DOCKGROUND resource for protein recognition studies

Keeley W. Collins,[1] Matthew M. Copeland,[1] Ian Kotthoff,[1] Amar Singh,[1] Petras J. Kundrotas[1]* and Ilya A. Vakser[1,2]*

[1]Computational Biology Program, The University of Kansas, Lawrence, Kansas 66045

[2]Department of Molecular Biosciences, The University of Kansas, Lawrence, Kansas 66045

*Corresponding authors: Ilya A. Vakser and Petras J. Kundrotas, Computational Biology Program, The University of Kansas, 2030 Becker Drive, Lawrence, Kansas 66045; Tel: (785) 864-1057, Fax: (785) 864-5558, Email: vakser@ku.edu and pkundro@ku.edu

Running title: DOCKGROUND resource

Keywords: protein interactions, protein docking, protein modeling, structure similarity, redundancy removal.

Number of manuscript pages: 24

Number of tables: 1

Number of figures: 9

Abstract

Structural information of protein-protein interactions is essential for characterization of life processes at the molecular level. While a small fraction of known protein interactions has experimentally determined structures, computational modeling of protein complexes (protein docking) has to fill the gap. The DOCKGROUND resource (http://dockground.compbio.ku.edu) provides a collection of datasets for the development and testing of protein docking techniques. Currently, DOCKGROUND contains datasets for the bound and the unbound (experimentally determined and simulated) protein structures, model-model complexes, docking decoys of experimentally determined and modeled proteins, and templates for comparative docking. The DOCKGROUND bound proteins dataset is a core set, from which other DOCKGROUND datasets are generated. It is devised as a relational PostgreSQL database containing information on experimentally determined protein-protein complexes. This report on the DOCKGROUND resource describes current status of the datasets, new automated update procedures and further development of the core datasets. We also present a new DOCKGROUND interactive web interface, which allows search by various parameters, such as release date, multimeric state, complex type, structure resolution, etc., visualization of the search results with a number of customizable parameters, as well as downloadable datasets with pre-defined levels of sequence and structure redundancy.

Significance

Proteins function by interacting with other molecules, including other proteins. Characterization of these interactions is important for understanding mechanisms of life processes and improving our ability to treat diseases. The DOCKGROUND public resource offers various datasets of protein structures needed for the development and testing of the techniques for computational modeling of protein interactions. In this report we present an update on the current DOCKGROUND release.

Introduction

Proteins do not function in isolation but by interacting with other molecules, including other proteins. Characterization of these interactions at the atomic level is essential for understanding biomolecular mechanisms, and for gaining insights into possible ways to modulate them. Protein-protein interactions (PPI) regulate many physiological processes and thus are important drug targets. The number of known PPIs has increased dramatically over the past several years. However, it is still difficult to experimentally determine the three-dimensional structure of protein complexes. Experimental methods do not scale well with the increasingly large sets of proteins and their interactions. Computational methods are needed to meet the demand of the large-scale determination of protein interactions.[1] Computational structure prediction of protein-protein complexes (protein docking) is an established way to accurately determine the structures of protein-protein complexes.[2]

Recent advances in deep learning-based prediction of protein complexes[3] have propelled the field to a new level. However, a number of significant challenges in the docking methodology development still remain, such as antigen-antibody complexes, multiprotein assemblies, and conformational ensembles of the interacting proteins. Current developments in the docking field further emphasize the need for protein-protein datasets that can be used to train and to validate data-driven and physics-based protein docking techniques. DOCKGROUND public resource (http://dockground.compbio.ku.edu) offers various interconnected datasets of experimentally determined and modeled structures suitable for the development and testing of various aspects of protein docking. In this report we present an update of the current DOCKGROUND release, focusing on the new developments of the basic data sets and user-oriented utilities.

Database Content and Description

The DOCKGROUND resource currently consists of five integrated and interconnected categories of the protein-protein datasets:

1. Bound set - experimentally determined protein-protein complexes. The core set from which other sets are derived.
2. Unbound sets - (a) experimentally determined and (b) simulated unbound protein structures determined outside of the protein-protein complex, corresponding to the structures in the Bound set. The sets are used for studies of conformational changes upon binding and for benchmarking of the docking routines. The simulated unbound proteins conformations deviate from the bound ones to the degree similar to that of the experimentally determined structures.
3. Model-model complexes - modeled protein structures from protein-protein complexes at different levels of structural accuracy, for benchmarking of the docking procedures designed for the modeled proteins.
4. Docking decoys (scoring benchmarks) - sets of (a) experimentally determined and (b) modeled protein structures in incorrect (decoys) and correct docking configurations for development and validation of the scoring functions/procedures capable of discriminating false-positive docking predictions.
5. Docking templates - experimentally determined structures of protein-protein complexes for comparative protein docking.

Figure 1 shows the general logic of the connections between these sets. User-friendly web interface provides a simple way to download current and previous versions of the pre-compiled datasets as well as options for advanced search/generation of custom datasets, and visualization of the search results.

Bound Proteins Set

The bound dataset of the experimentally determined protein-protein structures is the core set of the DOCKGROUND resource. Most other DOCKGROUND sets are derived from it. There are various other databases of experimentally determined protein-protein complexes.[4-11] The DOCKGROUND bound set aims to distinguish itself in comprehensiveness, customizability, and integration with various resources for the development and testing of docking techniques. Previously, the *ad hoc* updates of the bound dataset had to be performed semi-manually, on an irregular basis. As a result, many recent interesting structures have been missing in the DOCKGROUND releases, reducing its utility for the research community. In this report, we present a major DOCKGROUND development consisting of the comprehensive, fully automated update of the bound dataset, which follows the weekly Protein Data Bank (PDB)[12] updates. Since the previous comprehensive DOCKGROUND report,[13] the bound set has increased from 215,363 to 667,331 interfaces and from 169,295 to 363,240 protein chains. Statistics on the complexes in the current database are shown in Figure 2. New annotations have been made available for the protein complexes such as membrane localization, presence of the disulfide bonds or nucleic acids at the interface, and the initial release date. Protein complexes and the interacting residues can now be visualized on the DOCKGROUND website using JSmol.[14] Non-redundant datasets based on sequence or structure similarity criteria have been made available. The sets are updated on a weekly basis.

The DOCKGROUND resource now contains a separate set of membrane protein complexes. Despite their essential role in cellular mechanisms and significant progress in the structure determination of the membrane proteins, such complexes are still significantly underrepresented in PDB. Thus, computational approaches to prediction of membrane protein complexes are especially valuable. The generic protein docking approaches, developed primarily for the soluble proteins, are not well-suited for the membrane proteins because of the differences in physicochemical environment and the constraints on the docking space imposed by the

membrane.[15] The current DOCKGROUND resource contains a dataset of 456 non-redundant alpha helical binary interfaces,[16] which is significantly larger and more representative than the previously developed sets. The set will become the foundation for the development of docking and scoring benchmarks, similar to the ones for the soluble proteins.

Benchmark Sets and Tools for Docking

Datasets of unbound proteins corresponding to the complexes of bound proteins are important for the development and testing of the protein docking procedures. The DOCKGROUND unbound set is an integral part of the resource and the basis for docking decoy sets of the unbound proteins. The most recent Docking Benchmark set 4 contains 396 protein-protein complexes (223 complexes consisting of single-chain monomers, and the rest containing one or both protein subunits consisting of two or more chains).

The number of proteins determined experimentally in both bound and unbound conformation is relatively small. Thus, a significant expansion of such set can be achieved by computational simulation. DOCKGROUND set of such simulated structures contains 3,205 single protein chains from 1,918 complexes. Proteins were selected from the bound set and subjected to 1 ns Langevin dynamics simulation. The simulated unbound structures were selected according to criteria derived from comparison of experimentally determined unbound vs. bound proteins.[17]

DOCKGROUND also provides carefully curated sets of representative modeled structures of proteins with arrays of varying structural accuracy. The most recent comprehensive sets of protein models for the development and validation of protein docking[18] reflect the real case docking scenario where the accuracy of the protein models is assessed by the modeling procedure, without reference to the native structure, which would be unknown in practical applications. The protein models were generated by the Phyre modeling pipeline,[19] with

accuracy assessed by the Phyre ranks, for 171 and 963 binary protein complexes from the DOCKGROUND docking benchmark set 4 and GWIDD database,[20] respectively.

Scoring procedures, essential for docking, are assessed on scoring benchmarks - sets of docking poses, some of which are close to the native structure and the rest are false-positive matches (decoys). DOCKGROUND provides two such sets for the unbound experimentally determined proteins structures - the first set consisting of 99 non-native and one near-native match for 61 unbound complexes in the Docking Benchmark 2, and a larger decoy set derived from 396 unbound complexes in DOCKGROUND Docking Benchmark 4. In addition, DOCKGROUND provides a set based on protein models from the DOCKGROUND Model - Model Benchmark 2. The docking decoys were designed to reflect the reality of the real case docking applications with regard to the spatial distribution of matches and their energy balance.

Comparative (homology) protein docking relies on sequence or structure similarity of the docking targets to the available templates (experimentally determined structures of protein-protein complexes).[21-25] The key to successful template-based docking is the availability of high quality, diverse, non-redundant template libraries. Generating a high-quality template library is more complicated than simply selecting all pairwise protein-protein complexes from PDB. Although such selection would result in a full set of currently known structures, such structures would be highly redundant and thus, could bias the docking predictions, as well as decrease computational efficiency of the docking procedures. They also would include many erroneous, low-quality, and/or biologically irrelevant structures.[26,27] Generally, research groups developing comparative docking techniques generate their own template libraries by filtering the PDB for relevant interactions, which complicates a fair comparison of these methodologies. The DOCKGROUND resource contains sets of full and interface-only protein structures determined within a protein-protein complex for use as templates in homology docking. The input proteins were those with the structural resolution 3.5 Å and better (for X-ray and EM structures) and the

first models in the NMR structures. Pairs of chains were compared by MM-align program,[28] and redundancy removed with MM-score threshold 0.9.

Update Procedures and Development of the Core Dataset

In the current DOCKGROUND release the update procedures are implemented for the Bound set. These will be followed by the update procedures in the other DOCKGROUND sets in the future releases.

Details of the initial procedure for generation of the bound set were published earlier.[13,26,29] Here we provide a brief summary and a description of changes that have been made since then. The bound dataset is derived from the PDB biounit files, filtered, annotated, and inserted into a relational PostgreSQL database. The structures in the database are organized into pairwise complexes (binary combination of two chains within the same protein structure with a mean buried area > 250 $Å^2$ at the interface). We filtered out pairwise combinations that only contain alpha-carbon atoms, those with the resolution lower than 6.0 Å, with chains containing invalid residues, and/or have one or both chains with < 30 residues. Entries were annotated with relevant information such as presence of ligands, DNA or RNA at the interface, membrane localization, etc. Statistics on the protein entries in the DOCKGROUND annotated with various attributes are in Table 1. Users can download the whole dataset or generate a subset resulting from various search parameters on the "Bound -> Build Database" portion of the DOCKGROUND webpage (Figure 3). The sets can be downloaded in a tab delimited format. Figure 4 shows a fragment of the search results page. Changes have been made to the initial generation procedure for adding new structures. In addition to the structures derived from the X-ray crystallography, structures from electron microscopy and solution NMR can also be added to the custom sets. Structures are annotated and can be searched for by the initial release date. Each protein-protein complex and its interacting residues can be visualized using JSmol[14]

interface (Figure 5). The bound database updates on a weekly basis, following the weekly PDB update, and takes 2 - 3 hours to complete.

Redundancy Reduction

Redundancy can be removed between structures of the bound dataset using sequence and structure similarity criteria. On the search result page for the bound dataset, users can reduce the custom bound dataset by selecting the reduction method (by sequence or by structure similarity), and the redundancy cutoff. Currently, we offer precomputed datasets for three levels of the sequence redundancy and two levels of the structure redundancy. These reduced datasets are also updated on a weekly basis.

To generate the sequence-based non-redundant dataset, NCBI Blast[30] is used on all unique protein chains in the bound protein database. Chains with the sequence identity above a certain cutoff (30, 40, and 50%) are clustered using highly connected subgraphs (HCS).[31] Clusters of pairwise complexes are constructed from the chain clusters by grouping together pairs of chains that come from the same chain clusters (e.g., if chains A and B come from the chain cluster *N* and chains C and D come from the chain cluster *M*, then pairwise combinations AC, AD, BC and BD, if they exist, would belong to the pairwise cluster *NM*, while combinations AA and AB would belong to the pairwise cluster *NN*). Examples of pairwise clusters at a 30% sequence identity are shown in Figure 6. Most clusters have 1 - 10 members, but a few clusters have a very large occupancy. Relative frequency of cluster sizes, with the cutoff for redundancy removal 30% sequence identity, is shown in Figure 7.

To generate non-redundant datasets based on the structure similarity, we use Foldseek[32] first, to filter out structures with very low structural similarity between protein chains. Foldseek converts 3D structure into a one-dimensional “structural alphabet” to compare structures very quickly using a sequence-alignment-like algorithm. Similar chains are clustered using HCS[31] at a fraction identity score 0.30 (a metric used by Foldseek similar to the sequence identity), and

then sorted into clusters of pairwise complexes. Each member of the clusters of the pairwise complexes is validated using MM-align. Members of the cluster are removed and split into a new cluster if their structural similarity falls below the clustering cutoff. Representative sets are available with TM-score clustering cutoffs 0.6 and 0.9. An example of a pairwise cluster at 0.6 and 0.9 TM-score cutoffs is shown in Figure 8. Similar to the clusters from the sequence alignment, most clusters from the structure alignment have 1 - 10 members. A few clusters have a large occupancy, with the largest having > 500 members. The frequency of the cluster sizes with the structure similarity cutoff TM-score 0.6 is shown in Figure 9.

Concluding Remarks and Future Development

DOCKGROUND is a comprehensive public resource for studying protein-protein recognition and structural modeling of protein complexes. It contains integrated datasets for development and testing of the key aspects of the protein docking methodologies. The current datasets include a database of experimentally determined bound protein-protein complexes, datasets of experimentally determined and simulated unbound structures, model-model docking benchmark sets, docking decoys of experimentally determined and modeled protein structures, and sets of templates for comparative protein docking. All sets are available for the download through a user-friendly interface on the DOCKGROUND website at http://dockground.compbio.ku.edu.

The bound database is the core part of the DOCKGROUND from which the other datasets are derived. It updates automatically on a weekly basis with new and updated structures from the PDB. A number of the other datasets within DOCKGROUND are updated less frequently and have little or no automation. A major part of the future development will be creating automated procedures for the updates to the other DOCKGROUND datasets. The bound database can be searched in order to generate custom protein-protein datasets. It can be reduced using either sequence or structural similarity with several pre-defined cutoffs. In the future DOCKGROUND releases we plan to let users specify custom cutoffs for the redundancy reduction, and make

clusters of the redundant structures available, along with the datasets of representative non-redundant structures. This can serve, for example, as a tool for studying multiple protein conformations. We also plan to implement interfaces for generating custom sets for the other DOCKGROUND datasets, similar to those for the bound database.

Future development will involve larger sets of unbound (simulated) structures, and the sets of unbound and modeled structures for multimeric assembly (beyond the pairwise docking). A major expansion of the sets will be achieved by including structures predicted by AlphaFold. The docking decoy sets will be significantly increased in size to become more adequate to the real-case docking scenarios. We will develop automatic procedure for splitting of the sets into training and testing subsets, which would exclude similarity between the training and testing structures above certain sequence identity or structure similarity thresholds. We will add new datasets for modeling of protein complexes, such as libraries of rotamers and rotamer–rotamer transition probabilities.[33] The DOCKGROUND resource will expand to include other macromolecular complexes (protein-RNA, protein-DNA). It will be also integrated with other resources performing related functions, such as docking and mapping of the intermolecular energy landscapes.

Acknowledgements

Various people have contributed to the development of DOCKGROUND over the years. Dominique Douguet designed and generated the first bound database, Ying Gao built the original X-ray unbound sets, Shiyong Liu generated the first set of docking decoys, and Anatoly Ruvinsky and Tatsiana Kirys created the simulated unbound set, with the help of Deepak Singla. Andrey Tovchigrechko designed and implemented the original set of models for the model-model docking benchmarks, which eventually led to the development of the DOCKGROUND current model-model sets. Ivan Anishchenko generated the first set of docking templates. Sherman Choi helped to develop C++ functions used in the automated update

procedure for the bound database. This study was supported by NIH grant R01GM074255 and NSF grant DBI1917263.

References

1. Vakser IA. Low-resolution structural modeling of protein interactome. *Curr Opin Struct Biol.* 2013;23:198–205.
2. Vakser IA. Challenges in protein docking. *Curr Opin Struct Biol.* 2020;64:160-165.
3. Gao M, Nakajima An D, Parks JM, Skolnick J. AF2Complex predicts direct physical interactions in multimeric proteins with deep learning. *Nat Commun.* 2022;13:1744.
4. Basse MJ, Betzi S, Bourgeas R, et al. 2P2Idb: A structural database dedicated to orthosteric modulation of protein–protein interactions. *Nucl Acid Res.* 2013;41:D824–D827.
5. Davis FP, Sali A. PIBASE: A comprehensive database of structurally defined protein interfaces *Bioinformatics.* 2005;21:1901-1907.
6. Gong S, Park C, Choi H, et al. A protein domain interaction interface database: InterPare. *BMC Bioinformatics.* 2005;6:207.
7. Jefferson ER, Walsh TP, Roberts TJ, Barton GJ. SNAPPI-DB: A database and API of Structures, iNterfaces and Alignments for Protein–Protein Interactions. *Nucleic Acids Res.* 2007;35:D580-D589.
8. Keskin O, Tsai CJ, Wolfson H, Nussinov R. A new, structurally nonredundant, diverse data set of protein–protein interfaces and its implications. *Protein Sci.* 2004;13:1043-1055.
9. Meyer MJ, Das J, Wang X, Yu H. INstruct: A database of high-quality 3D structurally resolved protein interactome networks. *Bioinformatics.* 2013;29:1577–1579.
10. Teyra J, Doms A, Schroeder M, Pisabarro MT. SCOWLP: A web-based database for detailed characterization and visualization of protein interfaces. *BMC Bioinformatics.* 2006;7:104.
11. Xu Q, Dunbrack RL. ProtCID: A data resource for structural information on protein interactions. *Nature Comm.* 2020;11:711.
12. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res.* 2000;28:235-242.
13. Kundrotas PJ, Anishchenko I, Dauzhenka T, et al. DOCKGROUND: A comprehensive data resource for modeling of protein complexes. *Protein Sci.* 2018;27:172-181.
14. R. H. Jmol - a paradigm shift in crystallographic visualization. *J Appl Cryst.* 2010;43:1250-1260.
15. Hurwitz N, Schneidman-Duhovny D, Wolfson HJ. Memdock: An alpha-helical membrane protein docking algorithm. *Bioinformatics.* 2016;32:2444-2450.
16. Kotthoff I, Kundrotas PJ, Vakser IA. DOCKGROUND membrane protein-protein set. *PLoS One.* 2022;17:e0267531.
17. Kirys T, Ruvinsky AM, Singla D, Tuzikov AV, Kundrotas PJ, Vakser IA. Simulated unbound structures for benchmarking of protein docking in the DOCKGROUND resource. *BMC Bioinformatics.* 2015;16:243.
18. Singh A, Dauzhenka T, Kundrotas PJ, Sternberg MJE, Vakser IA. Application of docking methodologies to modeled proteins. *Proteins.* 2020;88:1180-1188.
19. Kelley LA, Sternberg MJE. Protein structure prediction on the web: a case study using the Phyre server. *Nature Prot.* 2009;4:363-371.
20. Kundrotas PJ, Zhu Z, Vakser IA. GWIDD: A comprehensive resource for genome-wide structural modeling of protein-protein interactions. *Human Genomics.* 2012;6:7.
21. Aloy P, Pichaud M, Russell RB. Protein complexes: Structure prediction challenges for the 21st century *Curr Opin Struct Biol.* 2005;15:15-22.
22. Dey F, Zhang QC, Petrey D, Honig B. Toward a ‘‘structural BLAST’’: Using structural relationships to infer function. *Protein Sci.* 2013;22:359-366.
23. Szilagyi A, Zhang Y. Template-based structure modeling of protein–protein interactions. *Curr Opin Struct Biol.* 2014;24:10–23.

24. Tuncbag N, Keskin O, Nussinov R, Gursoy A. Fast and accurate modeling of protein–protein interactions by combining template-interface-based docking with flexible refinement. *Proteins.* 2012;80:1239–1249.
25. Kundrotas PJ, Zhu Z, Janin J, Vakser IA. Templates are available to model nearly all complexes of structurally characterized proteins. *Proc Natl Acad Sci USA.* 2012;109:9438–9441.
26. Douguet D, Chen HC, Tovchigrechko A, Vakser IA. DOCKGROUND resource for studying protein-protein interfaces. *Bioinformatics.* 2006;22:2612–2618.
27. Kundrotas PJ, Vakser IA, Janin J. Structural templates for modeling homodimers. *Protein Sci.* 2013;22:1655—1663.
28. Mukherjee S, Zhang Y. MM-align: A quick algorithm for aligning multiple-chain protein complex structures using iterative dynamic programming. *Nucl Acids Res.* 2009;37:e83.
29. Kundrotas PJ, Kotthoff I, Choi SW, Copeland MM, Vakser IA. Dockground tool for development and benchmarking of protein docking procedures. *Methods Mol Biol.* 2020;2165:289-300.
30. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403-410.
31. Albirri ER, Sugeng KA, Aldila D. On the modification Highly Connected Subgraphs (HCS) algorithm in graph clustering for weighted graph. *J Phys.* 2018;1008:012037.
32. van Kempen M, Kim SS, Tumescheit C, Mirdita M, Söding J, Steinegger M. Foldseek: Fast and accurate protein structure search. *bioRxiv.* 2022;doi.org/10.1101/2022.02.07.479398.
33. Kirys T, Ruvinsky A, Tuzikov AV, Vakser IA. Rotamer libraries and probabilities of transition between rotamers for the side chains in protein-protein binding. *Proteins.* 2012;80:2089–2098.

TABLES

Table 1. Statistics on protein entries in DOCKGROUND annotated with various attributes

Attribute	Number of Entries	Fraction of Entries[a]
Disulfide Bond at Interface	7,360	0.011
DNA/RNA at Interface	125,033	0.187
Membrane	102,751	0.154
Ligand at Interface	452,751	0.678

[a] With respect to the total number of interfaces (667,331).

FIGURES

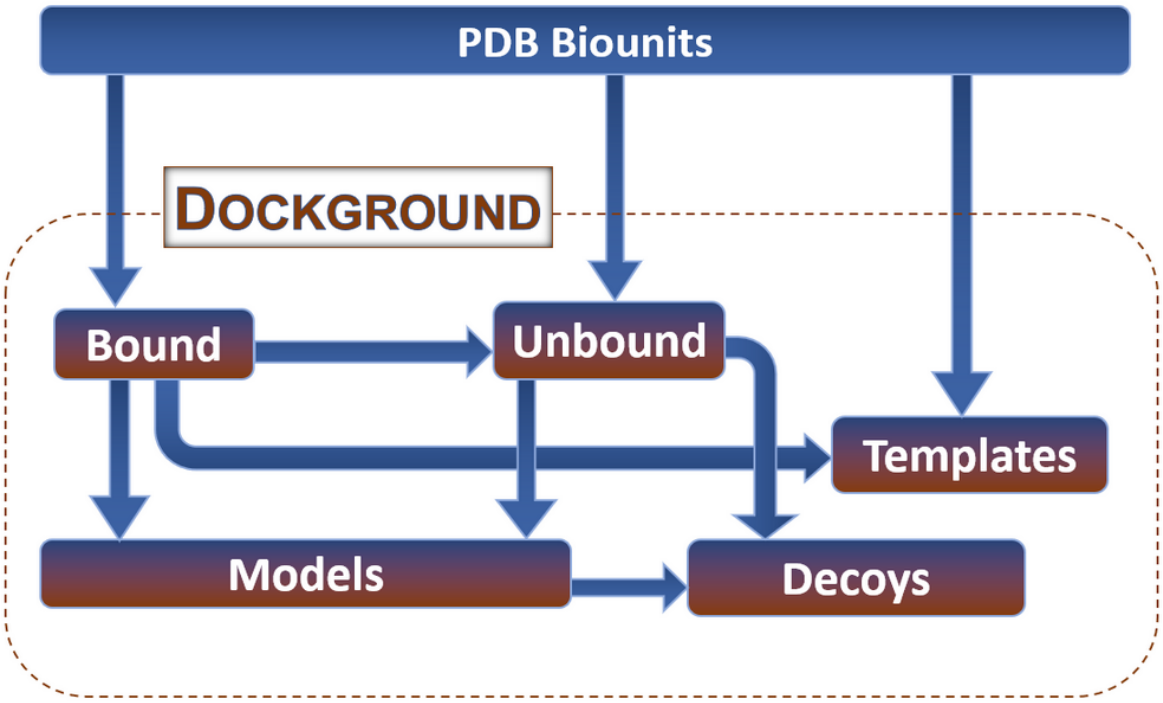

Figure 1. *DOCKGROUND datasets and connections between them.*

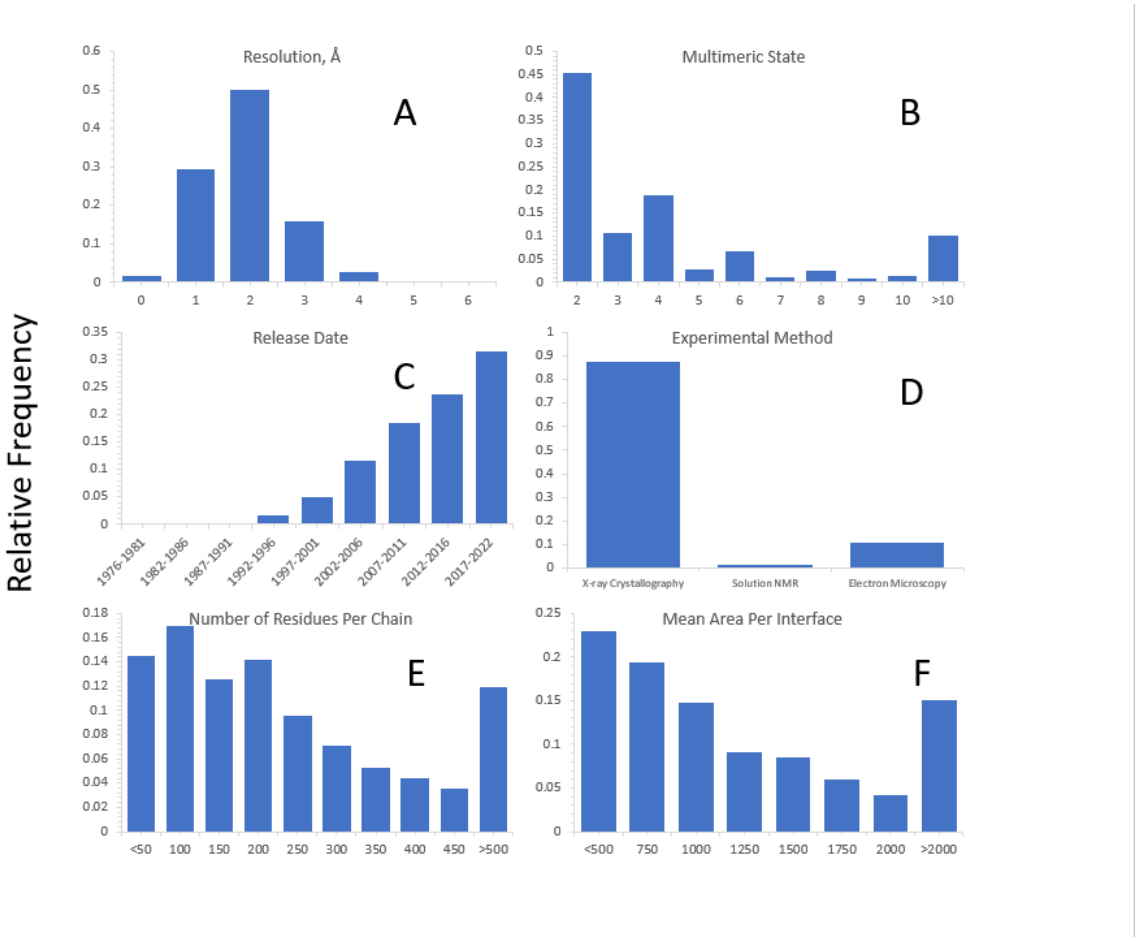

Figure 2. *Statistics on the contents of the bound database.* Normalization of the data is with respect to the total number of PDB biounit files (A - D), the total number of chains (E), and the total number of interfaces (F) in the database.

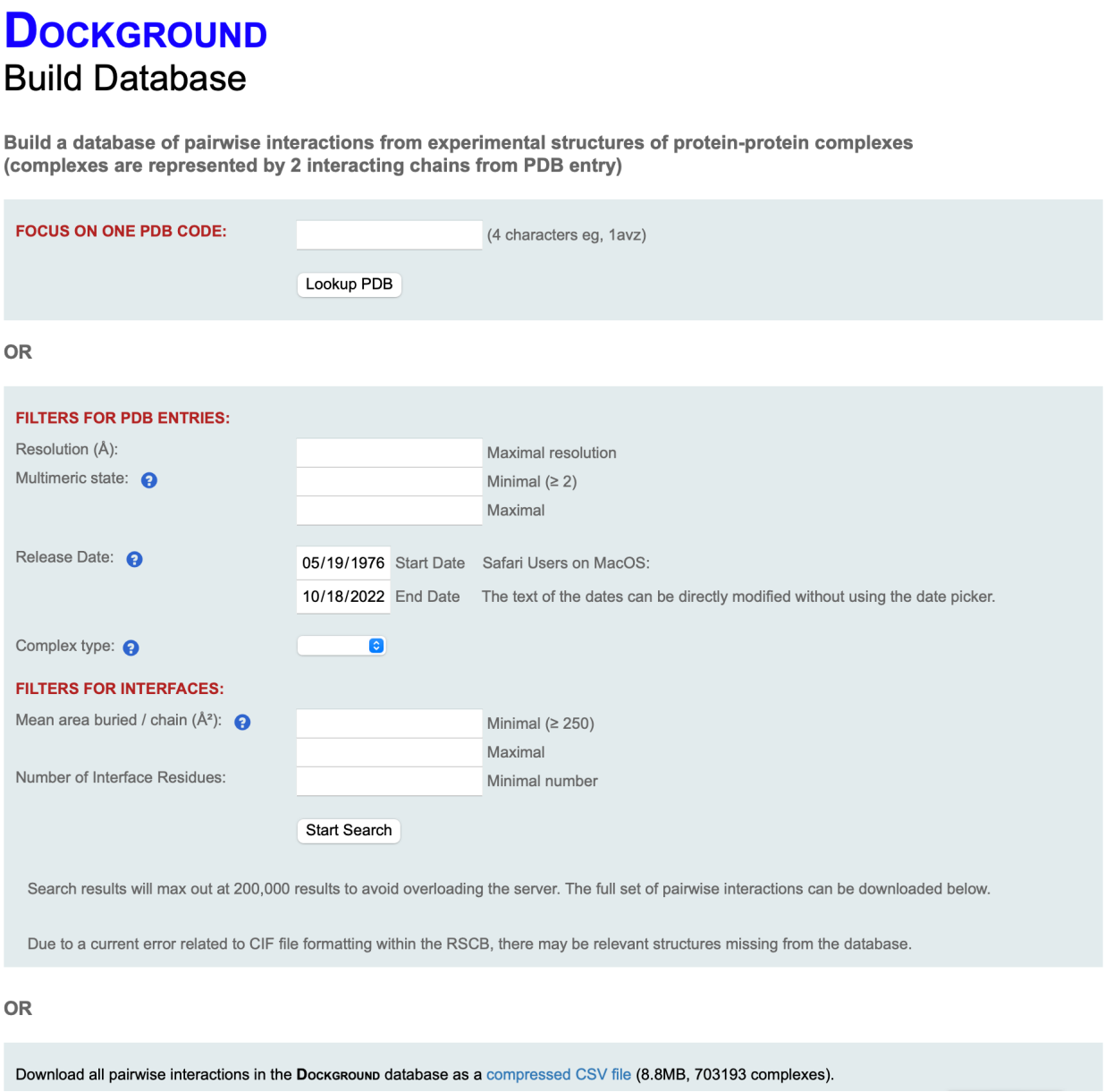

Figure 3. *The search form for the Bound database.* Users can search for a specific PDB code or search using various filters.

PDB CODE	Title	Complex Type	Biounit chain name (1)	Model number (1)	Biounit chain name (2)	Model number (2)	Mean area buried by each chain	Release Date	Visualize
1gx7	Best model of the electron transfer complex between cytochrome c3 and [Fe]-hydrogenase	HETERO	A	1	D	1	3866	2003-07-31	View
1gx7	Best model of the electron transfer complex between cytochrome c3 and [Fe]-hydrogenase	HETERO	A	1	E	1	775	2003-07-31	View
1gx7	Best model of the electron transfer complex between cytochrome c3 and [Fe]-hydrogenase	HETERO	D	1	E	1	431	2003-07-31	View
1h0d	Crystal structure of Human Angiogenin in complex with Fab fragment of its monoclonal antibody mAb 26-2F	HETERO	A	1	B	1	1925	2003-06-19	View
1h0d	Crystal structure of Human Angiogenin in complex with Fab fragment of its monoclonal antibody mAb 26-2F	HETERO	A	1	C	1	317	2003-06-19	View
1h0d	Crystal structure of Human Angiogenin in complex with Fab fragment of its monoclonal antibody mAb 26-2F	HETERO	B	1	C	1	441	2003-06-19	View
1j1o	Crystal Structure of HyHEL-10 Fv mutant LY50F complexed with hen egg white lysozyme	HETERO	H	1	Y	1	564	2003-01-14	View
1j1o	Crystal Structure of HyHEL-10 Fv mutant LY50F complexed with hen egg white lysozyme	HETERO	L	1	Y	1	385	2003-01-14	View
1j1o	Crystal Structure of HyHEL-10 Fv mutant LY50F complexed with hen egg white lysozyme	HETERO	L	1	H	1	758	2003-01-14	View
1j1p	Crystal structure of HyHEL-10 Fv mutant LS91A complexed with hen egg white lysozyme	HETERO	H	1	Y	1	567	2003-01-14	View
1j1p	Crystal structure of HyHEL-10 Fv mutant LS91A complexed with hen egg white lysozyme	HETERO	L	1	Y	1	400	2003-01-14	View
1j1p	Crystal structure of HyHEL-10 Fv mutant LS91A complexed with hen egg white lysozyme	HETERO	L	1	H	1	764	2003-01-14	View
1j1x	Crystal Structure of HyHEL-10 Fv mutant LS93A complexed with hen egg white lysozyme	HETERO	H	1	Y	1	572	2003-01-14	View
1j1x	Crystal Structure of HyHEL-10 Fv mutant LS93A complexed with hen egg white lysozyme	HETERO	L	1	H	1	759	2003-01-14	View
1j1x	Crystal Structure of HyHEL-10 Fv mutant LS93A complexed with hen egg white lysozyme	HETERO	L	1	Y	1	392	2003-01-14	View
1j34	Crystal Structure of Mg(II)-and Ca(II)-bound Gla Domain of Factor IX Complexed with Binding Protein	HETERO	A	1	B	1	1826	2003-07-08	View
1j34	Crystal Structure of Mg(II)-and Ca(II)-bound Gla Domain of Factor IX Complexed with Binding Protein	HETERO	A	1	C	1	579	2003-07-08	View

Figure 4. *A fragment of the search results of the Bound database*.

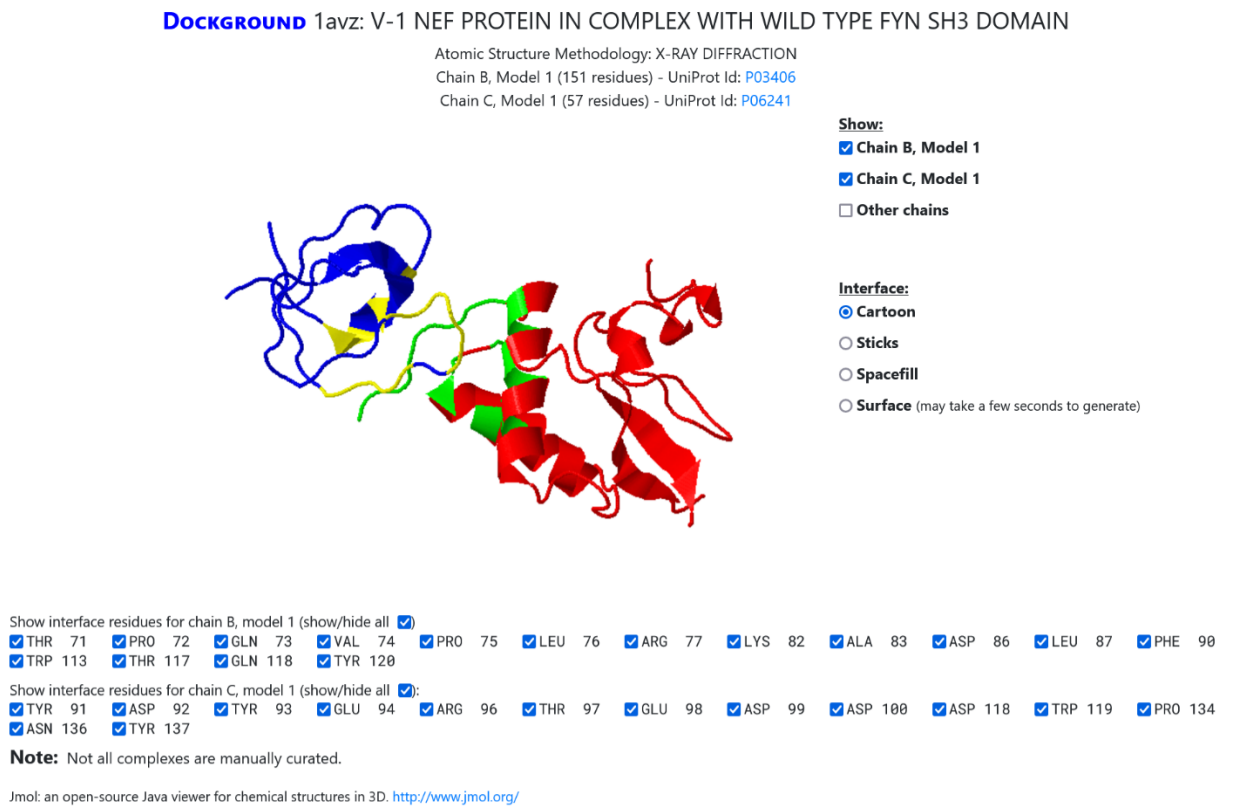

Figure 5. *An example of a protein-protein complex visualized on the* DOCKGROUND *website using JSmol.* The two chains are in blue and red, and their interacting residues are in yellow and green. Their interacting residues are listed at the bottom and can be displayed or hidden by checking the box next to their name.

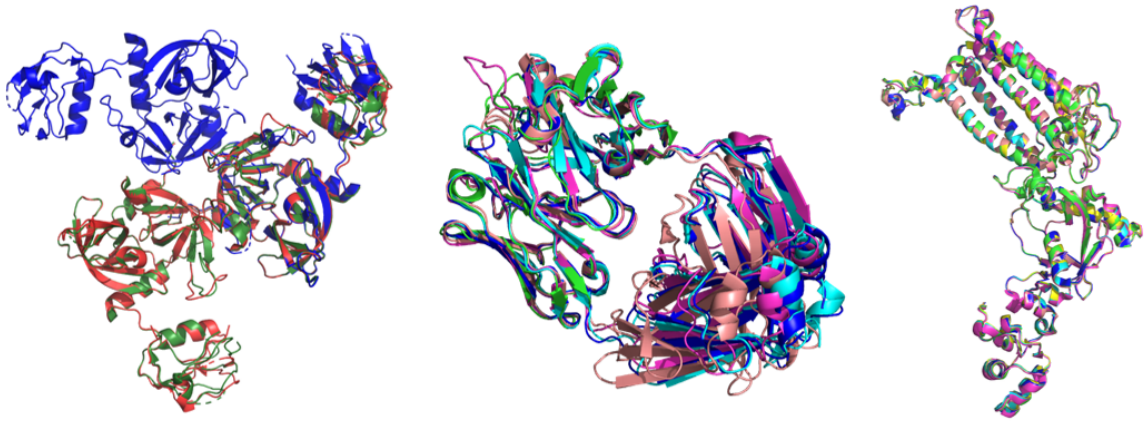

Figure 6. *Examples of clusters of redundant pairwise complexes.* The cutoff for the redundance removal is 30% sequence identity.

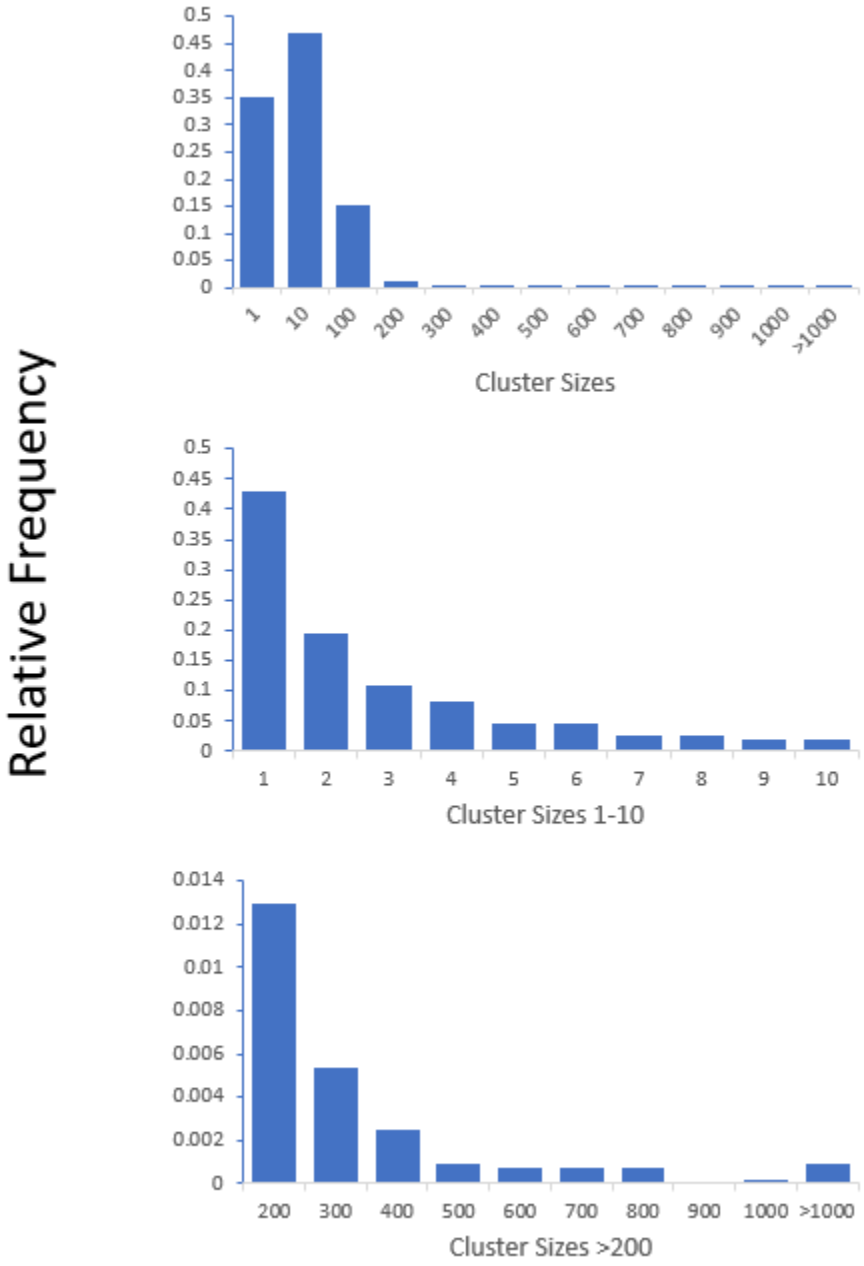

Figure 7. *Relative frequency of cluster sizes*. The cutoff for redundancy removal is 30% sequence identity.

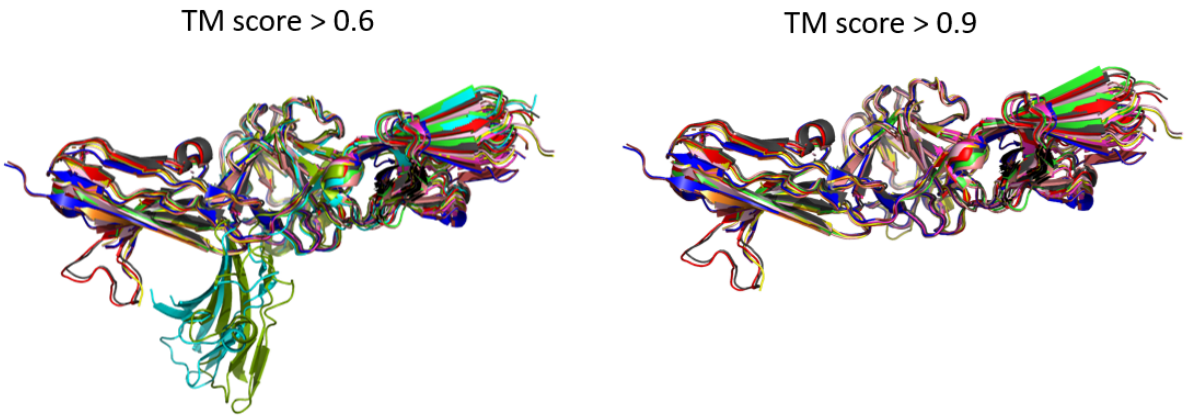

Figure 8. *Clusters of redundant structures with the same representative complex at different structure similarity cutoffs.*

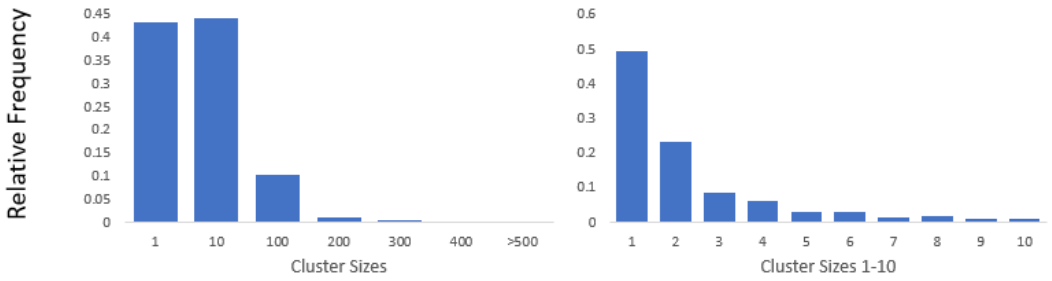

Figure 9. *Relative frequency of cluster sizes.* The structure similarity cutoff is TM-score 0.6.