PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

A benchmark testing ground for integrating homology modeling and protein docking

Tanggis Bohnuud,[1] Lingqi Luo,[1] Shoshana J. Wodak,[2,3,4] Alexandre M. J. J. Bonvin,[5] Zhiping Weng,[6] Sandor Vajda,[1,8]* Ora Schueler-Furman,[7]* and Dima Kozakov[1,9]*

[1] Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215

[2] VIB Structural Biology Research Center, VUB Pleinlaan 2, Brussels 1050, Belgium

[3] Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada

[4] Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada

[5] Bijvoet Center for Biomolecular Research, Faculty of Science, Chemistry, Utrecht University, Utrecht 3584CH, The Netherlands

[6] Program in Bioinformatics and Integrative Biology, Department of Biochemistry and Molecular Pharmacology University of Massachusetts Medical School, Worcester, Massachusetts

[7] Department of Microbiology and Molecular Genetics, Institute for Medical Research Israel-Canada, Hadassah Medical School Hebrew University, Jerusalem, Israel

[8] Department of Chemistry, Boston University, Boston, Massachusetts 02215

[9] Department of Applied Mathematics and Statistics, Stony Brook University, New York

ABSTRACT

Protein docking procedures carry out the task of predicting the structure of a protein–protein complex starting from the known structures of the individual protein components. More often than not, however, the structure of one or both components is not known, but can be derived by homology modeling on the basis of known structures of related proteins deposited in the Protein Data Bank (PDB). Thus, the problem is to develop methods that optimally integrate homology modeling and docking with the goal of predicting the structure of a complex directly from the amino acid sequences of its component proteins. One possibility is to use the best available homology modeling and docking methods. However, the models built for the individual subunits often differ to a significant degree from the bound conformation in the complex, often much more so than the differences observed between free and bound structures of the same protein, and therefore additional conformational adjustments, both at the backbone and side chain levels need to be modeled to achieve an accurate docking prediction. In particular, even homology models of overall good accuracy frequently include localized errors that unfavorably impact docking results. The predicted reliability of the different regions in the model can also serve as a useful input for the docking calculations. Here we present a benchmark dataset that should help to explore and solve combined modeling and docking problems. This dataset comprises a subset of the experimentally solved 'target' complexes from the widely used Docking Benchmark from the Weng Lab (excluding antibody–antigen complexes). This subset is extended to include the structures from the PDB related to those of the individual components of each complex, and hence represent potential templates for investigating and benchmarking integrated homology modeling and docking approaches. Template sets can be dynamically customized by specifying ranges in sequence similarity and in PDB release dates, or using other filtering options, such as excluding sets of specific structures from the template list. Multiple sequence alignments, as well as structural alignments of the templates to their corresponding subunits in the target are also provided. The resource is accessible online or can be downloaded at http://cluspro.org/benchmark, and is updated on a weekly basis in synchrony with new PDB releases.

Proteins 2017; 85:10–16.

Key words: protein–protein docking; method development; CAPRI docking experiment; protein structure prediction; user community.

Grant sponsor: NSF AF 1527292; Grant sponsor: NSF DBI 1458509; Grant sponsor: US Israel BSF 2009418; Grant sponsor: NIH R01 GM093147; Grant sponsor: NIH R01 GM061867; Grant sponsor: NIH R01 GM084884.
*Correspondence to: Dima Kozakov, Department of Applied Mathematics and Statistics, Stony Brook University, New York, USA. E-mail: midas@laufercenter.org; Ora Schueler-Furman; Department of Microbiology and Molecular Genetics, Institute for Medical Research Israel-Canada, Hadassah Medical School, Hebrew University, Jerusalem, Israel. E-mail: oraf@ekmd.huji.ac.il, or Sandor Vajda; Department of Biomedical Engineering, Boston University, 44 Cummington Street Boston MA 02215. E-mail: vajda@bu.edu
Received 10 July 2015; Accepted 8 May 2016
Published online 12 May 2016 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.25063

INRODUCTION

The performance of computational protein docking methods[1] has significantly improved in recent years, as witnessed by the results of CAPRI (Critical Assessment of Predicted Interactions), a community-wide experiment for objectively assessing progress in these methods.[2] Available docking procedures are increasingly successful in accurately predicting the structure of a protein complex starting form the known structures of the free monomers (unbound–unbound docking), when the latter undergo limited conformational changes upon association.[2]

However, experimentally determined 3D structures are currently available for <1 out of a thousand proteins with known amino acid sequence information. On the other hand, analyses of protein structures determined so far revealed that these structures represent variations around common themes, or proteins folds, that the repertoire of such folds is limited,[3] and that this repertoire has essentially been mapped out in recent years notably by structural genomics efforts.[4] This has brought to the forefront the role of homology modeling techniques that derive the 3D structure of a protein from its amino-acid sequence using as template the known structure of a related protein.[5] Results of CASP (Critical Assessment of Structure Prediction) have shown that these techniques have become increasingly reliable,[6,7] with protein structures predicted by these methods becoming sufficiently accurate for many applications in biomedical research.[8] Furthermore, it has been suggested that the landscape of known protein structures can also be exploited to computationally derive structural information on protein *complexes*.[9,10] This has opened up the attractive possibility of further populating the structural landscape of protein complexes, which has so far remained very poorly characterized by experimental methods.

Bootstrapping of structural information from known structures of individual proteins to complexes can be done using two main approaches. The most common approach, used increasingly often in CAPRI Rounds, is to predict the structures of the individual protein subunits by homology modeling techniques, and employ protein docking methods to derive the structure of the complex from the predicted subunit models.[11] Another approach, the so-called template-based modeling approach,[10,12] models the structure of the target complex directly on the basis of a template representing the known 3D structure of a complex whose components are related to the subunits of the target protein. It was recently suggested that this latter approach could be used for genome-scale predictions of protein–protein interactions, provided the template complexes available in the PDB are sufficiently similar to the target complex.[10] However, due to the paucity of known structures of protein complexes, the identified templates are often too distantly related to the target complex to enable accurate modeling of its structure. This currently hampers the large-scale applicability of template-based approaches. While the template-based approaches may ultimately represent a breakthrough that will play an important role in the future, almost 50% of the targets in CAPRI experiments are still best tackled by combining homology modeling of the individual subunits with docking methods to derive the structure of the complex. But even this approach remains challenging, as it yields results that can vary widely in accuracy. An analysis of CAPRI targets has recently shown that the global sequence identity between the template and target proteins is a good predictor of the achievable quality of the docking models,[13] in turn suggesting that the ability to build accurate homology models for the interacting components critically influences docking predictions. Another study suggested that in general, the structural differences between the target and template proteins may affect docking performance much more than moderate sequence diversity, even when a template represents an independently solved structure of the same protein.[14] Thus, it is presently unclear how the structural diversity of potential templates and its relation to the intrinsic flexibility of the protein or to evolutionary changes may be harnessed to improve docking predictions.

One possible solution to the above problem is simply combining the best available homology modeling and docking methods. However, the experience of CAPRI shows that this is not always the best approach. In fact, global measures of model quality such as the backbone root mean square deviation (RMSD) and the global distance test (GDT) do not properly account for localized errors that may unfavorably impact docking results. For example, the conformations of some side chains can be very important if they turn out to be in the interface, and hence their prediction may require special attention. It is also important to predict the accuracy of different regions in the models, as transferring this information to docking generally helps the selection of optimal interaction and tolerance parameters. While the specific requirements posed by docking do not reduce the value of good homology modeling tools, they show that further integration may substantially improve the predicted structures of protein–protein complexes.

A key requirement for methods developers who wish to address these questions is to have access to a dataset of experimentally determined target complexes, for which a large set of sufficiently diverse templates has been identified in the PDB for the components of each complex. The task at hand for developers would then be to test different homology modeling techniques in conjunction with various docking and refinement approaches, in order to determine the factors that critically influence the prediction performance. This is the idea underlying the benchmark dataset that we present in this study. Benchmark datasets play an important role in the development of computation methods. The protein docking benchmark,[15–19] has established a community-wide standard for the development and validation of docking methods.

This benchmark contains a list of protein complexes and corresponding individual (unbound) components whose structures have all been solved by X-ray crystallography. By using the unbound structures as input, the performance of docking protocols can be assessed by their ability to reproduce the solved crystal structure of the complex. This benchmark dataset has recently been extended to include association constants and binding free energies curated from the literature, enabling developers to benchmark scoring schemes that can be used to estimate these energies from structural information.[20,21] To further support method developments in the docking field, the DOCKGROUND resource[22,23] proposes a reference dataset for benchmarking the performance of docking methods that take as input subunit models built by homology. For each of the 165 experimentally determined target complexes in this dataset, six models are generated for each of the component proteins using one state-of-art homology modeling method.[23] These models display model-to-native Cα RMSD values ranging from 1 to 6 Å, enabling methods developers to assess the impact of the accuracy of subunit models on the docking results.

The dataset and resource described in this study offers the possibility of probing more deeply into the challenges of combining homology modeling with docking calculations. The dataset is built starting from complexes in version 4.0 of the Weng docking benchmark[15–18] that are in the enzyme inhibitor (*E*) and others (*O*) categories (considering the homology modeling of antibodies and antibody–antigen docking to represent a special challenge,[24,25] antibody–antigen complexes are not included). Entries in the present dataset consist of complexes where two single chain, single domain components contribute to the interface. This restriction is motivated by the general *modus operandi* of the CASP experiment that models single domain proteins. Applying this restriction reduces the dataset to 14 enzyme inhibitor (*E*), and 21 others (*O*) type complexes. For each complex, a large set of curated template structures (ranging between ~102 and 103 structures) is provided for both the receptor and the ligand protein, spanning a wide range in sequence similarity, and as a result, also encompassing a wide range in structural similarity. This dataset therefore enables to investigate how docking results will change when the user has the freedom of selecting not only the templates for the two proteins, but importantly also the homology modeling tools. The task at hand is hence more general than simply testing how docking performs on a predetermined set of homology built models, and enables to investigate a number of additional questions, such as: What are the best methods for building homology models for docking? Are current homology modeling methods optimal if the ultimate goal is docking? If multiple templates are available, how to select the ones for the best docking results? How to use the information on model variability obtained by superimposing multiple templates? What are the critical issues in homology modeling that may reduce the accuracy of the docking solutions? Are there any side chains or loops whose conformations substantially impact the docked structures and hence require specific attention? Should existing docking methods be modified to deal with homology models, or is there a need to develop totally new, more closely integrated approaches?

To facilitate such investigations, the benchmark dataset is accessible through a web interface that provides a range of useful features for constructing a set of homolog templates according to requirements set by the user. Provisions are made to frequently update the benchmark in synchrony with the periodic updates of the PDB, such that newly available templates are added to the database and obsolete entries are removed without human intervention. Snapshots of database versions of a given date range are also made available to enable comparative studies that rely on well-defined sets of structures. In addition, template structures can be selected based on any specified range of sequence similarity, thereby simulating different levels of difficulty to assess the performance of a docking algorithm. Templates can also be selected based on their date of release, enabling to simulate a realistic scenario of template availability. The possibility of excluding templates from the list is likewise provided to training and customizing algorithms, which build homology models based on all available PDB structures.[13] Our benchmark dataset, and the resource for its update and customization should serve as a useful tool for researchers developing methods that closely integrate template-based modeling of individual proteins with predicting their association modes by docking calculations.

METHODS

Building the benchmark dataset

The starting point for building the dataset is the list of complexes from the widely used docking benchmark version 4.0.[18] From this list we select only complexes of the enzyme–inhibitor (*E*) or others (*O*) category, since homology modeling of antibodies and antibody–antigen docking is considered to represent a special challenge. To ensure that homology modeling is restricted to the level of protein domains, we further trim the dataset to contain only complexes representing domain–domain associations. The domains are defined based on the CATH classification.[26]

For each of the complexes in the resulting list, we collect for both components of the complex all solved structures of related proteins deposited in the PDB. This is performed by running BLASTP[27] against the PDB SEQRES database using a high *e* value of 10. This lenient restriction guarantees a comprehensive coverage of the templates, which can then be easily filtered further to customize the benchmark for specific purposes. PDB entries containing a hit template for both components of the

complex (and hence representing a *homologous complex*) are stored in a separate list.

While BLASTP is fast and convenient for the preliminary identification of related PDB entries, further steps are necessary to verify that the alignment does indeed cover the protein domain of interest, and that the solved structure contains all residues in the PDB SEQRES records (discrepancies between the recorded sequence in SEQRES and that in the solved structure are not uncommon). To this end we retrieve the amino acid sequence from the atomic coordinates records of each BLASTP hit, and perform a pairwise *global* alignment of this sequence with that of the corresponding component of the target complex. This alignment is performed using the Needleman–Wunsch algorithm[28] as implemented in the Needle program of the EMBOSS[29] package with the BLOSUM62 scoring matrix. These alignments are then used to calculate the percentage of sequence identity and similarity by dividing the number of identical or positive matches by the length of the target sequence covered by the pairwise alignment. Two residues are defined to be a positive match (i.e., similar) if they have a positive score in the BLOSUM62 matrix.

The identified template structures are structurally aligned to the solved structure. The Root Mean Square Deviations (RMSDs) of the Cα atoms derived from these alignments, as well as the aligned templates structures are provided for each of the targets. These provide convenient and standardized starting structures for benchmarking of both homology modeling and docking algorithms.

Customized, dynamic generation of template sets

The set of templates for each of the complexes in the benchmark can be defined by the user, by specifying a range of sequence similarity, or sequence identity values. The retrieved set can be filtered further by experimental method (X-ray, NMR, or All) or release date. In addition, the user can generate a list of PDB chains to be excluded from the template set, e.g., PDB chains with sequence similarity higher than the higher cutoff in the specified range (e.g., higher than 60% if a range of 30–60% sequence similarity was specified). This is helpful for the optimization of modeling approaches that use templates with very low global sequence identity.[13] The customized benchmark can be viewed dynamically online, or downloaded in text format.

For each complex in the benchmark, template sets are provided for both the receptor and the ligand, as well as for the complex (i.e., structures of a complex formed by a homolog of the receptor and a homolog of the ligand), when available. The latter is helpful for the calibration of modeling tools aimed at refinement starting from a homologous interface. Thus, our benchmark allows for comparing the performance of modeling tools that use *free* or *bound* homolog structures, similar to earlier docking studies that used the bound or free monomer structures to assess the influence of binding on the structure of the individual partners, and the resulting effects on modeling accuracy.

Each template list contains a detailed description of the individual template structures, including the PDB ID (with link both to the PDB repository for download of the original structure, as well as a link to the structurally aligned template). In addition, information is provided on the protein chain(s), the name of the protein, the method used for structure determination, the resolution (if solved by X-ray crystallography), % sequence identity and similarity, the RMSD to the target structure, and the raw output from the pairwise sequence alignment. The table can be sorted by clicking on the header of any column. In addition to the pairwise sequence alignments between all target–template pairs, the benchmark resource also provides a link enabling dynamic calculation and display of multiple sequence alignments (MSA) of the shown templates by running MUSCLE.[30] Prior to generating these alignments, the redundancy of the template sequences is reduced by removing identical sequences.

While any arbitrary template sets can be dynamically generated using combinations of the above-mentioned filters, we also provide a Light Version of the benchmark. This version simply returns the best template structure available for each target (the highest resolution X-ray structure).

RESULTS AND DISCUSSION

The benchmark dataset

Characteristics of the full benchmark dataset of 35 single-domain complexes are listed in Table I. Templates in this list are structures of domains with 30–60% sequence similarity to individual components of each complex (the default similarity range). The table lists the number of retrieved templates, their sequence similarity, and Cα RMSD values of the template versus target structures after structural superimpositions (mean and standard deviation values). The number of templates retrieved for individual components tends to be rather large, ranging from slightly below 100 to as many as 1000. Interestingly, while the retrieved templates display on average similar sequence similarity levels of 35–45%, they feature a much larger range in structural similarity with the target complexes (from <2 Å to >11 Å RMSD). Analysis of the relationship between the sequence similarity and the RMSD values shows a weak anticorrelation between sequence similarity and RMSD (Pearson correlation coefficient $= -0.57$, among ~62,000 pairs of target/template with 30–70% sequence similarity). So in general, lower RMSDs do correspond to templates with high sequence similarity to the targets, but the correlation

Table I
Single-Domain Template Dataset Generated at 30–60% Sequence Similarity

Complex	N. Prot1[a]	Sequence similarity		RMSD[b] (Å)		N. Prot2[a]	Sequence similarity		RMSD[b] (Å)		N. homolog complexes
		Mean	SD	Mean	SD		Mean	SD	Mean	SD	
2SIC E:I	398	42.3	9.7	4	5.5	369	35.4	2.6	9	4.6	7
1AY7 A:B	399	35.1	2.6	5.1	5.1	624	35.9	3.9	9.2	5.2	101
1R0R E:I	387	42.1	9.5	3	4.2	326	37.8	5.3	4	3.8	7
1UDI E:I	136	41.1	10.2	4.4	5.2	617	37	4.2	8.2	4.4	150
1YVB A:I	559	45.2	8.2	2.6	4.5	330	34.9	4	6.7	5.5	16
2ABZ B:E	147	34.5	5.3	10.4	7.3	237	34.5	3.4	6.6	3.7	2
2J0T A:D	288	33.8	4	6.2	5.1	171	33.7	3.6	8	5.4	21
1MAH A:F	160	46.2	4.7	1.6	2.6	275	36.9	6.1	5.2	3.2	95
2SNI E:I	404	42.2	9.8	3.7	5.2	453	39.1	5.6	6.6	4.4	27
7CEI A:B	1252	37.1	3.9	8.4	4.5	309	33.6	2.4	8.4	4.9	208
1DFJ E:I	415	37.2	6.2	7.5	5.8	94	33.5	2.6	5.9	2.9	37
1M10 A:B	339	36.2	4.4	7.6	6.3	508	35.9	3.6	4.3	3.8	19
1NW9 B:A	569	38.1	6.7	3.5	4	556	36.7	5.9	6.7	4.3	8
1PXV A:C	102	32.1	2	11.8	3.8	345	34.5	2.2	9.1	5.1	5
1FFW A:B	753	41.5	6.8	4.9	4.9	389	38	3.7	7.2	4.1	29
1GCQ B:C	696	44.4	7.6	3.1	3.5	575	34.6	3.2	8.5	5.1	228
1GPW A:B	357	34.3	3	7.5	5	197	33.1	3.3	8	6.6	55
1IBR A:B	1089	42.7	7.7	3.4	4	86	35.6	4.7	10.8	4.6	17
1HE1 C:A	850	42.8	7.6	4.5	5.5	301	35.1	2.2	11	4.7	1
1J2J A:B	964	38.2	5.4	4.1	4.1	178	46.8	6	5	2.7	29
1KAC A:B	370	37.8	8.6	4.6	4.8	1182	35.2	3.2	6.4	4.5	369
1KTZ A:B	434	35.8	5.6	7.7	5.8	213	33.5	2.8	8.8	4.9	41
1QA9 A:B	598	35.5	2.8	5.8	5	447	35.5	3.9	7.1	5.2	52
1S1Q A:B	167	34	2.8	10.3	4.5	934	39.6	6.1	6.1	4.3	2
1XD3 A:B	172	34	4.9	9.8	6	983	39.1	6.3	6.4	4.7	61
2A9K A:B	959	44.3	10.1	5.1	5.4	141	37.1	6	8.6	5.7	11
2HLE A:B	96	34.7	7.6	8.8	6.8	358	33.4	2.3	8.4	5.5	114
1AK4 A:D	187	36.8	8.7	5.7	5.3	296	33.7	2.3	10.9	5.7	69
1E96 A:B	816	42.7	7.2	4.2	5.4	154	34.2	2.6	9.4	6.5	0
1I2M A:B	1107	42	7.2	4.1	4	90	35.2	6.2	6.9	5.7	0
1LFD B:A	794	45.5	9.5	4.7	5.3	624	36.5	3.7	7.9	4.1	196
1R6Q A:C	241	35.8	3.6	8.7	6	456	36.4	4.1	8.6	3.3	25
1SYX A:B	386	35	2.5	6.6	5.1	472	36.2	3.7	5.4	4.3	4
1EFN B:A	180	33.2	2.3	6.5	4	723	45.5	7.7	4	4.3	26
1H9D A:B	440	32.9	2.3	11.1	5.1	227	33.1	1.8	9.4	5	61

[a]Numbers are based on PDB released by November 18, 2015 and may change when a different date cutoff is used.
[b]RMSD values larger than 20 are removed before the calculation of mean and SD.

coefficient is not strong enough to justify selection based on sequence identity only. The last column in Table I lists the number of structures where homologs for both the receptor and ligand coexist in the same PDB. All the additional information about the retrieved templates, the corresponding multiple sequence alignments and superimposed structures are available on the benchmark website (http://cluspro.bu.edu/benchmark). Note that all but two of the complexes in Table I have homologs that can be used for testing template based docking methods. We emphasize that such complexes are available because the benchmark set includes heavily studied complexes. In contrast, very few CAPRI targets had homologous complexes suitable for accurate template based docking, and in most cases it was necessary to build and dock models of the component proteins. However, the standard docking methods frequently failed to provide accurate results, which contributed to the motivation for developing the benchmark described in this work.

An illustrative example: the complex of caspase-9 and XIAP-BIR3 (1NW9, entry 13 in Table I)

For this target complex, we identified 569 related structures for Caspase-9 (chain B), 555 corresponding templates for XIAP-BIR3 (chain A), and 8 related complexes containing both proteins (Fig. 1). A large number of structures have been solved for proteins related to caspase-9. Most of these are other caspases (caspase-3, −7, −8, and so forth) with high structural resemblance to the target, but rather low similarity in sequence. While the global high similarity may make it an easy case for modeling the structure of caspase-9, the low conservation

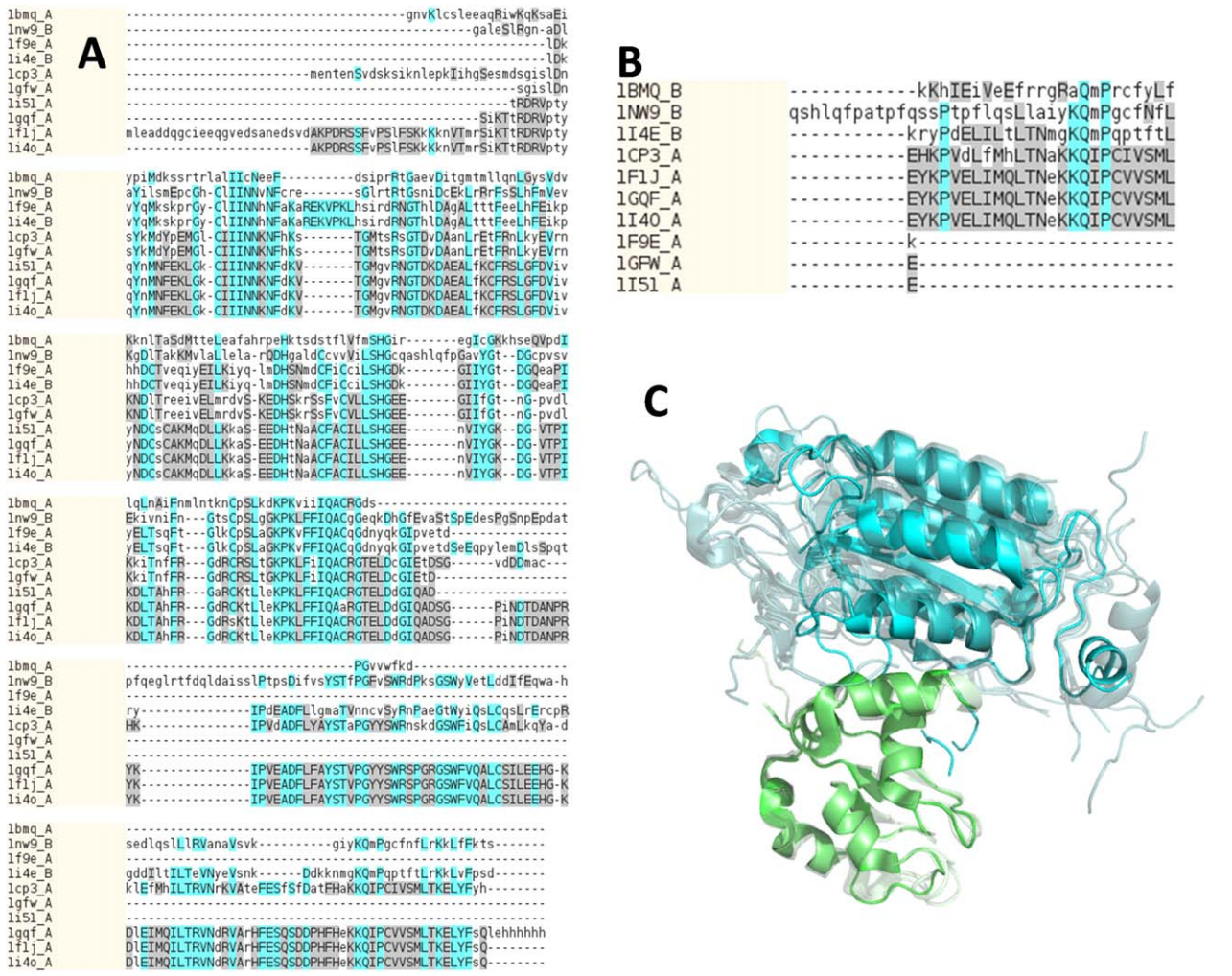

Figure 1

Complex 1NW9: interaction of Caspase-9 (chain B) with XIAP-BIR3 (chain A). **A**: Multiple sequence alignment of selected templates (around 30% sequence identity) of 1NW9_B, Caspase-9. **B**: Local sequence alignment of the same templates at the interface region of Caspase-9 and XIAP-BIR3. **C**: Target complex superimposed with each partner's template structures. Cyan and transparent blue, Caspase-9 and its templates; green and transparent green, XIAP-BIR3 and templates.

of the interface residues between the target and this selection of templates may represent a significant challenge in predicting, and accurately modeling, the interface of this complex.

CONCLUDING REMARKS

We built a resource aimed at facilitating the development and evaluation of combined homology modeling and protein–protein docking methods. The resource offers easy means for generating large, customized ensembles of potential template structures from the PDB, for the components of target complexes representing heterocomplexes of single domain proteins from the widely used docking benchmark V4. Using these template ensembles in conjunction with the corresponding targets, methods developers can investigate the influence of various factors, such as sequence and structural diversity, on the integrated process of template-based modeling and protein docking and its outcome. The resource tackles only the task of generating a curated set of templates for each target complex of the docking benchmark, and can thus serve as a starting point for the development and evaluation of any combination of template-based modeling, docking and refinement methods. Moreover, this curated set is not static over time but is dynamically updated on a weekly basis by automatically retrieving additional templates representing newly released PDB structures and eliminating those corresponding to obsolete PDB entries. In addition, new complexes added to the protein docking benchmark by the Weng lab will be manually introduced

to our resource on a regular basis, and complemented with the corresponding templates from the PDB. The benchmark dataset can be viewed online and downloaded from the web site, http://cluspro.org/benchmark.

ACKNOWLEDGMENTS

We thank Dr. Ilya Vakser for useful discussions in the course of developing the benchmark set.

REFERENCES

1. Park H, Lee H, Seok C. High-resolution protein–protein docking by global optimization: recent advances and future challenges. Curr Opin Struct Biol 2015;35:24–31.
2. Lensink MF, Wodak SJ. Docking, scoring, and affinity prediction in CAPRI. Proteins 2013;81:2082–2095.
3. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. EMBO J 1986;5:823–826.
4. Cormier CY, Mohr SE, Zuo D, Hu Y, Rolfs A, Kramer J, Taycher E, Kelley F, Fiacco M, Turnbull G, LaBaer J. Protein structure initiative material repository: an open shared public resource of structural genomics plasmids for the biological community. Nucleic Acids Res 2010;38:D743–749.
5. Webb B, Sali A. Comparative protein structure modeling using MODELLER. Current protocols in bioinformatics/editoral board. Andreas D Baxevanis 2014;47:5 6 1–5 6 32.
6. Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)—round x. Proteins 2014;82 Suppl 2:1–6.
7. Zhang Y, Skolnick J. The protein structure prediction problem could be solved using the current PDB library. Proc Natl Acad Sci USA 2005;102:1029–1034.
8. Schwede T, Sali A, Honig B, Levitt M, Berman HM, Jones D, Brenner SE, Burley SK, Das R, Dokholyan NV, Dunbrack RL, Fidelis K, Fiser A, Godzik A, Huang YJ, Humblet C, Jacobson MP, Joachimiak A, Krystek SR, Kortemme T, Kryshtafovych A, Montelione GT, Moult J, Murray D, Sanchez R, Sosnick TR, Standley DM, Stouch T, Vajda S, Vasquez M, Westbrook JD, Wilson IA. Outcome of a workshop on applications of protein models in biomedical research. Structure 2009;17:151–159.
9. Mosca R, Pons C, Fernandez-Recio J, Aloy P. Pushing structural information into the yeast interactome by high-throughput protein docking experiments. PLoS Comput Biol 2009;5:e1000490.
10. Kundrotas PJ, Zhu ZW, 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.
11. Lensink MF, Wodak SJ. Docking and scoring protein interactions: CAPRI 2009. Proteins 2010;78:3073–3084.
12. Gao M, Skolnick J. Structural space of protein–protein interfaces is degenerate, close to complete, and highly connected. Proc Natl Acad Sci USA 2010;107:22517–22522.
13. Rodrigues JPGLM, Melquiond ASJ, Karaca E, Trellet M, van Dijk M, van Zundert GCP, Schmitz C, de Vries SJ, Bordogna A, Bonati L, Kastritis PL, Bonvin AMJJ. Defining the limits of homology modeling in information-driven protein docking. Proteins Struct Funct Bioinform 2013;81:2119–2128.
14. Movshovitz-Attias D, London N, Schueler-Furman O. On the use of structural templates for high-resolution docking. Proteins 2010;78: 1939–1949.
15. Chen R, Mintseris J, Janin J, Weng Z. A protein–protein docking benchmark. Proteins 2003;52:88–91.
16. Mintseris J, Wiehe K, Pierce B, Anderson R, Chen R, Janin J, Weng Z. Protein–protein docking Benchmark 2.0: an update. Proteins 2005;60:214–216.
17. Hwang H, Pierce B, Mintseris J, Janin J, Weng Z. Protein–protein docking benchmark version 3.0. Proteins 2008;73:705–709.
18. Hwang H, Vreven T, Janin J, Weng Z. Protein–protein docking benchmark version 4.0. Proteins 2010;78:3111–3114.
19. Vreven T, Moal IH, Vangone A, Pierce BG, Kastritis PL, Torchala M, Chaleil R, Jimenez-Garcia B, Bates PA, Fernandez-Recio J, Bonvin AM, Weng Z. Updates to the integrated protein–protein interaction benchmarks: docking benchmark version 5 and affinity benchmark version 2. J Mol Biol 2015;427:3031–3041.
20. Kastritis PL, Bonvin AM. Are scoring functions in protein–protein docking ready to predict interactomes? Clues from a novel binding affinity benchmark. J Proteome Res 2010;9:2216–2225.
21. Kastritis PL, Moal IH, Hwang H, Weng Z, Bates PA, Bonvin AM, Janin J. A structure-based benchmark for protein–protein binding affinity. Protein Sci 2011;20:482–491.
22. Liu S, Gao Y, Vakser IA. DOCKGROUND protein–protein docking decoy set. Bioinformatics 2008;24:2634–2635.
23. Anishchenko I, Kundrotas PJ, Tuzikov AV, Vakser IA. Protein models docking benchmark 2. Proteins 2015;83:891–897.
24. Sivasubramanian A, Sircar A, Chaudhury S, Gray JJ. Toward high-resolution homology modeling of antibody Fv regions and application to antibody–antigen docking. Proteins 2009;74:497–514.
25. Sircar A, Gray JJ. SnugDock: paratope structural optimization during antibody–antigen docking compensates for errors in antibody homology models. PLoS Comput Biol 2010;6:e1000644.
26. Pearl FMG, Bennett CF, Bray JE, Harrison AP, Martin N, Shepherd A, Sillitoe I, Thornton J, Orengo CA. The CATH database: an extended protein family resource for structural and functional genomics. Nucleic Acids Res 2003;31:452–455.
27. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinform 2009;10:421.
28. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 1970;48:443–453.
29. Rice P, Longden I, Bleasby A. EMBOSS: The European molecular biology open software suite. Trends Genet 2000;16:276–277.
30. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004;32:1792–1797.