Electronic Journal of Biotechnology 20 (2016) 9–13

HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

Electronic Journal of Biotechnology

Research Article

The accuracy of protein structure alignment servers

Naeem Aslam [a,b], Asif Nadeem [a], Masroor Ellahi Babar [c], Muhammad Tariq Pervez [d,*], Muhammad Aslam [e], Nasir Naveed [d], Tanveer Hussain [f], Wasim Shehzad [a], Muhammad Wasim [a], Zhang Bao [g], Maryam Javed [a]

[a] *Institute of Biochemistry and Biotechnology, University of Veterinary and Animal Sciences, Lahore, Pakistan*
[b] *Department of Computer Science, NFC Institute of Engineering & Technology, Multan, Pakistan*
[c] *Department of Bioinformatics and Computational Biology, Virtual University of Pakistan*
[d] *Department of Computer Science, Virtual University of Pakistan*
[e] *Department of Computer Science and Engineering, University of Engineering and Technology, Lahore, Pakistan*
[f] *Department Molecular Biology, Virtual University of Pakistan*
[g] *Xi'an Jiaotong University, Xi'an 710061, China*

ARTICLE INFO

Article history:
Received 27 October 2015
Accepted 30 December 2015
Available online 6 February 2016

Keywords:
Comparative study
Protein structure alignment tools
Web based servers

ABSTRACT

Background: Protein structural alignment is one of the most fundamental and crucial areas of research in the domain of computational structural biology. Comparison of a protein structure with known structures helps to classify it as a new or belonging to a known group of proteins. This, in turn, is useful to determine the function of protein, its evolutionary relationship with other protein molecules and grasping principles underlying protein architecture and folding.
Results: A large number of protein structure alignment methods are available. Each protein structure alignment tool has its own strengths and weaknesses that need to be highlighted. We compared and presented results of six most popular and publically available servers for protein structure comparison. These web-based servers were compared with the respect to functionality (features provided by these servers) and accuracy (how well the structural comparison is performed). The CATH was used as a reference. The results showed that overall CE was top performer. DALI and PhyreStorm showed similar results whereas PDBeFold showed the lowest performance. In case of few secondary structural elements, CE, DALI and PhyreStorm gave 100% success rate.
Conclusion: Overall none of the structural alignment servers showed 100% success rate. Studies of overall performance, effect of mainly alpha and effect of mainly beta showed consistent performance. CE, DALI, FatCat and PhyreStorm showed more than 90% success rate.

1. Introduction

Protein structural alignment is one of the most fundamental and crucial areas of research in the domain of computational structural biology [1,2]. The true history of structural alignment begins from 1960 when Perutz et al. [3] used the approach of structural alignment and described that structures of myoglobin and hemoglobin are similar in spite of the fact that their sequences differ. Since then, structural biologists are more interested in structural similarity to detect the unknown function of a protein. Structural similarity is conserved more than sequence similarity; therefore, it can be used to trace the evolutionary history [1]. Systematic structural alignment started when Rossmann et al. [4,5,6] analyzed heme binding proteins and dehydrogenases.

Structural alignment is conducted among the known protein structures. It is based on the Euclidean distance between the residues being compared. The approaches of structural alignment are helpful in organizing and classifying known structures [7,8] and provide gold standard for sequence alignment [9,10]. A large number of protein structure alignment methods have been developed such as those described by Taylor and Orengo [11], Subbiah et al. [12] Holm and Sander [13], Holm and Park [14], Kleywegt [15], Shindyalov and Bourne [16], Kedem et al. [17], Yang and Honig [18] and Krissinel and Henrick [19].

Several comparative studies have been performed to evaluate functionality and performance of structural alignment methods. Most of these evaluation studies used CATH [7] or SCOP [20] repositories as gold standard. Sierk and Pearson [21] investigated receiver operating characteristic (ROC) curves to study the performance of various structural alignment tools to detect domains of the same topology. They used CATH as gold standard. Novotny et al. [22] evaluated functionality and performance of several structural alignment servers. They used CATH as the reference database and queried local database

[*] Corresponding author.
E-mail address: tariq_cp@hotmail.com (M.T. Pervez).
Peer review under responsibility of Pontificia Universidad Católica de Valparaíso.

http://dx.doi.org/10.1016/j.ejbt.2016.01.005
0717-3458/© 2016 Pontificia Universidad Católica de Valparaíso. Production and hosting by Elsevier B.V. All rights reserved.

of each server using seventy query structures. Leplae and Hubbard [23] used SCOP as the reference repository and deployed a server that assessed structural alignment programs through comparison of their ROC curves. Authors of structural alignment methods also evaluated the methods as part of their article such as Shindyalov and Bourne [8] evaluated CE to DALI, Gerstein and Levitt [2] compared Structural Alignments using SCOP, Shapiro and Brutlag [24] investigated FoldMiner, VAST and CE through the comparison of ROC curves.

This article presents comparative study of six structural alignment servers (SASs) as listed in Table 1. The comparison was performed using two steps. In the first step functionality of the SASs was evaluated and in the second step accuracy/performance of the SASs was evaluated. User friendliness of the interfaces and approach for presenting the results were the main functionality features compared for all SASs. To evaluate performance of SASs, several protein structures from each class of CATH were randomly selected for reporting the accuracy of each SAS.

2. Material and methods

For all the five SASs, web-based interfaces were used. The benefits of this strategy were to ensure the use of latest versions of the tools and databases with the best parameter settings according to each software's authors.

2.1. Functionality evaluation

Functionality of the SASs was investigated using user friendliness of the interfaces, presentation of results and performance/maintenance issues etc. Detail of the complete parameters used in functionality evaluation is provided in Table 2.

2.2. Performance evaluation

Identification of true positives is one of the popular approaches to investigate performance of the SASs. A true positive is the one that has similar structural composition (Class, Architecture, and Topology) to that of query structure. There are several protein structure classification systems, which can be used as standard-of-truth like FSSP[25,26], SCOP[2] and CATH[7]. In this study, CATHv4.0 was used as a benchmark. The CATH adopts both automatic and manual procedures.

CATH classifies protein structures downloaded from Protein Data Bank (PDB) into four major levels of similarity, namely, Class, Architecture, Topology and Homologous superfamily[7]. Class is the repository of structures whose secondary structure is similar (mainly α). The level of Architecture describes orientation of Secondary Structure Elements (SSEs). Topology is also called the fold family. At this level, structures are grouped based on both the overall shape and connectivity of SSEs. Homologous superfamily describes the structures that share a common ancestor and, therefore, have the similar structure and function. In this study, target protein structure is called as true positive if it has the structure (same class/architecture/topology) similar to the query protein structure.

Table 2
Functionality features of the SASs used for comparison.

1. User friendliness
   1. How much it is easy to understand/use the interface provided by SASs?
   2. Ways for provision of results (online vs. email notification/download and visualization of results)
   3. Online help/tutorials to use the server
   4. Elapsed time between request submission and result presentation
   5. Number of days to keep results on the server
   6. Provision of links to other tools/services
2. Presentation of results
   1. Provision of pairwise/multiple comparison
   2. Is 3D alignment of protein structures provided
   3. Connecting results to other services
   4. Provision of statistical significance of the results
   5. Provision of pre-calculated results
   6. Retrieval of results of a previous search
3. Performance/maintenance issues
   1. Whether SASs provide an option to a user to optimize results?
   2. Whether the server provides an option to select database?
   3. Updating frequency of databases

2.3. Test cases

A number of datasets were used to investigate the performance of SASs. Overall performance was measured by selecting structures from each of the four levels of CATH as given in Table 3.

3. Results

3.1. Functionality assessment

Although functionality evaluation was not as critical as performance investigation, however, knowledge of how easy are the interfaces to use, their features, options and how well documented/organized online help is available, can be useful in making decision which server to use. Table 4 displays the result of this part of the work. The symbol of '+' indicates high/good whereas '-' shows low/bad.

3.2. Performance evaluation

3.2.1. SAS evaluation: overall performance

Overall performance of each SAS was evaluated by counting number of true positives for all protein structures selected from each structural class (mainly-α, mainly-β, mixed α,-β and few SSEs) of CATH as elaborated in Table 3. CE, DALI, FatCat, VAST PDBeFold and PhyreStorm identified 432, 427, 414, 406. 281 and 427 true positives respectively whereas total entries in PDB (for all classes) were 456. Overall success rate of each SAS was computed as the percentage of the true positives identified (in all four classes) by an SAS. For example, overall success rate of CE 5is $(432/456 * 100) = 95\%$. It was observed that none of the SASs gave 100% success rate, however, CE and DALI and PhyreStorm outperformed other SASs as shown in Fig. 1. PDBeFold showed the poor performance.

Table 1
Protein structure alignment tools tested.

Program	URL	Database used
CE [16]	http://cl.sdsc.edu/jfatcatserver/	PDB
PhyreStorm [25]	http://www.sbg.bio.ic.ac.uk/phyrestorm/	PDB
DALI [26]	http://ekhidna.biocenter.helsinki.fi/dali_server/start	Default (PDB)
FatCat [27]	http://fatcat.burnham.org/fatcat/	PDB (90% non redundant set)
VAST [28]	http://www.ncbi.nlm.nih.gov/Structure/VAST/vastsearch.html	PDB
PDBeFold [29]	http://www.ebi.ac.uk/msd-srv/ssm/	PDB

Table 3
Test cases to test performance of SASs.

Class	CAT	No. of homologous	PDB entries	Name
Mainly-α	1.10.60	3	2DTR, 1BI2, 1DDN,1DPR, 1GY3, 1FWZ	Diphtheria Toxin Repressor; domain 2
Mainly-α	1.10.357	2	2TRT, 1A6I, 1ORK, 1QPI, 2VPR, 2VKE, 2XB5	Tetracycline Repressor; domain 2
Mainly-α	1.20.890	3	1RGS, 1NE4, 1NE6, 3IM4	cAMP-dependent Protein Kinase, Chain A
Mainly-β	2.80.10	1	1POS, 1PCP, 1E9T, 4I4S	Trefoil (Acidic Fibroblast Growth Factor, subunit A)
Mainly-β	2.50.10	1	4BCL, 2K37, 3ENI, 3VDI, 4TQ4, 4TQ6	Bacteriochlorophyll-a Protein
Mainly-β	2.115.10	2	1TL2, 3KIF, 3KIH	Tachylectin-2; Chain A
Mixed α–β	3.15.10	1	1BP1, 1EWF	Bactericidal permeability-increasing protein; domain 1
Mixed α–β	3.75.10	2	1JDW, 4JDW, 1JDX, 2JDX, 5JDW, 6JDW, 7JDW, 9JDW	L-Arginine/glycine Amidinotransferase; Chain A
Mixed α–β	3.80.30	2	1CM5, 1QHM, 1H16, 1MZO, 2F3O, 3CB8	Pyruvate-formate lyase-activating enzyme
Few SSEs	4.10.8	2	1LUC, 1LCI, 1BSL, 2D1Q, 2D1S, 2PSH, 3IEP	Luciferase; domain 5
Few SSEs	4.10.95	1	1OCC, 2CUA, 1OCZ, 1EHK, 1M56	Cytochrome C Oxidase; Chain G

3.2.2. SAS evaluation: effect of mainly-α

Seventeen protein structures from the class of mainly alpha were selected from CATH. These structures were investigated by counting their true positives obtained by each SAS. CE, DALI, FatCat, VAST, PDBeFold and PhyreStorm identified 240, 235, 232, 224, 155 and 235 true positives respectively whereas total entries with classification structure similar to query protein structures in PDB were 259. The results were consistent to the results acquired by the overall investigation study of protein structures. CE was on the top. DALI, FatCat and PhyreStorm were on the second and third positions respectively (Fig. 2). PDBeFold showed the least accuracy (Fig. 2).

3.2.3. SAS evaluation: effect of mainly-β

Performance of SASs was also evaluated using protein structures of the class of mainly beta. For this purpose, thirteen protein structures of this class from CATH were selected. CE, DALI, FatCat, VAST, PDBeFold and PhyreStorm identified 41, 40, 40, 39, 31 and 40 true positives respectively whereas total entries with the same classification structure in PDB were 44. All the five SASs showed some variation in case of this dataset. First of all, in contrast to the studies of overall performance and effect of mainly alpha all SASs gave good performance. Secondly, their performance was very close to each other. Results showed that CE was consistently on the top whereas DALI, FatCat and PhyreStorm showed the same performance. PDBeFold gave the lowest performance (Fig. 2).

3.2.4. SAS evaluation: effect of mixed alpha–beta

Effect of protein structures of the class of Alpha–Beta (mixed) on the performance of the selected SASs was evaluated using sixteen structures from CATH. True positives identified by each SAS were counted for this purpose. CE, DALI, FatCat, VAST and PDBeFold and PhyreStorm identified 25, 27, 23, 24, 21 and 27 true positives respectively whereas total entries with classification structure similar to query protein structures in PDB were 29. In contrast to other studies, CE, DALI and PhyreStorm showed different results. DALI and PhyreStorm outperformed (with success rate of 92%) other SASs and CE (88% success rate) was now on the second position. PDBeFold with 72% success rate showed the lowest performance (Fig. 2).

3.2.5. SAS evaluation: effect of few SSEs

To investigate the effect of few SSEs, twelve protein structures from the class of few SSEs were selected. Obtained true positives by CE, DALI, FatCat, VAST, PDBeFold and PhyreStorm were 125, 125, 119, 119, 75 and 125 respectively whereas total entries with same classification pattern in PDB 125 were respectively. Results showed that CE, DALI and PhyreStorm outperformed all other methods and gave 100% success rate. FatCat and VAST gave 95% success rate and were on the second position. PDBeFold showed lowest performance by giving 60% success rate only. (Fig. 2).

4. Discussion

This study was designed to measure functionality and performance of five most popular structure alignment servers. To investigate functionality of these servers, a list of parameters was designed. These parameters measured the functionality of the five servers through three major perspectives i.e. how much the servers are user friendly, what are the approaches to present results and what features are provided to the users by the servers to resolve performance issues. The results showed that all severs were user friendly, however

Table 4
Results of the functionality assessment of SASs (numbers in the first column represent the parameter number in Table 2).

	CE	DALI	FatCat	VAST	PDBeFold	PhyreStorm
User friendliness						
1 (Level of understandability)	+	+	+	-	+	+
2 (Provision of results)	Online	Email/Online	Email/Online	Online	Online/download	Download
3 (Online help)	+	+	+	+	+	+
4 (Time taken for provision of results)	Not mentioned	Not mentioned	Not mentioned	Not mentioned	Not mentioned	Less than 60 s
5 (No. of d to keep results on server)	Nil	2 w	Nil	1 w	4 h	Unlimited
6 (Links to other resources)	Yes	Yes	Yes	Yes	Yes	Yes
Presentation of results						
1 (Provision of pairwise/multiple comparison)	Yes	Yes	Yes	Yes	Yes	Yes
2 (3D alignment of protein structures)	Yes	Yes	Yes	Yes	Yes	Yes
3 (Connecting results to other services)	Yes	Yes	Yes	Yes	Yes	Yes
4 (Statistical significance)	No	Yes	Yes	No	Yes	Yes
5 (Provision of pre-calculated results)	Yes	Yes	Yes	Yes	Yes	Yes
6 (Retrieval of results of previous search)	Yes	Yes	Yes	Yes	Yes	Yes
Performance/maintenance issues						
1 (SASs result optimization provision)	Yes	No	No	No	Yes	Yes
2 (Option to select database)	No	No	Yes	No	Yes	No
3 (Frequency of updating databases)	Weekly	Weekly	Weekly	Weekly	Weekly	Weekly

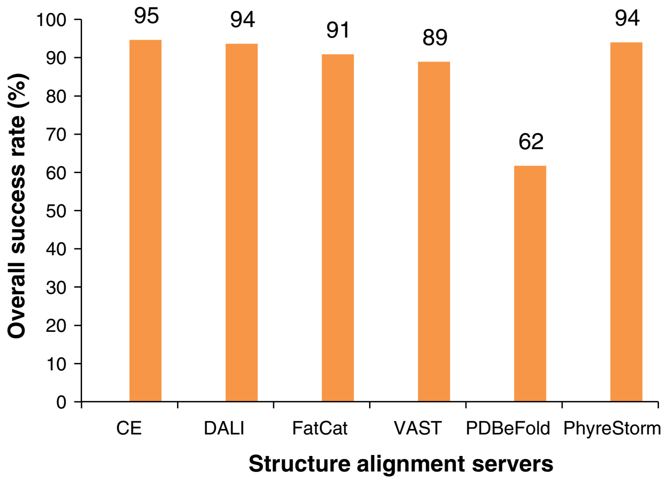

Fig. 1. Overall performance of SASs. CE was on the top of the tested SASs.

PDBeFold and VAST were more efficient. The major difference was in presenting results.

To evaluate performance of the SASs, CATH was used as the benchmark dataset. Results of the study of investigating overall performance of all SASs were similar to the results presented by the study performed by Novotny et al [22]. The results showed that CE and DALI were on top of the tested servers. Kolodny et al. [1] also showed that CE performed better than DALI. PDBeFold showed the least performance. However, none of the structure alignment servers showed 100% success rate. According to the overall performance investigation, the SASs can be divided into three classes: CE, DALI, FatCat and PhyreStorm showed more than 90% success rate, VAST gave more than 80% and PDBeFold showed less than 80% success rate. Results of the effect of mainly-α on the performance of the SASs were similar to the results obtained by the study of overall investigation of the SASs. The results showed that none of the SASs was 100% perfect. CE, DALI, FatCat and PhyreStorm gave more than 90% success rate while PDBeFold showed less than 80% success rate. Study of effect of main-β showed better performance of all SASs. All SASs showed higher success rates. PDBeFold was consistently on the bottom of list of the SASs. Performance of other four SASs was very close to each other, CE being on the top, DALI, FatCat and PhyreStorm on the second positions. Investigation of the effect of mixed alpha–beta showed different performance in contrast to the other studies. CE lost its first position which was captured by DALI and PhyreStorm. VAST was on the second position. CE and VAST gave more than 80% success rate. FatCat and PDBeFold gave less than 80% success rate. Evaluation of the effect of few SSEs showed much better performance of almost all SASs. CE, DALI and PhyreStorm gave 100% success rate. FatCat and VAST gave more than 90% success rate however PDBeFold showed very low performance. Performance of PhyreStorm was similar to DALI. The same was also claimed by authors of PhyreStorm [25].

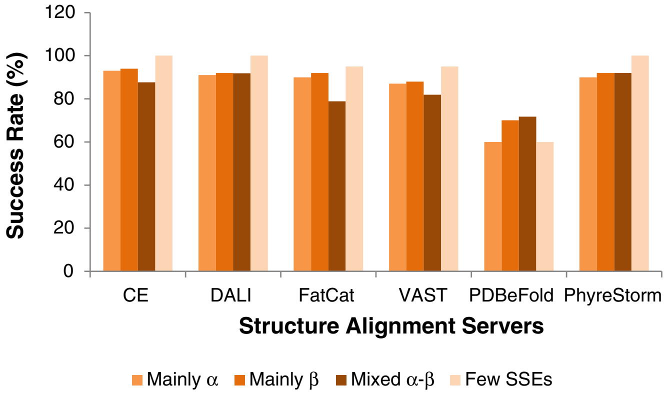

Fig. 2. Performance of SASs with respect to mainly-α, mainly-β, mixed α–β and few SSEs.

5. Conclusion

The study was aimed at the evaluation of functionality and performance of six most often used protein structure alignment servers. Functionality of all protein structure alignment servers was investigated using various parameters. Results showed that DALI, FatCat, PDBeFold and PhyreStorm showed results in more attractive and user friendly way. DALI keeps results for 2 weeks, VAST for one week and PDBeFold only for 4 h. CE and PDBeFold allow a user to optimize results. FatCat and PDBeFold provide the feature to change database. Performance of all SASs was investigated through five different ways. Overall none of the SASs showed 100% success rate. Studies of overall performance, effect of mainly alpha and effect of mainly beta showed consistent performance. CE, DALI, FatCat and PhyreStorm showed more than 90% success rate. VAST gave more than 80% while PDBeFold showed less than 80% success rate. In case of mixed alpha–beta study, CE lost the first position. DALI and PhyreStorm gave the highest performance. Study of effect of few SSEs showed 100% success rate for CE, DALI and PhyreStorm while FatCat and VAST showed similar performance.

Conflict of interest

The authors have declared that no competing interests exist.

References

[1] Kolodny R, Koehl P, Levitt M. Comprehensive evaluation of protein structure alignment methods: Scoring by geometric measures. J Mol Biol 2005;346: 1173–88.

[2] Gerstein M, Levitt M. Comprehensive assessment of automatic structural alignment against a manual standard, the Scop classification of proteins. Protein Sci 1998;7: 445–56. http://dx.doi.org/10.1002/pro.5560070226.

[3] Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, North ACT. Structure of haemoglobin: A three-dimensional Fourier synthesis at 5.5 Angstrom resolution, obtained by X-ray analysis. Nature 1960;185:416–22. http://dx.doi.org/10.1038/185416a0.

[4] Argos P, Rossmann MG. Structural comparisons of heme binding proteins. Biochemisty 1979;18:4951–60. http://dx.doi.org/10.1021/bi00589a025.

[5] Rossmann MG, Argos P. A comparison of the heme binding pocket in globins and cytochrome b_5. J Biol Chem 1975;250:7525–32.

[6] Rossmann MG, Liljas A, Branden CI, Banaszak LJ. Evolutionary and structural relationships among dehydrogenases. Enzymes 1975;11:61–102. http://dx.doi.org/10.1016/S1874-6047(08)60210-3.

[7] Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM. CATH—A hierarchic classification of protein domain structures. Structure 1997;5:1093–108. http://dx.doi.org/10.1016/S0969-2126(97)00260-8.

[8] Shindyalov IN, Bourne PE. An alternative view of protein fold space. Proteins Struct Funct Genet 2000;38:247–60. http://dx.doi.org/10.1002/(SICI)1097-0134(20000215)38:3%3C247::AID-PROT2%3E3.0.CO;2-T.

[9] Thompson JD, Plewniak F, Poch O. BAliBASE: A benchmark alignment database for the evaluation of multiple alignment programs. Bioinformatics 1999;15:87–8. http://dx.doi.org/10.1093/bioinformatics/15.1.87.

[10] Sauder JM, Arthur JW, Dunbrack RL. Large scale comparison of protein sequence alignment algorithms with structure alignments. Proteins Struct Funct Genet 2000;40:6–22. http://dx.doi.org/10.1002/(SICI)1097-0134(20000701)40:1%3C6::AID-PROT30%3E3.0.CO;2-7.

[11] Taylor WR, Orengo CA. Protein structure alignment. J Mol Biol 1989;208:1–22. http://dx.doi.org/10.1016/0022-2836(89)90084-3.

[12] Subbiah S, Laurents DV, Levitt M. Structural similarity of DNA-binding domains of bacteriophage repressors and the globin core. Curr Biol 1993;3:141–8. http://dx.doi.org/10.1016/0960-9822(93)90255-M.

[13] Holm L, Sander C. Protein structure comparison by alignment of distance matrices. J Mol Biol 1993;233:123–38. http://dx.doi.org/10.1006/jmbi.1993.1489.

[14] Holm L, Park J. DaliLite workbench for protein structure comparison. Bioinformatics 2000;16:566–7. http://dx.doi.org/10.1093/bioinformatics/16.6.566.

[15] Kleywegt GJ. Use of non-crystallographic symmetry in protein structure refinement. Acta Crystallogr, Sect D: Biol Crystallogr 1996;52:842–57. http://dx.doi.org/10.1107/S0907444995016477.

[16] Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng Des Sel 1998;11:739–47. http://dx.doi.org/10.1093/protein/11.9.739.

[17] Kedem K, Chew LP, Elber R. Unit-vector RMS (URMS) as a tool to analyze molecular dynamics trajectories. Proteins Struct Funct Genet 1999;37:554–64. http://dx.doi.org/10.1002/(SICI)1097-0134(19991201)37:4%3C554::AID-PROT6%3E3.0.CO;2-1.

[18] Yang AS, Honig B. An integrated approach to the analysis and modeling of protein sequences and structures. I. Protein structural alignment and a quantitative measure for protein structural distance. J Mol Biol 2000;301:665–78. http://dx.doi.org/10.1006/jmbi.2000.3973.

[19] Krissinel E, Henrick K. Protein structure comparison in 3D based on secondary structure matching (SSM) followed by C-alpha alignment, scored by a new structural similarity function. In: Kungl AJ, Kungl PJ, editors. Proceedings of the Fifth International Conference on Molecular Structural Biology; Vienna; September 3–7; 2003.

[20] Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: A structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–40. http://dx.doi.org/10.1016/S0022-2836(05)80134-2.

[21] Sierk ML, Pearson WR. Sensitivity and selectivity in protein structure comparison. Protein Sci 2004;13:773–85. http://dx.doi.org/10.1110/ps.03328504.

[22] Novotny M, Madsen D, Kleywegt GJ. Evaluation of protein-fold-comparison servers. Proteins Struct Funct Bioinf 2004;54:260–70. http://dx.doi.org/10.1002/prot.10553.

[23] Leplae R, Hubbard TJP. MaxBench: Evaluation of sequence and structure comparison methods. Bioinformatics 2002;18:494–5. http://dx.doi.org/10.1093/bioinformatics/18.3.494.

[24] Shapiro J, Brutlag D. FoldMiner: Structural motif discovery using an improved superposition algorithm. Protein Sci 2004;13:278–94. http://dx.doi.org/10.1110/ps.03239404.

[25] Mezulis S, Sternberg MJE, Kelley LA. PhyreStorm: A web server for fast structural searches against the PDB. J Mol Biol 2015. http://dx.doi.org/10.1016/j.jmb.2015.10.017 (in press).

[26] Holm L, Kääriäinen S, Rosenström P, Schenkel A. Searching protein structure databases with DaliLite v.3. Bioinformatics 2008;24:2780–1. http://dx.doi.org/10.1093/bioinformatics/btn507.

[27] Ye Y, Godzik A. Flexible structure alignment by chaining aligned fragment pairs allowing twists. Bioinformatics 2003;19:246–55. http://dx.doi.org/10.1093/bioinformatics/btg1086.

[28] Gibrat JF, Madej T, Bryant SH. Surprising similarities in structure comparison. Curr Opin Struct Biol 1996;6:377–85. http://dx.doi.org/10.1016/S0959-440X(96)80058-3.

[29] Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr, Sect D: Biol Crystallogr 2004;60:2256–68. http://dx.doi.org/10.1107/S0907444904026460.