

Comparative Computational Study to Augment UbiA prenyltransferases Inherent in Purple Photosynthetic Bacteria Isolated from Mangrove Microbial Mats in Qatar for Coenzyme Q₁₀ biosynthesis.

Drishya M. George^{a♦}, Ramya Ramadoss^{b♦}, Hamish R. Mackey^c, Annette S. Vincent^{d*}

^aCollege of Health and Life Sciences, Hamad bin Khalifa University, Qatar Foundation, Doha, Qatar.

Email: dgeorge@hbku.edu.qa

^bBiological Sciences, Carnegie Mellon University Qatar, Doha, Qatar.

Email: rramado2@andrew.cmu.edu

^cCollege of Health and Life Sciences, Hamad bin Khalifa University, Qatar Foundation, Doha, Qatar.

Division of Sustainable Development, College of Science and Engineering, Hamad bin Khalifa University, Qatar Foundation, Doha, Qatar.

Email: hmackey@hbku.edu.qa

^d College of Health and Life Sciences, Hamad bin Khalifa University, Qatar Foundation, Doha, Qatar.

Biological Sciences, Carnegie Mellon University Qatar, Doha, Qatar.

Email: annettev@andrew.cmu.edu , **Tel:** +974 4484852

*** Corresponding Author**

♦ These authors contributed equally to this work

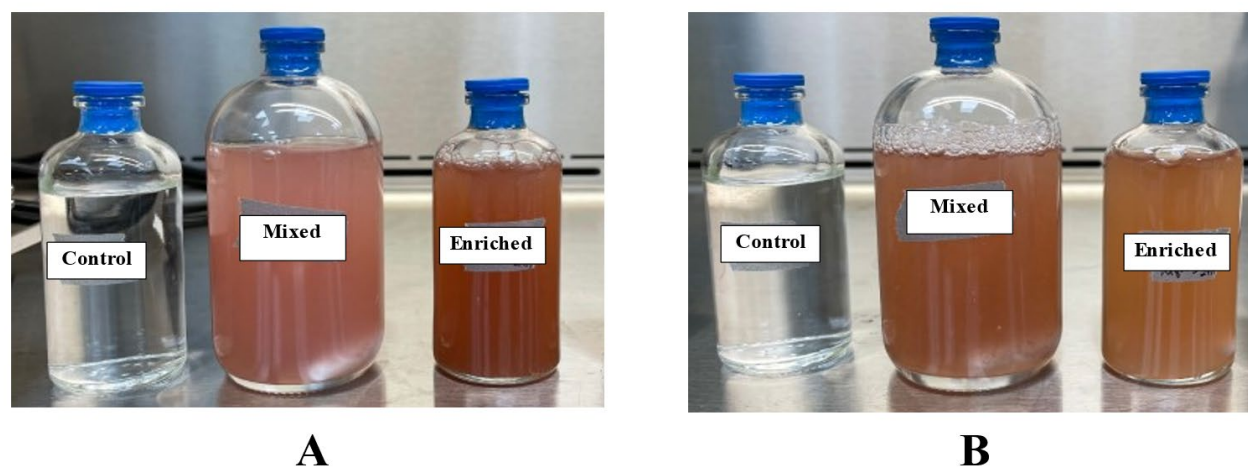


Fig. S1 Mixed (Pre-FACS) and enriched (Post-FACS) cultures at day 20, grown under (A) Low wave infra-red IR light conditions and (B) High wave IR light conditions.

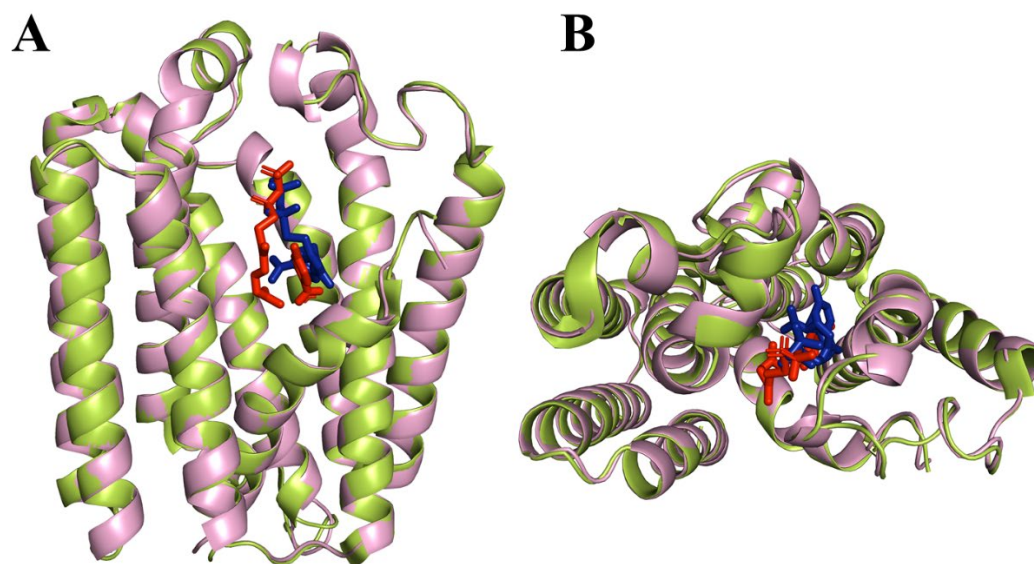


Figure S2: Superposed structure: The crystal structure of substrate-bound UbiA from *A. pernix* K1 deposited in protein data bank (PDB) under the PDB ID: 4OD5 was retrieved. The protein sequence of 4OD5 was retrieved from PDB. The protein sequence was modelled using I-TASSER [1] followed by MLSDS using MTiAutodock [2] with the same ligands in 4OD5. The substrate-bound UbiA crystal structure (protein-green and ligand-red) and the I-TASSER modelled, docked structure (protein-pink and ligand-dark blue) was superposed using CCP4 suite [3]. The superposed structure has a low RMSD score of 0.5 Å with the ligands bound at similar positions, revealing the efficiency of modelling and MLSDS studies as a preliminary step preceding experimental studies. (A) Front-view of superposed structure. (B) Top-view of superposed structure.

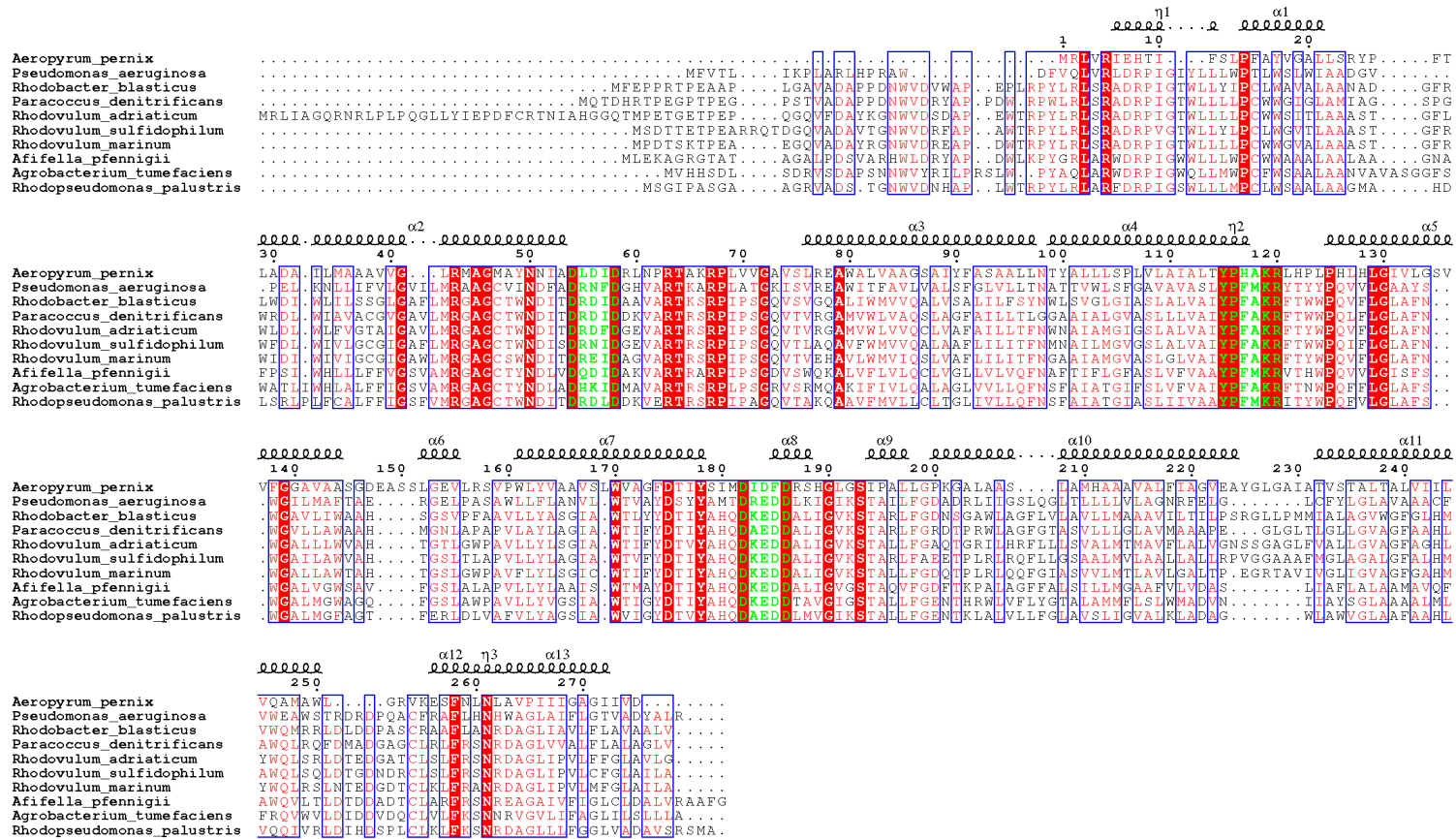


Figure S3: Multiple sequence alignment. Representation of multiple sequence alignment of UbiA sequence protein sequences from *A. pernix* K1, the candidate organisms – *R. adriaticum*, *R. marinum*, *R. blasticus* and *A. pfennigii* and the industrial/native producers of CoQ10 – *P. aeruginosa*, *P. denitrificans*, *A. tumefaciens*, *R. palustris* and *R. sulfidophilum* using ESPript [4]. The secondary structure of PDB structure 4OD5 is represented at the top of the alignment. The conserved motifs are highlighted in green.

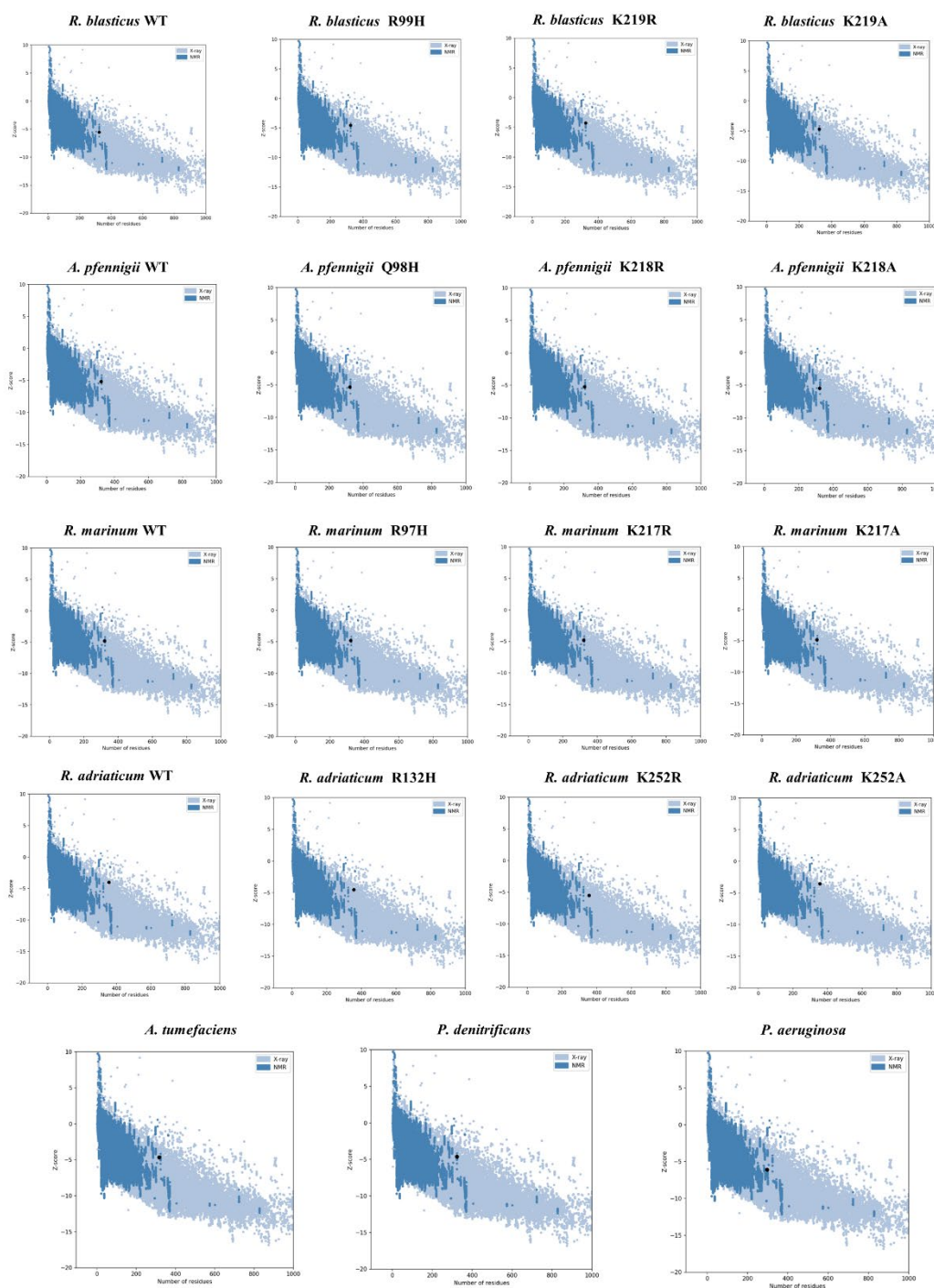


Figure S4: Z-score plot generated by ProSA: All the modelled structures predicted by I-TASSER that were used for MLSDS were validated using ProSA tool [5] for model quality assessment. All the predicted models had z-score well within the range of experimental protein structures.

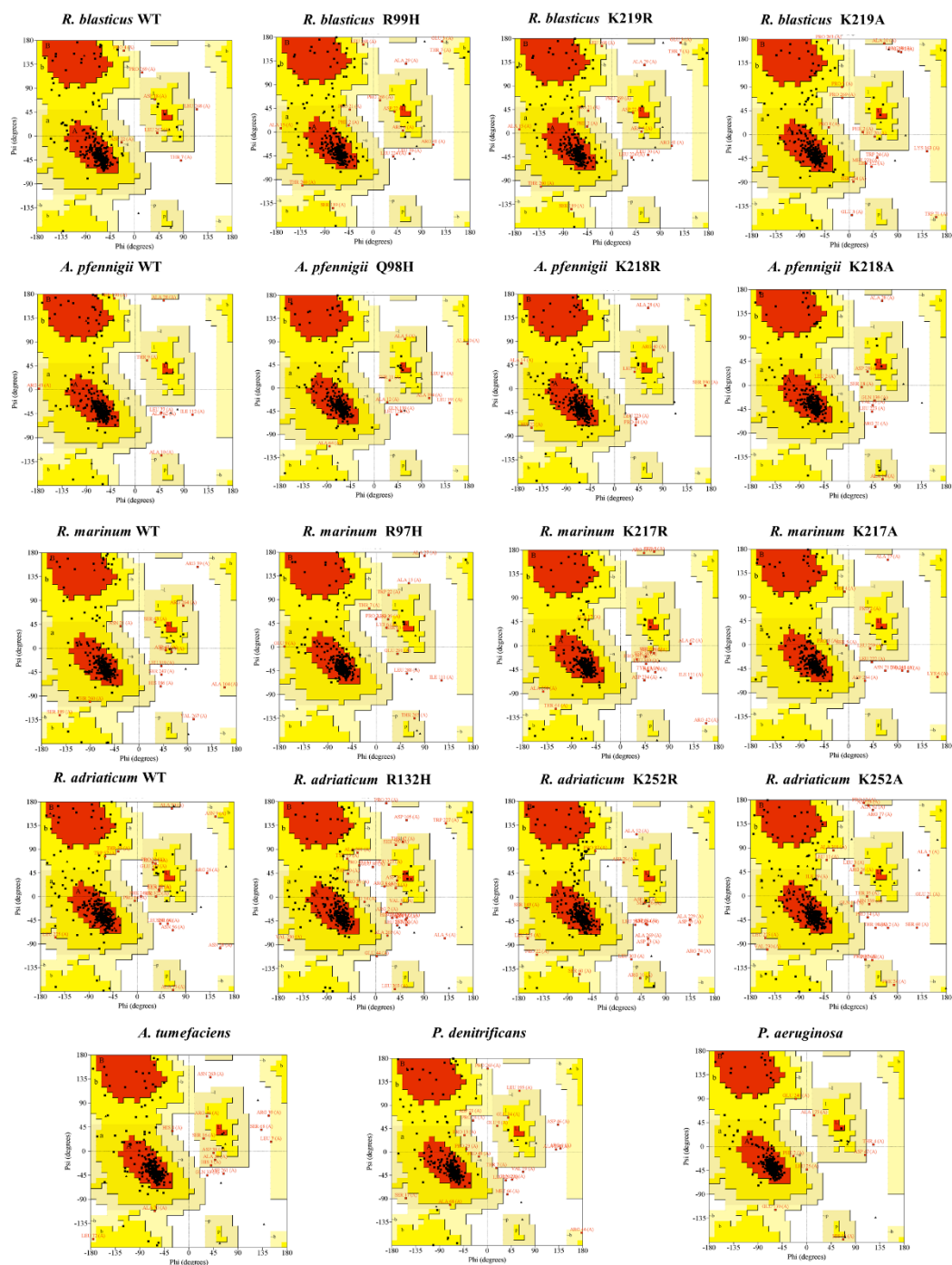


Figure S5: Ramachandran Plot: Ramachandran plots for all the modelled structures predicted by I-TASSER to be used for MLSDS were generated using PROCHECK [6] for model quality validation. More than 97% residues were present in the allowed region.

Table S1: Protein dataset.

Dataset of Protein Sequence entries of 4-Hydroxybenzoate octaprenyl transferase derived from UniProt database. This dataset was input to MMseqs2 tool for sensitive sequence search to cluster the protein sequences sharing 30% sequence identity and 50% minimum coverage.

Table S2: Preliminary clustering of protein sequences.

Largest Cluster, Cluster-19 identified during first step of clustering using MMseqs2 tool [7]. Protein Sequences share 30% sequence identity and 50% minimum coverage. Cluster-19 was further refined in next step by clustering the sequences that share 40% sequence identity and 80% minimum coverage.

Table S3: Refinement of largest cluster obtained from preliminary cluster analysis.

Largest Cluster, Cluster-35 identified during Second step of clustering using MMseqs2 tool. Protein Sequences share 40% sequence identity and 80% minimum coverage.

Table S4: Summary of protein-ligand interactions.

Summary of binding energy, hydrogen bond forming residues, interacting amino acid residues (residue environment) within 5Å of PHBA/GSPP ligand and the phosphorylation sites (predicted using GPS 5.0 webserver) of wild-type and mutated variants of the UbiA enzymes of organisms listed in **Table. 1**. The amino acid residues within the residue environment which were also predicted as possible phosphorylation sites are highlighted in red. The UbiA enzyme from native/industrial producers of CoQ₁₀ and the mutant from candidate organisms with least binding energy is highlighted in blue.

UbiA enzyme	Ligand	Binding energy (kcal/mol)	Hydrogen interaction	Residue environment	Predicted phosphorylation sites
<i>R. blasticus</i>					
Wild-type	PHBA	-4.48	R38 (1), S110 (1)	R38, L39, R41, R107, T108 , K109, S110 , R111, P112	T7, Y36, S40, T48, Y52, S78, S79, T92, T97, T108, S110, S115, S119, S134, S140, Y141, S145, S152, Y159, T166, S189, S191, Y200, S202, T207, Y209,
	GSPP	-2.37	R41 (2)	R38, L39, R41, R44, P45, T48 , L51, R87, C91, N94, D95, R107, T108 , K109, S110 , R111, P112	Y210, T212, Y214, S229, T230, S238, T260, S264, S298
R99H Mutant	PHBA	-5.19	P4 (1), R41 (2), N306 (1)	P4, P5, T7 , P8, R41, Y210 , F303, L304, N306, R307	

	GSPP	-2.82	R41 (2)	P4, P5, T7 , P8, R41, D43, Y210 , Y214 , F303, L304, N306, R307	
K219R Mutant	PHBA	-5.14	R87 (2)	M1, L50, L51, M86, R87, G90, V156, L173, G174, A176, F177, W206, T207 , Y210	
	GSPP	-2.36	R44 (1), C90 (1)	M1, P5, S40 , R41, R44, P45, G47, T48 , L50, L51, M86, R87, G90, C91, N94, V156, Y159 , L173, G174, A176, T207 , Y210 , D211, Y214 , F303	
K219A Mutant	PHBA	-5.38	E3 (1)	M1, F2, E3, P4, P5, R6, Y210 , Y214 , F303, L304, N306, R307, D308,	
	GSPP	-3.60	Q217 (1), R300 (1)	M1, F2, E3, P4, P5, R6, R107, L173, Y210 , D21, Y214 , Q217, R300, F303, L304, N306, R307, D308	
<i>A. pfennigii</i>					
Wild-type	PHBA	-4.72	R40 (2), A109 (1)	R40, D42, R43, A105, R106, T107 , R108, A109, R110, P111	T9, T11, S18, Y27, Y35, S71, S82, T91, Y92, T107, S114, S118, T144, S151,
	GSPP	-3.19	R40 (1), R43 (1), C90 (1), D94 (1), R110 (1)	R40, D42, R43, W47, M85, R86, G89, C90, N93, D94, A105, R106, T107 , R108, A109, R110, P111, V155, Y158 , P159, L172, S175	Y158, T165, S175, S177, S185, S190, Y199, S204, T206, Y209, T211, Y213, S228, T229, T237, S248, S262, T282, T285, 7290, S297

Q98H Mutant	PHBA	-4.87	R43 (1), F295 (1), R296 (1), R299 (1)	R43, G46, W47, L49, L50, Y209 , R294, F295, R296, S297 , N298, R299, E300	
	GSPP	-3.19	L172 (1), F176 (1)	R43, G46, W47, L49, L50, L51, S82, M85, R87, G89, C90, N93, V155, Y158 , P159, V171, L172, G173, S175 , F176, W178, G179, T206 , Y209 , R294, F295, R296, S297 , N298, R299, E300	
K218R Mutant	PHBA	-5.25	K4 (1), R43 (1)	K4, R43, G46, W47, L49, L50, W205, Y209 , F295, R296, N298, R299	
	GSPP	-4.32	R43 (3)	K4, R40, D42, R43, G46, W47, L49, L50, R106, R110, Y158 , M161, K162, P168, Q169, L172, W205, Y209 , D210, Y213 , A214, F295, R296, N298, R299	
K218A mutant	PHBA	-4.61	R86 (1)	R40, R43, W47, M85, R86, G89, C90, N93, V155, Y158 , S175	
	GSPP	-4.03	R86 (2), Y158 (2)	R40, R43, W47, L49, L50, M85, R86, G89, C90, N93, V155, Y158 , L172, S175 , F176, W205, T206 , Y209 , F295, R296, N298, R299	
<i>R. marinum</i>					
Wild-type	PHBA	-4.54	R42 (1),	W166, P167, Q168, V169, F170, L171,	T4, S5, T7, Y18, T31, Y34, S38, T46, S63,

			N303 (1), R304 (1)	G172, L173, A174, I202, I206,	T64, S90, T95, T106, T107, S108, S113, T117, S129, T138, S150, Y157, T164, Y165, T184, T187, S189, Y198, S200, T205, Y208, T210, Y212, S227, T228, S47, T252, T260, T265, Y283, S287, T290, T295
	GSPP	-2.64	R39 (3), N92 (1)	R39, D41, R42, G45, C89, N92, D93, D96, I99, D100, V103, R105, T106 , R109, Y157 , K161, L171, W204, Y208 , D209, Y212 , D216, V225, F300, N303, R304	
R97H Mutant	PHBA	-5.17	Y208 (1), N303 (2), R304 (1)	G45, L48, L49, P52, W204, Y208 , F300, N303, R304, D305, A306, G307, L308	
	GSPP	-4.45	R85 (1)	R39, G45, L48, L49, L50, P52, C53, A81, W82, M84, R85, G88, C89, V154, L171, A174, F175, G178, W204, Y208 , F300, N303, R304, D305, A306, G307, L308	
K217R Mutant	PHBA	-4.28	Y208 (1)	L48, L49, P52, F175, W204, T205 , Y208 , F277, F300, N303, R304, A306, G307, L308, P310	
	GSPP	-2.32	R39 (1)	R39, R42, G45, L48, L49, P52, M84, R85, G88, C89, N92, V154, Y157 , F158, L171, A174, F175, W177, W204, T205 , Y208 , F277, F300, N303, R304, A306, G307, L308, P310	
K217A Mutant	PHBA	-6.08	M1 (1), R301 (1), N303 (1)	M1, P2, T4 , G45, L49, R85, W204,	

				Y208 , F300, R301, N303, R304, D305	
	GSPP	-3.7	R39 (1)	M1, P2, T4 , R39, A40, D41, G45, L49, R85, N92, D93, R109, V154, Y157 , L171, A174, F175, W204, Y208 , F300, R301, N303, R304, D305	
<i>R. adriaticum</i>					
Wild-type	PHBA	-5.51	Q163 (2)	M1, R2, I4, A5, R10, F50, D51, A52, R77, P78, V117, L118, G121, W159, V162, Q163, V166	Y19, T27, T35, T39, T42, Y53, S60, T66, Y69, S73, T81, S91, T93, S98, T99, T112, T125, T130, T141, S143, S148, T152, T173, S185, Y192, T199, Y200, T222, Y233, S235, T240, Y243, T245, Y247, S262, T263, T271, S282, T287, S298, S299, Y318, S322, T326, T331, S334, S338
	GSPP	-2.99	R2 (2), D51 (1), G55 (1), W57 (1)	M1, R2, I4, A5, R10, F50, D51, A52, Y53, G55, N56, W57, T66, Y69, L70, R77, P78, V117, L118, G121, W159, V162, Q163, V166	
R132H Mutant	PHBA	-4.64	N339 (1)	L84, P87, F210, L232, S235, G236, W239, T240 , Y243 , M286, F313, N339, A342, G343, P346	
	GSPP	-2.33	Y243 (2)	L84, P87, M119, R120, G121, G123, C124, V189, Y192 , L206, G207, A209, F210, L232, S235, G236, W239, T240 , Y243 , D244, M286, F313, F336, N339, A342, G343, P346	
K252R Mutant	PHBA	-6.42	Q48 (1), F50 (1),	Q48, V49, F50, L83, L84, L86, P87, W239, N339, R340, D341,	

			R340 (2), L344 (1)	A342, G343, L344, V347	
	GSPP	-4.21	A306 (1)	Q48, V49, F50, L83, L84, L86, P87, C88, S91 , F210, A214, L232, Y233 , S235, G236, W239, T287 , V290, A306, G309, V310, F313, N339, R340, D341, A342, G343, L344, I345, P346, V347, F349	
K252A Mutant	PHBA	-6.57	R2 (1), S91 (2), L294 (1)	M1, R2, L3, S91, I92, L94, A95, L232, V290, L294, V295, S298 , S299 , G300, A301, G302, A306, P346, F349, F350	
	GSPP	-3.63	F101 (1), L102 (2)	M1, R2, L3, S91 , I92, L94, A95, T99 , G100, F101, L102, W103, L104, L232, V290, L294, V295, S298 , S299 , G300, A301, G302, A306, P346, F349, F350, A353, V354, L355, G356	
<i>A. tumefaciens</i>					
Wild- type	PHBA	-5.56	R87 (1)	L46, M47, C50, F79, G82, S83 , V84, M86, R87, V156, A176, F177, W179, G180, M183,	S5, S8, S12, S16, Y21, S27, Y31, S53, S64, S68, T71, S83, T92, Y93, T108, S110, S115, S119, S143, T148, S152, Y159, T166, S178, Y200, S203, T207, Y210, T212, Y214, T223, S229, T230, T238, Y246, T248, S256,
	GSPP	-2.85	N94 (1), Y159 (1)	R36, L46, M47, C50, F79, G82, S83 , V84, M86, R87, C91, N94, R107, V156, Y159 , P160, A162, K163, P169, Q170, F172,	

				L173, A176, F177, W179, G180, M183, Y210 , D211, Y214 , A215, D218, D222, I227, G228, S229	Y267, S268, S298, S313
<i>P. denitrificans</i>					
Wild-type	PHBA	-5.57	R306 (1), R307 (1)	R44, L54, W209, T210 , Y213 , L300, F303, R304, N306, R307	T3, T7, T12, S17, T18, Y31, S43, T51, S70, T95, T100, T111, S113, S118, T122, S134,
	GSPP	-3.13	V253 (1)	R44, L54, P57, C58, G61, I62, A65, F180, N181, V184, P199, A202, Y203 , A205, G206, W209, T210, Y213 , V253, L254, L256, G257, L258, V260, L274, V277, L300, F303, R304, N306, R307, G310, V313, F316	T143, S155, Y162, T169, Y203, T210, Y213, T215, Y217, S232, T233, T241, T250, S252, T271, S305
<i>P. aeruginosa</i>					
Wild-type	PHBA	-4.71	L34 (1), N182 (1)	L33, L34, L35, W36, P37, T38 , S41 , Y156 , I160, L179, N182, W185, L232, F245, L249, G282, I285	T4, Y32, T38, S41, T87, T94, S98, T105, S113, T120, T123, T124, S128, S136, Y138, Y144, T145,
	GSPP	-2.46	R 24 (2), R 27 (1)	R24, R27, I31, L33, L34, L35, W36, P37, T38 , S41 , R66, G69, C70, N73, D74, T87 , R90, P91, Y138 , P139, L152, Y156 , I160, L178, N182, W185, L232, F245, L249, G282, I285	Y146, Y147, Y156, S157, T165, S174, T186, Y189, S191, Y192, Y193, T196, S208, T209, S223, T228, Y246, S262, T263, T289, Y293

References

1. Roy, A. *et al.* (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5, 725-738. 10.1038/nprot.2010.5
2. Labbe, C.M. *et al.* (2015) MTiOpenScreen: a web server for structure-based virtual screening. *Nucleic Acids Res* 43, W448-454. 10.1093/nar/gkv306
3. Winn, M.D. *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D* 67, 235-242. doi:10.1107/S0907444910045749
4. Robert, X. and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42, W320-324. 10.1093/nar/gku316
5. Wiederstein, M. and Sippl, M.J. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research* 35, W407-W410. 10.1093/nar/gkm290
6. Laskowski, R. *et al.* (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* 26, 283-291. 10.1107/S0021889892009944
7. Steinegger, M. and Söding, J. (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology* 35, 1026-1028. 10.1038/nbt.3988