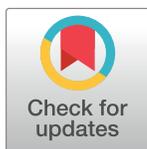


RESEARCH ARTICLE

Predicting drug targets by homology modelling of *Pseudomonas aeruginosa* proteins of unknown function

Nikolina Babic, Filip Kovacic¹*

Institute of Molecular Enzyme Technology, Heinrich-Heine Universität Düsseldorf, Forschungszentrum Jülich GmbH, Jülich, Germany

* f.kovacic@fz-juelich.de

Abstract

The efficacy of antibiotics to treat bacterial infections declines rapidly due to antibiotic resistance. This problem has stimulated the development of novel antibiotics, but most attempts have failed. Consequently, the idea of mining uncharacterized genes of pathogens to identify potential targets for entirely new classes of antibiotics was proposed. Without knowing the biochemical function of a protein, it is difficult to validate its potential for drug targeting; therefore, the functional characterization of bacterial proteins of unknown function must be accelerated. Here, we present a paradigm for comprehensively predicting the biochemical functions of a large set of proteins encoded by hypothetical genes in human pathogens to identify candidate drug targets. A high-throughput approach based on homology modelling with ten templates per target protein was applied to the set of 2103 *P. aeruginosa* proteins encoded by hypothetical genes. The >21000 homology modelling results obtained and available biological and biochemical information about several thousand templates were scrutinized to predict the function of reliably modelled proteins of unknown function. This approach resulted in assigning one or often multiple putative functions to hundreds of enzymes, ligand-binding proteins and transporters. New biochemical functions were predicted for 41 proteins whose essential or virulence-related roles in *P. aeruginosa* were already experimentally demonstrated. Eleven of them were shortlisted as promising drug targets that participate in essential pathways (maintaining genome and cell wall integrity), virulence-related processes (adhesion, cell motility, host recognition) or antibiotic resistance, which are general drug targets. These proteins are conserved in other WHO priority pathogens but not in humans; therefore, they represent high-potential targets for preclinical studies. These and many more biochemical functions assigned to uncharacterized proteins of *P. aeruginosa*, made available as PaPUF database, may guide the design of experimental screening of inhibitors, which is a crucial step towards the validation of the highest-potential targets for the development of novel drugs against *P. aeruginosa* and other high-priority pathogens.

OPEN ACCESS

Citation: Babic N, Kovacic F (2021) Predicting drug targets by homology modelling of *Pseudomonas aeruginosa* proteins of unknown function. PLoS ONE 16(10): e0258385. <https://doi.org/10.1371/journal.pone.0258385>

Editor: Abdelwahab Omri, Laurentian University, CANADA

Received: April 30, 2021

Accepted: September 24, 2021

Published: October 14, 2021

Copyright: © 2021 Babic, Kovacic. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting Information](#) files.

Funding: Both authors thank the Jürgen Manchot Stiftung for funding. The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Even after more than a half-century of development of antibiotics, bacterial infections remain a significant cause of morbidity and mortality worldwide for various reasons including unsuccessful antibiotic treatments. The rise in multidrug-resistant bacterial infections has been further accelerated by the overuse of antibiotics in present pandemic conditions [1]. It is evident that without new countermeasures against multidrug-resistant bacteria, so-called “superbugs”, routine medical procedures (cancer chemotherapy, organ transplantation, diabetes treatment) become very high risk, and the achievement of the Millennium Development Goals is endangered [2].

Attempts of the World Health Organization (WHO) to guide and promote research and development (R&D) to obtain new antibiotics failed [2, 3]. The hurdles faced by the current system for the R&D of antibiotics are tremendous, which is the reason why only a few solutions (32 antibiotics and ten biological treatments) the WHO priority pathogens are currently in the R&D pipeline (clinical phases 1 to 3) [2–4]. Generally, the most attractive targets for new classes of antibiotics are proteins involved in previously untargeted essential biochemical pathways that are conserved among pathogens and absent in humans [1, 5]. The targeted screening of such enzymes identified the reductase IspH, which is involved in isoprenoid biosynthesis [6]; the reductase FabI, which is involved in fatty acid biosynthesis [7]; and the peptidase LspA, which is involved in the posttranslational processing of lipoprotein [8]. The inhibition of these enzymes by C23.28–TPP (preclinical study) [6], isoniazid (approved drug) [7] and myxovirescin (clinical phase 3) [8], respectively, showed strong antibacterial activity.

New antibiotics represent only one solution with uncertain efficacy because of the fast evolution of resistance observed after each new antibiotic was introduced [9]. Another promising strategy is to counteract the nonessential functions mediated by bacterial virulence or antibiotic resistance factors, thereby blocking host-pathogen interactions, the progression of infection, or antibiotic expulsion without killing the bacteria or inhibiting their growth. Such antivirulence drugs, suggested as a promising supplement with the ability to enhance the effects of classical antibiotics, are thought to exert weak selective pressure, minimizing the risk of resistance emergence [10]. The main antivirulence approaches of drugs in clinical development include the inhibition of quorum-sensing systems, biofilm formation, and toxin production or function [7, 10].

Obviously, several proteins targeted by novel potential antibiotics or antivirulence drugs were identified in the last decade; however, many more novel essential proteins or important virulence factors with unknown functions in bacteria still need to be characterized, as it is difficult to set up an inhibitor screening for proteins of unknown function (PUFs) [1]. Therefore, understanding the function of these proteins is a prerequisite for the validation of targets for the development of antibacterial drugs with novel mechanisms of action.

P. aeruginosa, a critical priority pathogen according to the WHO, is a serious threat in hospitals, where multidrug-resistant strains cause severe acute infections, often with deadly outcomes [3]. Life-threatening chronic *P. aeruginosa* infections have been frequently observed in immunocompromised persons, e.g., cystic fibrosis and cancer patients [11]. *P. aeruginosa* infections are a concern because of the high resistance towards the best currently available antibiotics—carbapenems and the third generation of cephalosporins [3]. To cope with antibiotics, *P. aeruginosa* uses various intrinsic and acquired resistance mechanisms, including the expression of different classes of β -lactamases, the loss of outer membrane channel proteins to increase outer membrane impermeability, the overexpression of antibiotic efflux pumps, the synthesis of antibiotic modifying enzymes, and mutations in the primary targets of quinolone antibiotics [12]. Moreover, the formation of antibiotic-resistant biofilms [12], exceptional adaptability to infect various human tissues owing to multiple signalling pathways [13] and cell surface adhesins [14, 15], and the fast progress of infections as a result of the production of

a vast number of potent extracellular toxins [16] present challenges in the treatment of *P. aeruginosa* infections. All these features qualify *P. aeruginosa* as a model pathogen bacterium for the R&D of novel antibiotics.

Here, we aimed to assign biochemical functions to hypothetical proteins from the human pathogen *P. aeruginosa*. Using a homology modelling-based high-throughput approach, novel putative functions were reliably assigned to many proteins for which functions could not be previously predicted from the sequence. Several proteins with new putative biochemical functions could be linked to important infection-related processes, which suggests that these proteins are promising targets for the development of drugs against *P. aeruginosa*.

Materials and methods

Selection of target genes

The GUFs from *P. aeruginosa* PAO1 were defined as genes with “hypothetical protein” and “conserved hypothetical protein” annotations in the product name field of the *Pseudomonas* Genome Database (PGD) [17]. In the PGD, all genes were analysed by sequence alignment and classified into four classes. Class IV includes hypothetical proteins, which are defined as those proteins whose function could not be proposed based on conserved amino acid motifs, structural features, or limited sequence similarity with experimentally studied proteins. Gene and protein sequences were retrieved from the PGD (www.pseudomonas.com) and UniProt databases [18] (www.uniprot.org), respectively.

Comparative genomics

PGD annotations were used to find proteins with human homologues, proteins with homologues in other pathogens, and virulence factors. The list of virulence factors was obtained from the Virulence Factor Database (www.mgc.ac.cn/VFs) [19] and the Victors Virulence Factors Knowledge Database (www.phidias.us/victors) [20]. Comparative BLAST was performed using NCBI's BLAST suite (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

A Diamond BLASTp tool [21] from the PGD was used to search for groups of orthologues with >70% sequence identity and >95% alignment coverage among different *P. aeruginosa* strains isolated from humans with developed disease as indicated in the PGD and among other pathogenic *Pseudomonas* species, namely, *P. alcaligenes*, *P. mendocina*, and *P. otitidis*. NCBI BLAST was used to search for homologues with >35% sequence identity and >75% alignment coverage in 13 WHO priority pathogens.

Homology modelling

Homology modelling was performed using the batch processing feature of the Phyre2 web server [22, 23]. The obtained output file was imported to Microsoft Excel, and the information on the ten best structural models was analysed to identify reliable homology models. For this purpose, the following cut-off values were used: modelling confidence >75%, domain size >100 residues for transporters and enzymes, and domain size >50 residues for binding proteins. The functions of reliable templates were identified by text mining the information from the “PDB header” and “PDB molecule” fields in the Phyre2 output file. Text mining was performed automatically using Excel, and the keywords are listed in [S6 Table](#).

Selected models were energy minimized several times with the GROMOS96 43B1 force field as implemented in Swiss-PdbViewer [24]. The stereochemical parameters of the homology models were evaluated using PROCHECK v.3.5 [25]. The DALI database [26] (<http://ekhidna2.biocenter.helsinki.fi/dali/oldstyle.html>) was used to identify structures similar to the

homology models. The protein structures were analysed, aligned, and visualized using PyMOL (<http://www.pymol.org>).

Results

P. aeruginosa genes of unknown function are potential antibiotic targets

The development of efficient drugs against multiresistant bacteria relies on the identification of entirely new bacterial targets whose inhibition will hamper bacterial growth (antibiotics) or prevent the pathogen from being virulent (antivirulence drugs). Such targets are likely to be found among genes of unknown function (GUF), so-called genomic “dark matter” [1]. Interestingly, even 20 years after the *P. aeruginosa* PA01 genome was sequenced, nearly 40% of genes remain GUFs, which is true of the majority of sequenced bacterial genomes [27, 28].

To identify GUFs in *P. aeruginosa* PA01, we searched for functional class IV hypothetical genes in the manually curated *Pseudomonas* Genome Database (PGD) [17]. The results revealed 1470 hypothetical genes, which have no homologues in other species, and 633 conserved hypothetical genes, which have homologues without known function in other bacteria (Fig 1A and S1 Table). These 2103 genes of *P. aeruginosa* considered GUFs were analysed by integrating comparative sequence analysis and structural homology-based functional prediction.

Bacterial antibiotic targets that are absent in humans are particularly interesting, as their inhibition should not exert side effects on host physiology [1, 7]. PGD [17] annotation of *P. aeruginosa* proteins with human homologues revealed that even 96% of GUFs do not have human homologues (Fig 1C).

The essential genes, genes conserved only in pathogenic species and genes with demonstrated virulence function represent high-priority target groups for the R&D of antibiotics

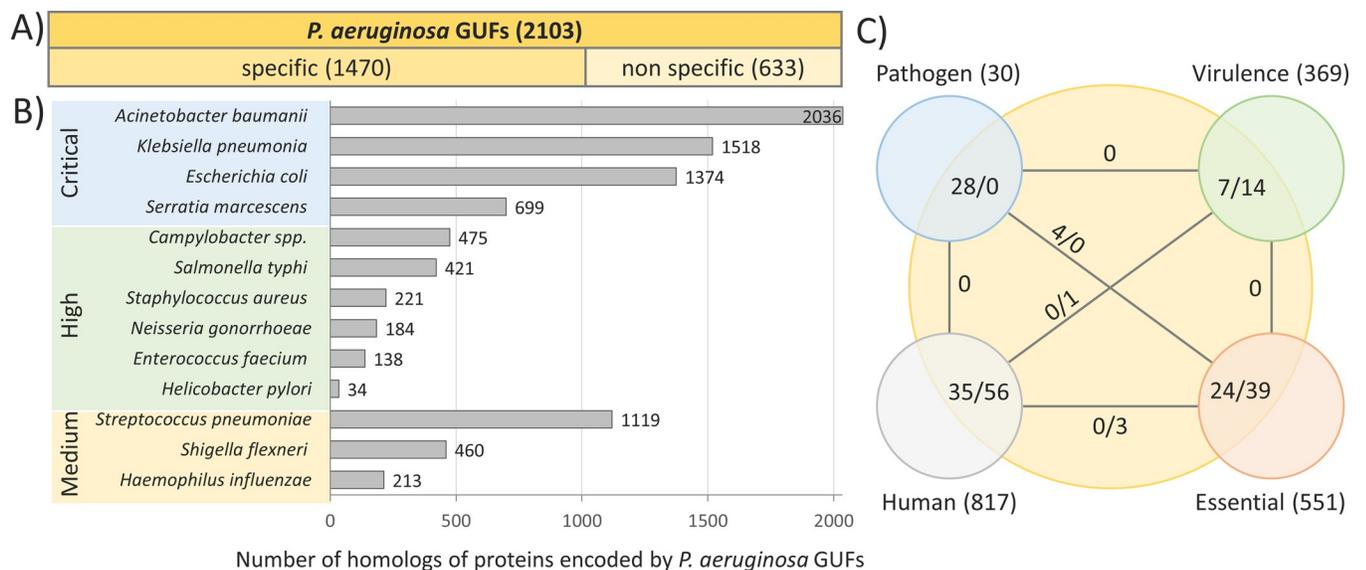


Fig 1. A large portion of the genomes of *P. aeruginosa* PA01 and WHO priority pathogens consist of GUFs. A) The majority of *P. aeruginosa* GUFs are found only in *Pseudomonas* species (specific), while some of them are conserved in other organisms (nonspecific). The analysis was performed using the PGD version 20.2 download on 5.11.2020. B) Conservation of *P. aeruginosa* genes of unknown function in WHO priority pathogens. *P. aeruginosa* GUFs were used as a query for a BLAST search on NCBI to find protein homologues (>35% sequence identity, >75% sequence coverage) in the following thirteen WHO priority pathogens. Taxonomic identifiers of organisms are listed in S2 Table. C) Pathogen-specific genes, essential genes, virulence factor genes and genes with human homologues among GUFs. Numbers in brackets indicate the number of genes in each group. A number of *P. aeruginosa* GUFs in overlapping sections is indicated as the number of specific genes/number of nonspecific genes. Numbers on lines indicate mutual GUFs among two linked groups.

<https://doi.org/10.1371/journal.pone.0258385.g001>

because they may participate in key cellular or virulence processes whose inhibition might be lethal or virulence attenuating [29]. We found that 63 GUFs of *P. aeruginosa* are essential genes (Fig 1C and S1 Table), defined as genes whose experimental mutation leads to cell death [30]. Furthermore, overlapping *P. aeruginosa* GUFs with *P. aeruginosa* genes that have homologues only in pathogenic microorganisms (pathogen-associated genes) [31] revealed 28 hits. Four of them (PA0442, PA0977, PA1369, PA2139) are particularly promising targets for a broad range antibiotics, as they were also annotated as generally essential genes (Fig 1B and S1 Table).

Furthermore, we analysed which GUFs might play a role in pathogenic processes by searching for virulence-associated GUFs listed in the virulence factor annotation of the PGD [17]. This list includes genes whose roles in virulence were experimentally proven by studying respective mutant strains or were inferred from sequence similarity with virulence factors from other pathogens. Thirteen identified GUFs were previously linked to virulence-related processes by studying their functions during the infection of model host organisms (e.g., *Caenorhabditis elegans*, *Drosophila melanogaster* or *Rattus norvegicus*), and an additional five were related to common virulence traits (flagella-mediated motility, pyoverdine-mediated iron uptake, type VI secretion of toxins) (S3 Table) [32–36]. Among these virulence-associated GUFs, eleven are conserved in other organisms (S3 Table), suggesting that their analysis might reveal novel insights into general virulence processes in bacteria.

Assigning putative biochemical functions to *P. aeruginosa* proteins of unknown function

To better understand the role of GUFs in *P. aeruginosa* infections, it is necessary to know the biochemical functions of proteins encoded by GUFs. We used 3D structure homology modelling as a tool for predicting the biochemical functions of PUFs. This approach relies on the fundamental principle that two homologous proteins with similar sequences have very similar structures. As protein structure is 3–10 times more conserved than the sequence [37], structural modelling can reliably detect homology between proteins with low sequence similarity (<20%) [22]. Knowing that structure determines protein function, one can assume that the functions of the modelled PUF and the template protein with a known structure are likely to be the same. This approach used the advantage of, in most cases, precisely assigned biochemical and even biological functions to proteins with known structures. A further advantage is that generated structural models of target PUFs are useful for the validation of functionally relevant residues, e.g., active site or ligand-binding residues. Additionally, several distinct functional domains in a single target PUF may be identified using different domains with solved structures, and these domains can be combined into a unique multidomain structure. Conclusively, homology modelling of PUF and its comparison with the template may reliably reveal the putative function of PUF.

Using protein sequences of PUFs obtained from the UniProt database, we performed homology modelling of 2103 PUFs using the batch homology modelling feature of the Phyre2 server [22]. Retrieving ten models per PUF, we compiled a set of data of approximately 21020 (PA2462, with a size of 573.2 kDa, was too large for homology modelling) homology modelled templates (S4 Table). These results were refined to select reliable models using an automated pipeline (Fig 2). A Phyre confidence score, which indicates a degree of homology between target proteins and the template, was used as the main criterion for the validation of structural prediction. According to this parameter, fairly reliable homology models (>80% conf.) were obtained for most PUFs (75%) (S1A Fig), although their sequence identities were rather low since almost 65% of modelled proteins had sequence identities below 20% compared with the corresponding templates (S2 Fig). Interestingly, 89 PUFs have very low confidences (<20%)

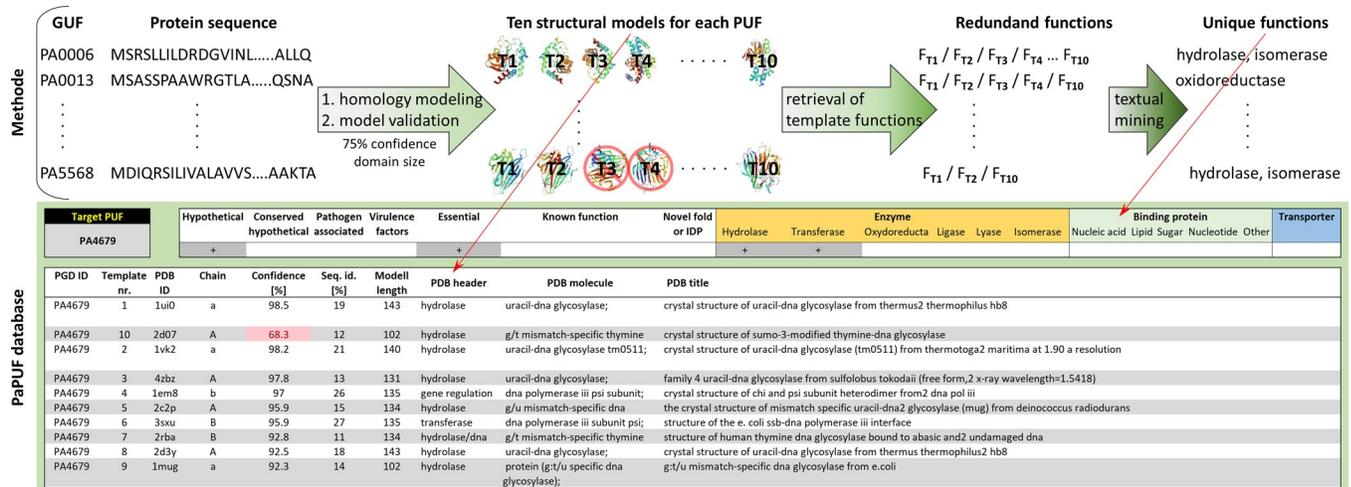


Fig 2. Pipeline for prediction of protein function based on homology modelling with multiple templates. Protein sequences of all 2103 *P. aeruginosa* GUFs were retrieved from the UniProt database and submitted to the Phyre2 server to model their structures. Ten structural models for each PUF were validated according to the confidence of homology modelling (>75%) and the size of the modelled domain (>100 residues for transporters and enzymes; >50 residues for binding proteins) to remove unreliable models. Information about the functions of reliable templates was retrieved from their PDB header and PDB molecule fields, and text mining was performed with keywords listed in S6 Table to assign biochemical functions to templates automatically. Redundant functions were removed to compile a list of unique functions of templates that were assigned as putative functions of PUFs. The results are freely available as a downloadable PaPUF database (www.iet.uni-duesseldorf.de/arbeitsgruppen/bacterial-enzymology) in which putative functions of *P. aeruginosa* PUFs and ten respective templates can be retrieved by entering a gene name into the “Target PUF” field (upper left corner).

<https://doi.org/10.1371/journal.pone.0258385.g002>

(S1 Table), which indicates that these PUFs do not have even remote structural homologues. Hence, these proteins might be intrinsically disordered or have a novel fold and therefore are particularly interesting from the structural biology point of view. Interestingly, we identified 65 PUFs with known structures, as their sequence identities with the template were >96%. Among them were 18 proteins with experimentally demonstrated functions (S5 Table); therefore, they were excluded from our further analysis, and the information is now uploaded to the PGD to improve its annotation.

To predict the biochemical functions of PUFs with the potential to be novel drug targets, we focused on enzymes, ligand-binding proteins and transporters, which perform indispensable functions for the bacterial cell cycle, infections and antibiotic resistance, e.g., primary metabolism, toxin-mediated host cell lysis, cell surface adhesion and antibiotic expulsion [15, 38, 39]. Consequently, their inhibition might lead to death or reduce the virulence of *P. aeruginosa*. To increase the reliability of the functional assignment, we eliminated homology models smaller than 100 residues for enzymes and transporters and smaller than 50 residues for ligand-binding domains. This is in agreement with a thermodynamic folding study that suggests that the optimal size of protein domains is 100 residues [40]. The cut-off for the rejection of binding domains was set to 50 residues, as helix-turn-helix zinc finger DNA-binding domains [41], five-stranded beta-barrel oligonucleotide and oligosaccharide-binding domains [42] and lipid-binding domains [43] of virulence-related proteins are usually not smaller than 50 residues. This comprehensive and accurately refined data set consisting of several reliable protein templates per PUF was a starting point for functional prediction using information about the biochemical functions of the templates. The general text mining algorithm based on the set of keywords (S6 Table), which precisely describe the protein functions studied here, resulted in the prediction of enzyme, binding or transport functions for 59% (1228) of PUFs (Fig 3A). Interestingly, 304 PUFs had two and 59 had three of the analysed functions (Fig 3B).

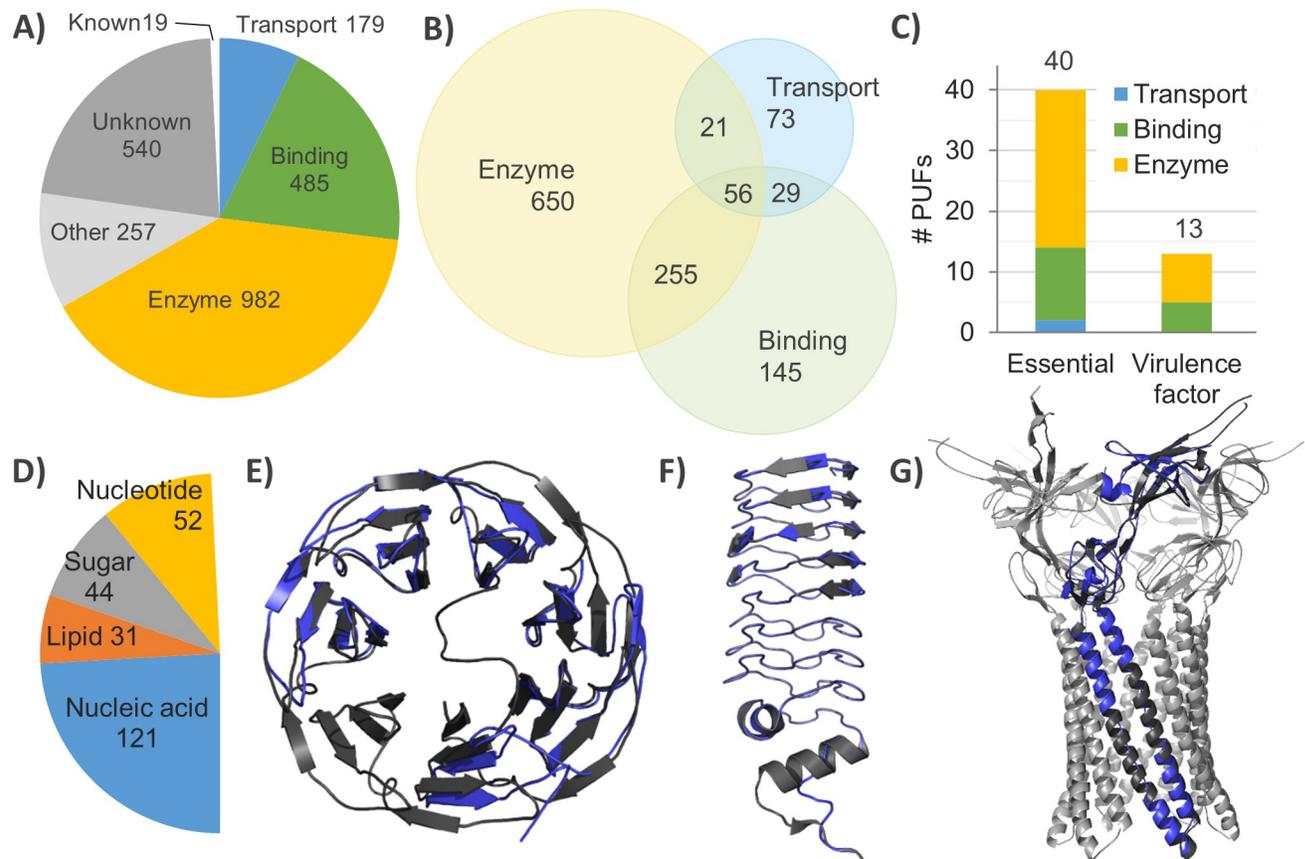


Fig 3. Predicted biochemical functions of *P. aeruginosa* proteins of unknown function. A) Number of putative enzymes, ligand-binding proteins, transporters, proteins with functions other than the previous three and proteins with known functions. B) Venn diagram of PUFs with enzyme, ligand-binding or transport functions. C) Distribution of essential and virulence factors among PUFs with the predicted enzyme, ligand-binding or transport functions. Gene identifiers of PUFs with predicted functions (Fig 3A–3C) are provided in S1 Table. D) Number of predicted ligand-binding proteins. E) Superimposition of a homology model of putative lectin PA5033 with *A. aegerita* lectin (PDB ID 4TQJ, chain A, 88.4% conf., 19% seq. id., 1.11 Å rmsd_{cc}, 92% coverage, 1.3% of disall. res.). F) Superimposition of a homology model of putative DNA gyrase binding protein PA1981 with the *Mycobacterium tuberculosis* MfpA protein (PDB ID 2BM5, chain A, 99% conf., 22% seq. id., 1.45 Å rmsd_{cc}, 80% coverage). G) Superimposition of a homology model of putative antibiotic efflux pump protein PA3304 with *E. coli* MacA (PDB ID 3FPP, chain A, 99% conf., 23% seq. id., 0.56 Å rmsd_{cc}, 81% coverage, 1.9% disall. res.). PA3304 is superimposed on one of six MacA protomers, which form a barrel-like functional protein. A homology model is indicated in blue, and the corresponding template grey. The protein structures were visualized using PyMOL software.

<https://doi.org/10.1371/journal.pone.0258385.g003>

Unfortunately, the functions of four proteins that belong to *Pseudomonas*-specific and essential protein groups (Fig 1C) were not predicted. Analysis of enzymes, transporters and binding proteins with newly predicted functions revealed that 41 PUFs might be considered potential drug targets, as they were found in essential or virulence-related groups (Fig 3C). Interestingly, comparative sequence analysis revealed that these proteins have 400 homologues (>35% seq. id.) in priority pathogens (S2 Table). Among them are only four proteins with human homologues (S2 Table). Thus, the assignment of putative functions presented here illustrates the global significance for R&D of potential bacteria-specific drug targets.

Virulence-related functions of putative ligand-binding proteins and transporters

Ligand-binding domains were predicted in 484 proteins, which showed a large overlap (310 proteins) with putative enzymes (Fig 3B). Among them, the proteins predicted to bind nucleic

acids were the most abundant (122), followed by nucleotide-binding (52), sugar-binding (44) and lipid-binding (33) proteins (Fig 3D). An interesting example of a virulence-related binding protein is the putative lectin PA5033 (Fig 3E), which was reliably modelled (88% conf.) with the structure of lectin from the mushroom *Agrocybe aegerita* [44]. The results revealed only a small portion (1.3%) of residues in disallowed regions (disall. res.) of the Ramachandran plot as revealed by the stereochemical validation of the energy-minimized PA5033 homology model with PROCHECK (S7 Table) [25]. Hence, most (~80%) high-quality X-ray structures have <5% residues in disallowed regions [45]. This protein, which has experimentally determined extracellular localization, may be involved in bacterial adhesion to host cells or in biofilm formation, as described for the other two (LecA and LecB) extracellular lectins of *P. aeruginosa*, which were suggested as targets of biofilm-specific antivirulence drugs [46]. The second interesting candidate is the putative DNA gyrase-binding protein PA1981 (Fig 3F), which was reliably (99% conf.) modelled using the *Mycobacterium tuberculosis* MfpA protein [47]. Proteins from this family bind to DNA gyrase, an antibiotic target with an essential function in the homeostatic control of chromosomal DNA supercoiling, which protects it against the antibiotic fluoroquinolone. Therefore, PA1981 is suggested to be implicated in resistance to fluoroquinolone.

Another example of an antibiotic resistance protein is the putative membrane-bound transporter PA3304, which was reliably modelled with the *E. coli* MacA adaptor protein of the efflux pump involved in antibiotic expulsion (Fig 3G and S7 Table) [48]. How many of 105 putative membrane transporters, among 179 predicted transporters in this study (Fig 3A), play a role in antibiotic resistance, tolerance to xenobiotics or the secretion of virulence factors remains to be elucidated.

Putative enzymatic activities of hypothetical proteins link them to important cellular processes

Putative enzymes are the largest group (982) of PUFs. Their functional predictions might help to identify potential novel drug targets involved in fundamental metabolic pathways of *P. aeruginosa*. The classification of enzymes using text mining revealed hydrolases as the most abundant class, followed by transferases and oxidoreductases (Fig 4A). Thirty-five PUFs with putative enzyme functions are particularly interesting drug targets, as they belong to a group of essential proteins or virulence factors (Fig 4B). Ten of them have two or even three predicted enzyme activities, which indicates profound substrate promiscuity among potential drug targets. However, the multiple predicted functions for these proteins may be caused by using up to ten templates; therefore, the predicted functions should be experimentally validated. An interesting candidate with an essential function is PA4679, a putative uracil-DNA glycosylase (UDG) whose function is predicted based on modelling with *Thermus thermophilus* UGD (Fig 4C). The functional prediction of the UDG activity of PA4579 is strengthened by the finding that the structure of the homology model is similar to that of several other enzymes with UDG activity (root mean square deviation of C_α atoms, rmsd_{C_α} < 1.7 Å), and the active site residues of *Tt*UDG that interact with the uracil molecule are conserved among them (Figs 4C and S3 and S8 Table). Its essentiality is consistent with the function of these enzymes in DNA base excision repair, a process essential for maintaining the stability of chromosomal DNA [49].

We predicted the biochemical functions of four PUFs that were previously experimentally determined to have roles in virulence (S3 Table). The putative hydrolase PA1009 was reliably modelled using a template of phosphoserine phosphatase SerB from *Mycobacterium avium*, which is involved in a pyrimidine salvage pathway that is important for genome integrity [50]

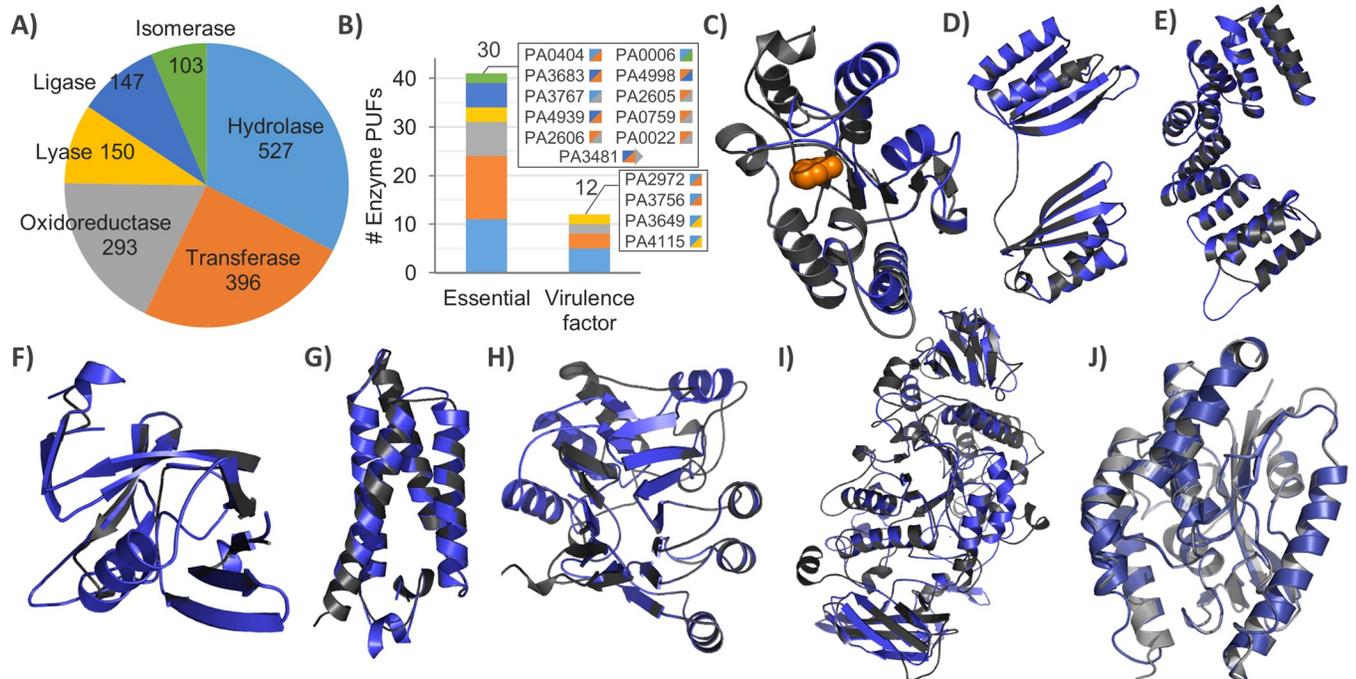


Fig 4. Putative enzymatic activities of *P. aeruginosa* proteins of unknown function. A) Distribution of putative enzymes among six EC groups. B) Distribution of essential genes and virulence factors among GUFs with putative enzyme functions. Gene identifiers of PUFs with predicted functions (Fig 4A and 4B) are provided in S1 Table. C) Superimposition of a homology model of putative uracil-DNA glycosylase PA4679 with *T. thermophilus* UGD (PDB ID 1UI0, chain A, 98.5% conf., 19% seq. id., $rmsd_{ca} = 0.62$ Å, 61% coverage, 0.9% disall. res.). The uracil molecule bound to the active site groove of *Tt*UGD is shown as an orange space-filling model. D) Superimposition of a homology model of the putative phosphoserine phosphatase PA1009 with *M. avium* SerB (PDB ID 3P96, chain A, 99% conf., 12% seq. id., $rmsd_{ca} = 0.50$ Å, 94% coverage, 1.2% disall. res.). E) Superimposition of a homology model of the putative Type IV pili biogenesis protein PA5441 with *N. meningitidis* PilW (PDB ID 2VQ2, chain A, 93% conf., 17% seq. id., $rmsd_{ca} = 0.29$ Å, 38% coverage, 0.5% disall. res.). Residues 1–130 of the template are hidden for clarity. F) Superimposition of a homology model of the putative L,D-transpeptidase PA3756 with *M. tuberculosis* LdtMt2 (PDB ID 3VYN, chain B, 100% conf., 18% seq. id., $rmsd_{ca} = 0.65$ Å, 83% coverage, 0.8% disall. res.). G) Superimposition of a homology model of putative flagellar chaperone PA1095 with *A. aeolicus* FliS (PDB ID 1ORJ, chain A, 100% conf., 28% seq. id., $rmsd_{ca} = 0.46$ Å, 94% coverage, 0% disall. res.). H) Superimposition of a homology model of the putative phosphorylcholine esterase PA2984 with *S. pneumoniae* Pce (PDB ID 1WRA, chain A, 100% conf., 16% seq. id., $rmsd_{ca} = 0.40$ Å, 32% coverage, 3% disall. res.). I) Superimposition of a homology model of the putative glycoside hydrolase PA2151 with *S. pneumoniae* SpuA (PDB ID 2YA0, chain A, 100% conf., 17% seq. id., $rmsd_{ca} = 0.54$ Å, 94% coverage, 1.9% disall. res.). J) Superimposition of a homology model of the putative patatin-like phospholipase PA1640 with *P. aeruginosa* outer membrane phospholipase PlpD (PDB ID 5FYA, chain A, 100% conf., 33% seq. id., $rmsd_{ca} = 0.36$ Å, 72% coverage, 1.4% disall. res.). The homology model is indicated in blue and the corresponding template in grey. The protein structures were visualized using PyMOL (<http://www.pymol.org>).

<https://doi.org/10.1371/journal.pone.0258385.g004>

(Fig 4D and S7 Table). The C-terminal part of the putative cell motility protein PA5441 was reliably modelled with the type IV pili biogenesis protein PilW from *N. meningitidis*, which plays an important role in the oligomerization and functionality of outer membrane-localized pores of pili [51] (Fig 4E). The putative transferase PA3756 was modelled with 100% confidence using *M. tuberculosis* L,D-transpeptidase (LdtMT2), which is a common target of β -lactam antibiotics [52] (Fig 4F and S7 Table). The putative function of PA1095 as a FliS-like chaperone, assigned based on homology with *Aquifex aeolicus* FliS (Fig 4G) [53], is involved in the assembly of the bacterial flagellum. The prediction of virulence functions for PA5441, PA1009, PA3756 and PA1095 is in agreement with the experimental determination of their virulence roles in the genetic screening of mutants in *Rattus norvegicus* virulence assays [54] or motility assays [36].

Furthermore, we predicted the functions of several PUFs based on homology to known virulence factors from other pathogens. Although these proteins were not previously known to be related to virulence in *P. aeruginosa*, their putative biochemical functions suggest that they might be novel virulence factors of *P. aeruginosa*. Among them are three enzymes, PA2984,

PA2151 and PA1640, with putative hydrolytic activities. The C-terminal part (residues 494–736) of PA2984 was reliably modelled (Fig 4H and S7 Table) with phosphorylcholine esterase Pce from *Streptococcus pneumoniae*, whose virulence function is related to adhesion to host cells [55]. The N-terminal domain of PA2984 could not be modelled with any known structure; however, analysis with Phyre2 predicted 12 putative transmembrane helices in this part (S2 Fig), suggesting that this putative hydrolase is a two-domain protein in which the catalytic domain is anchored to the membrane by an N-terminal domain. Interestingly, the *S. pneumoniae* template protein also has a second domain that exposes the esterase domain to the cell surface through the binding of surface choline. The structure of the putative hydrolase PA2151 was reliably modelled (Fig 4I and S7 Table) with the glycogen-degrading virulence factor SpuA from *Streptococcus pneumoniae*, which is involved in host-pathogen interactions [56]. The putative extracellular hydrolyse PA1640 was modelled (Fig 4J) with 100% confidence using the structure of type V secreted phospholipase PlpD, which is a putative virulence factor of *P. aeruginosa* [57, 58]. PlpD and the *P. aeruginosa* type III secretory toxin ExoU [59] belong to the patatin-like family of phospholipases. Patatin domain-containing proteins are very potent toxins due to their strong lytic activity, which they exert on host cells through, among other mechanisms, the hydrolysis of host membrane phospholipids [59].

The functions of the abovementioned enzymes could not be predicted from the sequence alignment, as they share less than 20% sequence identity with the template proteins (except PA1095, which share 28% seg. id.). However, the high confidence of homology modelling for many hypothetical proteins indicates that this comparative approach may be used to predict the biological processes in which they might be involved.

Conservation of promising drug targets in other pathogens

Protein conservation in multiple pathogens is one of the criteria used to prioritize targets for drug development. Broad-range antibiotics have an advantage over pathogen-specific drugs because of higher profits, easier development due to the uncomplicated recruitment of patients for phase III trials, and no requirement of specific diagnostic tools for pathogen detection [2]. Therefore, we searched orthologues of promising drug targets in clinical isolates of *P. aeruginosa*, in other pathogenic *Pseudomonas* species (*P. mendocina*, *P. otitidis* and *P. alcaligenes*), and in pathogens from the WHO priority list (Table 1). The results revealed that each of eleven promising targets with newly predicted ligand binding (Fig 3) or enzyme (Fig 4) functions are conserved (>70% sequence identity, >95% alignment coverage) in most clinical isolates of *P. aeruginosa* (Table 1). Furthermore, ten of eleven promising targets are conserved (>70% sequence identity, >95% alignment coverage) in at least two important pathogens from the *Pseudomonas* genus. Finally, most of these eleven proteins might be broad-range targets, as all of them have homologues (nine of eleven proteins were conserved in four or more pathogens) in other WHO priority pathogens (Table 1). Interestingly, seven of these potential drug targets have homologues in both Gram-negative and Gram-positive bacteria (Table 1), indicating that a drug binding these targets might show very broad-spectrum activity. However, this should be experimentally tested, as pathways essential in *P. aeruginosa* are not necessarily essential in other pathogens. These proteins whose function was previously unknown do not have human homologues (Table 1), which makes them even more attractive for the development of drugs to combat infections with *Pseudomonas* and other important bacterial pathogens.

Discussion

Comparative modelling of the target protein with sequence homologues with known structures is a useful tool for predicting the structures of hypothetical proteins [60–62]. Rarely,

Table 1. Promising drug targets of *P. aeruginosa* among proteins of unknown function.

Gene	Putative biochemical function	Putative virulence function	Phyre 2 confidence [%]	Conservation			
				WHO priority pathogens ^c	pathogenic <i>Pseudomonas</i> ^b	clinical <i>P. aeruginosa</i> ^a	humans
PA5033	lectin	cell adhesion	88	<i>Ab, Kp</i>	<i>Po, Pm</i>	2519	no
PA1981	DNA gyrase-binding protein	antibiotic resistance	99	<i>Ab, Sp, Ec, Sm, Sa, Kp, St</i>	<i>Pm</i>	261	no
PA3304	efflux pump adapter protein	antibiotic resistance	99	<i>Ab, Kp, Ec, Sf, Sm</i>	<i>Po, Pm, Pa</i>	2523	no
PA4679	uracil-DNA glycosylase	chromosome stability	99	<i>Ab, Kp, Ec, Sf, Sm</i>	<i>Po, Pm, Pa</i>	1963	no
PA5441	Type IV pili protein	cell motility	93	<i>Ab, Kp, Ec, Sm, Sp, Sa</i>	<i>Po</i>	2532	no
PA1009	phosphoserine phosphatase	pyrimidine salvage pathway, genome integrity	99	<i>Ab, Kp, Sp, Ec, Ef</i>	<i>Pm, Pa</i>	2535	no
PA3756	L,D-transpeptidases	peptidoglycan biosynthesis	100	<i>Ab, Kp, Ec, Sp</i>	<i>Po, Pm</i>	2516	no
PA1095	flagellar chaperon	cell motility	94	<i>Ab, Kp, Ec</i>	<i>Po, Pm</i>	1107	no
PA2984	phosphorylcholine esterase	cell adhesion	100	<i>Ab, Kp, Sp, Ec</i>	<i>Po, Pm</i>	1659	no
PA2151	glycogen-degrading protein	host-pathogen interactions	100	<i>Ab, Kp, Ef, Sp, Sm, Sa</i>	<i>Po, Pm</i>	2501	no
PA1640	patatin-like phospholipase A	secreted toxin	100	<i>Ab, Sp, Ec, Kp, Sf</i>	<i>Pm, Pa</i>	653	no

^aNumber of orthologues (>70% sequence identity, >95% alignment coverage) identified in strains isolated from humans with the known disease as indicated in the PGD.

^bOrthologs (>70% sequence identity, >95% alignment coverage) from *P. mendocina* (*Pm*), *P. otitidis* (*Po*) and *P. alcaligenes* (*Pa*), considered pathogenic strains from the *Pseudomonas* genus.

^cHomologs (>35% sequence identity, >75% alignment coverage) from 13 WHO priority pathogens listed in Fig 1. Gram-positive strains are indicated in bold.

<https://doi.org/10.1371/journal.pone.0258385.t001>

homology modelling is used for proteome-wide structural and functional annotation of hypothetical proteins [63, 64]. Here, we have demonstrated the applicability of multitemplate homology modelling for the global assignment of biochemical functions to hundreds of hypothetical proteins of the human pathogen *P. aeruginosa* to aid the identification of potential antibiotic targets. For that purpose, we used the Phyre2 server, which is among the most commonly used tools for the prediction of protein structure [22]. Phyre2 applies an advanced method for the detection of remote sequence homologues with high specificity and sensitivity to build an accurate model even for proteins with less than 20% sequence identity [22]. In contrast to functional prediction based on pairwise sequence alignment, which appears to be reliable only when sequence similarity is very high because homologous genes can diverge to have different functions [65, 66], structure-based functional predictions take advantage of the fact that the protein structure remains conserved long after sequences are substantially changed [67]. Conclusively, Phyre2 functional annotation shows an advantage over sequence-alone homology searches.

Our set of >21000 homology models for >2100 target *P. aeruginosa* PUFs generated using homologues with known structures and in most cases experimentally determined biochemical functions was refined to identify true homologues. The biochemical functions of true template homologues were automatically assigned as putative functions to hundreds of target proteins. The results of this functional prediction are available as the PaPUF database (www.iet.uni-duesseldorf.de/arbeitsgruppen/bacterial-enzymology). It is likely that for some PaPUFs, homology-modelling-based functional assignment is misleading because it is known that during evolution from one protein, several homologues may diverge to perform different functions [68]. Therefore, to reinforce the prediction of biochemical functions, we recommend verifying the biochemical functions by superimposing the homology model with the template homologues and inspecting active sites or ligand binding sites. However, this process is tedious and time consuming and therefore is not applicable for high-throughput analysis. Our approach revealed 360 PUFs with two or three putative biochemical functions, which were associated with different template structures. This is not surprising if we consider that many enzymes catalyse more than one reaction [69–71], that many enzymes and transporter domains have annotated ligand-binding domains [72–75], and that several distinct domains are often arranged into one functional protein [76, 77].

Although biochemical function is an important prerequisite for drug R&D, it is not sufficient alone, without biological function, for identifying potential drug targets [78, 79]. Therefore, the putative biological functions of the template homologues were used to predict the roles of PUFs in virulence and antibiotic resistance. During the selection of putative drug targets, we considered information obtained from large-scale screens [30, 32–36] about the relationship of target *P. aeruginosa* PUFs with virulence or essential processes, the conservation of PUFs in other high-priority relevant pathogens and the absence of PUF homologues in humans. This led to the classification of 107 PUFs (63 essential, 21 virulence factor, 28 pathogen-specific) into the highest priority group for the development of novel drugs against *P. aeruginosa*. Notably, 103 of these PUFs do not have a human homologue, which is a good starting point for the development of selective drugs. Forty-one of these top priority drug targets were assigned putative biochemical functions with high homology modelling confidence, query protein coverage and good stereochemical parameters. These functions allowed us, based on the biological processes ascribed to homology templates, to link them with maintaining genome integrity (putative phosphoserine phosphatase PA1009 and putative uracil-DNA glycosylase PA4679), peptidoglycan biosynthesis (putative L,D-transpeptidase PA3756), host-pathogen interactions (putative extracellular lectin PA5033, putative phosphorylcholine esterase PA2984, putative flagellar chaperone PA1095, putative Type IV pili biogenesis protein

PA5441 and putative glycogen-degrading protein PA2151), extracellular toxicity (patatin-like phospholipase PA1640) and antibiotic resistance (putative DNA gyrase binding protein PA1981 and putative efflux pump adapter protein PA3304) (Table 1). Interestingly, several of these drug targets were previously identified as virulence-related genes using a *Rattus norvegicus* virulence assay [54] or motility assay [36], although their functions were unknown.

All these processes are common drug targets; therefore, we believe that these proteins are promising targets for further experimental characterization and the design of inhibitors, which might represent novel antibiotics to treat not only *P. aeruginosa* but also other high-priority pathogens in which these proteins are largely conserved.

Supporting information

S1 Fig. Confidence score (A) and sequence identity (B) distributions for homology modelling of all 2102 *P. aeruginosa* PUFs obtained with the use of Phyre2 server. The only highest confidence score of each PUF was considered in the analysis.

(TIF)

S2 Fig. Topology of putative transmembrane helices predicted in PA2984 by Phyre2 server.

(TIF)

S3 Fig. Conservation of functional residues among PaPUFs and their structural homologs obtained by DALI database [1]. Literature reported active site and ligand binding residues are highlighted gray. Superimposition parameters are provided in S8 Table.

(TIF)

S1 Table. Classification of *P. aeruginosa* GUFs.

(XLSX)

S2 Table. Priority pathogens analyzed in this study.

(DOCX)

S3 Table. *P. aeruginosa* GUFs with experimentally demonstrated virulence function.

(DOCX)

S4 Table. Homology modelling of *P. aeruginosa* PUFs using Phyre2 server.

(XLSX)

S5 Table. PUFs with more than 96% of sequence identity to templates with known function.

(DOCX)

S6 Table. Keywords used for textual mining of enzyme, ligand binding and transporter functions of templates used for modelling of *P. aeruginosa* PUFs.

(DOCX)

S7 Table. Stereochemical validation of PaPUF homology models with PROCHECK [11].

(DOCX)

S8 Table. Structural alignments of PUF homology model with its corresponding template and additional proteins belonging to the same functional class, according to DALI database.

(DOCX)

S1 File. References specific to supporting information.

(DOCX)

Acknowledgments

Both authors thank Prof. Karl-Erich Jaeger from the Institute of Molecular Enzyme Technology (Heinrich Heine University Düsseldorf) for valuable discussions.

Author Contributions

Conceptualization: Filip Kovacic.

Data curation: Nikolina Babic, Filip Kovacic.

Formal analysis: Nikolina Babic, Filip Kovacic.

Funding acquisition: Filip Kovacic.

Investigation: Nikolina Babic.

Methodology: Filip Kovacic.

Project administration: Filip Kovacic.

Supervision: Filip Kovacic.

Validation: Nikolina Babic, Filip Kovacic.

Visualization: Nikolina Babic, Filip Kovacic.

Writing – original draft: Nikolina Babic, Filip Kovacic.

Writing – review & editing: Nikolina Babic, Filip Kovacic.

References

1. Brown ED, Wright GD. Antibacterial drug discovery in the resistance era. *Nature*. 2016; 529(7586):336–43. Epub 2016/01/23. <https://doi.org/10.1038/nature17042> PMID: 26791724.
2. Theuretzbacher U, Outterson K, Engel A, Karlén A. The global preclinical antibacterial pipeline. *Nature Reviews Microbiology*. 2020; 18(5):275–85. <https://doi.org/10.1038/s41579-019-0288-0> PMID: 31745331
3. Tacconelli E, Carrara E, Savoldi A, Harbarth S, Mendelson M, Monnet DL, et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *The Lancet Infectious diseases*. 2018; 18(3):318–27. Epub 2017/12/26. [https://doi.org/10.1016/S1473-3099\(17\)30753-3](https://doi.org/10.1016/S1473-3099(17)30753-3) PMID: 29276051.
4. Strathdee SA, Davies SC, Marcelin JR. Confronting antimicrobial resistance beyond the COVID-19 pandemic and the 2020 US election. *The Lancet*. 2020; 396(10257):1050–3. [https://doi.org/10.1016/S0140-6736\(20\)32063-8](https://doi.org/10.1016/S0140-6736(20)32063-8) PMID: 33007218
5. Aminov RI. A brief history of the antibiotic era: lessons learned and challenges for the future. *Frontiers in microbiology*. 2010; 1:134-. <https://doi.org/10.3389/fmicb.2010.00134> PMID: 21687759.
6. Singh KS, Sharma R, Reddy PAN, Vonteddu P, Good M, Sundarajan A, et al. IspH inhibitors kill Gram-negative bacteria and mobilize immune clearance. *Nature*. 2021; 589(7843):597–602. <https://doi.org/10.1038/s41586-020-03074-x> PMID: 33361818
7. Belete TM. Novel targets to develop new antibacterial agents and novel alternatives to antibacterial agents. *Human Microbiome Journal*. 2019; 11:100052. <https://doi.org/10.1016/j.humic.2019.01.001>.
8. Olatunji S, Yu X, Bailey J, Huang C-Y, Zapotoczna M, Bowen K, et al. Structures of lipoprotein signal peptidase II from *Staphylococcus aureus* complexed with antibiotics globomycin and myxovirescin. *Nature Communications*. 2020; 11(1):140. <https://doi.org/10.1038/s41467-019-13724-y> PMID: 31919415
9. Witzany C, Bonhoeffer S, Rolff J. Is antimicrobial resistance evolution accelerating? *PLoS Pathog*. 2020; 16(10):e1008905. <https://doi.org/10.1371/journal.ppat.1008905> PMID: 33091083
10. Dickey SW, Cheung GYC, Otto M. Different drugs for bad bugs: antivirulence strategies in the age of antibiotic resistance. *Nat Rev Drug Discov*. 2017; 16(7):457–71. Epub 2017/03/25. <https://doi.org/10.1038/nrd.2017.23> PMID: 28337021.

11. Filiatrault MJ, Picardo KF, Ngai H, Passador L, Iglewski BH. Identification of *Pseudomonas aeruginosa* genes involved in virulence and anaerobic growth. *Infection and immunity*. 2006; 74(7):4237–45. <https://doi.org/10.1128/IAI.02014-05> PMID: 16790798.
12. Taylor PK, Yeung AT, Hancock RE. Antibiotic resistance in *Pseudomonas aeruginosa* biofilms: towards the development of novel anti-biofilm therapies. *J Biotechnol*. 2014; 191:121–30. Epub 2014/09/23. <https://doi.org/10.1016/j.jbiotec.2014.09.003> PMID: 25240440.
13. Jimenez PN, Koch G, Thompson JA, Xavier KB, Cool RH, Quax WJ. The multiple signaling systems regulating virulence in *Pseudomonas aeruginosa*. *Microbiology and molecular biology reviews: MMBR*. 2012; 76(1):46–65. Epub 2012/03/07. <https://doi.org/10.1128/MMBR.05007-11> PMID: 22390972; PubMed Central PMCID: PMC3294424.
14. Prince A. Adhesins and receptors of *Pseudomonas aeruginosa* associated with infection of the respiratory tract. *Microb Pathog*. 1992; 13(4):251–60. Epub 1992/10/01. [https://doi.org/10.1016/0882-4010\(92\)90035-m](https://doi.org/10.1016/0882-4010(92)90035-m) PMID: 1363702.
15. Bucher KS, Babic N, Freichel T, Kovacic F, Hartmann L. Monodisperse Sequence-Controlled alpha-L-Fucosylated Glycooligomers and Their Multivalent Inhibitory Effects on LecB. *Macromol Biosci*. 2018; 18(12):e1800337. Epub 2018/11/09. <https://doi.org/10.1002/mabi.201800337> PMID: 30408344.
16. Blevès S, Viarre V, Salacha R, Michel GP, Filloux A, Voulhoux R. Protein secretion systems in *Pseudomonas aeruginosa*: A wealth of pathogenic weapons. *Int J Med Microbiol*. 2010; 300(8):534–43. Epub 2010/10/16. <https://doi.org/10.1016/j.ijmm.2010.08.005> PMID: 20947426.
17. Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA, Brinkman FS. Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. *Nucleic Acids Res*. 2016; 44(D1):D646–53. Epub 2015/11/19. <https://doi.org/10.1093/nar/gkv1227> PMID: 26578582; PubMed Central PMCID: PMC4702867.
18. UniProt C. UniProt: a hub for protein information. *Nucleic Acids Res*. 2015; 43(Database issue):D204–12. Epub 2014/10/29. <https://doi.org/10.1093/nar/gku989> PMID: 25348405; PubMed Central PMCID: PMC4384041.
19. Liu B, Zheng D, Jin Q, Chen L, Yang J. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic acids research*. 2019; 47(D1):D687–D92. <https://doi.org/10.1093/nar/gky1080> PMID: 30395255
20. Sayers S, Li L, Ong E, Deng S, Fu G, Lin Y, et al. Victors: a web-based knowledge base of virulence factors in human and animal pathogens. *Nucleic acids research*. 2019; 47(D1):D693–D700. <https://doi.org/10.1093/nar/gky999> PMID: 30365026
21. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nature methods*. 2015; 12(1):59–60. Epub 2014/11/18. <https://doi.org/10.1038/nmeth.3176> PMID: 25402007.
22. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015; 10(6):845–58. Epub 2015/05/08. <https://doi.org/10.1038/nprot.2015.053> PMID: 25950237; PubMed Central PMCID: PMC5298202.
23. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc*. 2009; 4(3):363–71. Epub 2009/02/28. <https://doi.org/10.1038/nprot.2009.2> PMID: 19247286.
24. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*. 1997; 18(15):2714–23. Epub 1998/03/21. <https://doi.org/10.1002/elps.1150181505> PMID: 9504803.
25. Laskowski RA, Macarthur MW, Moss DS, Thornton JM. (PROCHECK): a program to check the stereochemical quality of protein structures. *J Appl Cryst*. 1993; 26:283–91. doi: citeulike-article-id:1720734.
26. Holm L, Rosenstrom P. Dali server: conservation mapping in 3D. *Nucleic Acids Res*. 2010; 38(Web Server issue):W545–9. Epub 2010/05/12. <https://doi.org/10.1093/nar/gkq366> PMID: 20457744; PubMed Central PMCID: PMC2896194.
27. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, et al. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature*. 2000; 406(6799):959–64. Epub 2000/09/13. <https://doi.org/10.1038/35023079> PMID: 10984043.
28. Gerlt JA, Allen KN, Almo SC, Armstrong RN, Babbitt PC, Cronan JE, et al. The Enzyme Function Initiative. *Biochemistry*. 2011; 50(46):9950–62. Epub 2011/10/18. <https://doi.org/10.1021/bi201312u> PMID: 21999478; PubMed Central PMCID: PMC3238057.
29. Fields FR, Lee SW, McConnell MJ. Using bacterial genomes and essential genes for the development of new antibiotics. *Biochem Pharmacol*. 2017; 134:74–86. Epub 2016/12/08. <https://doi.org/10.1016/j.bcp.2016.12.002> PMID: 27940263.
30. Lee SA, Gallagher LA, Thongdee M, Staudinger BJ, Lippman S, Singh PK, et al. General and condition-specific essential functions of *Pseudomonas aeruginosa*. *P Natl Acad Sci USA*. 2015; 112(16):5189–94. Epub 2015/04/06. <https://doi.org/10.1073/pnas.1422186112> PMID: 25848053.

31. Ho Sui SJ, Fedynak A, Hsiao WW, Langille MG, Brinkman FS. The association of virulence factors with genomic islands. *PLoS One*. 2009; 4(12):e8094. Epub 2009/12/04. <https://doi.org/10.1371/journal.pone.0008094> PMID: 19956607; PubMed Central PMCID: PMC2779486.
32. Attila C, Ueda A, Cirillo SL, Cirillo JD, Chen W, Wood TK. *Pseudomonas aeruginosa* PAO1 virulence factors and poplar tree response in the rhizosphere. *Microb Biotechnol*. 2008; 1(1):17–29. Epub 2008/01/01. <https://doi.org/10.1111/j.1751-7915.2007.00002.x> PMID: 21261818; PubMed Central PMCID: PMC3864428.
33. Feinbaum RL, Urbach JM, Liberati NT, Djonovic S, Adonizio A, Carvunis A-R, et al. Genome-wide identification of *Pseudomonas aeruginosa* virulence-related genes using a *Caenorhabditis elegans* infection model. *PLoS Pathog*. 2012; 8(7):e1002813-e. Epub 2012/07/26. <https://doi.org/10.1371/journal.ppat.1002813> PMID: 22911607.
34. Romsang A, Atichartpongkul S, Trinachartvanit W, Vattanaviboon P, Mongkolsuk S. Gene expression and physiological role of *Pseudomonas aeruginosa* methionine sulfoxide reductases during oxidative stress. *Journal of bacteriology*. 2013; 195(15):3299–308. Epub 2013/05/17. <https://doi.org/10.1128/JB.00167-13> PMID: 23687271.
35. Potvin E, Lehoux DE, Kukavica-Ibrulj I, Richard KL, Sanschagrin F, Lau GW, et al. In vivo functional genomics of *Pseudomonas aeruginosa* for high-throughput screening of new virulence factors and anti-bacterial targets. *Environ Microbiol*. 2003; 5(12):1294–308. Epub 2003/12/03. <https://doi.org/10.1046/j.1462-2920.2003.00542.x> PMID: 14641575.
36. Dasgupta N, Wolfgang MC, Goodman AL, Arora SK, Jyot J, Lory S, et al. A four-tiered transcriptional regulatory circuit controls flagellar biogenesis in *Pseudomonas aeruginosa*. *Mol Microbiol*. 2003; 50(3):809–24. Epub 2003/11/18. <https://doi.org/10.1046/j.1365-2958.2003.03740.x> PMID: 14617143.
37. Illergård K, Ardell DH, Elofsson A. Structure is three to ten times more conserved than sequence—a study of structural response in protein cores. *Proteins*. 2009; 77(3):499–508. Epub 2009/06/10. <https://doi.org/10.1002/prot.22458> PMID: 19507241.
38. Sawa T. The molecular mechanism of acute lung injury caused by *Pseudomonas aeruginosa*: from bacterial pathogenesis to host response. *J Intensive Care*. 2014; 2(1):10. Epub 2014/12/19. <https://doi.org/10.1186/2052-0492-2-10> PMID: 25520826; PubMed Central PMCID: PMC4267601.
39. Richard S, Filip KD, Céline B-A, Susanne W, Jan T, Alain F, et al. The *Pseudomonas aeruginosa* patatin-like protein PlpD is the archetype of a novel Type V secretion system. *Environmental Microbiology*. 2010; 12(6):1498–512. <https://doi.org/10.1111/j.1462-2920.2010.02174.x> PMID: 20192961
40. Xu D, Nussinov R. Favorable domain size in proteins. *Folding and Design*. 1998; 3(1):11–7. [https://doi.org/10.1016/S1359-0278\(98\)00004-2](https://doi.org/10.1016/S1359-0278(98)00004-2) PMID: 9502316
41. Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM. The many faces of the helix-turn-helix domain: Transcription regulation and beyond*. *FEMS Microbiology Reviews*. 2005; 29(2):231–62. <https://doi.org/10.1016/j.fmrre.2004.12.008> PMID: 15808743
42. Murzin AG. OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *The EMBO journal*. 1993; 12(3):861–7. PMID: 8458342.
43. Stahelin RV. Lipid binding domains: more than simple lipid effectors. *J Lipid Res*. 2009; 50 Suppl (Suppl):S299–304. Epub 2008/11/15. <https://doi.org/10.1194/jlr.R800078-JLR200> PMID: 19008549; PubMed Central PMCID: PMC2674730.
44. Ren X-M, Li D-F, Jiang S, Lan X-Q, Hu Y, Sun H, et al. Structural Basis of Specific Recognition of Non-Reducing Terminal N-Acetylglucosamine by an *Agroclybe aegerita* Lectin. *PLoS One*. 2015; 10(6):e0129608. <https://doi.org/10.1371/journal.pone.0129608> PMID: 26114302
45. Kleywegt GJ, Jones TA. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr D Biol Crystallogr*. 1994; 50(Pt 2):178–85. Epub 1994/03/01. <https://doi.org/10.1107/S0907444993011333> PMID: 15299456.
46. Meiers J, Zahorska E, Röhrig T, Hauck D, Wagner S, Titz A. Directing Drugs to Bugs: Antibiotic-Carbohydrate Conjugates Targeting Biofilm-Associated Lectins of *Pseudomonas aeruginosa*. *J Med Chem*. 2020; 63(20):11707–24. <https://doi.org/10.1021/acs.jmedchem.0c00856> PMID: 32924479
47. Hegde SS, Vetting MW, Roderick SL, Mitchenall LA, Maxwell A, Takiff HE, et al. A Fluoroquinolone Resistance Protein from *Mycobacterium tuberculosis* That Mimics DNA. *Science*. 2005; 308(5727):1480–3. <https://doi.org/10.1126/science.1110699> PMID: 15933203
48. Greene NP, Kaplan E, Crow A, Koronakis V. Antibiotic Resistance Mediated by the MacB ABC Transporter Family: A Structural and Functional Perspective. 2018; 9(950). <https://doi.org/10.3389/fmicb.2018.00950> PMID: 29892271
49. Hoseki J, Okamoto A, Masui R, Shibata T, Inoue Y, Yokoyama S, et al. Crystal Structure of a Family 4 Uracil-DNA Glycosylase from *Thermus thermophilus* HB8. *Journal of Molecular Biology*. 2003; 333(3):515–26. <https://doi.org/10.1016/j.jmb.2003.08.030> PMID: 14556741

50. Abendroth J, Gardberg AS, Robinson JI, Christensen JS, Staker BL, Myler PJ, et al. SAD phasing using iodide ions in a high-throughput structural genomics environment. *Journal of Structural and Functional Genomics*. 2011; 12(2):83–95. <https://doi.org/10.1007/s10969-011-9101-7> PMID: 21359836
51. Trindade MB, Job V, Contreras-Martel C, Pelicic V, Dessen A. Structure of a Widely Conserved Type IV Pilus Biogenesis Factor that Affects the Stability of Secretin Multimers. *Journal of Molecular Biology*. 2008; 378(5):1031–9. <https://doi.org/10.1016/j.jmb.2008.03.028> PMID: 18433773
52. Li WJ, Li DF, Hu YL, Zhang XE, Bi LJ, Wang DC. Crystal structure of L,D-transpeptidase LdtMt2 in complex with meropenem reveals the mechanism of carbapenem against *Mycobacterium tuberculosis*. *Cell research*. 2013; 23(5):728–31. Epub 2013/04/17. <https://doi.org/10.1038/cr.2013.53> PMID: 23588382; PubMed Central PMCID: PMC3641605.
53. Evdokimov AG, Phan J, Tropea JE, Routzahn KM, Peters HK, Pokross M, et al. Similar modes of polypeptide recognition by export chaperones in flagellar biosynthesis and type III secretion. *Nat Struct Biol*. 2003; 10(10):789–93. Epub 2003/09/06. <https://doi.org/10.1038/nsb982> PMID: 12958592.
54. Potvin E, Lehoux DE, Kukavica-Ibrulj I, Richard KL, Sanschagrin F, Lau GW, et al. In vivo functional genomics of *Pseudomonas aeruginosa* for high-throughput screening of new virulence factors and antibacterial targets. *Environmental microbiology*. 2003; 5(12):1294–308. <https://doi.org/10.1046/j.1462-2920.2003.00542.x> PMID: 14641575
55. Garau G, Lemaire D, Vernet T, Dideberg O, Di Guilmi AM. Crystal Structure of Phosphorylcholine Esterase Domain of the Virulence Factor Choline-binding Protein E from *Streptococcus pneumoniae*: new structural features among the metallo- β -lactamase superfamily. 2005; 280(31):28591–600. <https://doi.org/10.1074/jbc.M502744200> PMID: 15908436
56. Lammerts van Bueren A, Ficko-Blean E, Pluvinage B, Hehemann JH, Higgins MA, Deng L, et al. The conformation and function of a multimodular glycogen-degrading pneumococcal virulence factor. *Structure*. 2011; 19(5):640–51. Epub 2011/05/14. <https://doi.org/10.1016/j.str.2011.03.001> PMID: 21565699.
57. da Mata Madeira PV, Zouhir S, Basso P, Neves D, Laubier A, Salacha R, et al. Structural basis of lipid targeting and destruction by the type V secretion system of *Pseudomonas aeruginosa*. *J Mol Biol*. 2016; 428(9 Pt A):1790–803. Epub 2016/03/26. <https://doi.org/10.1016/j.jmb.2016.03.012> PMID: 27012424.
58. Salacha R, Kovacic F, Brochier-Armanet C, Wilhelm S, Tommassen J, Filloux A, et al. The *Pseudomonas aeruginosa* patatin-like protein PlpD is the archetype of a novel Type V secretion system. *Environ Microbiol*. 2010; 12(6):1498–512. Epub 2010/03/03. <https://doi.org/10.1111/j.1462-2920.2010.02174.x> PMID: 20192961.
59. Sawa T, Hamaoka S, Kinoshita M, Kainuma A, Naito Y, Akiyama K, et al. *Pseudomonas aeruginosa* Type III Secretory Toxin ExoU and Its Predicted Homologs. *Toxins*. 2016; 8(11). Epub 2016/10/30. <https://doi.org/10.3390/toxins8110307> PMID: 27792159; PubMed Central PMCID: PMC5127104.
60. Goswami AM. Structural modeling and in silico analysis of non-synonymous single nucleotide polymorphisms of human 3β -hydroxysteroid dehydrogenase type 2. *Meta Gene*. 2015; 5:162–72. <https://doi.org/10.1016/j.mgene.2015.07.007> PMID: 26288759
61. Schindler BD, Patel D, Seo SM, Kaatz GW. Mutagenesis and modeling to predict structural and functional characteristics of the *Staphylococcus aureus* MepA multidrug efflux pump. *J Bacteriol*. 2013; 195(3):523–33. Epub 2012/11/24. <https://doi.org/10.1128/JB.01679-12> PMID: 23175649; PubMed Central PMCID: PMC3554018.
62. Nikolaev DM, Shtyrov AA, Panov MS, Jamal A, Chakchir OB, Kochemirovsky VA, et al. A Comparative Study of Modern Homology Modeling Algorithms for Rhodopsin Structure Prediction. *ACS Omega*. 2018; 3(7):7555–66. Epub 2018/07/09. <https://doi.org/10.1021/acsomega.8b00721> PMID: 30087916.
63. Mao C, Shukla M, Larrouy-Maumus G, Dix FL, Kelley LA, Sternberg MJ, et al. Functional assignment of *Mycobacterium tuberculosis* proteome revealed by genome-scale fold-recognition. *Tuberculosis*. 2013; 93(1):40–6. <https://doi.org/10.1016/j.tube.2012.11.008> PMID: 23287603
64. Houston S, Lithgow KV, Osbak KK, Kenyon CR, Cameron CE. Functional insights from proteome-wide structural modeling of *Treponema pallidum* subspecies *pallidum*, the causative agent of syphilis. *Bmc Structural Biology*. 2018;18. <https://doi.org/10.1186/s12900-018-0094-3> PMID: 30554569 PubMed PMID: WOS:000432719900001.
65. Pandey G, Kumar V, Steinbach M, editors. *Computational Approaches for Protein Function Prediction: A Survey* 2006.
66. Higdon R, Louie B, Kolker E. Modeling sequence and function similarity between proteins for protein functional annotation. *Proc Int Symp High Perform Distrib Comput*. 2010; 2010:499–502. <https://doi.org/10.1145/1851476.1851548> PMID: 25101328.
67. Whisstock JC, Lesk AM. Prediction of protein function from protein sequence and structure. *Q Rev Biophys*. 2003; 36(3):307–40. Epub 2004/03/20. <https://doi.org/10.1017/s0033583503003901> PMID: 15029827.

68. Acharya KR, Shapiro R, Allen SC, Riordan JF, Vallee BL. Crystal structure of human angiogenin reveals the structural basis for its functional divergence from ribonuclease. *Proc Natl Acad Sci U S A*. 1994; 91(8):2915–9. Epub 1994/04/12. <https://doi.org/10.1073/pnas.91.8.2915> PMID: 8159679; PubMed Central PMCID: PMC43485.
69. Lescic Asler I, Ivic N, Kovacic F, Schell S, Knorr J, Krauss U, et al. Probing enzyme promiscuity of SGNH hydrolases. *Chembiochem*. 2010; 11(15):2158–67. Epub 2010/10/12. <https://doi.org/10.1002/cbic.201000398> PMID: 20931591.
70. Martinez-Martinez M, Coscolin C, Santiago G, Chow J, Stogios PJ, Bargiela R, et al. Determinants and Prediction of Esterase Substrate Promiscuity Patterns. *ACS Chem Biol*. 2018; 13(1):225–34. Epub 2017/11/29. <https://doi.org/10.1021/acscchembio.7b00996> PMID: 29182315.
71. Rauwerdink A, Kazlauskas RJ. How the Same Core Catalytic Machinery Catalyzes 17 Different Reactions: the Serine-Histidine-Aspartate Catalytic Triad of alpha/beta-Hydrolase Fold Enzymes. *ACS Catal*. 2015; 5(10):6153–76. Epub 2015/10/02. <https://doi.org/10.1021/acscatal.5b01539> PMID: 28580193; PubMed Central PMCID: PMC5455348.
72. Nocek B, Stein AJ, Jedrzejczak R, Cuff ME, Li H, Volkart L, et al. Structural Studies of ROK Fructokinase YdhR from *Bacillus subtilis*: Insights into Substrate Binding and Fructose Specificity. *Journal of Molecular Biology*. 2011; 406(2):325–42. <https://doi.org/10.1016/j.jmb.2010.12.021> PMID: 21185308
73. Schlundt A, Buchner S, Janowski R, Heydenreich T, Heermann R, Lassak J, et al. Structure-function analysis of the DNA-binding domain of a transmembrane transcriptional activator. *Sci Rep*. 2017; 7(1):1051. Epub 2017/04/23. <https://doi.org/10.1038/s41598-017-01031-9> PMID: 28432336; PubMed Central PMCID: PMC5430869.
74. Gushchin I, Melnikov I, Polovinkin V, Ishchenko A, Yuzhakova A, Buslaev P, et al. Mechanism of transmembrane signaling by sensor histidine kinases. *Science*. 2017; 356(6342). Epub 2017/05/20. <https://doi.org/10.1126/science.aah6345> PMID: 28522691.
75. Amikam D, Galperin MY. PilZ domain is part of the bacterial c-di-GMP binding protein. *Bioinformatics*. 2006; 22(1):3–6. Epub 2005/10/27. <https://doi.org/10.1093/bioinformatics/bti739> PMID: 16249258.
76. Sawma P, Roth L, Blanchard C, Bagnard D, Cremel G, Bouveret E, et al. Evidence for new homotypic and heterotypic interactions between transmembrane helices of proteins involved in receptor tyrosine kinase and neuropilin signaling. *J Mol Biol*. 2014; 426(24):4099–111. Epub 2014/10/16. <https://doi.org/10.1016/j.jmb.2014.10.007> PMID: 25315821.
77. Wilhelm S, Tommassen J, Jaeger KE. A novel lipolytic enzyme located in the outer membrane of *Pseudomonas aeruginosa*. *J Bacteriol*. 1999; 181(22):6977–86. Epub 1999/11/13. <https://doi.org/10.1128/JB.181.22.6977-6986.1999> PMID: 10559163; PubMed Central PMCID: PMC94172.
78. Emmerich CH, Gamboa LM, Hofmann MCJ, Bonin-Andresen M, Arbach O, Schendel P, et al. Improving target assessment in biomedical research: the GOT-IT recommendations. *Nature Reviews Drug Discovery*. 2020. <https://doi.org/10.1038/s41573-020-0087-3> PMID: 33199880
79. Kraljevic S, Stambrook PJ, Pavelic K. Accelerating drug discovery. *EMBO Rep*. 2004; 5(9):837–42. <https://doi.org/10.1038/sj.embor.7400236> PMID: 15470377.