	H2Q638	Pan troglodytes
117	SMUG1_RAT	Q811Q1	Rattus norvegicus
118	G7PHW5_MACFA	G7PHW5	Macaca fascicularis
119	SMUG1_MOUSE	Q6P5C5	Mus musculus
120	A0A096MT05_PAPAN	A0A096MT05	Papio anubis
121	A0A0D9QZG3_CHLSB	A0A0D9QZG3	Chlorocebus sabaues
122	F6PUI2_MACMU	F6PUI2	Macaca mulatta
123	G3SPZ8_LOXAF	G3SPZ8	Loxodonta africana
124	A0A087X7G5_POEFO	A0A087X7G5	Poecilia formosa
125	G1L1H7_AILME	G1L1H7	Ailuropoda melanoleuca
126	M3ZJJ0_XIPMA	M3ZJJ0	Xiphophorus maculatus
127	M3Y384_MUSPF	M3Y384	Mustela putorius furo
128	G3QCU0_GORGO	G3QCU0	Gorilla gorilla gorilla
129	I3M4F6 ICTTR	I3M4F6	Ictidomys tridecemlineatus
130	G3N660_GASAC	G3N660	Gasterosteus aculeatus
131	F7I3E8_CALJA	F7I3E8	Callithrix jacchus
132	S9WWW6_CAMFR	S9WWW6	Camelus ferus
133	L5L7F0_PTEAL	L5L7F0	Pteropus alecto
134	L5MAG6_MYODS	L5MAG6	Myotis davidii
135	W5LR77_ASTMX	W5LR77	Astyanax mexicanus
136	G1NTF2_MYOLU	G1NTF2	Myotis lucifugus
137	A0A0F8ADX5_LARCR	A0A0F8ADX5	Larimichthys crocea
138	S7N0S2_MYOBR	S7N0S2	Myotis brandtii
139	H2MKL2_ORYLA	H2MKL2	Oryzias latipes
140	I3LJI9_PIG	I3LJI9	Sus scrofa
141	G3IGK5_CRIGR	G3IGK5	Cricetulus griseus
142	SMUG1_BOVIN	Q59I47	Bos taurus
143	G3WEA4_SARHA	G3WEA4	Sarcophilus harrisii
144	F1Q3F6_CANLF	F1Q3F6	Canis lupus familiaris
145	F1Q241_CANLF	F1Q241	Canis lupus familiaris
146	M3WCS1_FELCA	M3WCS1	Felis catus
147	Q5SKC5_THET8	Q5SKC5	Thermus thermophilus (strain HB8 / ATCC 27634 / DSM 579)
148	F2NQC7_MARHT	F2NQC7	Marinithermus hydrothermalis (strain DSM

			14884 / JCM 11576 / T1)
149	D7BHW7_MEISD	D7BHW7	Meiothermus silvanus (strain ATCC 700542 / DSM 9946 / VI-R2)
150	E4U889_OCEP5	E4U889	Oceanithermus profundus (strain DSM 14977 / NBRC 100410 / VKM B-2274 / 506)
151	D3PTB3_MEIRD	D3PTB3	Meiothermus ruber (strain ATCC 35948 / DSM 1279 / VKM B-1258 / 21)
152	G9PXF0_9BACT	G9PXF0	Synergistes sp. 3_1_syn1
153	A8F661_PSELT	A8F661	Pseudothermotoga lettingae (strain ATCC BAA-301 / DSM 14385 / NBRC 107922 / TMO)
154	E8U9Y5_DEIML	E8U9Y5	Deinococcus maricopensis (strain DSM 21211 / LMG 22137 / NRRL B-23946 / LB-34)
155	D7CWJ9_TRURR	D7CWJ9	Truepera radiovictrix (strain DSM 17093 / CIP 108686 / LMG 22925 / RQ-24)
156	A6LJD3_THEM4	A6LJD3	Thermosipho melanesiensis (strain DSM 12029 / CIP 104789 / BI429)
157	A0A124FHE1_9BACT	A0A124FHE1	Acetothermia bacterium 64_32
158	A0A073IQ13_9BACT	A0A073IQ13	Synergistes jonesii
159	F7YUP4_9THEM	F7YUP4	Pseudothermotoga thermarum DSM 5069
160	I4BVU8_ANAMD	I4BVU8	Anaerobaculum mobile (strain ATCC BAA-54 / DSM 13181 / NGA)
161	A0A0G2ZB04_9BACT	A0A0G2ZB04	Kosmotoga pacifica
162	A0A101FM96_9BACT	A0A101FM96	Synergistales bacterium 54_24
163	A7HJD1_FERNB	A7HJD1	Fervidobacterium nodosum (strain ATCC 35602 / DSM 5306 / Rt17-B1)
164	D2Z3F4_9BACT	D2Z3F4	Dethiosulfovibrio peptidovorans DSM 11002
165	C5CIT9_KOSOT	C5CIT9	Kosmotoga olearia (strain TBF 19.5.1)
166	A0A101IYY8_9BACT	A0A101IYY8	Synergistales bacterium 54_9
167	D1BA65_THEAS	D1BA65	Thermanaerovibrio acidaminovorans (strain ATCC 49978 / DSM 6589 / Su883)
168	A0A101EEF6_9BACT	A0A101EEF6	bacterium 42_11
169	Q1IY33_DEIGD	Q1IY33	Deinococcus geothermalis (strain DSM 11300)
170	Q31K46_SYNE7	Q31K46	Synechococcus elongatus (strain PCC 7942)
171	A0A098TNB4_9CHRO	A0A098TNB4	Neosynechococcus sphagnicola sy1
172	D5EEM5_AMICL	D5EEM5	Aminobacterium colombiense (strain DSM 12261 / ALA-1)
173	I2F485_9BACT	I2F485	Mesotoga prima MesG1.Ag.4.2
174	K9SMC9_9CYAN	K9SMC9	Pseudanabaena sp. PCC 7367
175	E3D056_9BACT	E3D056	Aminomonas paucivorans DSM 12260
176	D1Y4R2_9BACT	D1Y4R2	Pyramidobacter piscolens W5455
177	U5QJR1_9CYAN	U5QJR1	Gloeobacter kilaeuensis JS1
178	K9ZWH2_DEIPD	K9ZWH2	Deinococcus peraridilitoris (strain DSM 19664 / LMG 22246 / CIP 109416 / KR-200)
179	H0UK71_9BACT	H0UK71	Jonquetella anthropi DSM 22815

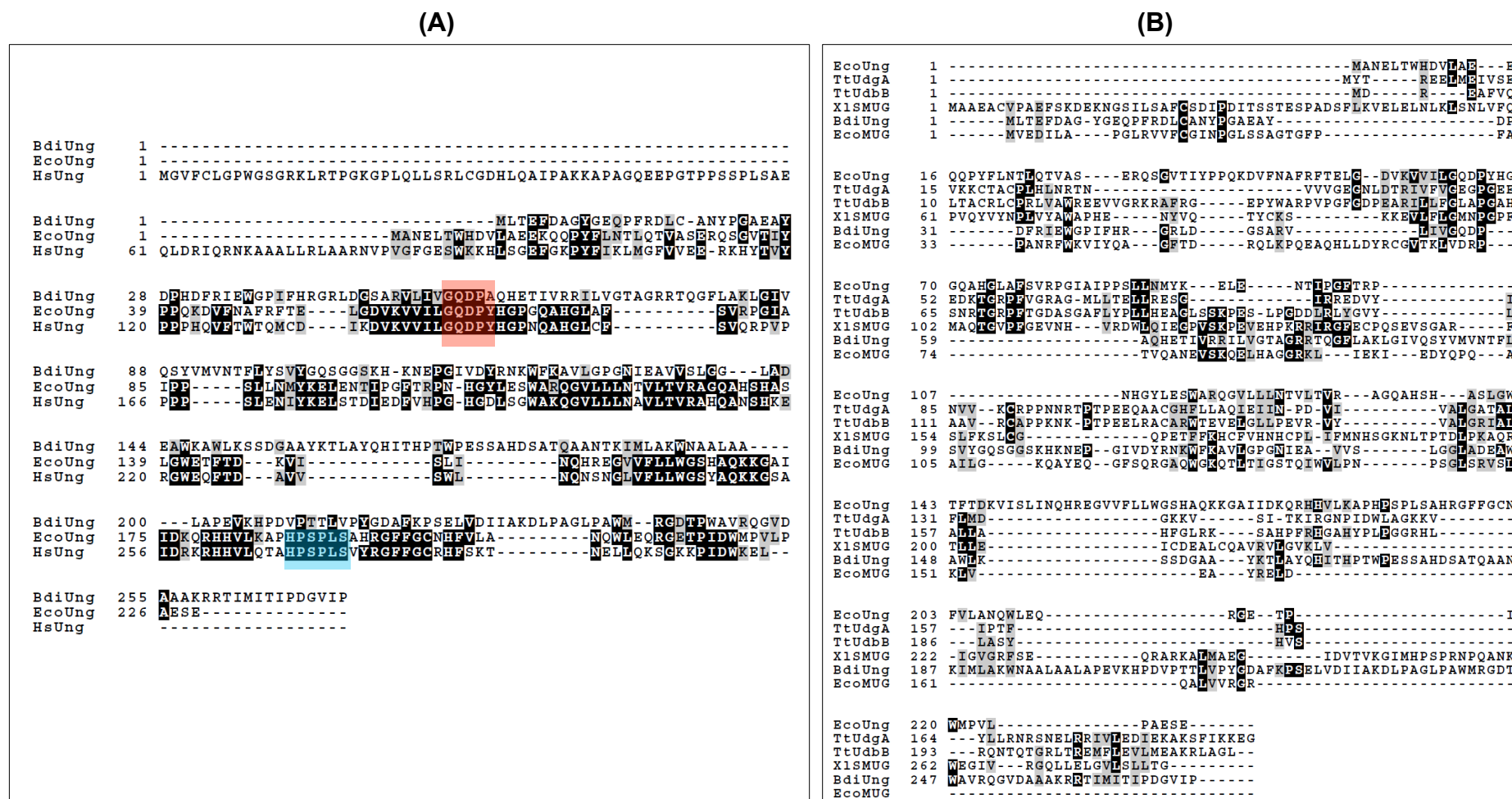
180	A0A124G0Z9_9BACT	A0A124G0Z9	Synergistales bacterium 58_81
181	A0A101HYT4_9BACT	A0A101HYT4	candidate division TA06 bacterium 34_109
182	A0A0C1Y5L3_9CYAN	A0A0C1Y5L3	Lyngbya confervoides BDU141951
183	A0A101HQ32_9BACT	A0A101HQ32	Mesotoga prima
184	Q8DHJ3_THEEB	Q8DHJ3	Thermosynechococcus elongatus (strain BP-1)
185	K9WJ37_9CYAN	K9WJ37	Microcoleus sp. PCC 7113
186	A0A081C383_9BACT	A0A081C383	bacterium UASB270
187	F0RKU2_DEIPM	F0RKU2	Deinococcus proteolyticus (strain ATCC 35074 / DSM 20540 / JCM 6276 / NBRC 101906 / NCIMB 13154 / VKM Ac-1939 / CCM 2703 / MRP)
188	F4XUN6_9CYAN	F4XUN6	Moorea producens 3L
189	B8HRZ5_CYAP4	B8HRZ5	Cyanotheca sp. (strain PCC 7425 / ATCC 29141)
190	Q7NHX7_GLOVI	Q7NHX7	Gloeobacter violaceus (strain PCC 7421)
191	K9SC25_9CYAN	K9SC25	Geitlerinema sp. PCC 7407
192	A0A073CM55_PLAAG	A0A073CM55	Planktothrix agardhii NIVA-CYA 126/8
193	K9P5R0_CYAGP	K9P5R0	Cyanobium gracile (strain ATCC 27147 / PCC 6307)
194	G7VA85_THELD	G7VA85	Thermovirga lienii (strain ATCC BAA-1197 / DSM 17291 / Cas60314)
195	L8N552_9CYAN	L8N552	Pseudanabaena biceps PCC 7429
196	A0A101EU31_9THEM	A0A101EU31	Thermotoga sp. 50_1627
197	Q5SJ65_THET8	Q5SJ65	Thermus thermophilus (strain HB8 / ATCC 27634 / DSM 579)
198	F2NKD5_MARHT	F2NKD5	Marinithermus hydrothermalis (strain DSM 14884 / JCM 11576 / T1)
199	D3PN59_MEIRD	D3PN59	Meiothermus ruber (strain ATCC 35948 / DSM 1279 / VKM B-1258 / 21)
200	E4U9L6_OCEP5	E4U9L6	Oceanithermus profundus (strain DSM 14977 / NBRC 100410 / VKM B-2274 / 506)
201	D7BEP3_MEISD	D7BEP3	Meiothermus silvanus (strain ATCC 700542 / DSM 9946 / VI-R2)
202	L0A585_DEIPD	L0A585	Deinococcus peraridilitoris (strain DSM 19664 / LMG 22246 / CIP 109416 / KR-200)
203	A0A0P6XLM9_9CHLR	A0A0P6XLM9	Levilinea saccharolytica
204	D7CQJ9_TRURR	D7CQJ9	Truepera radiovictrix (strain DSM 17093 / CIP 108686 / LMG 22925 / RQ-24)
205	Q1IRD4_KORVE	Q1IRD4	Koribacter versatilis (strain Ellin345)
206	A0A0P6XV96_9CHLR	A0A0P6XV96	Ornatilinea apprima
207	A0A0P6WNG6_9CHLR	A0A0P6WNG6	Leptolinea tardivitalis
208	C1F2C9_ACIC5	C1F2C9	Acidobacterium capsulatum (strain ATCC 51196 / DSM 11244 / JCM 7670 / NBRC 15755 / NCIMB 13165 / 161)
209	A0A101H4T6_9CHLR	A0A101H4T6	Anaerolineae bacterium 49_20
210	G8NT74_GRAMM	G8NT74	Granulicella mallensis (strain ATCC BAA-

			1857 / DSM 23137 / MP5ACTX8)
211	Q091T5_STIAD	Q091T5	Stigmatella aurantiaca (strain DW4/3-1)
212	E8WW97_GRATM	E8WW97	Granulicella tundricola (strain ATCC BAA-1859 / DSM 23138 / MP5ACTX9)
213	A0A085WRF4_9DELT	A0A085WRF4	Hyalangium minutum
214	A0A0C1RIV5_9ACTN	A0A0C1RIV5	marine actinobacterium MedAcidi-G2A
215	A0A0C1UUF1_9ACTN	A0A0C1UUF1	marine actinobacterium MedAcidi-G1
216	F8I6V1_SULAT	F8I6V1	Sulfobacillus acidophilus (strain TPY)
217	A9GFE5_SORC5	A9GFE5	Sorangium cellulosum (strain So ce56)
218	Q2IFN8_ANADE	Q2IFN8	Anaeromyxobacter dehalogenans (strain 2CP-C)
219	A0A0N8GM42_9CHLR	A0A0N8GM42	Bellilinea caldifistulae
220	E8V8B5_TERSS	E8V8B5	Terriglobus saanensis (strain ATCC BAA-1853 / DSM 23119 / SP1PR4)
221	H0EAB6_9ACTN	H0EAB6	Patulibacter medicamentivorans
222	Q1D6E4_MYXXD	Q1D6E4	Myxococcus xanthus (strain DK 1622)
223	A0A0C1RIT9_9ACTN	A0A0C1RIT9	marine actinobacterium MedAcidi-G2B
224	H8MFC3_CORCM	H8MFC3	Coralloccoccus coralloides (strain ATCC 25202 / DSM 2259 / NBRC 100086 / M2)
225	Q0S418_RHOJR	Q0S418	Rhodococcus jostii (strain RHA1)
226	B3DVE5_METI4	B3DVE5	Methylacidiphilum inferorum (isolate V4)
227	A0A0R2XDU5_9BACT	A0A0R2XDU5	Verrucomicrobia subdivision 6 bacterium BACL9 MAG-120924-bin69
228	A0A0K2SIW3_9FIRM	A0A0K2SIW3	Limnochorda pilosa
229	G2LFM3_CHLTF	G2LFM3	Chloracidobacterium thermophilum (strain B)
230	M5AIS6_9ACTN	M5AIS6	Ilumatobacter coccineus YM16-304
231	Q01X21_SOLUE	Q01X21	Solibacter usitatus (strain Ellin6076)
232	A0A0G2ZPH4_9DELT	A0A0G2ZPH4	Archangium gephyra
233	W6K3D0_9MICO	W6K3D0	Tetrasphaera australiensis Ben110
234	Q67MX9_SYMTH	Q67MX9	Symbiobacterium thermophilum (strain T / IAM 14863)
235	A0A077MCB3_9MICO	A0A077MCB3	Tetrasphaera jenkinsii Ben 74
236	E8N6F1_ANATU	E8N6F1	Anaerolinea thermophila (strain DSM 14523 / JCM 11388 / NBRC 100420 / UNI-1)
237	H5XAX7_9PSEU	H5XAX7	Saccharomonospora marina XMU15
238	R4L555_9ACTN	R4L555	Actinoplanes sp. N902-109
239	Q1ARW9_RUBXD	Q1ARW9	Rubrobacter xylanophilus (strain DSM 9941 / NBRC 16129)
240	A0A0S7BFS6_9CHLR	A0A0S7BFS6	Longilinea arvoryzae
241	G2G6Q2_9ACTN	G2G6Q2	Streptomyces zinciresistens K42
242	A0A060HT19_9ARCH	A0A060HT19	Nitrososphaera viennensis EN76
243	A0A0P6XX76_9CHLR	A0A0P6XX76	Thermanaerotherix daxensis
244	A0A0K1Q7U9_9DELT	A0A0K1Q7U9	Labilithrix luteola
245	D9XQ44_9ACTN	D9XQ44	Streptomyces griseoflavus Tu4000

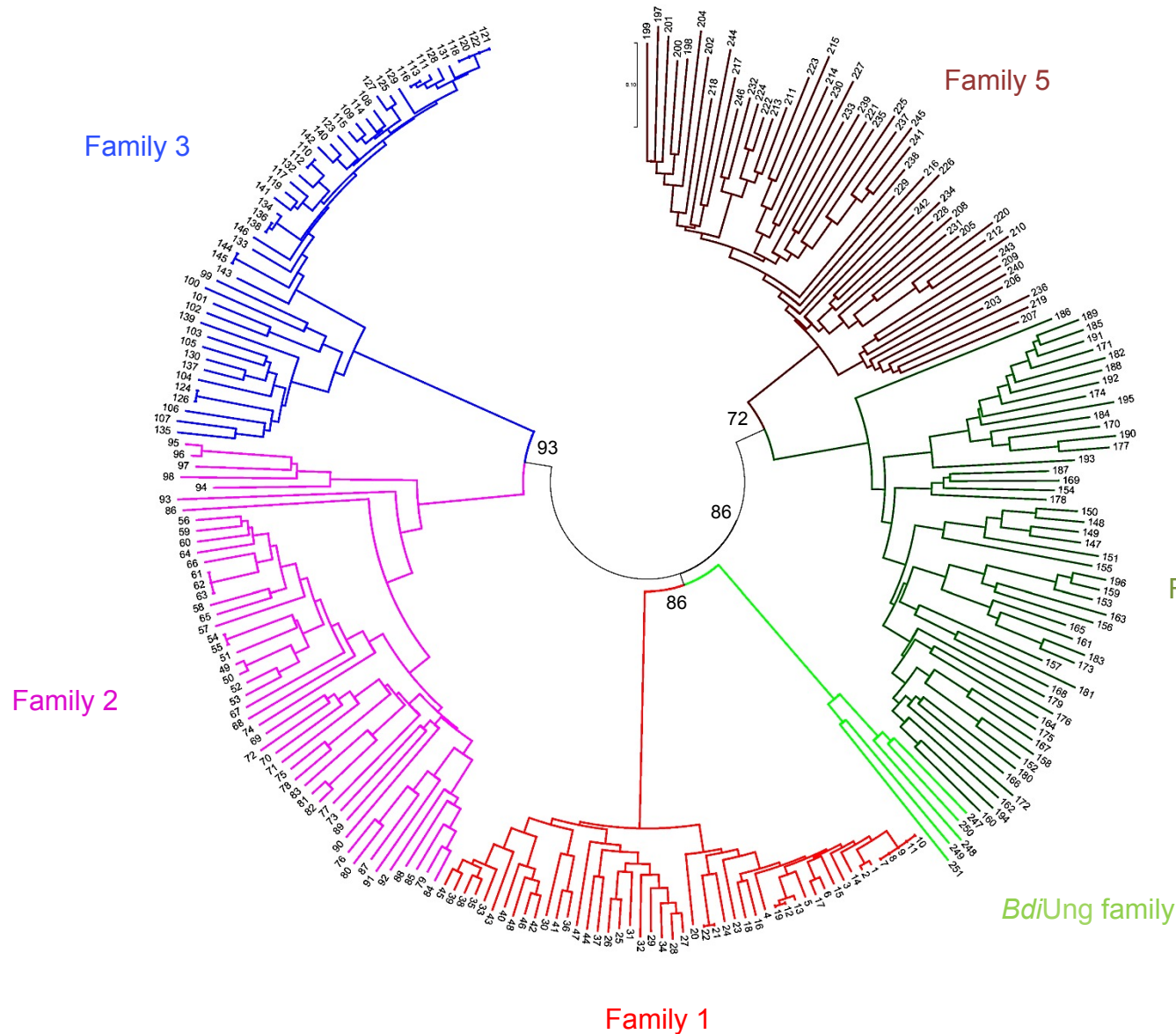
246	S9Q4Y3_9DELT	S9Q4Y3	Cystobacter fuscus DSM 2262
247	E3J5Z2_FRASU	E3J5Z2	Frankia sp. (strain Eu11c)
248	SrUDG	WP_013427486 (pubmed ID)	Streptacidiphilus rugosus
249	A0A0C3NQE5_9PROT	A0A0C3NQE5	Nitrospira sp. NpAV
250	KaUDG	WP_035839389 (pubmed ID)	Kitasatospora azatica
251	Q89XR0_BRADU	Q89XR0	Bradyrhizobium diazoefficiens (strain JCM 10833 / IAM 13628 / NBRC 14792 / USDA 110)



**Figure S1:** *E. coli* CC102 $\Delta ung::kan$  knockout confirmation. **(A)** PCR amplification of *ung* locus (lane 2) and  $\Delta ung::kan$  (lane 3) using EcoUng-up-Fp and EcoUng-dn-Rp primers resolved on 1% agarose gel adjacent to a 100 bp ladder (lane 1). **(B)** Uracil excision assay using a ssU9 substrate with cell free extracts (1  $\mu$ g) confirming the Ung<sup>-</sup> phenotype.

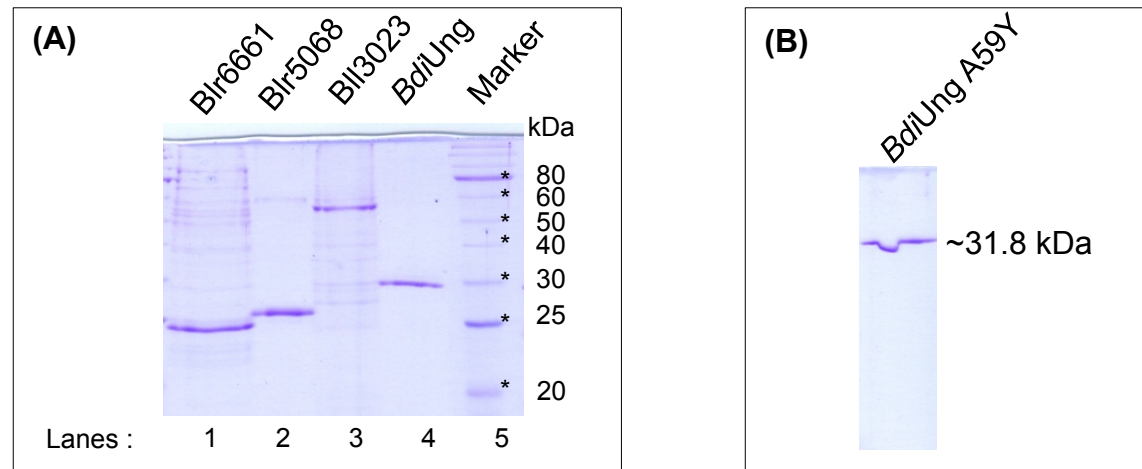


**Figure S2:** (A) Multiple sequence alignment of *BdiUng* with *EcoUng* and *HsUng* (family 1 UDGs from *E. coli* and *Homo sapiens*, respectively). Conserved regions corresponding to motif A (GQDPY) and motif B are highlighted in red and blue respectively. (B) Multiple sequence alignment of *BdiUng* with representative members from other families of UDGs – namely *EcoUng*, *EcoMUG*, *XlSMUG*, *TthUdgA* and *TthUdgB* belonging to families 1 to 5, respectively). Sequences were aligned using clustal omega and schematic representation was obtained by using BOXSHADE ([www.ch.embnet.org](http://www.ch.embnet.org)). Identical residues are shaded in black while similar residues in grey.

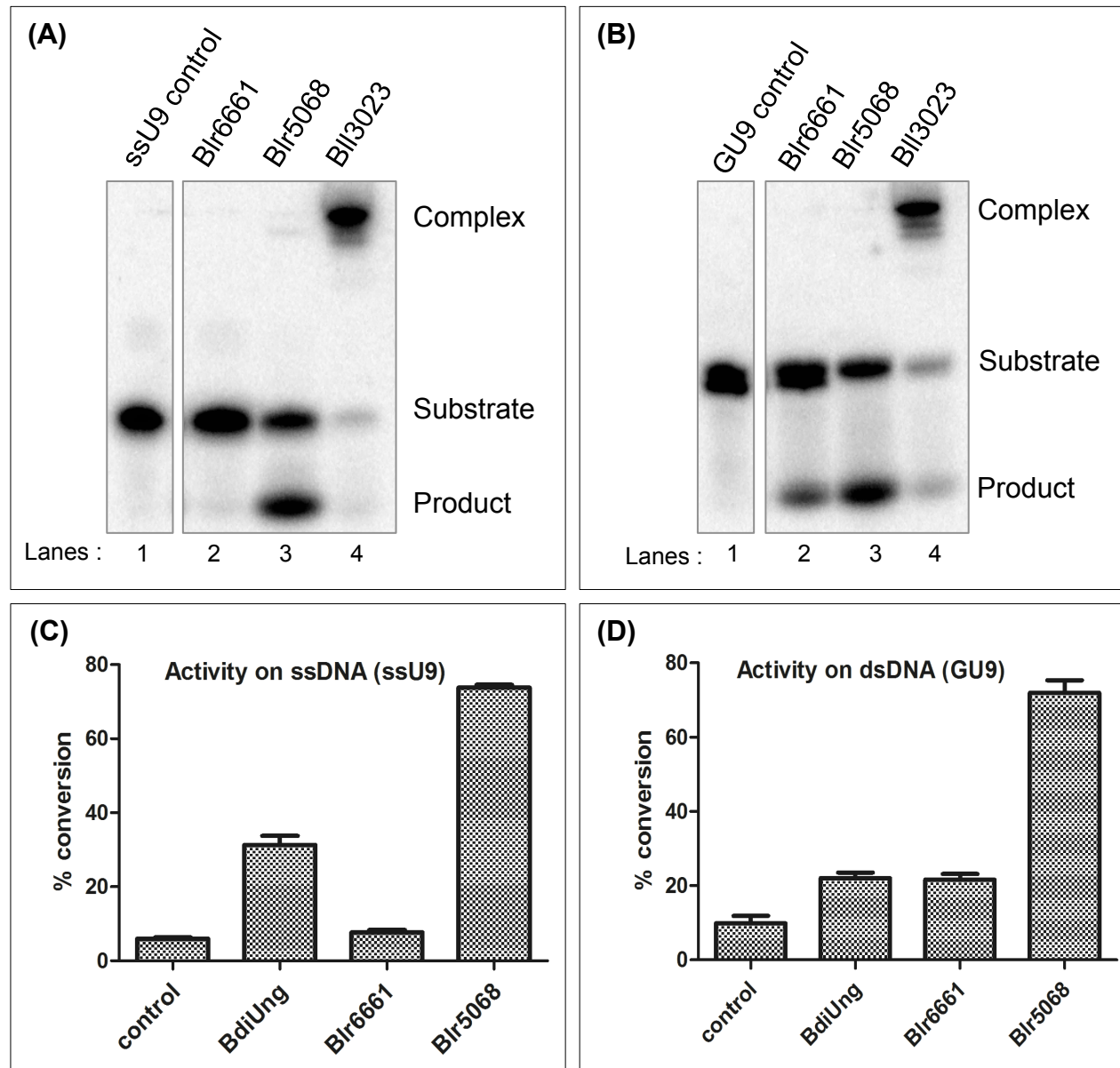


**Figure S3:** The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1978). The optimal tree with the sum of branch length = 27.05401915 is shown. The confidence probability (multiplied by 100) that the interior branch length is greater than 0, as estimated using the bootstrap test (1000 replicates) is shown next to the selected branches (Dopazo, 1992; Rzhetsky and Nei, 1992). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method (Nei and Kumar, 2000) and are in the units of the number of amino acid differences per site. The analysis involved 251 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 93 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar et. al., 2016). Details of protein sequences considered for constructing the tree are provided in Table S4.

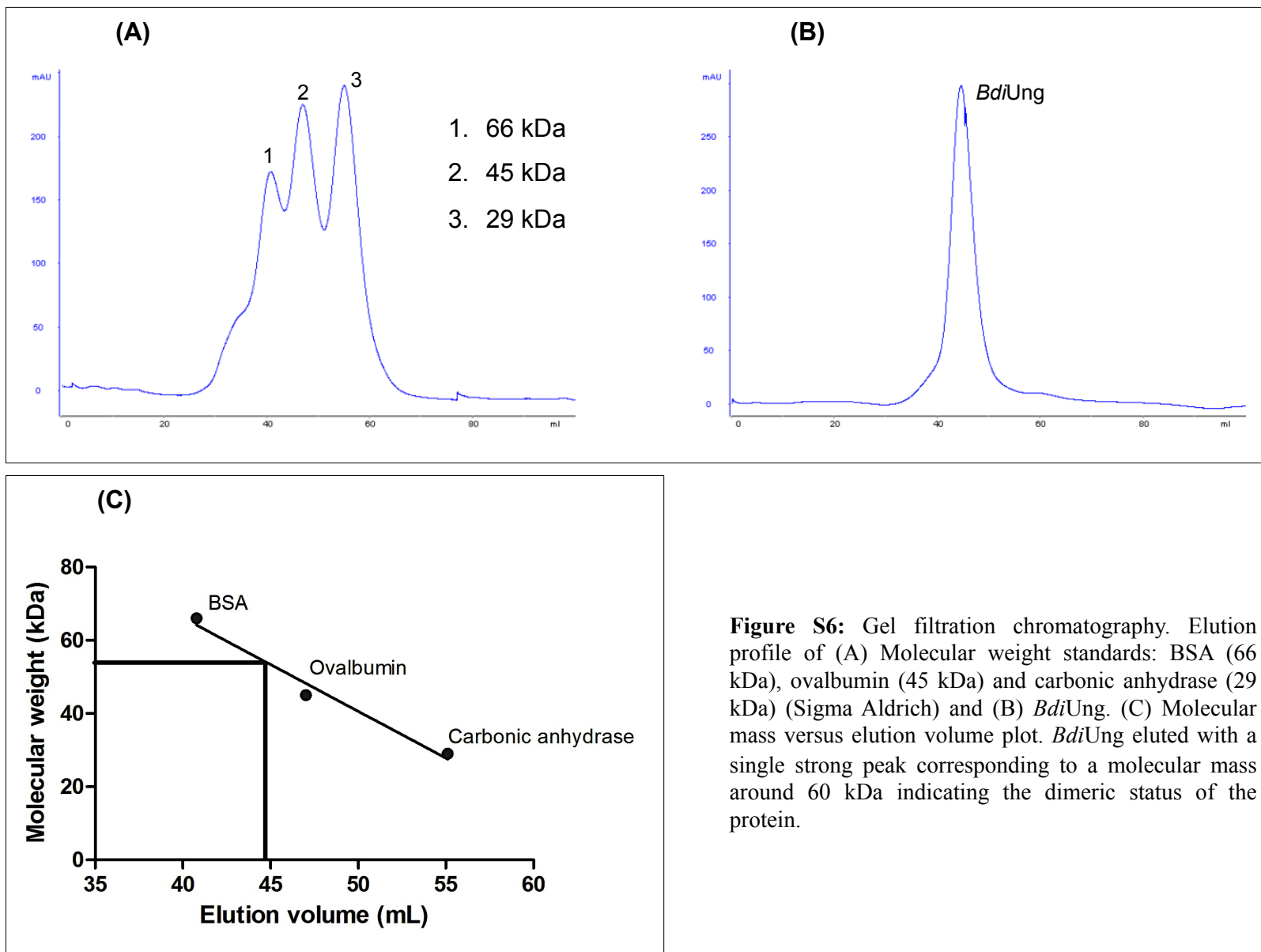




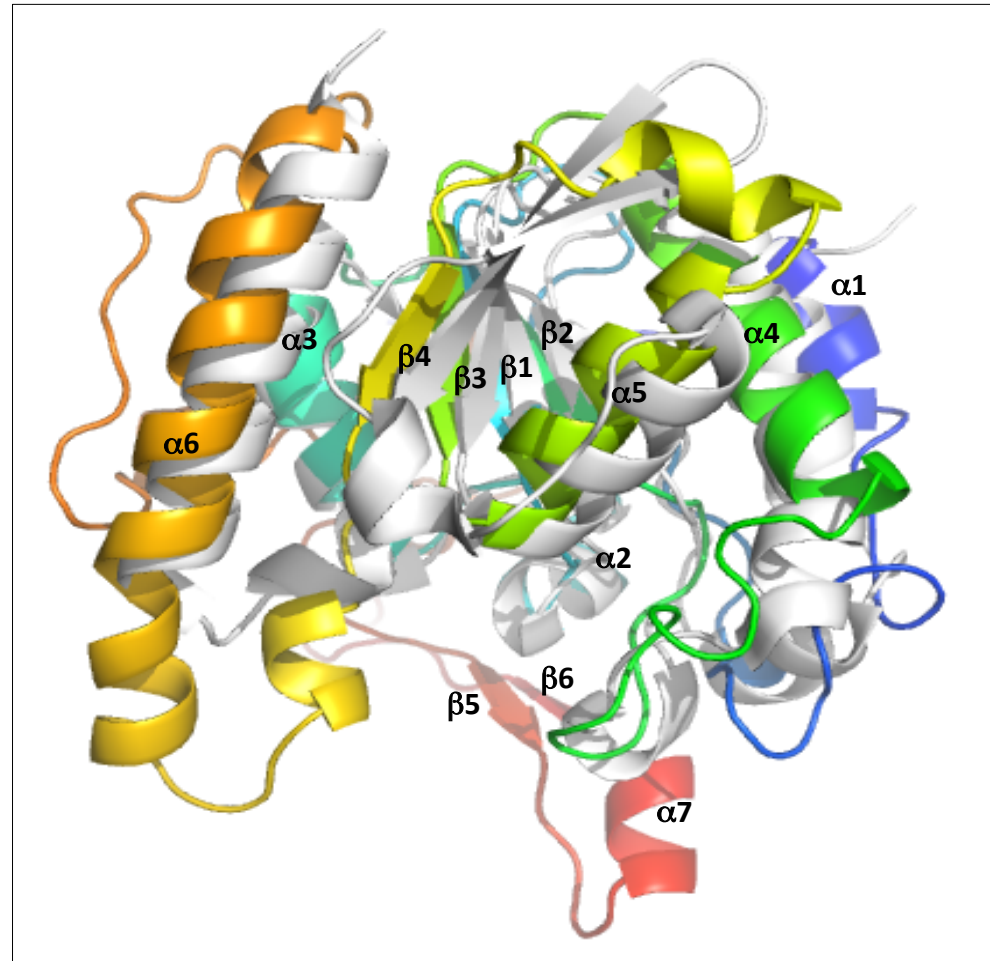
**Figure S4:** (A) Purified proteins (0.5  $\mu$ g) visualized on 15% SDS PAGE. Lanes: 1, Blr6661; 2, Blr5068; 3, Bll3023; 4, *BdiUng*; and 5, protein size markers. (B) Purified *BdiUngA59Y* mutant (1  $\mu$ g) analysed on 15% SDS PAGE.



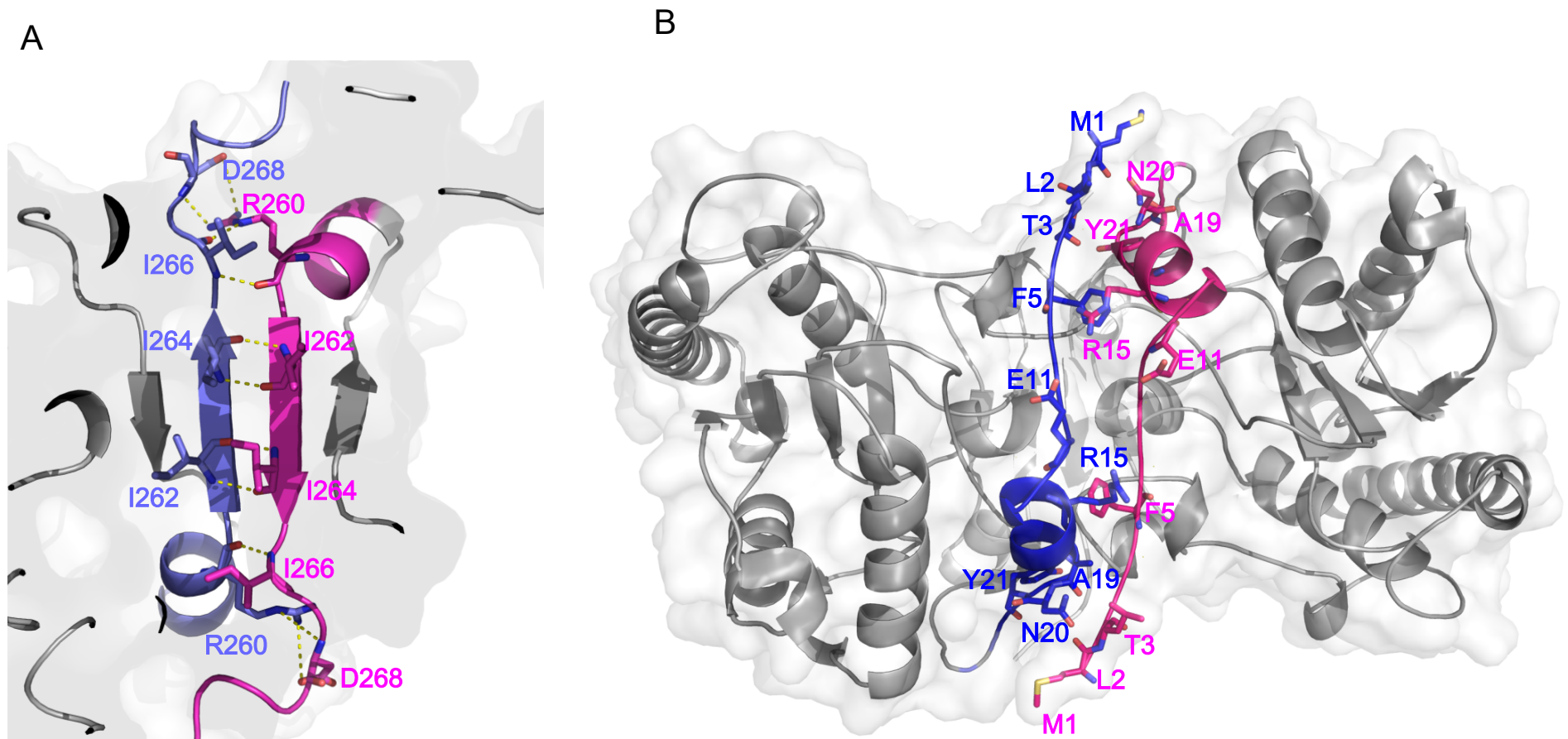
**Figure S5:** Uracil excision assays with different UDGs (100 ng each) identified from *B. diazoefficiens* on (A) ssU9 and (B) GU9. Products of excision assay were resolved on 8 M urea PAGE (15%). (C) and (D) show estimated % conversions as a mean of three independent assays with standard deviations for ssU9 (with 100 ng protein) and GU9 (with 500 ng protein), respectively in comparison to *BdiUng* activity. Note: Bll3023 is an ortholog of UdgX from *M. smegmatis* published earlier (Sang, P.B., Srinath, T., Patil, A.G., Woo, E.J. and Varshney, U., 2015, A unique uracil-DNA binding protein of the uracil DNA glycosylase superfamily. *Nucleic Acids Research*, **43**, 8452-8463 )



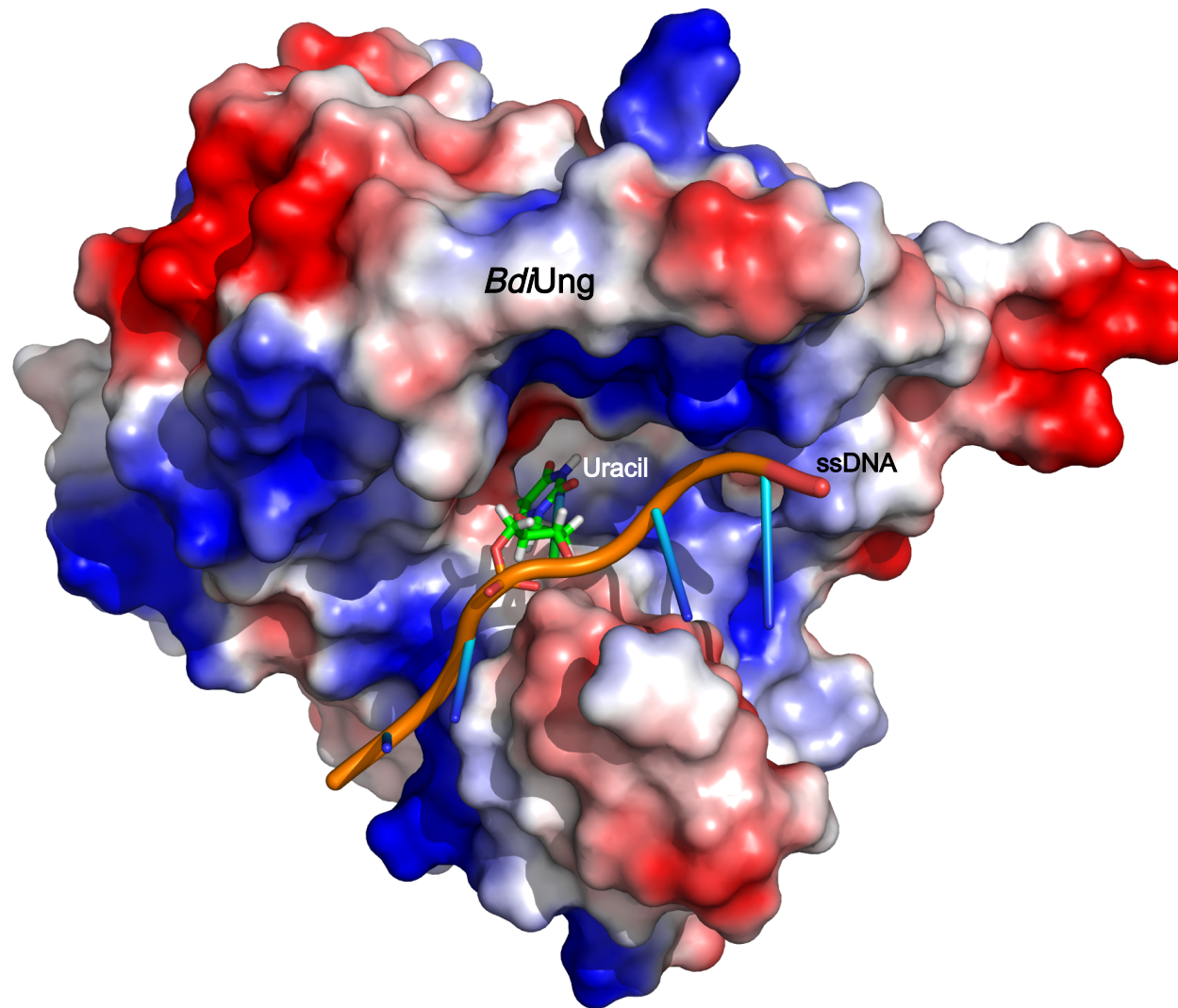
**Figure S6:** Gel filtration chromatography. Elution profile of (A) Molecular weight standards: BSA (66 kDa), ovalbumin (45 kDa) and carbonic anhydrase (29 kDa) (Sigma Aldrich) and (B) *BdiUng*. (C) Molecular mass versus elution volume plot. *BdiUng* eluted with a single strong peak corresponding to a molecular mass around 60 kDa indicating the dimeric status of the protein.



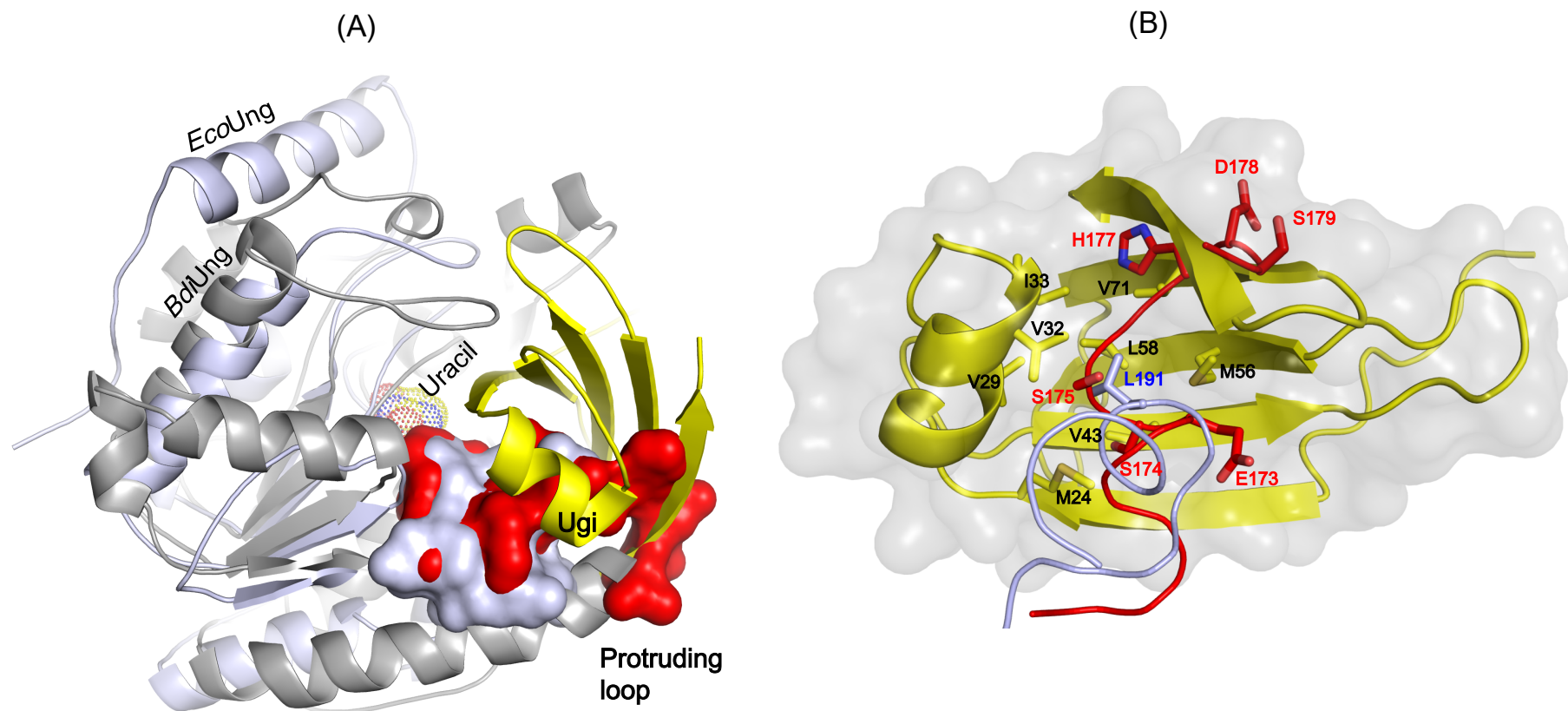
**Figure S7:** Overlap of *BdiUng* and uracil-DNA glycosylase from *Sulfolobus tokodaii* (PDB ID 4ZBZ) showing common  $\alpha/\beta/\alpha$  sandwich fold of UDG family 4 without [4Fe-S] cluster.



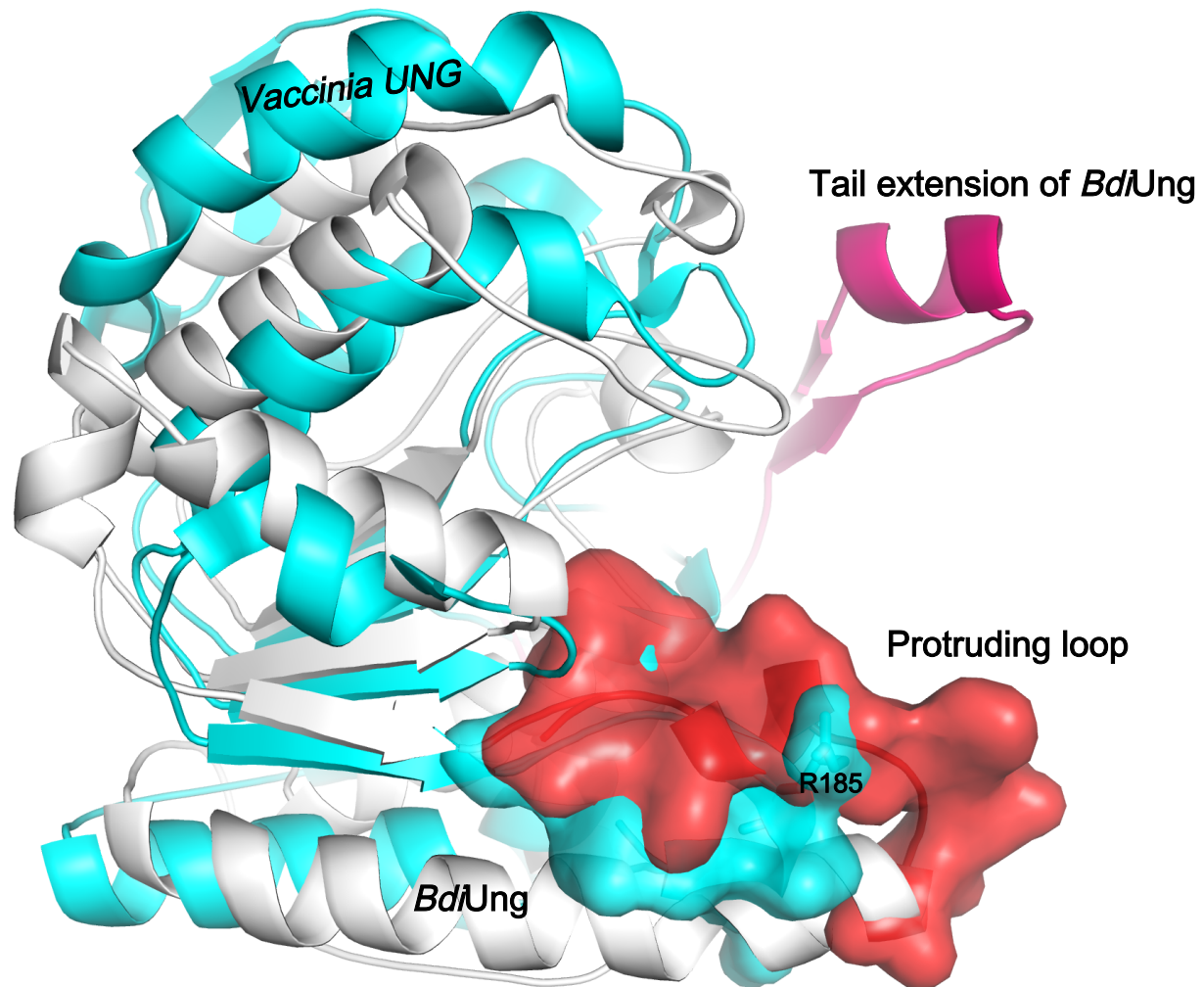
**Figure S8:** The dimer interface of *BdiUng*. (A) The interactions (yellow dash lines) between the residues in C-terminal tail extensions of each monomer involved in dimer formation. (B) The interactions between the N-terminal residues of monomers involved in dimer formation.



**Figure S9:** The DNA binding surface of *BdiUng*. The model of ssDNA bound *BdiUng* with an electrostatic charge distribution on the surface (Blue represents positive charge and negative charge is represented by red color)

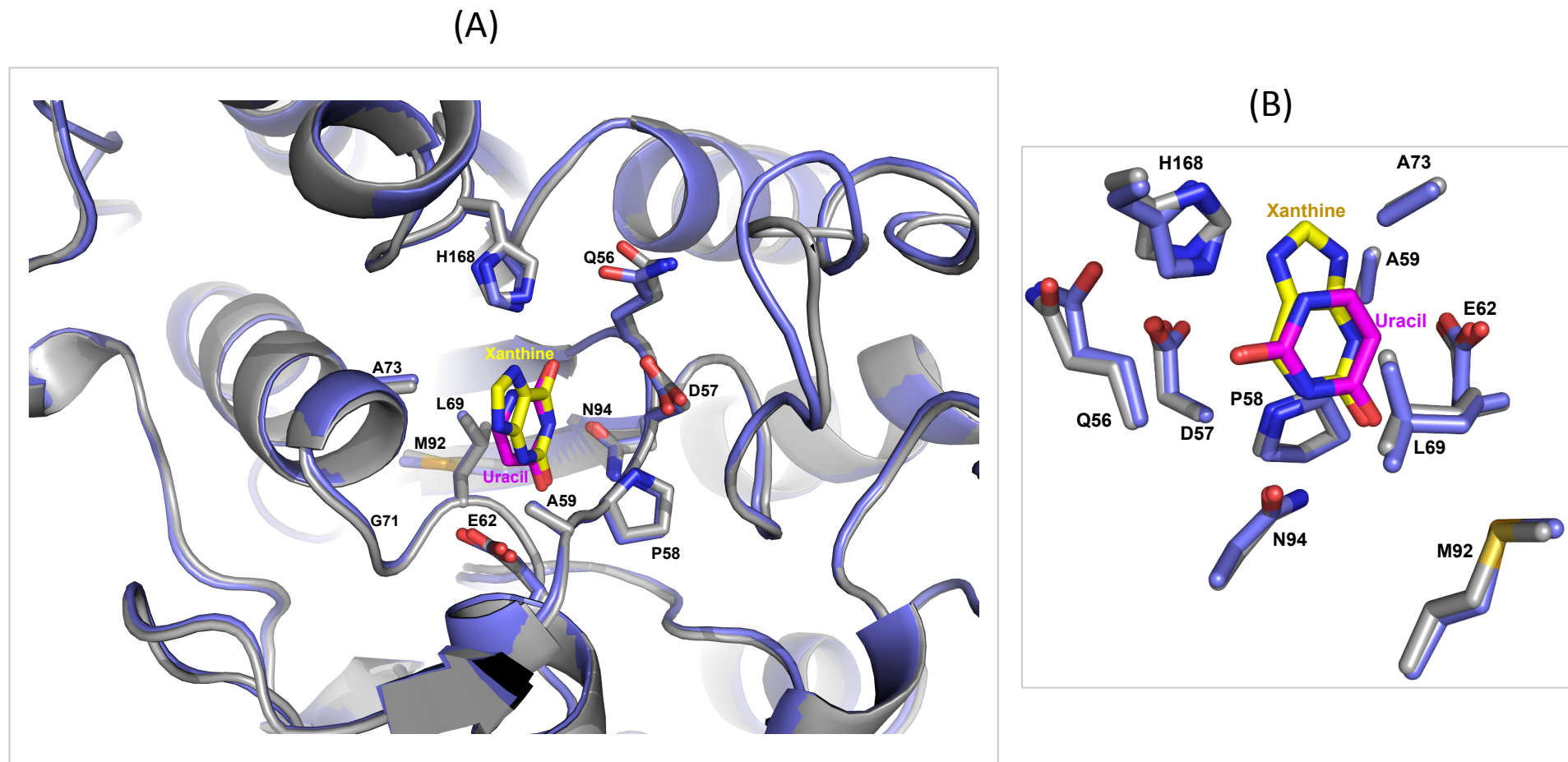


**Figure S10:** The protruding loop makes *BdiUng* resistant to Ugi. (A) The overlap of *BdiUng* (grey cartoon) and *E. coli* Ung (pale blue cartoon)-Ugi (yellow) complex. The protruding loop of *BdiUng* is shown as red while corresponding region in *E. coli* is shown as pale blue surface diagram. (B) The detailed overview of the overlap of protruding loop of *BdiUng* (red) and corresponding region of *E. coli* Ung (pale blue)-Ugi (yellow) complex. The L191 of *E. coli* Ung interacts with hydrophobic residues (shown as yellow sticks M24, V29, V32, I33, V43, M56, L58 and V71) of Ugi. The corresponding residues of *BdiUng* have been shown as red sticks.

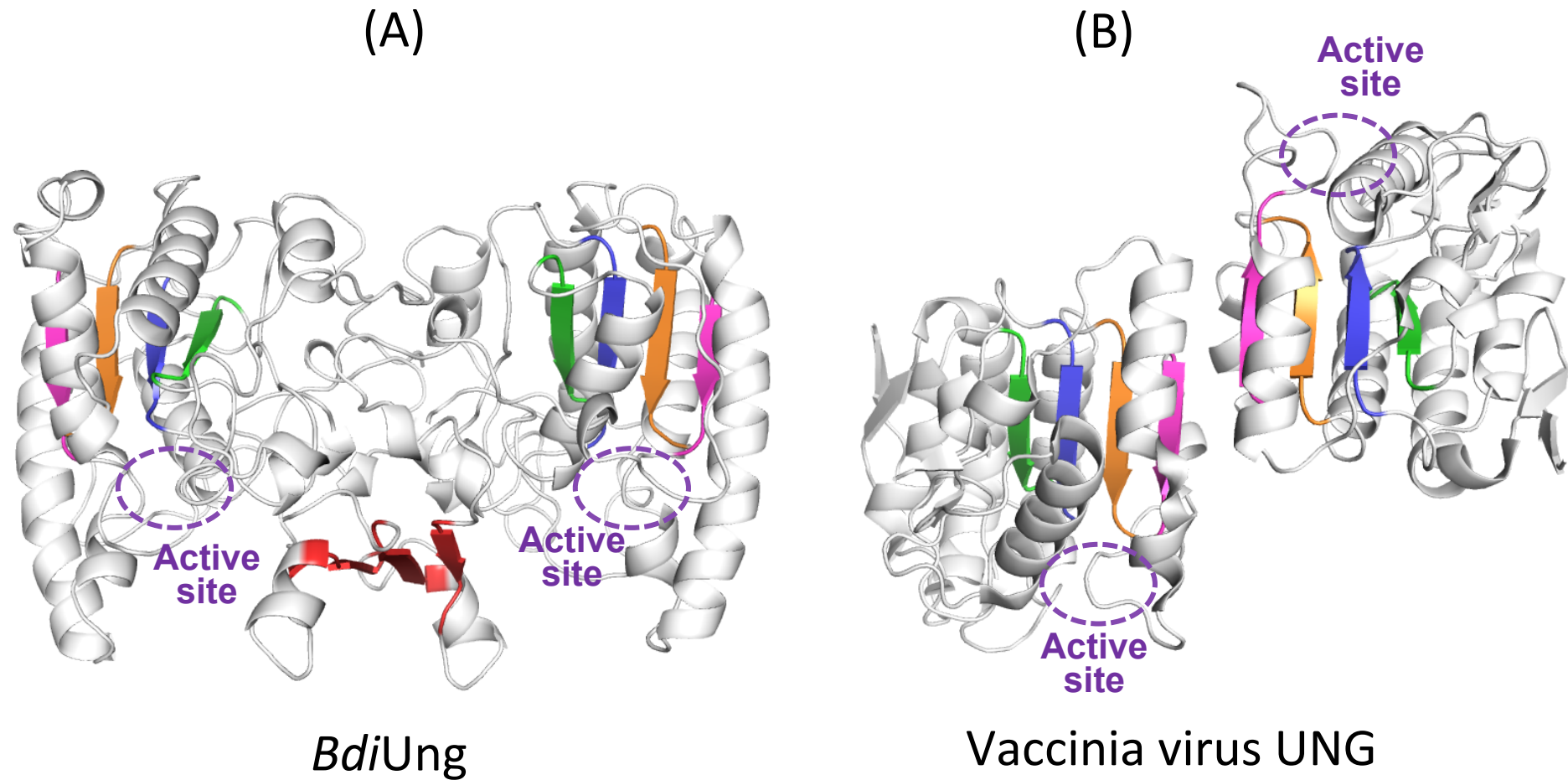


**Figure S11:** Comparison of unique features of *BdiUng* and vaccinia virus UNG. (A) The overlap of *BdiUng* (grey) with vaccinia virus UNG (Vaccinia UNG, cyan). The protruding loop is shown as red surface while corresponding region in vaccinia UNG as cyan surface with arginine represented as a stick. The tail extension of *BdiUng* is shown as magenta.





**Figure S12:** The larger cavity of active site of *BdiUng* is capable of accommodating diverse substrates. (A) An Overlap of *BdiUng* bound to xanthine (yellow) and uracil (magenta). The residues in active sites of *BdiUng*-Xanthine and *BdiUng*-Uracil complex are shown as sticks in grey and blue colours, respectively. (B) The detailed structure of active site of *BdiUng*-Xanthine and *BdiUng*-Uracil complex. Critical residues are shown as sticks in grey and blue colours respectively



**Figure S13:** The structures of *BdiUng* (A) and vaccinia virus UNG (PDB ID 5JX3, B) are compared to highlight the different dimeric interfaces. The core  $\beta$ -strands of the main  $\alpha/\beta/\alpha$  sandwich folds in each monomer are colored blue, green, orange and pink in spectrum from N- to C-terminal. The additional  $\beta$ -sheet formed by dimerization in *BdiUng* is colored red. The active site pockets are depicted by dotted circles.