Trends in Biotechnology

CellPress REVIEWS

Scientific Life

The Biofilms Structural Database

Rita P. Magalhães [1,4], Tatiana F. Vieira [1,4], Henrique S. Fernandes [1], André Melo [2], Manuel Simões [3] and Sérgio F. Sousa [1,*]

The Biofilms Structural Database (BSD) is a collection of structural, mutagenesis, kinetics, and inhibition data to understand the processes involved in biofilm formation. Presently, it includes curated information on 425 structures of proteins and enzymes involved in biofilm formation and development for 42 different bacteria. It is available at www.biofilms.biosim.pt.

Biofilms in Health

Biofilms are heterogeneous and complex structures of microorganisms, typically adhered to a surface and presenting sophisticated singular and collective behaviors. Bacteria in biofilms are embedded in a self-produced matrix of extracellular polymeric substances (EPS) composed of (glyco)proteins, (glyco)lipids, mono- or poly-saccharides, extracellular DNA, minerals, and water [1]. This protected mode of growth allows bacteria to survive in hostile environments. In the health context, some diseases and adverse medical conditions are now recognized to be the result of a biofilm infection, and it has been estimated that more than 65% of all microbial infections in humans are caused by biofilms [2]. Accordingly, there has been a great amount of research to understand better biofilm development and to identify improved control strategies [3]. However, the structure, composition, and physiology of microbial biofilms have become inexorably linked with our failure to control them by antimicrobial treatments that are effective against planktonic bacteria. In fact, once set, biofilm-associated infections are very difficult to treat because the colonizing bacteria are 10–1000 times less susceptible to antimicrobials than their planktonic counterparts. This greatly limits the therapeutic options for the treatment of many infections [4].

Strategies for Inhibiting Biofilm Formation

Antibacterial strategies focusing on inhibiting cellular growth often induce antimicrobial resistance by imposing a strong selective pressure on bacteria. By contrast, alternatives that target the processes associated with biofilm formation and development, which include bacterial motility, cell adhesion, biofilm dispersion, the synthesis of EPS, and also cell-to-cell communication or quorum sensing (QS), can be more efficient [5,6], as these processes are less amenable to induce antimicrobial resistance. Throughout the years, many signals, regulatory pathways, and autoinducer molecules have been identified, but their intricate network remains unclear at the atomic and molecular level [7]. Understanding the mechanisms involved in these processes is essential for the development of new therapies against biofilms.

The Shift from the Cellular to the Molecular Level

Over the past few years, there has been an immense increase in the number of 3D structures (particularly X-ray) of proteins and enzymes associated with biofilm formation and development. The data associated with these structures, together with the vast body of work dispersed throughout the scientific literature – including kinetic and mutagenesis data, amino acid sequences, and inhibitory activity of known molecules – offers new opportunities to understand biofilms at a structural and atomic level. This recent development has allowed researchers to move from the cellular to the molecular level, inducing a whole new dimension in the development of antibiofilm therapies. In particular, the new structures available offer an appealing alternative for the rational design of new drugs and for the application of techniques, such as docking, virtual screening, quantitative structure–activity relationship (QSAR) models, and molecular dynamics.

The BSD was designed to organize all the available structural information on these promising targets, presenting up-to-date atomic information on the proteins and enzymes involved in biofilm formation and development. It is the first database dedicated to the biochemistry of biofilms from a structural and molecular perspective of the proteins associated. It is a platform that can be used not only by those who wish to study the proteins responsible for the mechanisms of biofilm formation, but also by structural and computational biologists who want to study the molecular details of substrate recognition. It helps to elucidate the mechanisms of action of several key proteins and enzymes and opens the door to rational development of new compounds with antibiofilm activity.

Data Selection

The main focus of the BSD is at the molecular level. However, biofilm research integrates several different scales. Therefore, the database was designed to create a comprehensive repository of structural data on biofilm research that could integrate with the information available in other databases, beyond the molecular level, to link different fields. An intensive literature search was performed to extract the relevant structural information about the proteins and enzymes involved in biofilm formation. The Institute for Scientific Information (ISI) Web of Science, Scopus, and PubMed databases were searched using combinations of biofilm-

related keywords, such as 'biofilm', 'quorum sensing', 'motility', 'dispersion', 'matrix protein', and other relevant terms, with the keywords 'structure', 'X-ray', 'NMR', and 'cryo-EM' (cryogenic electron microscopy). In addition, queries on the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) [8] were also performed. Additional queries on specific proteins known to be involved in biofilm formation and development were also conducted. The overall information was manually checked and cross-validated with the literature available for each protein, and the structures were visualized and evaluated.

Other databases, such as ChEMBL [9], BindingDB [10], ExPASy [11], KEGG [12], and UniProt [13], were cross-referenced to provide additional information for each protein. The BSD contains direct links for other drug-like molecules that have already been tested against that biofilm protein, information on kinetic and binding affinity data on mutant variants, and genomic and proteomic data regarding each specific protein. As investigations into biofilm progress and new structures are made available on the PDB database, the BSD will continue to be expanded, updated, and maintained to ensure that it remains a useful and scientifically accurate tool to study biofilms from a molecular perspective. The database will be maintained and regularly updated by the BioSIM research group at UCIBIO/REQUIMTE, Faculty of Medicine, University of Porto, and any researcher can contribute with new entries by filling the available form at www.biofilms.biosim.pt.

Organization of the Database

The BSD is organized in a simple and intuitive manner. Each entry in the database corresponds to one atomic structure with a unique PDB code. A total of 425 PDB entries are currently included, corresponding to a total of 133 unique proteins (Figure 1A). Each entry is identified through the PDB code associated with the corresponding structure in the PDB and by the name of the protein. Entries are classified by protein category and mechanism. The term 'category' refers to the general and main function of the protein. The term 'mechanism' refers to the mechanism by which the protein is associated with biofilm formation and development. These include QS, quorum quenching, motility, EPS production, dispersion, and signaling. Database entries are also listed by bacteria, specific strain, and by Gram type. Entries are also classified by the existence of a ligand in the structure and by its designation.

Mutagenesis data is important to identify amino acid residues with a special role in protein function, including ligand binding, enzymatic activity, etc. For each protein, the availability of mutagenesis information is indicated. Links to the mutants section in BRENDA (Braunschweig Enzyme Database) [14] were included, providing information on the effect on kinetics and/or K_M data of the corresponding mutations.

Information about the biological and metabolic pathways in which each protein is involved is also presented (KEGG). This allows the user to quickly visualize the whole mechanism of action and consider including other intervening proteins in the study. The BSD also lists information about the method by which the atomic structure of the protein was obtained (X-ray crystallography, solution NMR, or cryo-EM), resolution, year of deposition on the PDB, and DOI of the corresponding publication.

Each entry contains direct links for structure retrieval from the PDB and for direct access to the entries associated with each protein in the ChEMBL, BindingDB, ExPASy, KEGG, and UniProt databases, when available (Figure 1B). The ChEMBL and BindingDB databases contain information on collections of molecules previously shown to exhibit experimentally confirmed activity against the corresponding protein. These links enable direct download of all known inhibitors or activators for a specific protein, with the corresponding activity and properties for drug development efforts.

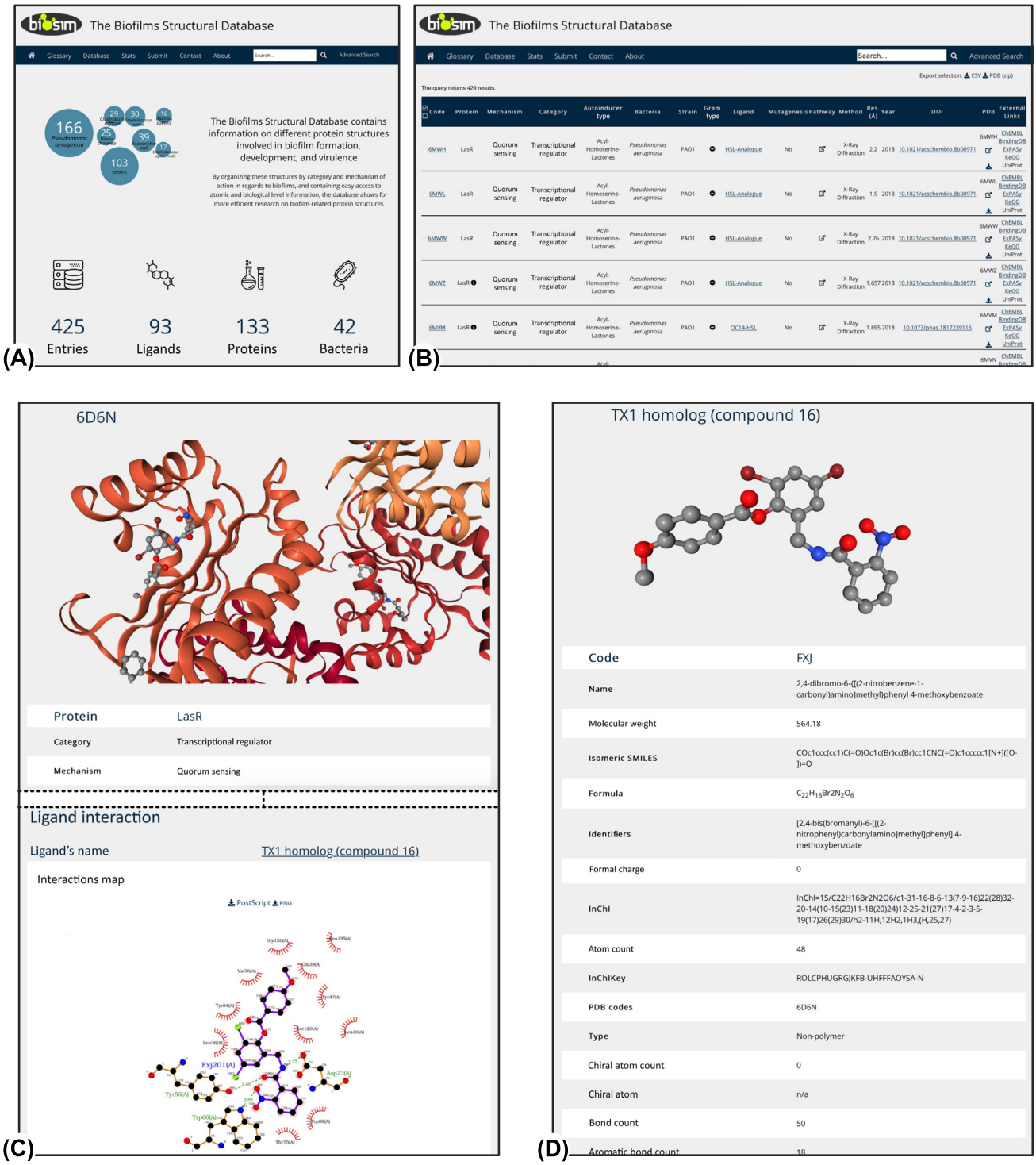
Each BSD entry contains a link to a dedicated webpage containing a summary of all the information indicated previously, including an interactive 3D representation of its structure (Figure 1C). An interactions map of the ligand, obtained with LigPlot+ [15], is also displayed. This is complemented with an interactive 3D representation of the ligand and surrounding amino acid residues and main interactions, which allows users to quickly visualize the 3D structure of the protein–ligand complex at the binding pocket without having to use a molecular visualization software.

Entries are also organized by ligand (Figure 1D). A total of 93 unique ligands are presently available, including substrates, autoinducers, and inhibitors. For each ligand, the corresponding PDB code, systematic name, molecular weight, isomeric SMILES (simplified molecular-input line-entry system, molecular formula, and several other molecular properties are included. An interactive image of the ligand's structure is also displayed.

The database enables queries through a variety of options, including queries by protein, ligand, category, mechanism, autoinducer type, bacteria, strain, and Gram type. Selections can be easily downloaded as a CSV (comma-separated values) file or as a collection of PDB structures.

Concluding Remarks

The BSD is a free-access catalogue containing the structures of all known



Trends in Biotechnology

Figure 1. The Biofilms Structural Database (BSD). (A) BSD main page. (B) List of database entries in BSD. (C) Entries by structure for each Protein Data Bank (PDB) code with indication of the main information available for each protein/structure and interactive 3D representation of the protein. (D) Entries by ligand with main information and interactive 3D representation.

proteins and enzymes involved in biofilm formation. This database is a tool that helps to visualize, explore, and understand biofilm targets to design and develop new and effective antibiofilm drugs and to understand the structure and activity of proteins and enzymes involved in biofilm development. The interface is easy to use and accessible to anyone wishing to start their work in this field.

Acknowledgments

This work was supported by national funds from Fundação para a Ciência e a Tecnologia (grant numbers: SFRH/BD/137844/2018, SFRH/BD/115396/2016, IF/00052/2014, UID/Multi/04378/2019, UIDB/04378/2020, PTDC/QUI-QIN/30649/2017, UID/QUI/50006/2019, and NORTE-01-0145-FEDER-000011) and the Interreg SUDOE NanoDesk (SOE1/P1/E0215; UP). This work was also financially supported by: Base Funding – UIDB/00511/2020 of LEPABE, funded by national funds through the FCT/MCTES (PIDDAC), projects POCI-01-0145-FEDER-030219, POCI-01-0247-FEDER-035234, POCI-01-0247-FEDER-033298, POCI-01-0145-FEDER-028397, and POCI-01-145-FEDER-006939, funded by FEDER funds through COMPETE2020 (POCI), and by national funds (PIDDAC) through SFRH/BSAB/150379/2019.

[1]UCIBIO/REQUIMTE, BioSIM, Departamento de Biomedicina, Faculdade de Medicina da Universidade do Porto, Alameda Prof. Hernâni Monteiro, 4200-319 Porto, Portugal
[2]LAQV/REQUIMTE, Departamento de Química e Bioquímica, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal
[3]LEPABE, Department of Chemical Engineering, Faculty of Engineering, University of Porto, Porto, Portugal
[4]These authors contributed equally to this work.

*Correspondence:
sergiofsousa@med.up.pt (S.F. Sousa).

https://doi.org/10.1016/j.tibtech.2020.04.002

References

1. Flemming, H.-C. and Wingender, J. (2010) The biofilm matrix. *Nat. Rev. Microbiol.* 8, 623–633
2. Jamal, M. *et al.* (2018) Bacterial biofilm and associated infections. *J. Chinese Med. Assoc.* 81, 7–11
3. Simões, M. *et al.* (2010) A review of current and emergent biofilm control strategies. *LWT Food Sci. Technol.* 43, 573–583
4. Roy, R. *et al.* (2018) Strategies for combating bacterial biofilms: a focus on anti-biofilm agents and their mechanisms of action. *Virulence* 9, 522–554
5. Koo, H. *et al.* (2017) Targeting microbial biofilms: current and prospective therapeutic strategies. *Nat. Rev. Microbiol.* 15, 740–755
6. Kamaruzzaman, N.F. *et al.* (2018) Targeting the bacterial protective armour; challenges and novel strategies in the treatment of microbial biofilm. *Materials (Basel)* 11, 1705
7. Guilhen, C. *et al.* (2017) Biofilm dispersal: multiple elaborate strategies for dissemination of bacteria with unique properties. *Mol. Microbiol.* 105, 188–210
8. Berman, H.M. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242
9. Gaulton, A. *et al.* (2017) The ChEMBL database in 2017. *Nucleic Acids Res.* 45, D945–D954
10. Liu, T. *et al.* (2007) BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. *Nucleic Acids Res.* 35, D198–D201
11. Artimo, P. *et al.* (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* 40, W597–W603
12. Kanehisa, M. and Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30
13. Bateman, A. (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515
14. Schomburg, I. *et al.* (2004) BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res.* 32, D431–D433
15. Wallace, A.C. *et al.* (1995) Ligplot: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng. Des. Sel.* 8, 127–134