

## ORIGINAL ARTICLE

# Functional Annotation of Hypothetical Proteins related to Antibiotic Resistance in *Pseudomonas Aeruginosa* PA01

Alariqi Reem<sup>1</sup>, Zhao-Hua Zhong<sup>1</sup>, Wadee A. AL-Shehari<sup>1</sup>, Fadhl Al-Shaebi<sup>2</sup>, Gehad A. Amran<sup>3</sup>,  
Yahya A. G. Moeed<sup>4</sup>, Rokayya Sami<sup>5</sup>, Ebtihal Khojah<sup>5</sup>, Ahmad El Askary<sup>6</sup>

<sup>1</sup> Department of Microbiology, Harbin Medical University, Harbin, China

<sup>2</sup> Department of Immunology, Hebei Medical University, Shijiazhuang, Hebei, China

<sup>3</sup> Department of Software Engineering, Software College, Northeastern University, Shenyang, China

<sup>4</sup> Department Maternal and Child Health, Guangxi Medical University, Nanning, China

<sup>5</sup> Department of Food Science and Nutrition, College of Sciences, Taif University, Taif, Saudi Arabia

<sup>6</sup> Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, Taif University, Taif, Saudi Arabia

## SUMMARY

**Background:** *Pseudomonas aeruginosa* is a Gram-negative bacteria that causes a large range of human infections such as lung infection (cystic fibrosis) and urinary tract infection. Even worse, antibiotic resistant bacteria have become a serious health care problem throughout the last decade, and there is a need for a clear approach to regulate and prevent the spread of *pseudomonas aeruginosa* resistance.

**Methods:** A complete analysis of *Pseudomonas aeruginosa* proteomics data showed that 25% of proteins are hypothetical proteins (HPs) whose function is not precisely defined. HP gene sequence analysis offers a framework for defining sequence-function relationships with a deeper understanding of organisms' molecular mechanisms at the system level. In the current research, we used the power of different bioinformatics tools to assign the potential roles for the HPs based on protein family association, amino acid function, motifs, and pathway analysis.

**Results:** The current findings show that 30 HPs have well-defined functions and are classified as enzymes, DNA binding, periplasmic binding protein, transport, etc. Seven HPs showed virulence characteristics that is to be expected to be essential for *Pseudomonas aeruginosa* and pathogenesis survival.

**Conclusions:** This study's findings may encourage a better understanding of virulence mechanisms, drug resistance, pathogenesis, and drug discovery to treat *Pseudomonas aeruginosa* infections.

(Clin. Lab. 2021;67:xx-xx. DOI: 10.7754/Clin.Lab.2021.210536)

---

### Correspondence:

Zhao-Hua Zhong  
Department of Microbiology  
Harbin Medical University  
Harbin  
China  
Email: Zhongzh@hrbmu.edu.cn

## Supplemental Tables and Figures

Table S1 provides a summary of this study's results.

Table S1. List of expected physiochemical parameters for hypothetical proteins from *Pseudomonas aeruginosa* by ExPasy's ProtParam tool.

S. No.	Uniprot ID	Molecular weight, Mw (Da)	Theoretical PI	Extinction coefficient (M-1 cm-1)	Instability Index Aliphatic index computed	Classification	Aliphatic index	Grand average of hydropathicity (GRAVY)
1	Q9HTH1	49,556.47	5.84	78,505	37.93	stable	82.18	-0.371
2	Q9HTA4	32,809.35	9.38	54,680	135.89	unstable	40.66	0.928
3	Q9I1V6	14,890.01	5.86	3,015	44.50	unstable	98.26	-0.060
4	Q9HTA1	27,809.23	5.72	26,470	31.6	stable	128.98	0.807
5	HMUUV_PSEAE	27,524.73	6.21	14,105	35.08	stable	114.75	0.017
6	G3XD06	12,598.43	5.13	2,980	45.04	unstable	117.42	0.142
7	G3XDB1	9,221.93	10.29	7,240	54.44	unstable	103.41	0.073
8	G3XD38	17,538.62	4.93	26,470	28.34	stable	73.17	-0.263
9	G3XD68	12,870.43	5.06	9,970	46.73	unstable	84.87	-0.391
10	G3XD56	9,159.48	6.97	15,470	17.91	stable	95.60	-0.131
11	G3XD57	63,717.11	8.88	63,940	24.46	stable	77.17	-0.194
12	G3XCW1	12,684.22	9.51	26,470	24.95	stable	78.50	-0.430
13	Q9I2G2	8,402.36	8.16	8,480	29.38	stable	61.84	-0.433
14	Q9HTC4	10,6371.08	6.58	10,8205	51.86	unstable	110.42	-0.087
15	Q9I0Y4	21,721.62	5.48	8,605	26.58	stable	111.67	0.049
16	Q9I1Y2	31,896.77	9.48	32,345	46.70	unstable	100.42	-0.246
17	Q9I1Y1	20,596.58	6.38	6,990	40.22	unstable	85.53	-0.245
18	Q9I1X9	19,213.77	5.48	20,970	51.40	unstable	97.07	-0.219
19	Q9I1X8	19,313.33	5.56	8,605	50.81	unstable	110.73	0.014
20	Q9HUR9	19,034.77	7.09	19,605	26.33	stable	130.26	1.054
21	Q9HWS9	47,313.03	6.00	58,455	27.96	stable	84.63	-0.291
22	Q9HZR9	45,934.61	7.11	68,410	63.67	unstable	82.87	-0.690
23	Q9HZR8	21,501.55	10.12	17,085	35.13	stable	124.85	0.897
24	Q9I6W1	13,501.53	5.56	8,855	52.73	unstable	76.67	-0.046
25	Q9I6W0	19,800.49	6.52	10,555	50.56	unstable	94.50	-0.254
26	Q9I0F8	74,012.20	6.52	98,905	43.83	unstable	93.00	-0.206
27	Q9I0F4	37,347.31	8.45	6,085	39.22	stable	104.38	0.011
28	Q9I0F3	32,028.72	9.16	23,045	42.83	unstable	91.29	-0.235
29	Q9I0F6	19,800.49	6.52	10,555	50.56	unstable	94.50	-0.254
30	Q9I0F2	17,487.18	4.53	7,450	41.49	unstable	117.52	0.334
31	Q9I120	68,931.95	4.78	15,470	45.34	unstable	96.06	-0.121
32	Q9I1I9	13,148.19	8.96	9,970	39.69	stable	88.68	-0.209
33	Q9I1J2	33,632.66	6.98	36,900	31.16	stable	106.73	0.230
34	Q9HUP5	34,364.33	9.58	40575	38.37	stable	95.53	-0.204
35	Q9I3S8	145,820.60	6.17	231,200	41.33	unstable	76.90	-0.549
36	Q9HY89	81,652.49	5.90	108,220	37.79	stable	76.25	-0.514
37	Q9I1U3	45,495.08	5.37	51,715	40.09	unstable	94.26	0.023
38	Q9I1S5	32,665.28	9.08	43,430	44.52	unstable	95.86	-0.181

**Table S1.** List of expected physiochemical parameters for hypothetical proteins from *Pseudomonas aeruginosa* by Expasy's ProtParam tool (continued).

S. No.	Uniprot ID	Molecular weight, Mw (Da)	Theoretical PI	Extinction coefficient (M-1 cm-1)	Instability Index Aliphatic index computed	Classification	Aliphatic index	Grand average of hydropathicity (GRAVY)
39	Q9I1U5	7,488.90	10.78	17,990	52.65	unstable	132.90	0.906
40	Q9I1T5	51,396.99	5.51	73,465	58.43	unstable	88.34	-0.386
41	Q9I0U0	26,862.69	8.64	33,140	49.33	unstable	84.85	-0.346
42	Q9I1Q9	18,778.14	9.06	43,555	41.18	unstable	115.45	0.874
43	Q9I4A6	73,496.79	10.29	106,630	43.07	unstable	111.28	0.350
44	G3XCX9	28,012.57	11.11	49,960	45.21	unstable	100.87	-0.237
45	Q9I6P6	70,216.36	5.04	61,810	40.08	unstable	94.10	0.114
46	Q9I2M4	38,411.44	6.11	37,595	54.69	unstable	111.03	0.050
47	Q9HUC7	25,079.77	10.19	22,460	47.26	unstable	47.26	-0.535
48	Q9HVG1	16,827.52	6.90	21,095	50.26	unstable	90.20	-0.151
49	Q9HUC7	25,079.77	10.19	22,460	47.26	unstable	72.77	-0.535
50	Q9HU46	27,915.01	5.17	36,900	16.09	unstable	96.44	-0.225

Table S2 summarizes the results of those predictions.

**Table S2.** List of selected hypothetical proteins in *pseudomonas aeruginosa* predicted by the sub cell localization.

S. No.	Uniprot ID	Sub-cellular localization Psort	Signal Peptide	Secretory Protein (SecretomeP)	Transmembrane helices prediction	
					CCTOP	TMHMM
1	Q9HTH1	cytoplasm	no	no	NIL	1TMH
2	Q9HTA4	cytoplasm	yes	no	10TMH	9TMH
3	Q9I1V6	cytoplasm	no	no	NIL	NIL
4	Q9HTA1	cytoplasm	yes	no	7TMH	7 TMH
5	HMUV_PSEAE	cytoplasm	no	no	NIL	NIL
6	G3XD06	periplasmic	yes	no	1TMH	1TMH
7	G3XDB1	periplasmic	yes	no	1TMH	NIL
8	G3XD38	cytoplasm	no	no	1TMH	NIL
9	G3XD68	cytoplasm	no	no	1THM	NIL
10	G3XD56	cytoplasm	yes	no	1TMH	NIL
11	G3XD57	cytoplasm	no	no	1THM	NIL
12	G3XCW1	cytoplasm	no	no	1THM	NIL
13	Q9I2G2	cytoplasm	no	no	1THM	NIL
14	Q9HTC4	cytoplasm	yes	no	6THM	5TMH
15	Q9I0Y4	periplasmic	yes	no	NIL	NIL
16	Q9I1Y2	cytoplasm	no	no	1THM	NIL
17	Q9I1Y1	cytoplasm	yes	no	1THM	NIL
18	Q9I1X9	cytoplasm	yes	no	1THM	1THM
19	Q9I1X8	periplasmic	no	no	NIL	NIL
20	Q9HUR9	cytoplasm	yes	no	6THM	7THM
21	Q9HWS9	cytoplasm	yes	no	4THM	NIL

Table S2. List of selected hypothetical proteins in *pseudomonas aeruginosa* predicted by the sub cell localization (continued).

S. No.	Uniprot ID	Sub-cellular localization Psort	Signal Peptide	Secretory Protein (SecretomeP)	Transmembrane helices prediction	
					CCTOP	TMHMM
22	Q9HZR9	cytoplasm	no	no	1THM	NIL
23	Q9HZR8	cytoplasm	no	no	6THM	NIL
24	Q9I6W1	cytoplasm	no	no	1THM	NIL
25	Q9I6W0	cytoplasm	yes	no	2THM	NIL
26	Q9I0F8	cytoplasm	no	no	1THM	NIL
27	Q9I0F4	cytoplasm	no	no	1THM	NIL
28	Q9I0F3	periplasmic	no	no	1THM	NIL
29	Q9I0F6	cytoplasm	no	no	1THM	NIL
30	Q9I0F2	periplasmic	yes	no	1THM	NIL
31	Q9I120	periplasmic	yes	no	1THM	NIL
32	Q9I1I9	periplasmic	yes	no	3THM	NIL
33	Q9I1J2	cytoplasm	yes	no	3THM	NIL
34	Q9HUP5	cytoplasm	no	no	NIL	NIL
35	Q9I3S8	cytoplasm	no	no	2THM	NIL
36	Q9HY89	cytoplasm	no	no	NIL	NIL
37	Q9I1U3	periplasmic	no	no	3THM	NIL
38	Q9I1S5	cytoplasm	no	no	1THM	NIL
39	Q9I1U5	periplasmic	yes	no	2THM	NIL
40	Q9I1T5	cytoplasm	no	no	5 THM	NIL
41	Q9I0U0	cytoplasm	no	no	1THM	NIL
42	Q9I1Q9	cytoplasm	no	no	4THM	NIL
43	Q9I4A6	cytoplasm	no	no	11THM	NIL
44	G3XCX9	periplasmic	yes	no	1THM	NIL
45	Q9I6P6	cytoplasm	no	no	2THM	NIL
46	Q9I2M4	cytoplasm	no	no	4THM	NIL
47	Q9HUC7	cytoplasm	no	no	1THM	NIL
48	Q9HVG1	cytoplasm	no	no	1THM	NIL
49	Q9HUC7	cytoplasm	no	no	1THM	NIL
50	Q9HU46	periplasmic	yes	no	1THM	NIL

Table S3, S4 summarizes the results of those predictions.

Table S3. Function annotation of selected hypothesized *pseudomonas aeruginosa* proteins predicted using CATH, Superfamily, PFAM, and CDART.

S. No.	Uniprot ID	Pfam	Superfamily	CATH	CDART
1	Q9HTH1	Pyridine nucleotide-disulphideoxidoreductase	FAD/NAD(P)-binding domain	Uncharacterized oxidoreductaseCzcO	MSMEG_0569 family flavin-dependent
2	Q9HTA4	Membrane transport protein	NIL	NIL	AEC family transporter
3	Q9I1V6	CBS domain	CBS domain Pair	CBS domain	MULTISPECIES: CBS domain-containing

**Table S3. Function annotation of selected hypothesized *pseudomonas aeruginosa* proteins predicted using CATH, Superfamily, PFAM, and CDART (continued).**

S. No.	Uniprot ID	Pfam	Superfamily	CATH	CDART
4	Q9HTA1	Integral membrane protein TerC family	NIL	NIL	TerC family protein
5	HMOV_PSEAE	ABC transporter	P-loop containing nucleoside triphosphate hydrolases	ABC transporter, ATP-binding subunit	heme ABC transporter
6	G3XD06	NIL	NIL	NIL	NIL
7	G3XDB1	NIL	NIL	NIL	Local query sequence
8	G3XD38	Phage tail tube, TTP, lambda-like	NIL	NIL	Local query sequence
9	G3XD68	Phage tail assembly chaperone, TAC	NIL	NIL	Local query sequence
10	G3XD56	Catechol dioxygenase N	NIL	NIL	Local query sequence
11	G3XD57	Lambda phage tail tape-measure	NIL	NIL	Local query sequence
12	G3XCW1	Phage minor tail protein	NIL	NIL	Local query sequence
13	Q9I2G2	Replicasepolyprotein 1a N-terminal domain	NIL	NIL	NIL
14	Q9HTC4	EAL domain	PYP-like Sensor domain (PAS domain)	Diguanylatecyclase/phosphodiesterase with PAS/PAC sensor	NIL
15	Q9I0Y4	Ankyrin repeats	ankyrin repeat domain-containing	Probable serine/threonine-protein kinase DDB_G0267566	ankyrin repeat domain-containing
16	Q9I1Y2	EAL domain	EAL domain-like	Diguanylatecyclase/phosphodiesterase with PAS/PAC sensor	EAL domain-containing protein taxonomy span: cellular organisms
17	Q9I1Y1	Domain of unknown function	NIL	NIL	Local query sequence
18	Q9I1X9	NIL	NIL	NIL	Local query sequence
19	Q9I1X8	Response regulator receiver domain	CheY-like	NIL	Response regulator
20	Q9HUR9	HupE/UreJ protein	NIL	NIL	Urease accessory protein UreJ
21	Q9HWS9	Carbohydrate-selective porin, OprB family	NIL	NIL	Hypothetical protein
22	Q9HZR9	NIL	NIL	NIL	Local query sequence
23	Q9HZR8	LysE type translocator	NIL	NIL	LysE family translocator
24	Q9I6W1	MCM P-loop domain	NIL	NIL	Local query sequence
25	Q9I6W0	Stage II sporulation protein E (SpoIIE)	HAMP domain-like	Phosphoserine phosphatase RsbU	HAMP domain
26	Q9I0F8				Probable quinoneoxidoreductase
27	Q9I0F4	RHS Repeat	PDB	NIL	NIL
28	Q9I0F3	Phage late control gene D protein	PDB	Rhs element Vgr protein	type VI secretion protein ImpA

**Table S3. Function annotation of selected hypothesized *pseudomonas aeruginosa* proteins predicted using CATH, Superfamily, PFAM, and CDART (continued).**

S. No	Uniprot ID	Pfam	Superfamily	CATH	CDART
29	Q9I0F6	Dienelactone hydrolase family	Alpha/beta-Hydrolases	Putative carboxy methylene butenolidase	Carboxy Methylene butenolidase
30	Q9I0F2	Putative esterase	alpha/beta-Hydrolases	Bacillibactinprilactone hydrolase BesA	Hypothetical protein
31	Q9I120	Hemagglutination activity domain	Pectin lyase-like Superfamily DmpA/ArgJ-like Superfamily	NIL	filamentous hemagglutinin N-terminal
32	Q9I1I9	Peptidase family M48	NIL	NIL	Peptidase family M48
33	Q9I1J2	Peptidase family M48	NIL	NIL	Peptidase family M48
34	Q9HUP5	Isochorismatase family	Isochorismatase-like hydrolases	pyrazinamidase/nicotinamidase	Nicotinamidase /pyrazinamidase
35	Q9I3S8	FAD linked oxidases	FAD linked oxidases	Glycolate dehydrogenase	FAD-binding oxidoreductase
36	Q9HY89	Heavy-metal-associated domain	NIL	NIL	Heavy-metal-associated domain
37	Q9I1U3	M42 glutamylaminopeptidase	M42 glutamylaminopeptidase	M42 glutamylaminopeptidase	M42 glutamylaminopeptidase
38	Q9I1S5	KGG Stress-induced bacterial acidophilic repeat motif	NIL	NIL	KGG Stress-induced bacterial acidophilic repeat motif
39	Q9I1U5	Phospholipase_D-nuclease N-terminal	NIL	NIL	Phospholipase_D-nuclease N-terminal
40	Q9I1T5	Iron-containing redox enzyme	NL	NIL	Iron-containing redox enzyme
41	Q9I0U0	Acetyltransferase (GNAT) domain	Acetyltransferase (GNAT) domain	Acetyltransferase (GNAT) domain	Acetyltransferase (GNAT) domain
42	Q9I1Q9	Tripartite tricarboxylate transporter TctB	NIL	NIL	Tripartite tricarboxylate transporter TctB
43	Q9I4A6	Fusaric acid resistance protein family	NIL	NIL	Fusaric acid resistance protein family
44	G3XCX9	DnaJ domain	Chaperone J-domain	DnaJ-like protein DjIA	co-chaperone DjIA
45	Q9I6P6	4-hydroxyphenylpyruvate dioxygenase and related hemolysins	Hydroxyphenylpyruvate dioxygenase	Dihydroxybiphenyl dioxygenase	Probable 4-hydroxyphenylpyruvate dioxygenase oxidoreductase
46	Q9I2M4	CorA-like Mg <sup>2+</sup> transporter protein	NIL	CorA-like Mg <sup>2+</sup> transporter protein	CorA-like Mg <sup>2+</sup> transporter protein
47	Q9HUC7	Sporulation related domain	Sporulation related repeat	Cell division protein	Sporulation related domain
48	Q9HVG1	SnoaL-like polyketidecyclase	NTF2-like	NIL	NTF2-like
49	Q9HUC7	Sporulation related repeat	Sporulation related repeat	NIL	Cell division protein FtsN
50	Q9HU46	Periplasmic binding protein-like II,	Periplasmic binding protein-like II	NIL	ABC-type amino acid transport/signal transduction system, periplasmic component/

**Table S4. Virulence protein annotations of selected hypothesized *pseudomonas aeruginosa* proteins predicted using VICM pred.**

<b>Cellular Process</b>	<b>G3XD68, G3XD57, G3XCW1, Q9I2G2, Q9I1Y2, Q9I1Y1, Q9I1X9, Q9I1X8, Q9HZR8, Q9I6W1, Q9I0F6, Q9HUP5, Q9I1S5, Q9I1T5, Q9I6P6, Q9HUC7, Q9HUC7, Q9HU46, Q9I1Q9</b>
<b>Information Molecule</b>	<b>Q9I1V6, Q9HTC4, Q9I0F4, Q9I120, Q9I1I9, Q9I4A6</b>
<b>Metabolism</b>	<b>Q9HTH1, Q9HTA1, Q9HTA4, HMUV_PSEAE, G3XD06, G3XDB1, G3XD56, Q9HUR9, Q9I6W0, Q9I0F3, Q9I1J2, Q9I3S8, Q9HY89, Q9I1U5, Q9I0U0, G3XCX9, Q9I2M4,</b>
<b>Virulence factors</b>	<b>G3XD38, Q9I0Y4, Q9HWS9, Q9HZR9, Q9I0F8, Q9I0F2, Q9I1U3</b>