

Supplementary information

C-Glycoside metabolism in the gut and in nature: Identification, characterization, structural analyses and distribution of C-C bond-cleaving enzymes

Takahiro Mori^{1,2,3,#}, Takuto Kumano^{4,5,#}, Haibing He^{1,#}, Satomi Watanabe^{4,#}, Miki Senda⁶, Toshio Moriya⁶, Naruhiko Adachi⁶, Sanae Hori⁴, Yuzu Terashita⁴, Masato Kawasaki⁶, Yoshiteru Hashimoto^{4,5}, Takayoshi Awakawa^{1,2}, Toshiya Senda^{6,7,*}, Ikuro Abe^{1,2,*}, Michihiko Kobayashi^{4,5,*}

¹ Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

² Collaborative Research Institute for Innovative Microbiology, The University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113-8657, Japan

³ PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan

⁴ Graduate School of Life and Environmental Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8572, Japan

⁵ Microbiology Research Center for Sustainability, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8572, Japan

⁶ Structural Biology Research Center, Institute of Materials Structure Science, High Energy Accelerator Research Organization (KEK), 1-1 Oho, Tsukuba, Ibaraki 305-0801, Japan

⁷ Faculty of Pure and Applied Sciences, University of Tsukuba, 1-1-1 Tennodai, Ibaraki 305-8571, Japan

These authors contributed equally to this work.

* Corresponding authors

Supplementary Tables

Supplementary Table 1. Sequence similarity of *Pu*CGD homolog enzymes.

A. Homolog enzymes of α -subunit of *Pu*CGD

Organism	Accession No. (Enzyme name)	Length (aa)	Identity (%)	Similarity (%)	E-value
PUE strain	BBG22495.1 (DgpC; α -subunit of <i>Pu</i> CGD)	324	100	100	0
<i>Eubacterium cellulosolvens</i>	WP_004604489.1 (DfgA; α -subunit of <i>Eu</i> CGD)	290	34	50	$3e^{-38}$
<i>Klebsiella pneumoniae</i>	WP_023290705.1	337	35	53	$9e^{-22}$
<i>Enterococcus durans</i> BDGP3	WP_081133590.1	323	48	70	$7e^{-117}$
<i>Microbacterium</i> sp. 5-2b	LC387599 (CarB; α -subunit of <i>Mi</i> CGD)	348	29	49	$2e^{-41}$
<i>Arthrobacter globiformis</i> NBRC12137	WP_003803559.1 (AgCarB1: α -subunit of <i>Ag</i> CGD1)	367	26	45	$5e^{-31}$
	WP_003803556.1 (AgCarB2: α -subunit of <i>Ag</i> CGD2)	352	26	43	$8e^{-33}$
<i>Microbacterium trichothecenolyticum</i> NBRC15077	WP_003803556.1 (MtCarB: α -subunit of <i>Mt</i> CGD)	352	27	45	$8e^{-33}$
<i>Arthrobacter</i> sp. Rue61a	WP_014921466.1	366	27	45	$4e^{-31}$
<i>Microbacterium pygmaeum</i> DSM23142	WP_091486435.1	369	27	44	$1e^{-31}$
<i>Agromyces flavus</i> CPCC202695	WP_092669325.1	357	27	43	$5e^{-32}$
<i>Sphingobium yanoikuyae</i> S72	WP_010338599.1	335	30	49	$3e^{-47}$
<i>Leifsonia</i> sp. 98AMF	WP_089877376.1	323	29	49	$8e^{-43}$

B. Homolog enzymes of β -subunit of *Pu*CGD

Organism	Accession No. (Enzyme name)	Length (aa)	Identity (%)	Similarity (%)	E-value
PUE strain	BBG22494.1 (DgpB; β -subunit of <i>Pu</i> CGD)	142	100	100	$6e^{-109}$
<i>Eubacterium cellulosolvens</i>	WP_004604488.1 (DfgB; β -subunit of <i>Eu</i> CGD)	147	38	56	$9e^{-23}$
<i>Klebsiella pneumoniae</i>	WP_023290704.1	130	25	44	$5e^{-38}$
<i>Enterococcus durans</i> BDGP3	WP_081133591.1	143	43	63	$2e^{-45}$
<i>Microbacterium</i> sp. 5-2b	LC387600 (CarC; β -subunit of <i>Mi</i> CGD)	132	30	52	$2e^{-15}$
<i>Arthrobacter globiformis</i> NBRC12137	WP_003803560.1 (AgCarC1; β -subunit of <i>Ag</i> CGD1)	131	27	50	$4e^{-10}$
	WP_003803554.1 (AgCarC2; β -subunit of <i>Ag</i> CGD2)	132	30	50	$2e^{-14}$
<i>Microbacterium trichothecenolyticum</i> NBRC15077	WP_045298609.1 (MtCarC; β -subunit of <i>Mt</i> CGD)	131	29	48	$2e^{-11}$
<i>Arthrobacter</i> sp. Rue61a	WP_014921465.1	131	27	50	$3e^{-10}$
<i>Microbacterium pygmaeum</i> DSM23142	WP_091486431.1	131	32	51	$4e^{-15}$
<i>Agromyces flavus</i> CPCC202695	WP_092669324.1	134	28	44	$1e^{-10}$
<i>Sphingobium yanoikuyae</i> S72	WP_069336345.1	130	34	54	$1e^{-24}$
<i>Leifsonia</i> sp. 98AMF	WP_089877379.1	138	34	54	$2e^{-16}$

Supplementary Table 2. SEC-MALS analyses of *Pu*CGD, *Eu*CGD, *Mi*CGD, *Ag*CGD1, *Ag*CGD2, and *Mt*CGD.

	Mw (calcd: α -subunit+ β -subunit)	Mw (Obvious)	Solution state
<i>Pu</i> CGD (DgpB-C complex)	52,133	214,400 ($\pm$ 0.52%)	heterooctomer
<i>Eu</i> CGD (DfgA-B complex)	54,432	216,600 ($\pm$ 1.29%)	heterooctomer
<i>Mi</i> CGD (CarB-C complex)	55,130	49,030 ($\pm$ 0.54%)	heterodimer
<i>Ag</i> CGD1 (<i>Ag</i> CarB1-C1 complex)	56,209	58,100 ($\pm$ 2.05%)	heterodimer
<i>Ag</i> CGD2 (<i>Ag</i> CarB2-C2 complex)	56,663	54,170 ($\pm$ 5.63%)	heterodimer
<i>Mt</i> CGD (<i>Mt</i> CarB1-C1 complex)	56,043	57,610 ($\pm$ 2.34%)	heterodimer

Supplementary Table 3. Data collection and refinement statistics of X-ray crystallographic data.

	<i>Pu</i> CGD-apo (PDB 7EXZ)	<i>Eu</i> CGD -apo (PDB 7EXB)	<i>Ag</i> CGD2 apo (PDB 7DNM)	<i>Ag</i> CGD2 with homoorientin (PDB 7DNN)
Data collection				
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 6 ₄ 22	<i>P</i> 2 ₁ 2 ₁ 2	<i>P</i> 2 ₁ 2 ₁ 2
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	86.5, 156.7, 158.9	139.0, 139.0, 227.0	71.7, 100.6, 136.3	73.9, 102.5, 136.1
α , β , γ (°)	90, 90, 90	90, 90, 120	90, 90, 90	90, 90, 90
Resolution (Å)	46.9-2.50 (2.55-2.50)*	47.1-2.40 (2.47-2.40)*	47.2-2.30 (2.42-2.30)*	47.97-2.25 (2.37-2.25)*
<i>R</i> _{merge} (%)	12.2 (81.5)	14.2 (144.9)	4.2 (26.5)	5.4 (85.2)
$\langle I / \sigma I \rangle$	12.0 (2.1)	27.4 (2.8)	28.2 (7.7)	17.5 (2.0)
Completeness (%)	100.0 (100.0)	100.0 (99.9)	96.6 (91.6)	99.8 (99.9)
Multiplicity	6.9 (5.9)	36.9 (22.8)	6.8 (6.6)	4.2 (4.2)
Refinement				
Resolution (Å)	46.9-2.50	47.1-2.40	47.2-2.30	36.2-2.25
No. reflections	75,414	51,231	81,711	94,646
<i>R</i> _{work} / <i>R</i> _{free}	0.182/0.233	0.178/0.203	0.180/0.223	0.196/0.247
No. atoms				
Protein	14584	3,567	7,372	7,319
Ligand/ion	183	21	4	36
Water	475	329	178	173
<i>B</i> -factors (Å ²)				
Protein	41.6	41.6	40.6	55.8
Ligand/ion	47.3	57.6	35.1	46.2
Water	38.5	45.1	39.2	49.0
R.m.s. deviations				
Bond lengths (Å)	0.008	0.007	0.008	0.008
Bond angles (°)	0.948	0.899	0.996	0.935

*One crystal was used for collection. Values in parentheses are for highest-resolution shell.

Supplementary Table 4. Data collection and refinement statistics of cryo-EM data.

	<i>Pu</i> CGD (EMDB-30808) (PDB 7DRD)	<i>Eu</i> CGD (EMDB-30809) (PDB 7DRE)
Data collection and processing		
Microscope	Talos Arctica	Talos Arctica
Voltage (kV)	200	200
Detector	Falcon 3EC	Falcon 3EC
Magnification	120,000	150,000
Pixel size (Å) (calibrated)	0.88 (-)	0.69 (0.676)
Automation software	EPU	EPU
Total exposure (e-/Å ²)	50	50
Exposure rate (e-/Å ² frame)	1.02	0.81
No. of frames	49	62
Defocus range (µm)	-1 to -2.5	-0.6 to -1.5
Symmetry imposed	D2	D2
No. of collected micrographs	2,122	1,664
No. of selected micrographs	1,963	1,614
No. of particles for 2D classification	857,817	330,866
No. of particles for 3D classification	597,534	232,605
No. of particles for 3D refinement	56,924	60,587
Map resolution (Å)	2.85	2.54
FSC threshold	0.143	0.143
Refinement		
Initial model used (PDB code)	none	none
Model composition		
Non-hydrogen atoms	12,709	12,604
Protein residues	1604	1572
Ligands	0	0
<i>B</i> factors (min/max/mean, Å ²)		
Protein	28.6/86.9/48.8	18.0/51.9/31.2
Ligand	0/0/0	0/0/0
R.m.s. deviations from ideal values		
Bond lengths (Å)	0.004	0.005
Bond angles (°)	0.598	0.554
Validation		
MolProbity score	1.24	1.36
Clashscore	4.64	4.6
Poor rotamers (%)	1.48	1.49
Ramachandran plot		
Favored (%)	98.2	98.7
Allowed (%)	1.8	1.3
Disallowed (%)	0	0

Supplementary Table 5. Primers used in this study.

Primers used for cloning of <i>MiCGD</i>, <i>AgCGD1</i>, <i>AgCGD2</i> and <i>MtCGD</i>, and construction of mutants	
<i>CarB</i> Fw	TAAGAAGGAGAGATATACATATGACCGCCCTCGCCCTGCCCCGCCTCG
<i>CarB</i> Rv	TTGTCGACGGAGCTCGAATTCTCATCGCACGCCCGCCTCGGCCTCG
<i>CarC</i> Fw	TAAGAAGGAGAGATATACATATGTTTCAGAGAACGCATCATCGTTCAGGACTCG
<i>CarC</i> Rv	TTGTCGACGGAGCTCGAATTCTCATCGGGTGAGCTCCTTGACGTCCATG
<i>MiCGD</i> Fw	TAAGAAGGAGAGATATACATATGTTTCAGAGAACGCATCATCGTTCAGGACTCG
<i>MiCGD</i> Rv	GTGGTGGTGGTGGTGTCTCGAGTCGCACGCCCGCCTCGGCTCGA
<i>AgCGD1</i> Fw	TAAGAAGGAGAGATATACATATGATCGCCGACCGCATAATCGAGCAC
<i>AgCGD1</i> Rv	CTCGAGTGCGGCCGCAAGCTTGCCGCGGACGGGCTGGGCGA
<i>AgCGD2</i> Fw	TAAGAAGGAGAGATATACATATGTCAGAAGGCATCGCCGGCTCG
<i>AgCGD2</i> Rv	CTCGAGTGCGGCCGCAAGCTTTGCGAGAACGACCTTGCGCTCG
<i>MtCGD</i> Fw	TAAGAAGGAGAGATATACATATGATCCCCGATCGCATCATCGAGCAGG
<i>MtCGD</i> Rv	CTCGAGTGCGGCCGCAAGCTTTGCGTGTGCTCCTTCGGTGTGTG
<i>MiCGD</i> ^α E145A-F	ATGGGCATGGcGATCCACAGC
<i>MiCGD</i> ^α E145A-R	GGTCAGATCGAGCTTCTC
<i>MiCGD</i> ^α H147A-F	CATGGAGATCgcCAGCCCGCAC
<i>MiCGD</i> ^α H147A-R	CCCATGGTCAGATCGAGC
<i>MiCGD</i> ^α D177A-F	TTCATCCCGGcCTGGGGCGCC
<i>MiCGD</i> ^α D177A-R	TCCGAGATGCGGGGATCCG
<i>MiCGD</i> ^α H275A-F	GTGCACGTGgcCGGCAAGTTC
<i>MiCGD</i> ^α H275A-R	GACCCAGGGCATGATCGC
<i>MiCGD</i> ^α E311A-F	ATTTCAGTGcGTGGGAGGGCTGG
<i>MiCGD</i> ^α E311A-R	GTAGCCGGTGTAGCCCGC
<i>MiCGD</i> ^α E313A-F	AGTGAGTGGGcGGGCTGGCAC
<i>MiCGD</i> ^α E313A-R	GGAAATGTAGCCGGTGTAGC
Primers used for cloning of <i>dgpA</i>, <i>PuCGD</i>, <i>EuCGD</i> and <i>dfgE</i>, and construction of mutants	
<i>DgpA</i> -F	AAGGAGATATACATATGTCCAAGCTGAAAATTG
<i>DgpA</i> -R	CTCGAGTGCGGCCGCAAGCTTTCAGAACTTAATG
<i>PuCGD</i> ^β -F	GAAGGAGATATACATATGGGCTTAGCGTTACG
<i>PuCGD</i> ^β -R	CTTTTATCTCCTCCCTAATTCTGCACATTC
<i>PuCGD</i> ^α -F	AGGGAGGAGATAAAAAGATGTCTAACGTGAAAC
<i>PuCGD</i> ^α -R	GTGCGGCCGCAAGCTTGTCCACAAAGTTGTG
<i>EuCGD</i> ^α -F	GAAGGAGATATACATATGTATCGTTATGAGAAG
<i>EuCGD</i> ^α -R	TTATCTCCTCCCTAGTCTCCTAATAAGTTC
<i>EuCGD</i> ^β -F	CTAGGGAGGAGATAAAAAGATGGAGAAACAGG
<i>EuCGD</i> ^β -R	GTGCGGCCGCAAGCTTCACGATCAACAG
<i>DfgE</i> -F	GAAGGAGATATACATATGGCGCAGGACATTCG
<i>DfgE</i> -R	GTGCGGCCGCAAGCTTAAAGCTTCTCGAAGTC
<i>PuCGD</i> ^α E141A-F	CGTCAAAGTTGGTATTgccGTGCATAATCCG
<i>PuCGD</i> ^α E141A-R	GGATTATGCACggcAATACCAACTTTGACG
<i>PuCGD</i> ^α H143A-F	GTTGGTATTGAGGTGgccAATCCGGAAACC
<i>PuCGD</i> ^α H143A-R	GGTTTCCGGATTggcCACCTCAATACCAAC
<i>PuCGD</i> ^α D173A-F	GGTTGATCCCCgccTTTGGCTGTTTTGCCAAC
<i>PuCGD</i> ^α D173A-R	GTTGGCAAAACAGCCAAAggcGGGGATCAACC
<i>PuCGD</i> ^α H263A-F	CCATATTGCATCCACATGgcgGGCAAATATCACTACATG
<i>PuCGD</i> ^α H263A-R	GTAGTGATATTTGCCcgcCATGTGGATGCAATATGGAATC
<i>PuCGD</i> ^α K265A-F	GCATCCACATGCATGGCgccTATCACTACATGTACG
<i>PuCGD</i> ^α K265A-R	CGTACATGTAGTGATAggcGCCATGCATGTGGATGC
<i>PuCGD</i> ^α E301A-F	CGTGAGCGAATACgcgGAATACAACTCAGG
<i>PuCGD</i> ^α E301A-R	GAGTTGTATTCgcGTATTCGCTCACGATATAGC

The underlined letters represent the restriction sites.

Supplementary Table 6. Sequences of enzymes.

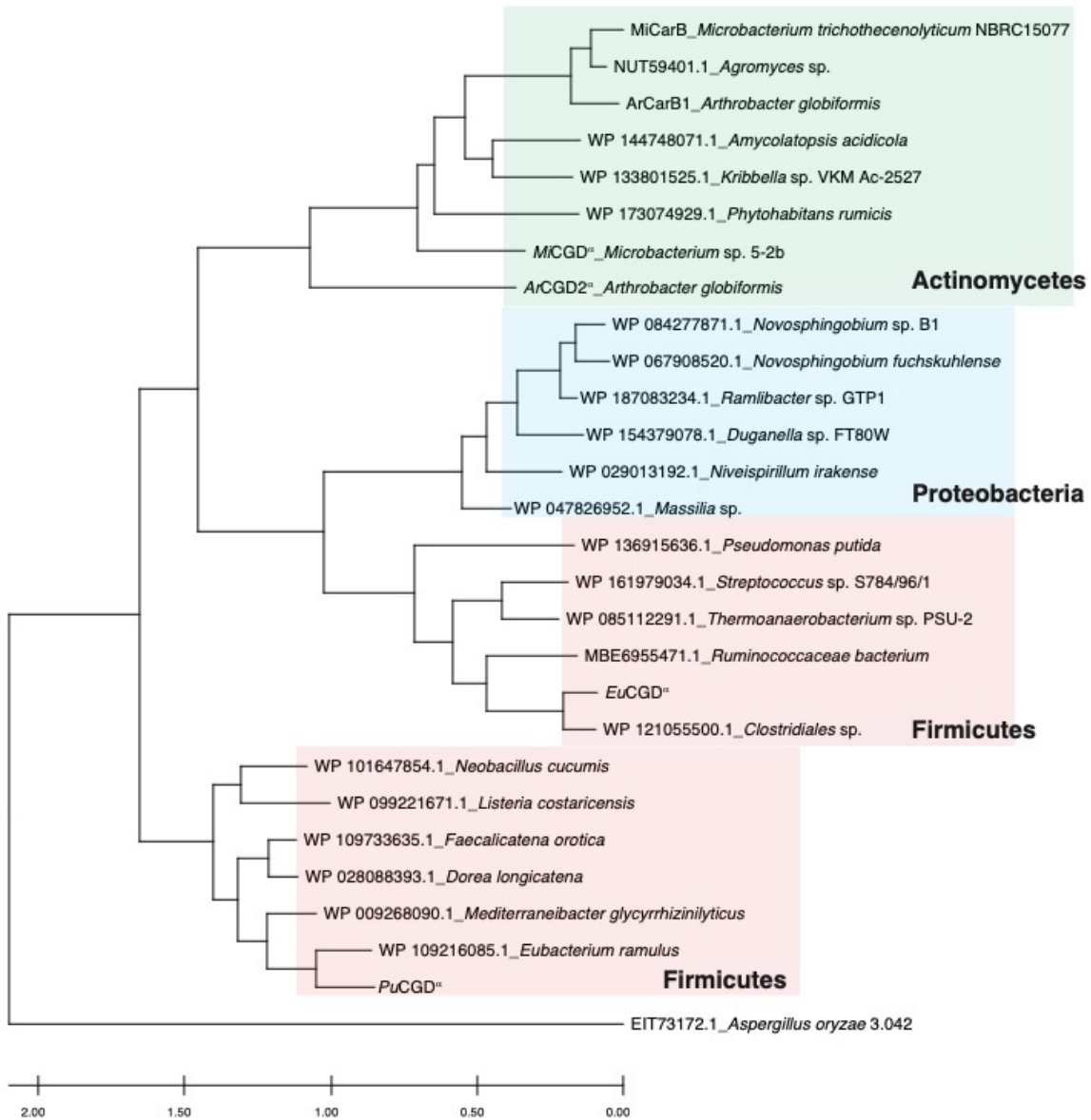
Nucleotide sequences	
<i>MiCGD^α</i>	1 ATGACCGCCCTCGCCCTGCCCCGCCTCGGCGTGACGCTGTACAGCTTCACGCCCTACTACCACGCGGTGAGTAC 76 TCCTTCGAGGACCTCATCCGCATCGCCGGAGAGCGGACCTCGGCCCGGCCTCGAGATCGTCGGCTACCAGAGC 151 ATCAAGGGCTTCCCGAAGCTACCTGACGGCTTCGTGAAGGACTTCGCCCGCGGGTCGACGAGGCCGGTCTCGAA 226 CTGAGCGCGATGGGCGCAACGCCGACGCGGGCATCCCGCACGACCGGTGCTGAACGAGGACGAGCTACCGAG 301 TACATGGCCACACGCTGCACACCGCCAAGGAGCTCGGCTTCCCGATCGTCCGCGTGACGACTCGGTACGCCC 376 GACCTCATGGAGCGCTGCTCCCCCTCGCGGAGAAGCTCGATCTGACCATGGGATGGAGATCCACAGCCCGCAC 451 AGCGTGACACCCGAAGATCCAGGCTCTCGTCGAGCGCTACGAGAAGCTCGGATCCCGCATCTCGGATTATC 526 CCGGACTGGGGCGCCAGCTCACCCGCTGCCCCGAGCGCGCTGCAGACCTATGCCGCGGCCGACGTACCGCGG 601 GAGCTGCTGACGCCTACGACCGGCAGTGGGAGCTGTTCCACGCCGAGGGCGTCATACCACCGACGCCGAGCAG 676 GGCGCCAGTTCCGTGCGATGCGCGAGCTCAACGAGCGGTTCCGGGGTGACGACGTCTCGGTGCGCATCGGCACG 751 AACGCGGTGGGCTGTTCCGCCACCGCGGCCAGAGGACTGGTCGGCGATCATGCCCTGGGTCTGTCACGTGCAC 826 GGCAAGTTCTACGGCATCGACGAGAAGCGGAGAGCCCTCCGTGCCGCACGGCTGCTGCTGCCAGCTCGTG 901 GACGCGGCTACACCGGTACATTTCCAGTGAGTGGGAGGGCTGGCACTGGAACACGACCGACGACCCGTTTCGAG 976 ATGGTCGCCTGGCAGCACCGGCTCATGCGCAGGATCTTCGGTGAGATCGAGGCCGAGGCGGGCGTGCGATGA
<i>MiCGD^β</i>	1 GTGTTACAGAAACGCATCATCGTTCAGGACTCGCTCGTCGCCGACGAGTCGGGATCCGCCTTCCGGATCCGCCTG 76 CCCTGGTACCGCGCGCTGCCGCTCAGCACGATCGAGGAGCTCTCGGTACGGTCGACGGCACGGCCTTCGACCCT 151 GCACGGCTACGGATCGCGGTGAACGACGGCGAGTGGGCCCTGGCCGAGGCGCAGCTGCGCACGGACGACGTCTGG 226 TTCGTCTGGACGATGCGACCGTGGCGCTGCCCGCCTCGTGCTCGACGCCGGTGCACGAAAGTGCAGGCCACG 301 CTCTCGATGCGCATCCCCTATCTGCCGGTCGCCGCAAGCCGCTCTCGATGGCCGAGACCGACCAGAAGCGCATG 376 GACGTCAAGGAGCTACCGCATGA
Protein sequences	
<i>MiCGD^α</i>	1 MTALALPRLGVTLYSFTPPYHAREYSFEDLIRIAGERDLGPGLIIVGYQSIKGFPKLPDGFVKDFRRRVDEAGLE 76 LSAMGANADAGIPHDRLLNEDELTEYMAHQLHTAKELGFPPIVRVQHSVTPDLMERLLPLAEKLDLTMGMEIHSPH 151 SVHHPKIQALVERYEKLGSPLHGFIPDWGASLTRLPPSALQTYAAADVPRELLDAYDRQWDVFHAEGVITTDAAEQ 226 GAQFRMRRELNERFGDDVSVRIGTNAVGLFGHQRPEWDSAIMPWVHVHGKFGYIDENGEEPSVPHGLLLRQLV 301 DAGYTGYSSEWEGWHWNTTDDPFEMVAWQHRLMRRIFGEIEAEAGVR
<i>MiCGD^β</i>	1 VFRERIIVQDSLVADESGSAFRIRLPWYRALPLSTIEELSVTVDGTAFFDPA RLRIAVNDGEWALAEQLRTDDVW 76 FVLDDATVRLPGLVLDAGAHEVQATLSMRIPYLPVAGKPLSMAETDQKRM DVKELTA

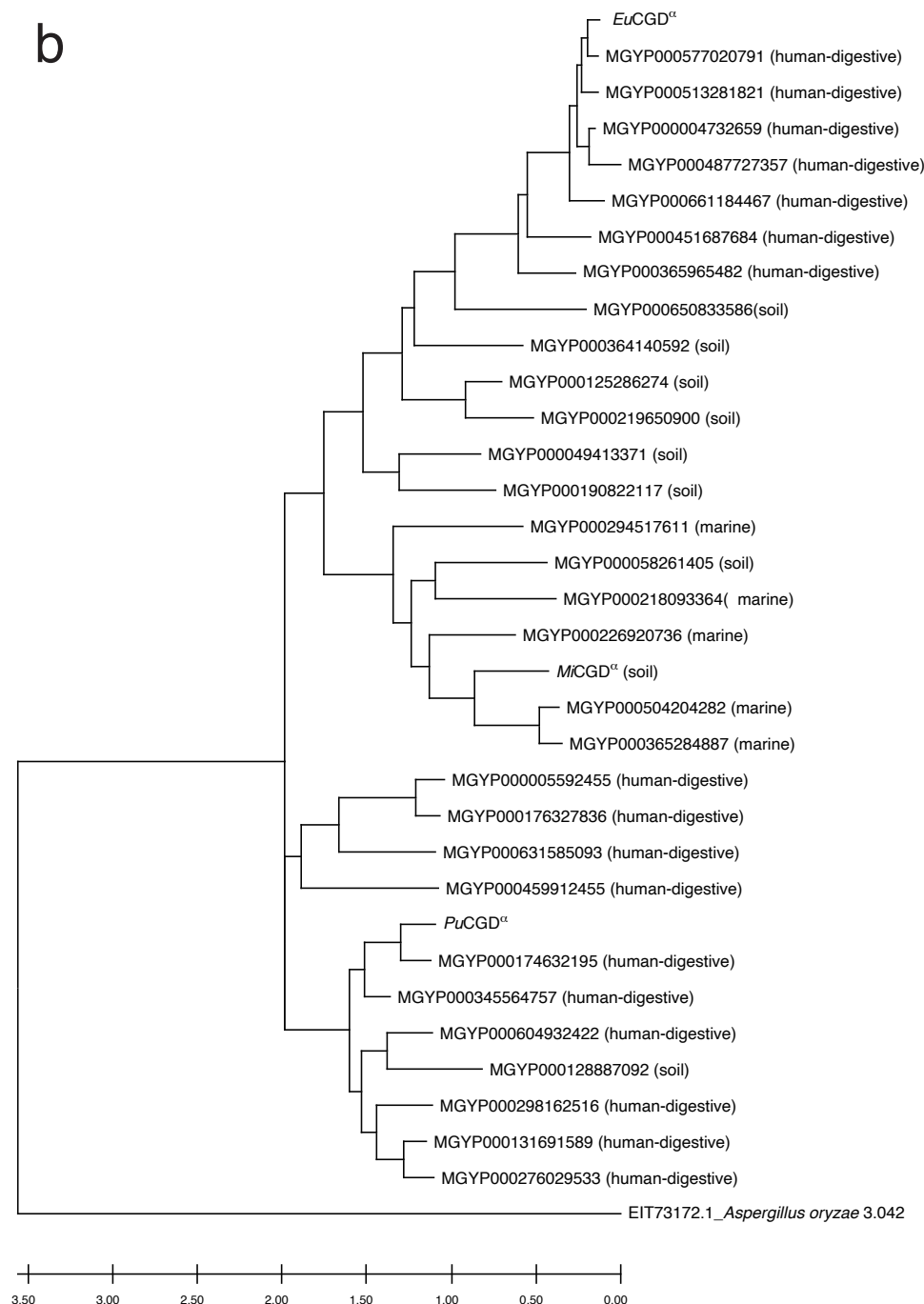
codon optimized nucleotide sequences for expression in <i>E. coli</i>	
<i>PuCGD^α</i>	ATGTCTAACGTGAAACTTGGCGTAACCCTGTATAGCTTCAGTACGGAGTATTGTCAGGGTAAAAT GACCTTAGAGGATTGTATTCGGACAGCGAAAGAACTGGGTGCAGCTGGTTTCGAGATTGTCGCC ACACAGATGATCCGTCCTATCCGTATGTGTCGGACAAATTCCTGGGTGAGCTGAAAAGCATTG CCAGTACTATGACATGGAACCTGTTTGCTATGGGGCTAACTGTGATCGCGGTCTGCGTGGCGATC GCAATCTGACGGGTGATGAAATGGTCGCAATGGCGGTACGCGATATCAAGAATGCGCATAAAAT GGGCTGCAAGGTTGTTTCGCGAACAGTGGCTCATGGGTCCGGAACCTTTGCCAAATTAGCGCCT TTTGCCGAACACTATGGCGTCAAAGTTGGTATTGAGGTGCATAATCCGGAACCCCGATTACGCA ATCCACCAAAGACTACATTGCCGCTATCGATAAACTGGCAGTAAATACCTTGGGTTGATCCCCGA TTTTGGCTGTTTTGCCAACAAACCGAACAAGATGAATTGGGATAATGCGCTCGCAGATGGAGCG GATAAGAACTGCTGGAAATGGCCCGTGATATGAAATACGATAATGTACCTATGACGAAGCGGT TAAACGCCTGACTGCAGCAGGGCGAAGAAAGTGAATTGACGACCATGCGCGATATGTATACC TTTCTGACCTTCAAGAAAGACGTGTCTGCTGAGCTCCAAGGACTGAAAGATATGATTCCATATTG CATCCACATGCATGGCAAATATCACTACATGTACGAAACTTACAGGAAGCGGCCATTCCATACGA TGACATCATGAAAATCGTGTGAGAATCGGACTATGACGGCTATATCGTGAGCGAATACGAGGAAT ACAACTCAGGCCATAGCATTGAAATGTTGCGTCGTCATCTGAAAATGATGCACAACTTTGTGGAC
<i>PuCGD^β</i>	ATGGGCTTAGCGTTACGCCTGAACCTTTGTGGATGTCGTTTGCGATGACAGCTTGAAGAACTTTTG GGCGAATGGGAAGAAAATCGGCTATCAGTTCGATGTTTCGTCTGAGTTACTATCGCGGTCATTTCC TGTCGACCATGACGAAATTGGCGTCAAAGTGGATGGTGTGGATGTACCGGCAGAGAACATTTTC GCTGTGTCTTGATGGGAAGAGTATGGCGTAGCTGAACTGCATGATCTGGTTAACGTCTTTTGGC CGATTATCGAACCAGCCACTATCAAAGTGTTCACCTGGTGGTTTGAGCGAAGAAGAACACGA TGTTGACTTTACGCTGTACTTTCTAGTCCGTACATGGCCCTCTCAGAAACCGAGTATCAGTCCAT TGACTCTTGCGGAAGCAAACGCCTGAATGTGAGAAT
<i>dgpA</i>	ATGTCCAAGCTGAAAATTGGTATTATCGGTTGCGGAGGCATTGCTAACCAGAAACACTTCCCTGC GTTAAAGAACAATGCTGATCTGAATGAGATTGTAGCCTTTTGTGACATCCAGATCGACCGCGCGG AGAAGGCTGCTGCGGAGTTTGGTGCAGAGGGAGCTCAGGTGACAGCTGACTATAAAGAATTGC TCGCGAACCCGGAAGTCGAAGTTGTACACGTCTGCACACCAAACGTTTCCCACTCTGAAATTACC ATCGCCGCTTTTGAAGCTGGGAACACGTATACTGTGAGAAGCCCATGAGTCACTCAACAGAGG AAGCCGAGAAGATGGTTGAGGCCTGGAAGAAATCCGGTAAACAATTTACAATTGGTTACCAAAA CCGTTTTCTGTGAGGAAGTCATGAACCTCAAGAAAAGCTGCGATAAAGGGGAACTCGGTGAGAT CTACTACGGTAAAGCGCATGCAGTACGTCGTCGTCGCGGTTCCGACTTGGGGTGTGTTTATGGACA AGGAGGCTCAGGGAGGCGGACCGCTCATCGACATTGGCACGCATGCCTTGATATCACCTTATG GTGCATGAACAACTACGATGTTGACAGCGTGACGGGCTCTGTGTTCTATAAACTGGGACAGAAA GAGAACGGACCGGAAGGAACTTGTTCGGCCCGTGGGACCCGAAAACATTTGAAGTGGAGGA TAGTGCTGTGGGATTCTGTAATGAAGAATGGAGCTACTATTGGCCTTGAAGCGTCCTGGGCA ATCAACATGCTCGATTGCGCGAAGCATCTACGACTCTGTGTGGGACCGAGGCTGGGGCAGAGA TTCATAGTGGCATGTCATACCCAAAGAACGAGTTGATCTACAATCGCGCACGTAATAACCAGCTTA TGGAAGAAACCTCAGTTCGGTGGGTTGATTGCTTACTTTGCGGGAGGAGCGGGCGAAGAAG GTACGGTTGACAATCGCCAGTGGTTGGAGGCGATCCAAAACGGGACTGAGCCCTTGTGAAAC CCGAAGAGGCCCTGGCAGTTACAAAGATCTTAGACGCCATCTATAAAGCGCAAAGACGAATGA GACCATTAAGTTC
<i>EuCGD^α</i>	ATGTATCGTTATGAGAAGAAGGGCCCTAAGCGCGGAGTTGCGTTATACTCGTACTCAGCCGAATT TGGGTAACTAAGACGTTGGAAGATTGTTTCGAGGATCTTCACGATATGGGTGCTCACGGTATCG AGATTTTAGCAAACACTCACATTGAAAACATCCATACCCGACGGATGAATGGGTGGAGAAATGG TGCGCCTTGTGTGACAAATACGAAATCGTTCCAGTGGAGTACGGAACTGGATCGACTCTCATGT ACTGGGAGATCGTGACCTGACTACCGAGGAGTCCGTTGAAATGCTCAAACGCGACATTGCTTG GCGCACCGTCTCGGTTTACCGTCATGCGTACCAAGATGCCGTCATTAATGACCTCTTAGAGCC GGTAGAAAACCTGGAAGGAAATCATCAAAGGTGCCCTGCCTCTCGCAGAAGAGTTAGGCATCAA GATGTGTCCGGAGATCCATACGCCATCAAATTTGAAGGGAAAGCTGGTGAATGACTTCGTAGAG TTTATCAAGGAGACAGGAACGAAGAACTTTGGTCTCAACATCGATTTTAGCGTCTTTCGCACCG TTTTGCTGAAGGTGAATGGGTGGACCCAAATTACACACCGAATAAGCCAGAAGACATTATCCGC TGCTGCCCTATGTGTATTGCTGTCATGCTAAATTTATCCACATGAGTGACGATTTTAAAGAAACCA

	CATTCCCTACGAGGAGGTAGTCAAGACCATGGAAGATAACGGTTACGAGGGCTATTTGTTATCTG AATACGAAGGCGCTGATAAATACGATGAGGGATACGAGGTAGGCCAGACACTCCGTAAACATCA TATCTTACTCAAGAACTTATTAGGAGAC
<i>EuCGD^β</i>	ATGGAGAAACAGGTAATTCAGTCTGTAGGCTTCCGTAACATTAAGAACGGTAATGGCGAGATTAC AGGCTTTCAGTTCAAGGTTAACTTCCGTATTATCGCGGCGTCTTCTTATCCCAAATCCGCCCAGG GACTTTGTTCTGTCGACGGCCAGAAAATCGAGAAAGACCAATTACCTGGACCATCAATGGCGAG GAGTACACGAACCAGGAAATGCGTGGTGATTTCAGACCCACTGGGCCACAACCAAACCCGCG GTAAGTAAAGTAAAATGCCAGGTGGGTAGCACAAGGGTATCATGACCTGAAATACGGATTTT GTTTTACCAGCTCGTATATGCCGCCGATTATCCAAGATGGCCTGGACCCTGATAAAGAATCAATGG TGACATGCCAGAATTTGGGCATCACGTGAATGAACGCCGTCTGTTGATCGTG
<i>dfgE</i>	ATGGCGCAGGACATTCGCATCGCCATTGTGGGCCACGGTTTCATGGGACACGAGCATGAGAAAA TGCTGTATGATTTTCCCGGTATTCGTCTGGTAGGTTTTAGCGATAAAGATCCTGCGCAATTAGAAG ACGTTAAAGAGGGTCTTAAACGCTATCAATCGAATGAGGAGTTATTCGCCGATCCAGAAGTGGA CGTCGTCATCATTGCGGCGAACAACAATCAGCATCGTGAAGTGGTGATTAGGCGCGAATGCCG GTAAGAATATCATTTGTGAAAAGCCGGTGGCGATGTCTCTGCAGGAGCTGGAAGAGATGGAAG AGGCTGTAAAGCGCAATGGAGTAAAATTTACGGTACATCAACAGCGTCGCTTTGATCCTGACTTT CGTACCGCGAAAGCCGTTTATGATTCGGGCACATTGGGTGAAGTGTATGCCATCAAGAACCAACT GTACGGCTTCAATGGGAACATGCACGACTGGCACGTGTTTGTAGAAGAGGGTGGTGGAAATGCTT TATGACTGGGGTGTCCACCTGTTGGATCAGGTGTGTTTCATGTTCCCGGTGCGAAGTTAAAGTC AGTATTTGCGGACATTCGTAACGTGATCAATGATGAAGTCGATGACTACTTCAATATTATGTTACGC TTCGACAATCAGGTGACAGCCACCGTCGAGCTGGGTACATACTACTTGGCGGATAAAGTGCAGG ACAAATGGTTCGAACGCCACTGGTTCGTAGCTGGCAACAAGGGAACAGCTTACGTAGATGGGTT CTTTCCTGAAGGCAAATTTGTACGTACTACGCACTTGCTTACTAATGTGGGCGGTAAACGTACTAT GACTGCAGCTGGCCCGACGCGTAGCTTCGACCCGCCCGGAAGGGACTATCGTGACTGAGCC TCTCCGGAGGTGCATACCGAGCATCGCAATTACTTTGAGAATTATATCCGTGCGTACTATGGCGA AGAGGAGTTTTATGTTAAAATTTCCCGAGGTTAAGCGCGTACTTAAGTTAATGGACGCCGTTTCGTG AAAGCGGCCGCACGGGCAAATCAATTGACTTCGAG

Supplementary Figures

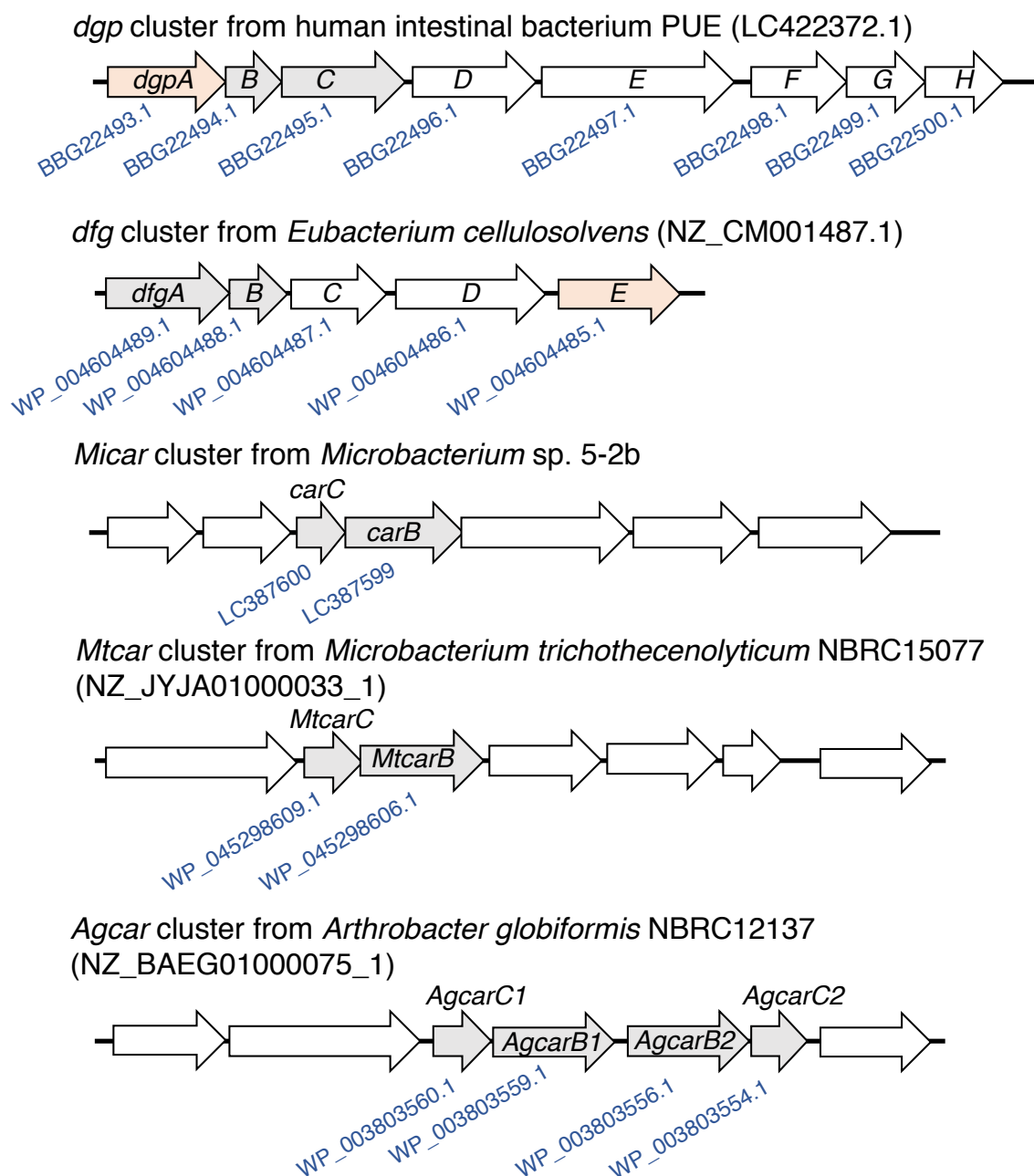
a





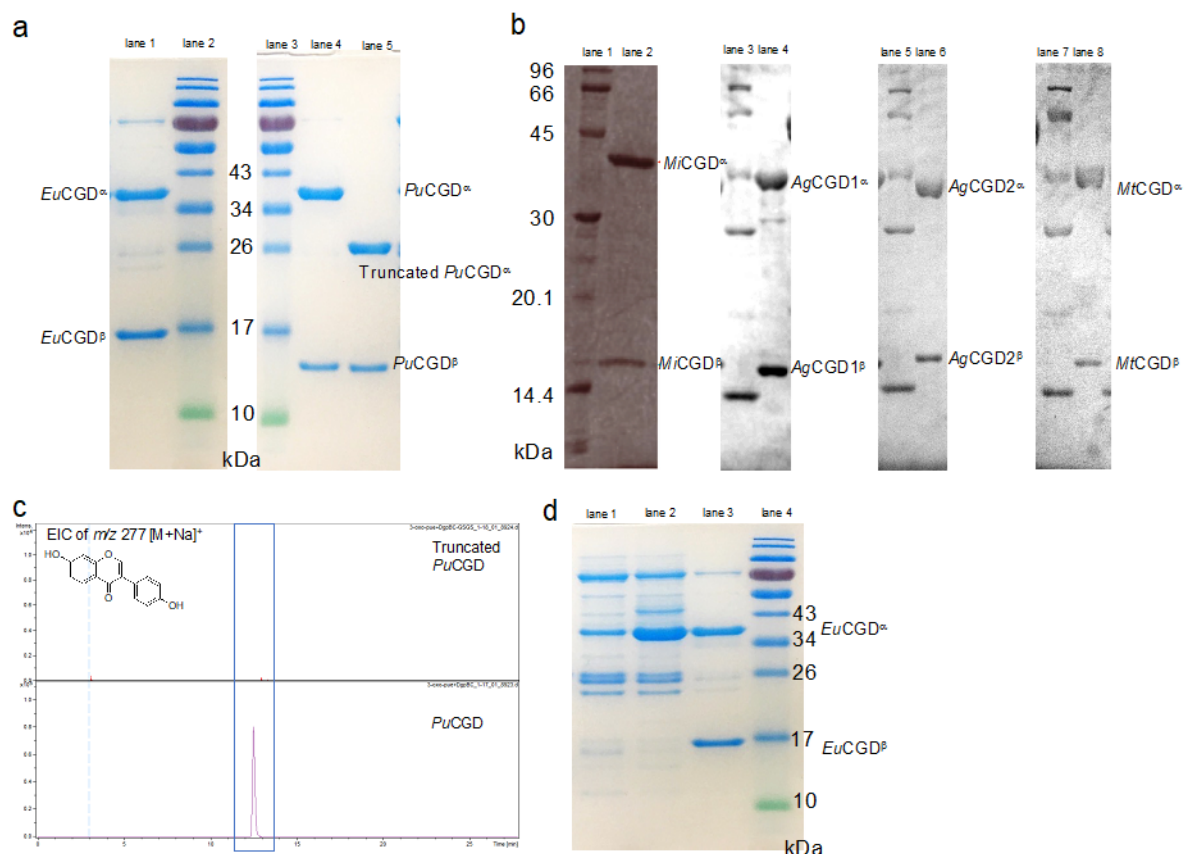
Supplementary Figure 1. Phylogenetic analyses of *PuCGD* and its homologs.

(a) Phylogenetic analysis with the enzymes identified in a search of the National Center for Biotechnology Information (NCBI) database. The amino acid sequences were prepared by BLAST analyses, using *PuCGD*^α, *EuCGD*^α, and *MiCGD*^α as queries. The proteins with more than 45% similarity with *PuCGD*^α were used for the phylogenetic analysis. (b) Phylogenetic analysis with the enzymes from metagenomic data. The amino acid sequences were prepared by a similarity analysis from the MGnify protein sequence database (<https://www.ebi.ac.uk/metagenomics/>), using *PuCGD*^α, *EuCGD*^α, and *MiCGD*^α, as queries. Sugar isomerase from *Aspergillus oryzae* was used as an outer group. The *PuCGD* homologs are widely distributed in both aerobic and anaerobic Actinobacteria, Proteobacteria, and Firmicutes species from soil, marine environments, and the human digestive system.

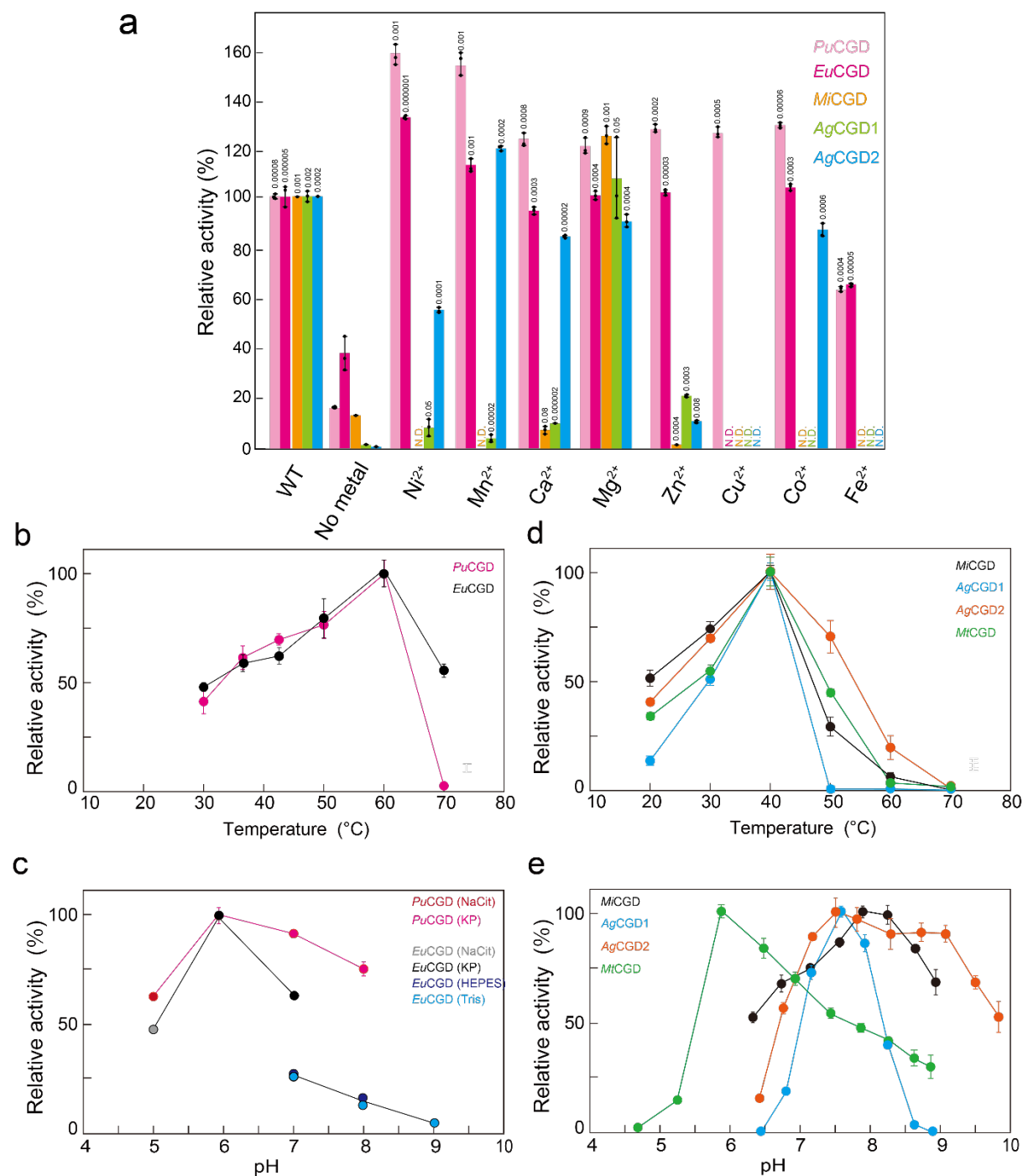


Supplementary Figure 2. Gene cluster analysis of *Pu*CGD homologs.

Analyses of gene clusters of various bacteria that contain a DgpB (*Pu*CGD^a) homologous enzyme gene. The genes of C-deglycosylases are colored gray and the genes of oxidoreductase enzymes are colored orange. The homologous genes of DgpC are located next to DgpB homologous genes in the genomes.

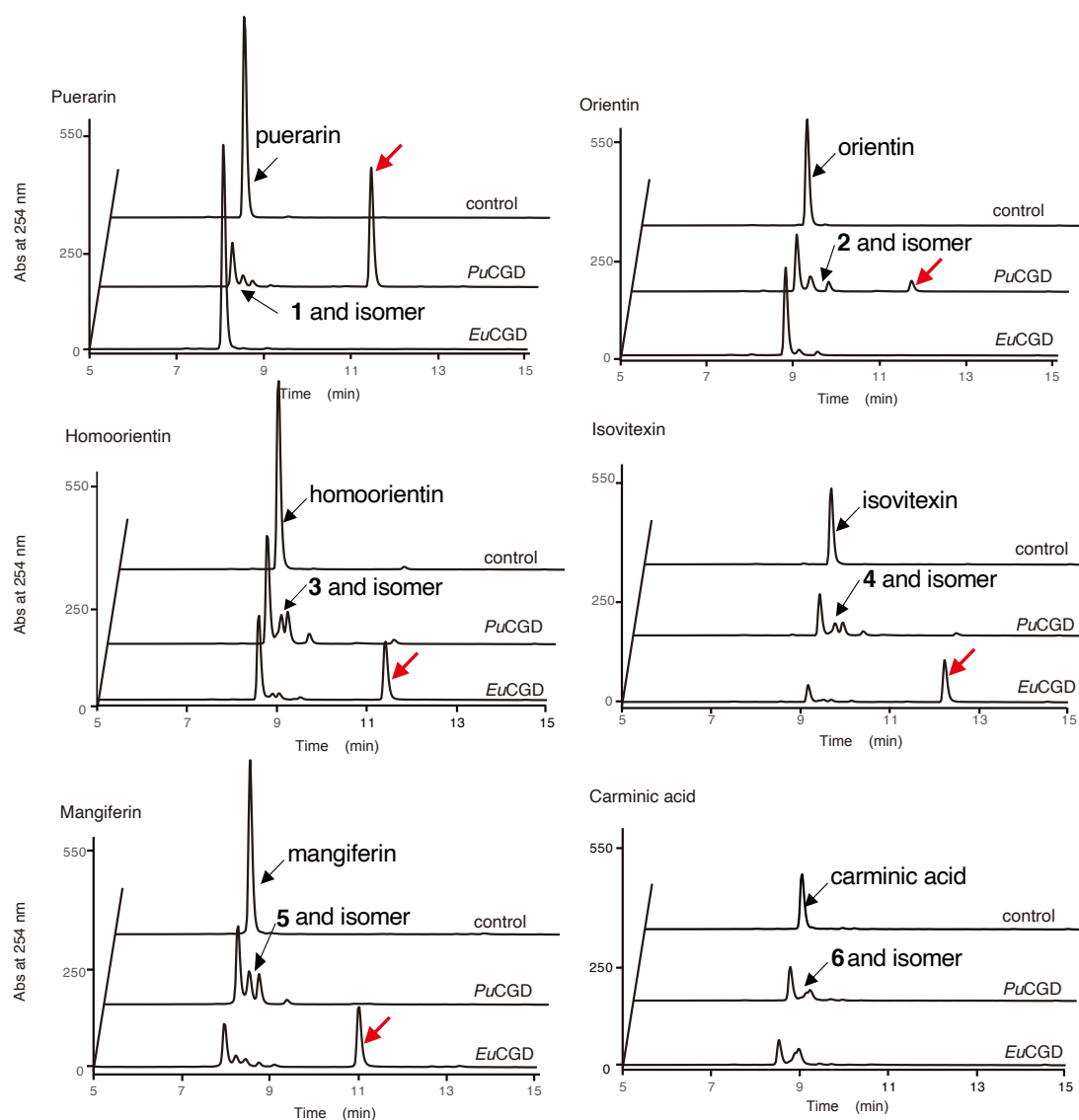


Supplementary Figure 3. Purification and gel-filtration analysis of C-deglycosylation enzymes. (a) SDS-PAGE analysis of purified *EuCGD* (DgpB-C), *PuCGD* (DfgA-B), and truncated *PuCGD* (lane 1, purified *EuCGD*; lanes 2 and 3, standard markers (PageRuler Prestained Protein Ladder (Thermo Fisher Scientific)); lane 4, purified *PuCGD*; and lane 5, N179-D244 lid domain truncated *PuCGD*). (b) SDS-PAGE of purified *MiCGD* (*MiCGD* $^{\alpha}$: CarB, and *MiCGD* $^{\beta}$: CarC), *AgCGD1*, *AgCGD2*, and *MtCGD* (lanes 1, 3, 5 and 7; standard markers (LMW Marker Kit (GE Healthcare)); lane 2, purified *MiCGD*; lane 4, purified *AgCGD1*; lane 6, *AgCGD2*; and lane 8, *MtCGD*). (c, d) Activity of the N179-D244 lid domain-truncated *PuCGD*. (c) Extracted ion chromatograms ($m/z=277$ [M+Na] $^{+}$) of the reaction mixture with truncated *PuCGD* and the reaction mixture incubated with wild type *PuCGD*. (d) SDS-PAGE of swapped complexes (lane 1, *PuCGD* $^{\alpha}$ -*EuCGD* $^{\beta}$; lane 2, *EuCGD* $^{\alpha}$ -*PuCGD* $^{\beta}$; lane 3, *PuCGD*; and lane 4, standard markers). Similar results of SDS-PAGE were obtained at least three times.



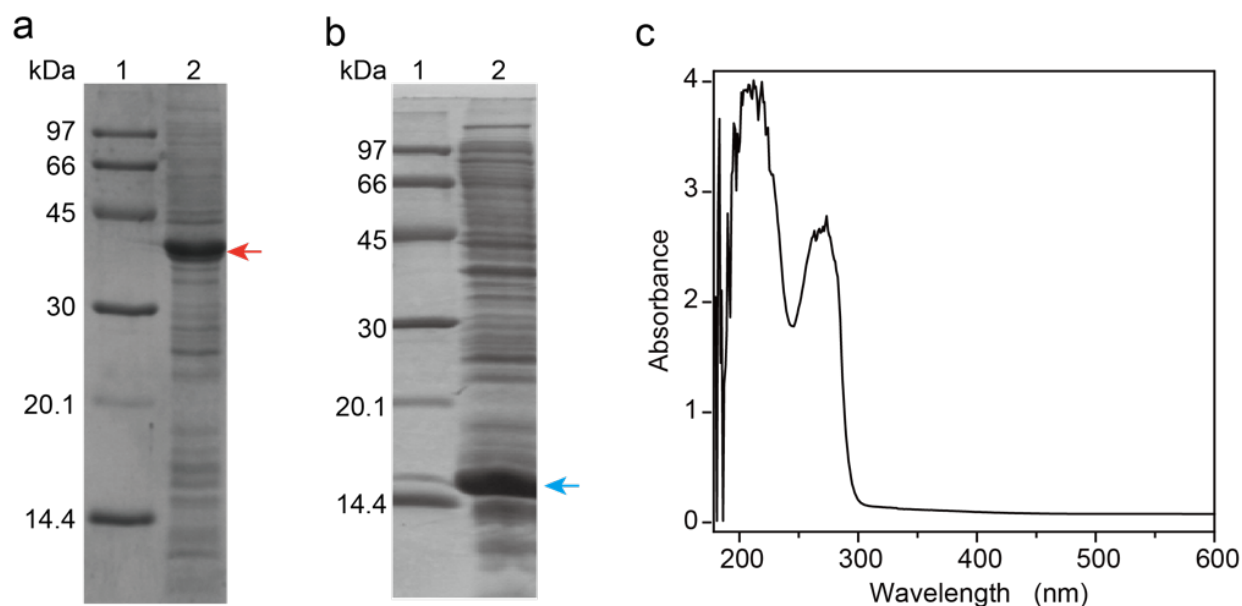
Supplementary Figure 4. Metal-dependency and optimization of C-deglycosylation reaction.

(a) Effects of metal ions on the enzyme reactions of C-deglycosylation enzymes. (b-e) Temperature and pH optimization of (b, c) the enzymes from intestinal bacteria and (d, e) the enzymes from soil bacteria. Statistical comparisons between means for the “No metal”, which is the chelate reagent-treated enzyme and WT (wild type) and each mutant were performed by Student’s t-test (2-tailed): *p*-values were shown over each bar. The bars are means of *n* = 3 independent experiments and error bars indicate standard deviations. Data are presented as mean values $\pm$ SD. All experiments were repeated independently three times with similar results.



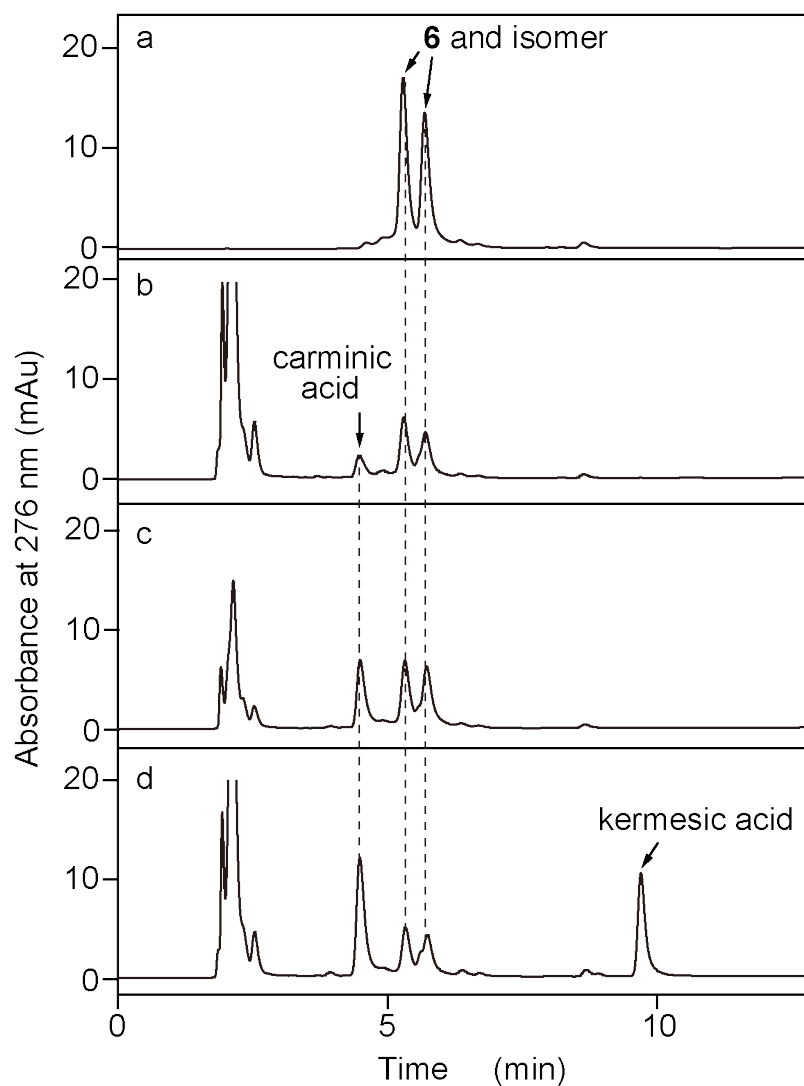
Supplementary Figure 5. HPLC analysis of the enzyme reaction products of CGDs from intestinal bacteria.

HPLC chromatograms of the C-deglycosylation reactions performed with enzymes from intestinal microorganisms. Red arrows indicate the production of aglycones.



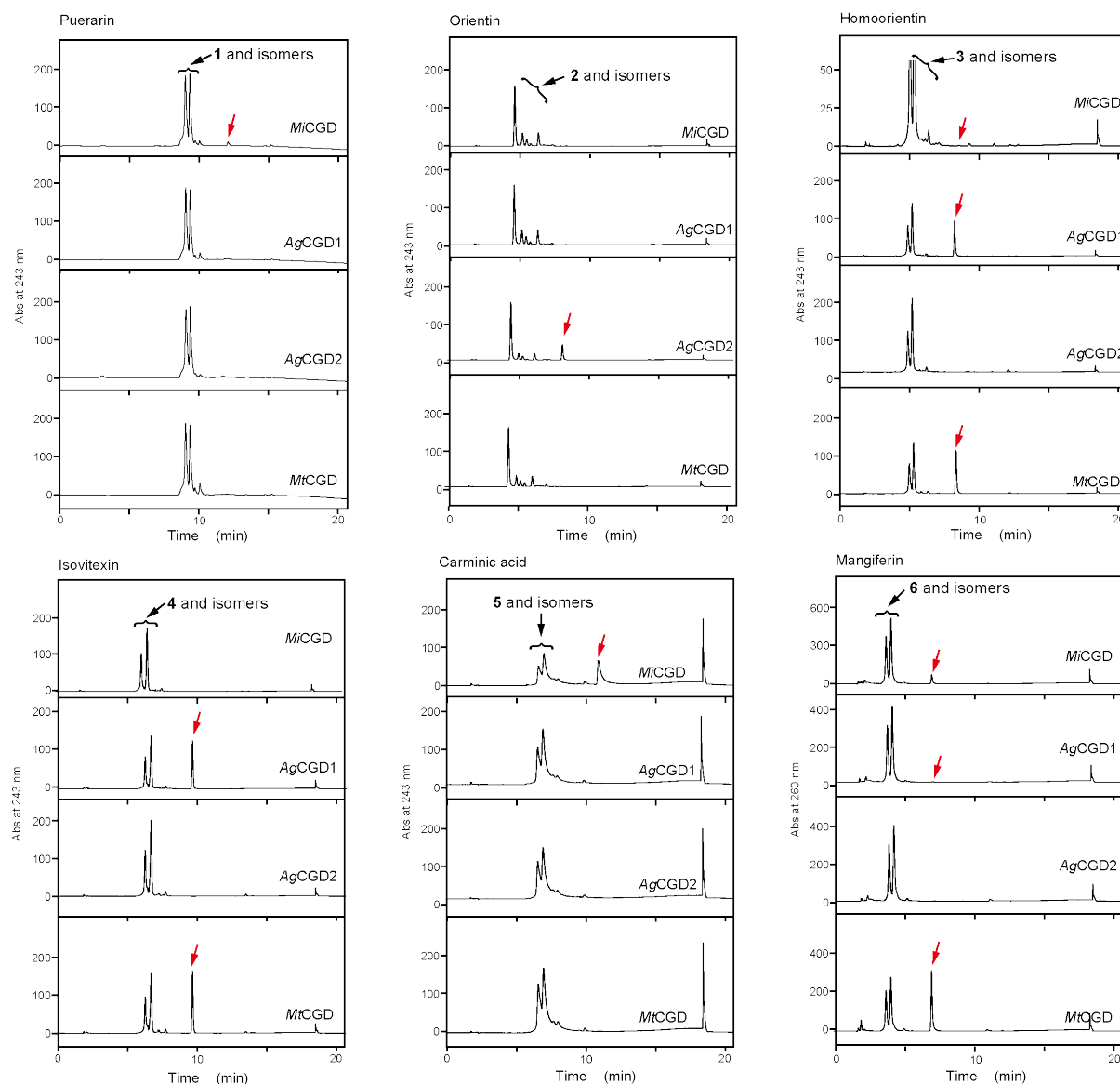
Supplementary Figure 6. SDS-PAGE of CarB (α - and β -subunit of *MiCGD*) and UV-Vis spectrum of *MiCGD* expressed in *E. coli* Rosetta2 (DE3).

(a) Lane 1, protein markers; lane 2, cell-free extract of *E. coli* Rosetta2 (DE3)-expressed CarB. The red arrow indicates CarB. (b) Lane 1, protein markers; lane 2, the cell-free extract of *E. coli* Rosetta2 (DE3)-expressed CarC. The light blue arrow indicates CarC. (c) UV-Vis spectrum of the purified CarB-C complex, which was co-expressed in *E. coli* Rosetta2 (DE3). Similar results of SDS-PAGE were obtained at least three times.



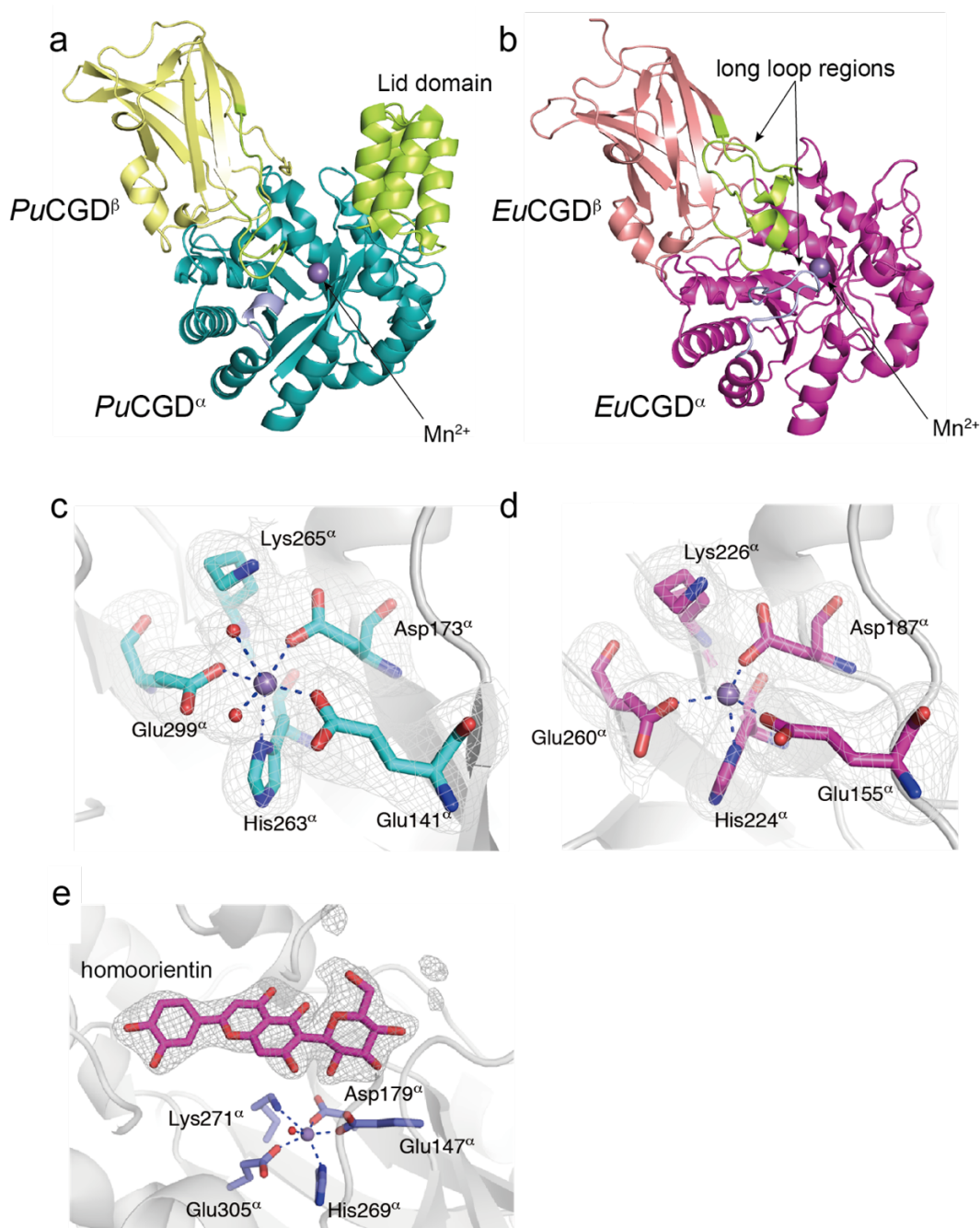
Supplementary Figure 7. HPLC chromatograms of reaction mixtures after incubation of CarB and/or CarC with 3'-oxo-carminic acid.

3'-oxo-carminic acid (**a**), and the 3'-oxo-carminic acid incubated with CarB (**b**), with CarC (**c**), and with CarB and CarC (**d**). The arrows indicate carminic acid, 3'-oxo carminic acid **6** and its isomer, kermesic acid.



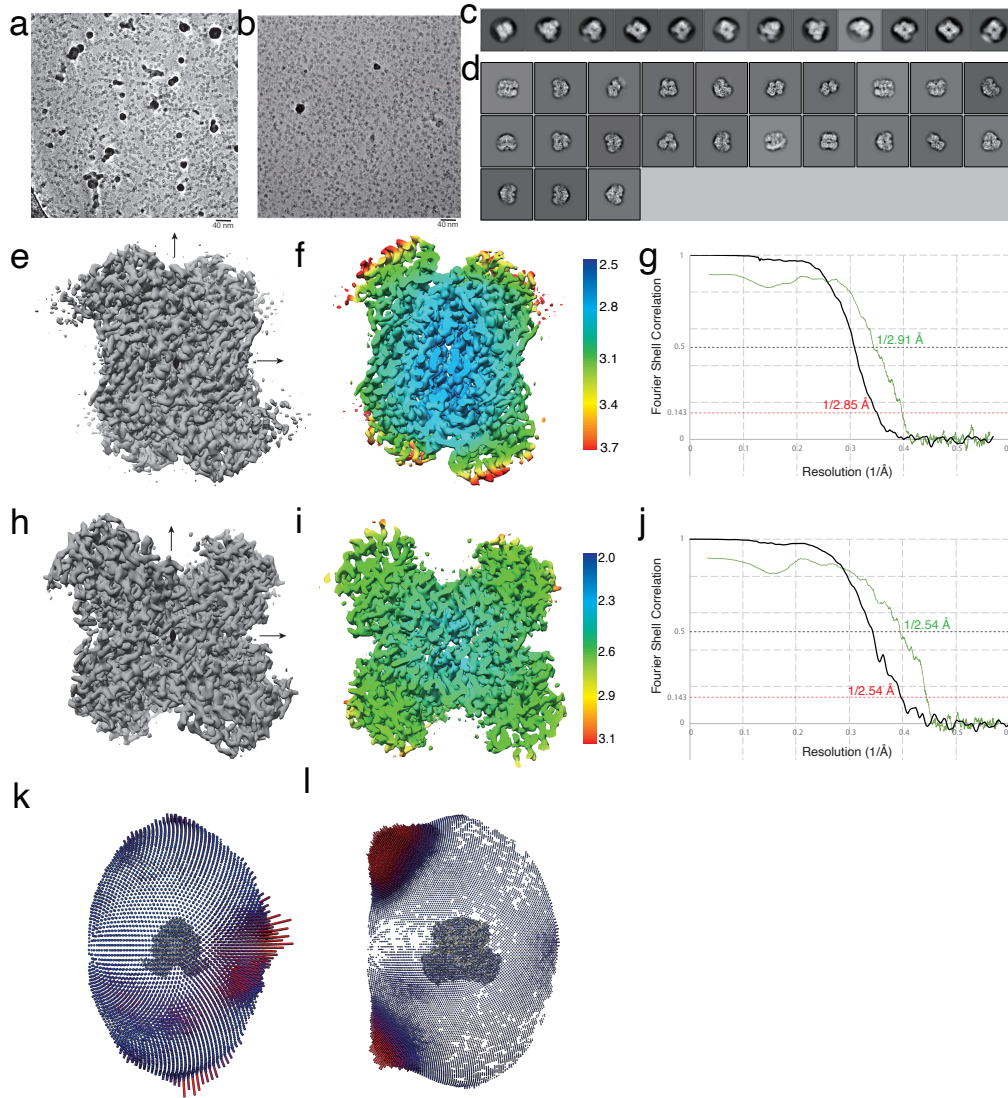
Supplementary Figure 8. HPLC analysis of the enzyme reaction products of CGDs from soil bacteria.

HPLC chromatograms for the C-deglycosylation reactions performed with enzymes from soil microorganisms. Red arrows indicate the production of aglycones.



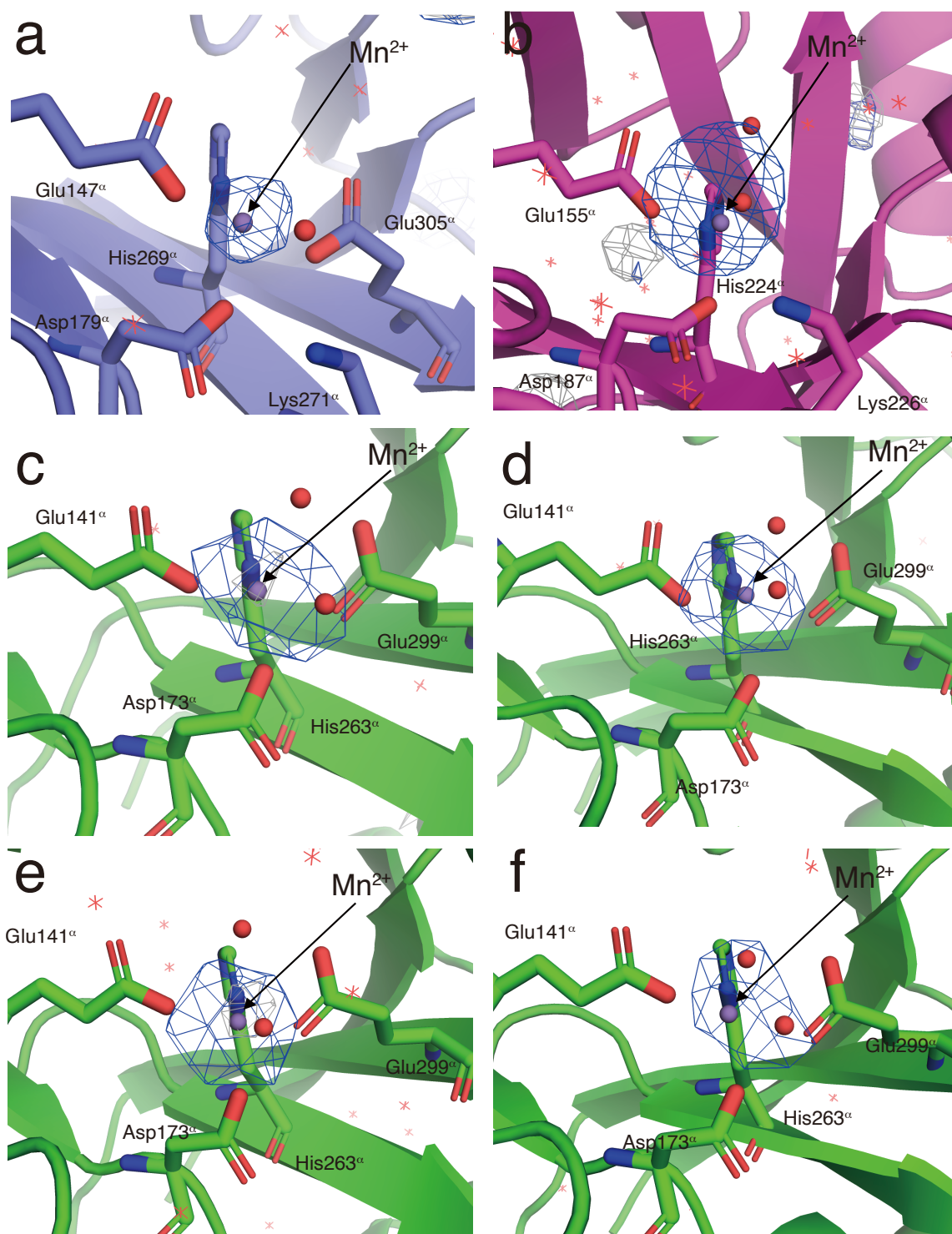
Supplementary Figure 9. Structures of *PuCGD*, *EuCGD*, and *AgCGD2* (related to Figure 2).

(a, b) The heterodimer structures of (a) *PuCGD* and (b) *EuCGD*. β -Sandwich structures of *PuCGD* $^{\beta}$ and *EuCGD* $^{\beta}$, shown in light yellow and pink, respectively, and TIM barrel structures of *PuCGD* $^{\alpha}$ and *EuCGD* $^{\alpha}$, shown in cyan and magenta, respectively. The lid domains (*PuCGD* $^{\alpha}$) and long loop regions (M121-N133 in *EuCGD* $^{\alpha}$ and T110-V139 in *EuCGD* $^{\beta}$) of *EuCGD* are colored light green and light blue, respectively. Corresponding regions in *PuCGD* are shown in the same colors. (c, d) Fo-Fc polder omit maps of the metal-binding sites of (c) *PuCGD* and (d) *EuCGD*. The electron density maps of the ligands are represented by a gray mesh, contoured at +4.0 sigma. The coordination of the manganese ion is represented by dashed blue lines. (e) Fo-Fc polder omit map of homoorientin in the structure of *AgCGD2*. The electron density map of the ligands is represented by a gray mesh, contoured at +5.0 sigma.



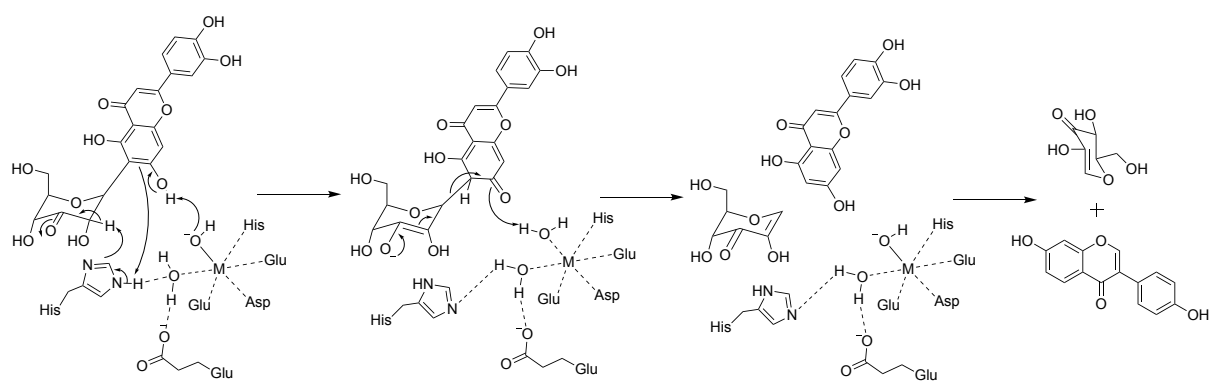
Supplementary Figure 10. Cryo-EM analysis of *PuCGD* and *EuCGD*.

(a and b) Representative cryo-EM micrographs of (a) *PuCGD* and (b) *EuCGD* particles. Similar micrographs were obtained two times. (c and d) 2D class averages of (c) *PuCGD* and (d) *EuCGD* particle images. (e) Cryo-EM map of *PuCGD*. (f) Local resolution of the reconstructed density map of *PuCGD*. (g) Half-map FSC (black) and model-map FSC (green) curves of the *PuCGD* cryo-EM maps. The horizontal dashed lines represent the 0.5 (black) and 0.143 (red) cutoff values, respectively. The gold-standard FSC curves between the two half maps with the indicated resolutions at FSC = 0.143 are colored red. The FSC curves between the atomic model and the final map with the indicated resolution at FSC = 0.5 are colored green. (h) Cryo-EM map of *EuCGD*. (i) Local resolution of the reconstructed density map of *EuCGD*. (j) Half-map FSC (black) and model-map FSC (green) curves of *EuCGD* cryo-EM maps. The horizontal dashed lines represent the 0.5 (black) and 0.143 (red) cutoff values, respectively. The gold-standard FSC curves between the two half maps with the indicated resolution at FSC = 0.143 are colored red. The FSC curves between the atomic model and the final map with the indicated resolution at FSC = 0.5 are colored green. The complete workflow for cryo-EM data-processing of *PuCGD* and *EuCGD* is shown in the *Supplementary Methods*. (k and l) The sharpened density map of the final reconstruction and the angular distributions of (k) *PuCGD* and (l) *EuCGD*.

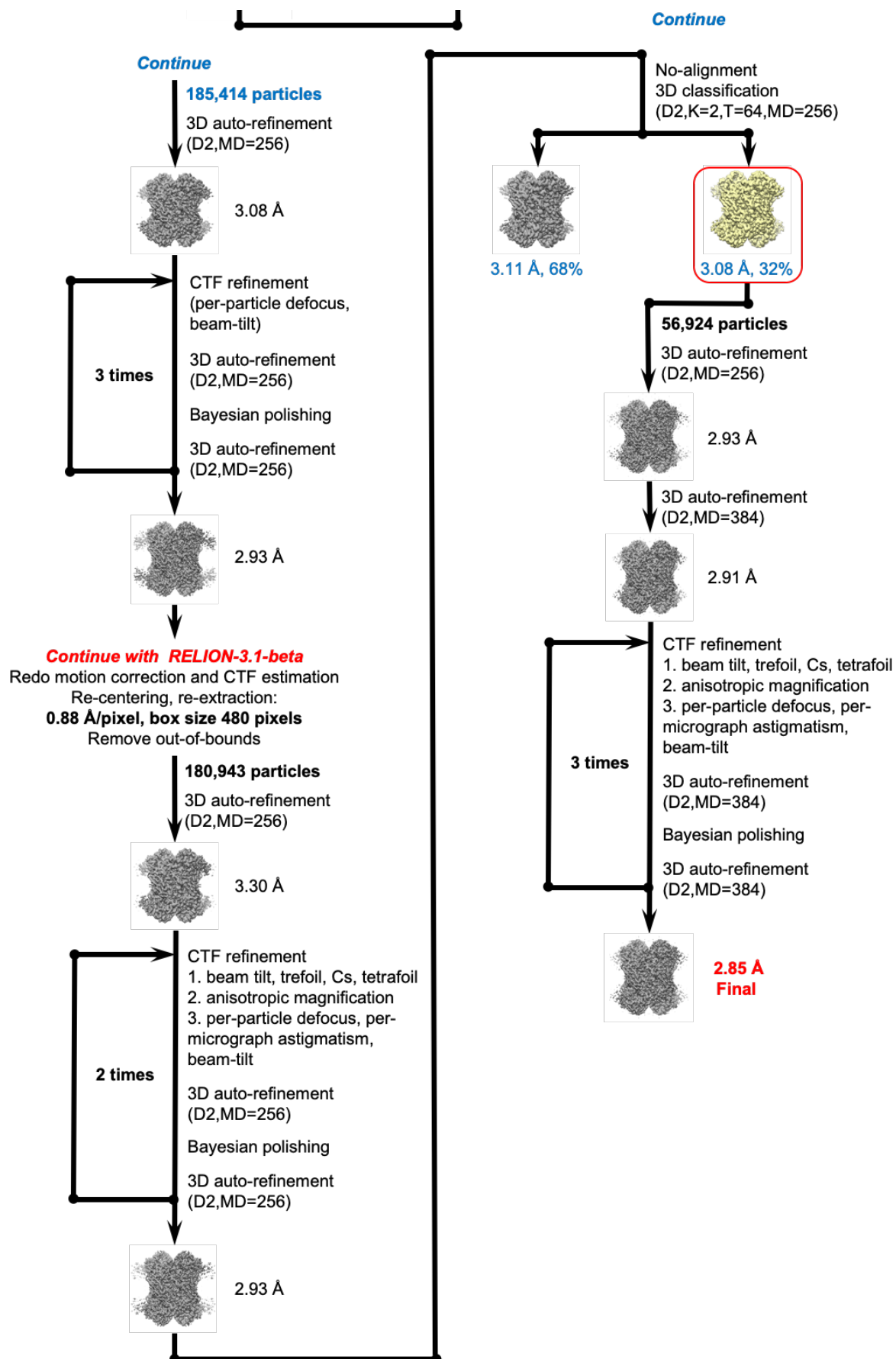


Supplementary Figure 11. Anomalous difference Fourier map.

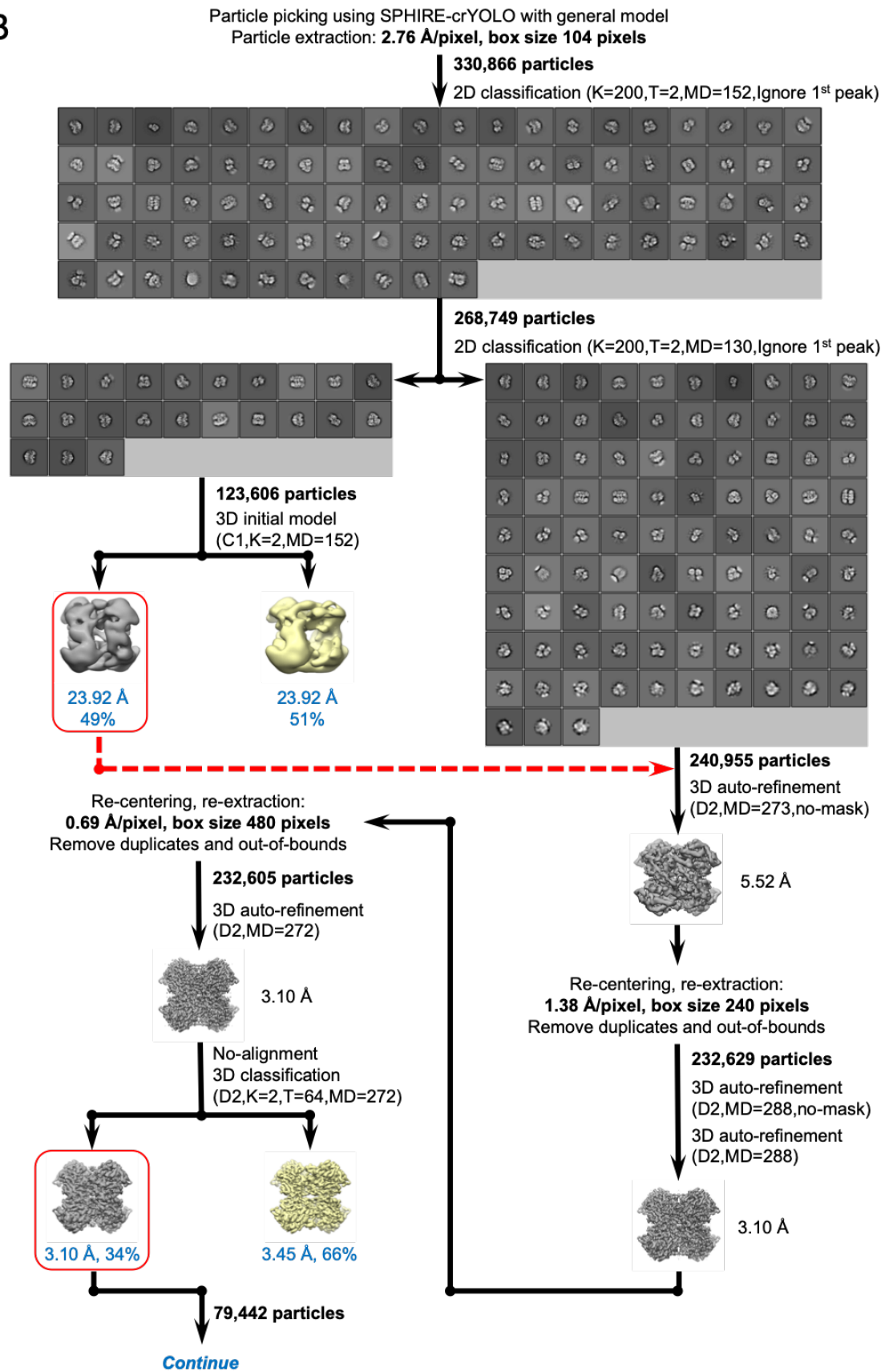
Anomalous difference Fourier map for the Mn ion of (a) *AgCGD2*, (b) *EuCGD*, (c-f) *PuCGD* (monomers A (c), C (d), E (e), and G (f) of *PuCGD*). This map was calculated using the diffraction data collected at the peak wavelength (1.8900 Å) of manganese. The map is contoured at 5 σ level.

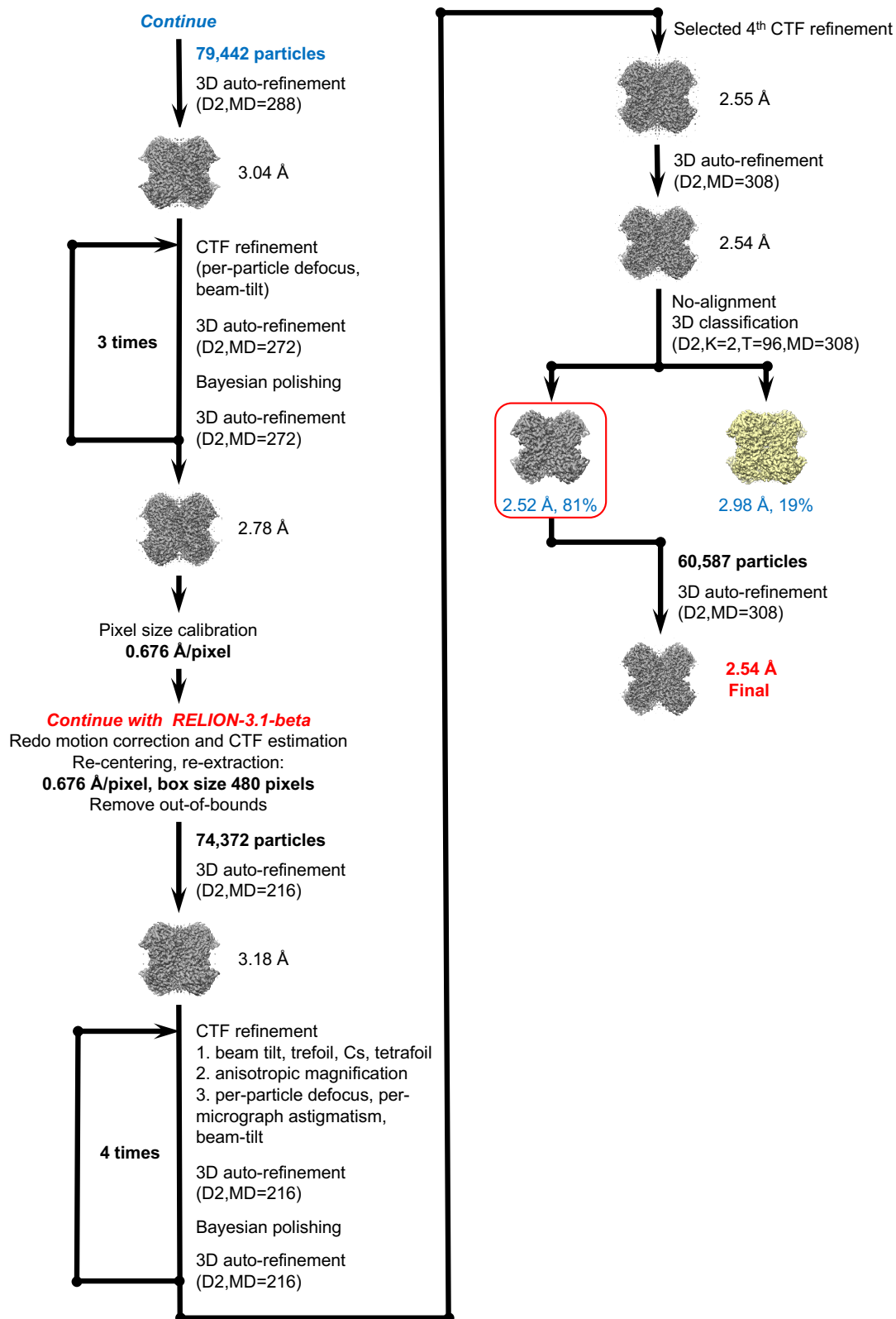


Supplementary Figure 14. Proposed mechanism of C-deglycosylation reaction for C6-glucosides.
M represents a metal ion.



B





Supplementary Figure 15. Workflow for cryo-EM data-processing of DgpB-C and DfgA-B, and data statistics.

Workflow for (A) *PuCGD* and (B) *EuCGD* processings.

Supplementary Methods

Cryo-EM data processing of *PuCGD* and *EuCGD* (Supplementary Fig. 15)

The movie frames were aligned, dose-weighted, and averaged using Motioncor2, on 5×5 tiled frames with a B-factor of 200 applied, to correct for beam-induced specimen motion and to account for radiation damage by applying an exposure-dependent filter. The micrographs with total accumulated motion larger than 100 Å were discarded. For *PuCGD*, the non-weighted movie sums were used for CTF estimation (512-pixel box size, 30 Å minimum resolution, 5 Å maximum resolution, 0.1 amplitude contrast) with Gctf, while the dose-weighted sums were used for all subsequent steps of image processing. Micrographs showing an obvious indication of ice crystallization (i.e. strong ice ring in the Fourier space) were manually discarded. The particles were picked fully-automatically using SPHIRE-crYOLO with the general model, by setting a 137-pixel box size and a selection threshold of 0.1. Micrographs with fewer than 130 picks were excluded. For *EuCGD*, the non-weighted movie sums were used for CTF estimation (512-pixel box size, 30 Å minimum resolution, 3 Å maximum resolution, 0.10 amplitude contrast) with Gctf and CTFFIND4, while the dose-weighted sums were used for all subsequent steps of image processing. The images with Gctf's CTF max resolution better than 4 Å were selected. Micrographs showing an obvious indication of ice crystallization (i.e., strong ice ring in the Fourier space) were manually discarded. The particles were picked fully-automatically using SPHIRE-crYOLO with the general model, using the selection threshold of 0.3. Micrographs that contained fewer than 94 picks were excluded.

RELION3 was used in the following processing. For *PuCGD*, a stack of 857,817 particle images was extracted from 1,963 dose-weighted sum micrographs while rescaling to 3.52 Å/pixel with a 64-pixel box size, and then subjected to two consecutive runs of reference-free 2D classification (1st run: 200 expected classes, 155 Å mask diameter; 2nd run: 200 expected classes, 127 Å mask diameter). The 597,534 particles corresponding to the best 12 classes of the 2nd run that displayed secondary-structural elements and multiple views of *PuCGD* were selected for *ab initio* reconstruction (asymmetry, 4 expected classes, 155 Å mask diameter). The generated *ab initio* map was imposed with D2 symmetry, low-pass filtered to 25 Å, and used as an initial model for the subsequent 3D classification (4 expected classes). The volume of the best 3D class with the highest resolution was then rescaled to 0.88 Å/pixel with a 320-pixel box size, low-pass filtered to 25 Å, and used as an initial reference map for the subsequent 3D refinement. The two 3D classes with obviously bad images containing non-targeted objects were removed, and the selected particle images were re-centered and re-extracted using the same rescale settings. The particle images that became duplicated as a result of alignments and were not fully inside of the micrograph boundary anymore due to the changed box size were excluded. A total of 508,886 selected particles were 3D auto-refined (D2 symmetry, 178 Å mask diameter, no padding) twice, with the 1st run without a 3D mask and the 2nd with a soft-edged 3D mask created from the initial 3D reference of the 1st run (5-pixel extension, 10-pixel soft cosine edge). The 2nd run

resulted in the resolution of 3.31 Å. To refine the per-particle defocus, beam tilt, and beam-induced motion corrections, the cycle of CTF refinement and Bayesian polishing was repeated three times. To measure the degree of the improvement, 3D refinement (D2 symmetry, 178 Å mask diameter, no padding), with the volume of the previous run as the initial 3D reference, a soft-edged 3D mask created from the initial reference (5-pixel extension, 10-pixel soft cosine edge) and the solvent-flattened FSCs options, was used after each CTF refinement and Bayesian polishing step. The 3D refinement after the 3rd Bayesian polishing run yielded the resolution of 3.16 Å.

At this point, 457,708 selected particles were again re-centered and re-extracted with a 486-pixel box size (0.88 Å/pixel), followed by the removal of duplicated and out-of-bounds particles. Accordingly, the density map obtained by the 3D refinement after the 3rd Bayesian polishing run was also padded to the same box size, low-pass filtered to 15 Å, and used as an initial reference map for the subsequent 3D refinement (D2 symmetry, 256 Å mask diameter, no padding) with a soft-edged 3D mask (5-pixel extension, 10-pixel soft cosine edge). To improve the homogeneity of the particle stack, no-alignment 3D classification was conducted by setting the expected classes to 2 and the regularization parameter T to 8 (D2 symmetry, 256 Å mask diameter, no padding) with a soft-edged 3D mask (5-pixel extension, 10-pixel soft cosine edge), and 185,414 particles were selected by choosing the 3D class with higher resolution. The subsequent 3D refinement (D2 symmetry, 256 Å mask diameter, no padding) with a soft-edged 3D mask (5-pixel extension, 10-pixel soft cosine edge) generated 3.08 Å resolution. Three cycles of CTF refinement and Bayesian polishing were then executed as above, and the resolution improved to 2.93 Å.

For *EuCGD*, a stack of 330,866 particle images was extracted from 1,614 dose-weighted sum micrographs while rescaling to 2.76 Å/pixel with a 104-pixel box size, and subjected to two consecutive runs of reference-free 2D classification ignoring first peak of CTFs (1st run: 200 expected classes, 152 Å mask diameter; 2nd run: 200 expected classes, 130 Å mask diameter). For the *ab initio* reconstruction, 123,606 particles corresponding to the best 23 classes of the 2nd run that displayed secondary-structural elements and multiple views of *EuCGD* were selected (asymmetry, 2 expected classes, 152 Å mask diameter), and for the subsequent 3D refinement, 240,955 particles were selected with more relaxed criteria. A generated *ab initio* map with no obvious signs of structural damage was selected, imposed with D2 symmetry, low-pass filtered to 15 Å, and used as an initial reference map for the 3D refinement (D2 symmetry, 273 Å mask diameter, with padding).

The refined volume was then rescaled to 1.38 Å/pixel with a 240-pixel box size, and low-pass filtered to 15 Å for the subsequent 3D refinement. Accordingly, the particle images were also re-centered and re-extracted using the same rescale settings, followed by the removal of duplicated and out-of-bounds particles. Afterwards, 232,629 selected particles were 3D auto-refined (D2 symmetry, 288 Å mask diameter, no padding) twice, with the 1st run without the 3D mask and the 2nd with a soft-edged 3D mask created from the refined 3D map of the 1st run (5-pixel extension, 10-pixel soft cosine edge). The resolution of the 2nd run was 3.10 Å.

The refined volume was rescaled to 0.69 Å/pixel with a 480-pixel box size, and low-pass filtered to 15 Å for the subsequent 3D refinement. Accordingly, 232,605 selected particles were again re-centered and re-extracted using the same rescale settings, followed by the removal of duplicated and out-of-bounds particles, and were 3D auto-refined (D2 symmetry, 272 Å mask diameter, no padding) with a soft-edged 3D mask (5-pixel extension, 10-pixel soft cosine edge). To improve the homogeneity of the particle stack, no-alignment 3D classification was conducted by setting the expected classes to 2 and the regularization parameter T to 64 (D2 symmetry, 272 Å mask diameter, no padding) with a soft-edged 3D mask (5-pixel extension, 10-pixel soft cosine edge), and 79,442 particles were selected by choosing the 3D class with higher resolution. The subsequent 3D refinement (D2 symmetry, 288 Å mask diameter, no padding) with a soft-edged 3D mask (5-pixel extension, 10-pixel soft cosine edge) generated 3.04 Å resolution. Three cycles of CTF refinement and Bayesian polishing were then executed, and the resolution improved to 2.78 Å.

To correct higher-order aberrations and anisotropic magnification, in addition to per-particle defocus, beam tilt, and beam-induced motion, RELION 3.1.0-beta was used from this point forward. The movie frames were aligned, dose-weighted, and averaged using RELION's own implementation of a MotionCor2-like algorithm on 5×5 tiled frames, with a B-factor of 200 applied. The CTF parameters of each non-weighted movie sum were estimated with CTFFIND4 with a 512-pixel box size, 30 Å minimum resolution, 3 Å maximum resolution and 0.1 amplitude contrast. The images with CTF max resolutions better than 4 Å were selected.

For *PuCGD*, from 1,879 dose-weighted sum micrographs, 180,943 particles associated with the 2.93 Å refined map were re-centered and re-extracted with a 480-pixel box size. The refined map was low-pass filtered to 15 Å, and used as an initial reference map for the subsequent 3D refinement (D2 symmetry, 256 Å mask diameter, no padding) with a soft-edged 3D mask (5-pixel extension, 10-pixel soft cosine edge). The cycle of CTF refinement and Bayesian polishing was then repeated twice. Following the RELION 3.1.0 tutorial, CTF refinement at each cycle was executed three times: first to correct higher-order aberrations (beam tilt, trefoil astigmatism, Cs, and tetrafoil astigmatism), second to correct anisotropic magnification, and third for per-particle defocus and per-micrograph astigmatism. The 3D refinement after the 2nd Bayesian polishing run yielded the resolution of 2.93 Å.

To further improve the homogeneity of the particle stack, no-alignment 3D classification was again conducted by setting the expected classes to 2 and the regularization parameter T to 64 (D2 symmetry, 256 Å mask diameter, no padding) with a soft-edged 3D mask (5-pixel extension, 10-pixel soft cosine edge). By choosing the 3D class with higher resolution, 56,924 particles were selected. The selected volume was 3D auto-refined (D2 symmetry, no padding, with a soft-edged 3D mask) twice: the 1st run with a 256 Å mask diameter, and the 2nd with a 384 Å mask diameter. The 2nd run resulted in the resolution of 2.91 Å. Again, three cycles of CTF refinement and Bayesian polishing were executed as above and generated the final result of 2.85 Å.

For *EuCGD*, a stack of 74,372 particle images associated with the 2.78 Å refined map was re-centered and re-extracted from 1,598 dose-weighted sum micrographs, while rescaling to the calibrated pixel size of 0.676 Å/pixel with a 480-pixel box size. Accordingly, the refined map was also rescaled with the same settings, low-pass filtered to 15 Å, and used as an initial map for the subsequent 3D refinement (D2 symmetry, 216 Å mask diameter, no padding) with a soft-edged 3D mask (5-pixel extension, 10-pixel soft cosine edge). The cycle of CTF refinement and Bayesian polishing was then repeated four times, as above. The 3D refinement after the 4th CTF refinement step yielded the best resolution of 2.55 Å during the cycles. The particle stack was further 3D auto-refined using a larger mask diameter (D2 symmetry, 308 Å mask diameter, no padding) with a soft-edged 3D mask (5-pixel extension, 10-pixel soft cosine edge), and the resolution of 2.54 Å was obtained. To improve the homogeneity of the particle stack, no-alignment 3D classification was conducted by setting the expected classes to 2 and the regularization parameter T to 96 (D2 symmetry, 308 Å mask diameter, no padding) with a soft-edged 3D mask (5-pixel extension, 10-pixel soft cosine edge), and 60,587 particles were selected by choosing the 3D class with higher resolution. The last 3D refinement (D2 symmetry, 308 Å mask diameter, no padding) with a soft-edged 3D mask (5-pixel extension, 10-pixel soft cosine edge) generated the final result of 2.54 Å.

The local resolution of the final reconstruction was estimated using RELION-3's own implementation.

The initial *PuCGD* and *EuCGD* models were built with the Map to Model program in PHENIX¹, using the cryo-EM maps and amino acid sequences. The model was then manually modified with Coot² and refined by using Real-space refinement in PHENIX³. The cryo-EM maps of *PuCGD* and *EuCGD* have been deposited in the Electron Microscopy Data Bank (EMDB-30808 and EMD-30809, respectively). The coordinates of *PuCGD* and *EuCGD* have been deposited in the Protein Data Bank (7DRD and 7DRE, respectively). The statistics of EM data processing and refinement are summarized in Supplementary Table 3. The local resolution maps were produced with the ResMap program⁴.

Supplementary References:

1. Terwilliger, T. C., Adams, P. D., Afonine, P. V. & Sobolev, O. V. A fully automatic method yielding initial models from high-resolution cryo-electron microscopy maps. *Nat. Methods* **15**, 905-908 (2018).
2. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126-2132 (2004).
3. Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **74**, 531-544 (2018).
4. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps, *Nat. Methods.* **11**, 63-65 (2014).