Evaluation of Protein Fold Comparison Servers

Marian Novotny[1], **Dennis Madsen,**[2] **and Gerard J. Kleywegt**[1*]
[1]*Department of Cell and Molecular Biology, Uppsala University, Biomedical Centre, Uppsala, Sweden*
[2]*Novo Nordisk Park, Department of Scientific Computing, Måløv, Denmark*

ABSTRACT **When a new protein structure has been determined, comparison with the database of known structures enables classification of its fold as new or belonging to a known class of proteins. This in turn may provide clues about the function of the protein. A large number of fold comparison programs have been developed, but they have never been subjected to a comprehensive and critical comparative analysis. Here we describe an evaluation of 11 publicly available, Web-based servers for automatic fold comparison. Both their functionality (e.g., user interface, presentation, and annotation of results) and their performance (i.e., how well established structural similarities are recognized) were assessed. The servers were subjected to a battery of performance tests covering a broad spectrum of folds as well as special cases, such as multidomain proteins, Cα-only models, new folds, and NMR-based models. The CATH structural classification system was used as a reference. These tests revealed the strong and weak sides of each server. On the whole, CE, DALI, MATRAS, and VAST showed the best performance, but none of the servers achieved a 100% success rate. Where no structurally similar proteins are found by any individual server, it is recommended to try one or two other servers before any conclusions concerning the novelty of a fold are put on paper. Proteins 2004;54:260–270.**

Key words: protein structure; protein fold; fold classification; structure comparison; structure superposition; structural bioinformatics

INTRODUCTION

In the past 10 years, major progress has been made in both the speed and the quality of the protein structure determination process. At present, about 65 structures are released by the Protein Data Bank (PDB)[1,2] every week. The growth of the PDB offers unprecedented possibilities for data mining. However, that same growth has made it impossible for human beings (with the possible exception of Alexei Murzin) to remember all the folds of all the proteins. In response to this problem, several groups have developed computer programs to carry out (semi-) automatic structure comparison (reviewed in Refs. 3–6).

Structural comparisons are of interest for several reasons. They may provide clues about the function of a protein based on structural similarity, even in the absence of detectable sequence similarity. Furthermore, comparisons may reveal unexpected evolutionary relationships and may contribute to our collective understanding of the principles underlying protein architecture and folding. Finally, such comparisons aid in the creation of multiple-sequence alignments, in the delineation of structural cores for sequence-to-structure alignment, and in the evaluation of structure predictions.

As the structural database grows, there is an urgent need for good tools for structure comparison that operate without a great deal of human intervention. Unfortunately, structure comparison is anything but a straightforward problem. There are many ways to superimpose two structures, and no optimal superposition exists if the proteins are not identical or extremely similar in sequence and structure. Apart from the choice of algorithm, there are usually subjective choices (e.g., parameter settings) to be made in the process as well. Several partially or fully automatic programs for structure comparison have been developed in recent years. They are based on a variety of different methods: alignment of secondary structure elements (SSEs),[7,8] distance-measure matrices,[9] environmental profiles,[10] or a combination of these methods.[11] The aim of all these methods is to facilitate and accelerate structure comparison so that it can be accomplished in reasonable times (even when a structure is screened against an entire database of structures) and with a minimum of user input. To this end, many of these programs have been made available to the community through Web-based interfaces. However, these programs have never been subjected to a critical evaluation and comparison of their functionality and performance. The only comparison that we are aware of[11] involved a small subset of available programs and was conducted in-house (i.e., not using any Web interfaces) and with only a few test cases.

The objective of this work was to evaluate the functionality and the performance of Web-based programs for structure comparison. With respect to functionality, we assessed the user interface, the presentation of results, and

Grant sponsor: Ministry of Education, Youth, and Sports of the Czech Republic; Grant number: 113100003; Grant sponsor: Socrates II/Erasmus Programme of the European Union; Grant sponsor: Knut and Alice Wallenberg Foundation.

*Correspondence to: Gerard J. Kleywegt, Department of Cell and Molecular Biology, Uppsala University, Biomedical Centre, Box 596, SE-751 24, Uppsala, Sweden. E-mail: gerard@xray.bmc.uu.se

Received 18 March 2003; Accepted 11 June 2003

TABLE I. Fold Comparison Programs Tested

Program	URL	Database used	References
CE	http://cl.sdsc.edu/	Structure representatives	[11]
DALI	http://www2.ebi.ac.uk/dali	Default	[9,42]
DEJAVU	http://portray.bmc.uu.se/dejavu	<100% Sequence identity	[8,12,43]
LOCK	http://gene.stanford.edu/LOCK	Largest database (582 proteins)	[14]
MATRAS	http://bongo.lab.nig.ac.jp/~takawaba/cgi-bin/Matras/LibMatForm.pl.cgi	<40% Sequence identity	[16]
PRIDE	http://hydra.icgeb.trieste.it/pride	CATH database	[17,18]
SSM	http://www.ebi.ac.uk/msd-srv/ssm	PDB or SCOP	—
TOP	http://bioinfo1.mbfys.lu.se/TOP	4220 structures in SCOP	[19]
TOPS	http://tops.ebi.ac.uk/tops/compare1.html	CATH and SCOP databases	[20,44]
TOPSCAN	http://www.rubic.rdg.ac.uk/~andrew/bioinf.org/topscan	Probe database	[22]
VAST	http://www.ncbi.nlm.nih.gov/Structure/VAST/vastsearch.html	Nonredundant subset of PDB	[7]

TABLE II. Fold Comparison Programs That Could Not Be Tested in This Study

Program	URL	Method	Reference
CATH server	http://www.biochem.ucl.ac.uk/bsm/cath/server	Double dynamic programming	[45]
SHEBA	http://lily.nci.nih.gov/~jung/index.html	Hierarchical alignment with profiles	[10]
STRUCTAL	http://bioinfo.mbb.yale.edu/align/server.cgi	Double dynamic programming	[46]
3D SEARCH	http://gene.stanford.edu/3dSearch	Hash table alignment	—
GRATH	http://www.biochem.ucl.ac.uk/cgi-bin/cath/Grath.pl	Graph matching	[25]

several implementation issues. To assess the performance, test structures that covered a broad spectrum of folds were selected. In addition, a number of potentially difficult cases were used, including multidomain proteins, Cα-only structures, NMR-based models, and a number of nontrivial structural similarities.

MATERIALS AND METHODS

Programs

The aim of this work was to assess and compare the functionality and performance of the programs listed in Table I. Although most of these programs can be installed on local computers, doing so tends to require a lot of work, such as installation and maintenance of software and databases. Hence, we chose to test only programs with a Web-based interface that were available through the Internet between October 2001 and June 2003. An additional benefit of this approach is that one is presumably using the most recent version of the software and the database, as well as the best default parameter settings according to each program's authors. We identified several other programs (Table II), but their Web-based servers were either not functioning during our study, or, as in the case of GRATH, there were technical reasons that prevented their inclusion in our study. To access the servers, we used standard Web browsers (Internet Explorer 5.5 and Netscape Communicator 4.5) on PCs under Windows NT and Macintoshes under OS X or, in a few cases, on SGI workstations under Irix 6.5.

In the following, the principles and algorithms of the evaluated programs are briefly described. Here, and throughout, the programs will be listed in alphabetical order.

CE (Combinatorial Extension of the optimal pathway)[11] attempts to find the best possible alignment of two structures by combinatorial extension of the path of aligned fragment pairs that satisfy certain criteria regarding structural similarity. The evaluation of structural similarity is based on interresidue distances and the root-mean-square distance (RMSD) of the matched atoms after rigid body superpositioning. Gaps are allowed, but the maximum size of a gap is restricted. A Z score is used as the significance measure, and it is calculated for the best alignment of two structures. This is done by evaluating the probability of finding an alignment of the same length when comparing two random structures.

DALI[9] calculates residue–residue distance matrices from three-dimensional (3D) coordinates of proteins. The distance matrices are first divided into hexapeptide fragments to simplify the alignment task in later stages. DALI attempts to find common local patterns in two fragments of the distance matrices. Such fragments are paired, stored, and combined into larger overlapping segments. The alignment of fragments within the segment is further optimized by a Monte Carlo method, which is not guaranteed to converge to the globally optimal solution. Therefore, several alignments are optimized in parallel, which yields the best, second best, and so on solutions. The method is fully automatic and allows sequence gaps of any length, reversal of chain direction, and free topological connectivity of aligned segments.

DEJAVU[8,12] uses SSEs, represented as vectors, to detect structural similarity between the query and database structures. The structural similarity is defined with regard to the number of SSEs, their lengths, mutual distances and angles, and, optionally, connectivity and directionality. Results from the SSE-based search are first refined by RMSD minimization (based on Cα atoms) and then by a dynamic programming procedure. The hits are sorted according to their significance, expressed as a *P* value and a *Z* score as defined by Levitt and Gerstein.[13]

LOCK[14] is a hierarchical structure-superposition method that attempts to minimize the RMSD of two structures at three levels. The RMSD is minimized by the "absolute orientation of corresponding points" algorithm.[15] The starting superposition is obtained by aligning SSEs, represented as vectors, with use of dynamic programming. The RMSD is minimized for corresponding Cα atoms in the second step. In the third step, the core of the structure is defined and an RMSD minimization is once more applied to this core.

MATRAS[16] uses a Markov transition model of evolution to measure protein structure similarity. The similarity score of two structures i and j is calculated as log $P(j \rightarrow i)/P(i)$, where $P(j \rightarrow i)$ is the probability of a structural change from j to i during evolution and $P(i)$ is the probability that structures appear by random processes. The Markov transition model has been created in a similar fashion as Dayhoff's amino acid substitution matrices. Three types of structural similarity scores are used: an environmental score (a combination of local structure and solvent accessibility), a residue–residue distance score, and an SSE score. The program uses a hierarchical alignment algorithm. The first alignment is obtained by comparing SSEs using a branch-and-bound method. This initial alignment is further improved by using more detailed environmental and residue–residue distance scores. The significance of the results is expressed as a *Z* score.

PRIDE[17,18] describes protein structures by a set of distributions of $C\alpha(i)$-$C\alpha(i+n)$ distances, where i is a residue number in the protein chain and n is an integer ranging from 3 to 30. Structural similarity is evaluated as the similarity of the 28 distance distributions and is expressed as a score that varies between 0 and 1 (where a value of 1 indicates identical structures).

SSM represents SSEs as vectors that form a 3D graph and uses a rapid graph-matching algorithm to match the SSE graphs of query and database structures. Subsequently, the Cα atoms of matched SSEs plus some nearby atoms are superimposed. A target function that depends on the number of matched atoms and their RMSD is minimized. The significance of the hits is evaluated with a *P* value and a *Z* score.

TOP[19] aligns subsets of SSEs in two proteins, and their similarity is measured by the angles between aligned SSEs, the distances between matched SSEs and the RMSD of the superimposed coordinates. If the number of matched SSEs for a structure exceeds a certain fraction of all its SSEs, TOP considers these two structures to be structurally similar. It then proceeds with the second comparison stage, which entails detecting the matching residues.

TOPS[20] compares topology diagrams of protein structures instead of the more conventional approaches involving SSEs and Cα coordinates or distances. A TOPS diagram[21] is a simplified representation of protein structure as a string of SSEs that preserves connectivity and directionality and also contains information about hydrogen bonds and chirality. In pictorial representations, helices are represented as circles and strands as triangles. The query structure is converted into a so-called TOPS pattern, which is basically a TOPS diagram in which gaps for insertion of SSEs are allowed. The TOPS pattern is compared to a database of TOPS diagrams, which is a fast way to do a symbolic structure comparison.

TOPSCAN[22] is a rapid, but approximate, method for protein structure comparison. It was developed as a prescreen method for the more sophisticated, but slow, SSAP algorithm.[23] TOPSCAN translates structure information into topology strings. Topology strings are defined at two levels: the primary topology represents the state of the secondary structure for each residue (helix or strand), and the secondary topology contains additional information about length, direction, proximity, and accessibility of SSEs. The topology strings are aligned by a global dynamic programming algorithm, and a similarity score is calculated.

VAST[7] uses a graph theory-based approach[24] to align SSEs. Pairs of SSEs (one from each structure) are represented as nodes if they are of the same type. These nodes are connected by an edge if the angles and the distances between corresponding SSEs from the two proteins do not violate certain constraints. The graph shows the correspondence of SSE pairs based on type, relative orientation, and connectivity. The significance of the results is indicated by a *P* value, which is defined as the probability of obtaining the results by chance alone, multiplied by the number of possible, alternative substructure alignments for the given pair of structures.

GRATH[25] is a graph-based structure comparison program. Protein structures are described as graphs composed of nodes and edges, where nodes represent the SSEs and edges correspond to the geometrical relationships (chirality, distances, and angles) between the SSEs. GRATH could not be used in this study because it is intimately coupled to CATH which is the classification standard used in this work. Moreover, GRATH always returns just a single hit for every topology class in CATH, which makes it unsuitable for the performance tests that we have conducted (see below).

Functionality Assessment

Functionality assessment involved a set of criteria that could be divided into three groups: operation, presentation of results, and implementation issues, as listed in Table III. This table was used as a checklist, and the various items were assessed or investigated by using a simple structure comparison test case (PDB entry 1cbs).

Performance Assessment

To assess the performance of fold comparison programs, one has to decide how to identify true positives (i.e., structures that are known to have a fold similar to a query structure). A number of structure classification systems that can be used as benchmarks are available, including SCOP[26,27] CATH,[28] and FSSP.[29,30] In this work, CATH was taken as the "standard-of-truth" (with a few caveats). CATH is based on an automatic procedure, which is

TABLE III. Criteria Considered in the Functionality Assessment of the Fold Comparison Servers

1. Operation
   1.1. User interface, on-line help, choice of parameters and databases
   1.2. Interactive vs e-mail results or notification
   1.3. Documentation, access to literature references
   1.4. Confidentiality
   1.5. Elapsed time between submission and results
   1.6. How long are results kept on the server?
2. Presentation of results
   2.1. What information is reported for the hits?
   2.2. Which measures of goodness-of-fit and statistical significance are provided?
   2.3. Is a structure-based sequence alignment produced?
   2.4. Can PDB files with superimposed structures be retrieved?
   2.5. Is there an interface with visualisation software?
   2.6. Are the results hyperlinked to other Web-based resources (e.g., PDB)?
3. Implementation issues
   3.1. Are there parameters the user can set to optimise the results?
   3.2. Is there a choice of databases? How much of the structural universe do they cover?
   3.3. How often are the databases updated?

corrected by manual intervention. Therefore, CATH is midway between the manually curated classification of SCOP and the automatic classification of FSSP. Moreover, the authors of SCOP use literature information for structural classification, whereas the structure comparison programs can only use information that is contained within the structures themselves. The FSSP classification is generated automatically by DALI, which is one of the programs included in our study. Hence, using it as a reference would have introduced a strong bias in favor of DALI. However, in practice, we used CATH (version 2.0 and 2.4) to identify sets of true positives. Subsequently, we consulted SCOP and FSSP and performed BLAST[31] searches against the sequences in the PDB to check the true positives and to identify any "invisible" structures. The latter are structures that are considered to have the same fold as a query structure but that have not yet been incorporated into CATH. Such structures were not counted as either positives or negatives. This procedure avoids many of the pitfalls involved in the use of a single classification system.[32,33]

CATH is a hierarchical system that distinguishes four main levels of structural similarity: Class, Architecture, Topology, and Homologous superfamily.[28] Class is the most general level of structural similarity, and structures within the same class have similar secondary structure content (e.g., mainly α). Architecture is described by the orientation of SSEs, regardless of their connectivity, and Topology is derived from connections and numbers of SSEs. A homologous superfamily contains proteins with highly similar structure and function. In this work, we adopted the convention that a database structure retrieved by a server was called a true positive if it had the same Class/Architecture/Topology classification within CATH as the query structure without being the query structure itself. The query structure itself was never counted as a true positive because in real-life situations one cannot expect that a newly determined structure is already in the database. Moreover, the problem of finding an identical structure is far too easy because the similarity between query and hit is perfect (i.e., it is both exact and complete).

Simple counting of true positives was not feasible because of the many differences that exist between the various servers (e.g., different databases and scoring systems). Most of the programs' databases do not contain all entries of the PDB, but rather some subset of structure representatives. Another complication lies in the different scoring systems and in the fact that some programs do not specify which database hits are deemed significant and which are accidental.

To circumvent these problems, we used a simple binary scoring system: at least one true positive either was or was not found in the list of significant hits (for servers that did not indicate the significance of the hits, we examined up to 100 hits). We worked with groups of proteins that share the same topology (T-level) in the CATH classification. For example, to assess the performance on mainly-β proteins, one of the cases we selected was the cyclophilin fold (CATH classification 2.40.100) with nine members: 1a33, 1awg, 1cyn, 1dyw, 1ihg, 1lop, 1qng, 1qoi, and 2rmc. Each of these structures was submitted to all the servers, and we checked if any of the structures, other than the query structure itself, were found as true positives. For each server, we then counted in how many of the nine cases such a result was obtained. A perfect result would be a score of 9 out of 9. However, some programs used small databases that contain perhaps only one representative for the cyclophilin family and which was also a member of our set of nine proteins. In such a case, the maximum attainable score was 8 out of 9. It should be noted here that, in essence, we tested the "real-life" performance of the combination of software, its default parameter settings (as defined by the program's authors), and the most comprehensive database available for that program. It may well be that a program that fails to produce a true positive could improve its performance if parameters were tweaked or its database extended, but this is not an issue that we could address here.

In addition to the number of cases in which a true positive other than the query structure was retrieved, we also recorded if the structure itself was retrieved (this is not considered a true positive result because the two structures are identical), what the rank of the first false positive result was, and how much time elapsed between submission of the query and receipt of the results. To exemplify this, Table IV lists the results obtained for the cyclophilin test case using the TOP server. Clearly, the TOP database contains seven of the nine proteins, and in all cases TOP succeeds in producing a true positive result. Table V shows a summary of the results obtained for the cyclophilin case using all servers. The table reveals that many of the programs attain a perfect score (9 out of 9).

TABLE IV. Results of the Cyclophilin Test Case for the TOP Server (see text)

Query	Found itself?	Other true positives?	Rank of first false positive	Elapsed time (min)
1awg	yes	yes (6 of them)	8	9
1a33	yes	yes (6)	8	9
1cyn	yes	yes (6)	none found	9
1qoi	yes	yes (6)	8	7
1lop	yes	yes (1)	none found	10
1qng	no	yes (5)	6	8
2rmc	yes	yes (6)	8	9
1dyw	yes	yes (6)	8	9
1ihg	no	yes (3)	none found	8

LOCK uses a small (and out-of-date) database of <600 proteins that obviously contains only one cyclophilin structure; therefore, it scores 8 out of 9.

Test Cases

To assess the performance of the servers, we selected several test cases. To get an impression of the overall performance, structural families were chosen from each of the four main structural classes defined in CATH (see Table VI). The class with few SSEs was also included to test the sensitivity of the programs in cases with low secondary structure contents. From this class, the kringle domain was selected, even though it is actually classified as mainly-β in CATH, despite its low secondary structure content (<15%).

In a second experiment, we used 10 nontrivial structural similarities that were taken from Fischer et al.[34] and that also have been studied by Shindyalov and Bourne.[11] In addition, the structure of ribosome antiassociation factor IF6 (PDB code 1g61) was included in this experiment. This protein was initially described as having a new fold[35] because at the time of publication, the authors failed to find any structural similarity in the PDB using DALI. However, it turned out that its fold is similar to that of the amidino-transferases.[36,37] Thus, the 11 difficult cases were 1bgeB (similarity with 2gmfA), 1cewI (1molA), 1cid (2rhe), 1crl (1ede), 1fxiA (1ubq), 1ten (3hhrB), 1tie (4fgf), 2azaA (1paz), 2sim (1nsbA), 3hlaB (2rhe), and 1g61 (1jdw). In these tests, success was defined as the ability of a program to retrieve the target structure or a close relative (mutant, complex, or homologue). We also conducted a test with the target and query structure swapped to investigate if the programs were symmetric in their ability to retrieve hits and in the details of the structural alignment (number of aligned residues and RMSD). For this test, we used 1molA (1cewI) for most servers. For TOPSCAN, we used 1nsbA (2sim) and for DEJAVU and SSM, we used 1ubq (1fxiA) because these servers failed to retrieve 1molA when queried with 1cewI. Because PRIDE failed for all 11 difficult cases, we used 1dyw (1lop) to assess the symmetry of its results.

In a third and final experiment, some special cases were used to test various aspects of the programs, including their sensitivity to small variations in a structure, and their ability to cope with multidomain proteins, to handle Cα-only models, and to identify new folds.

Many proteins contain more than one domain, and we wanted to investigate how well the programs handled such multidomain proteins. Two members of the Src protein kinase family (Src kinase, 2src, and Hck kinase, 2hckA) were chosen. Their structures contain four distinct domains of different lengths (62–175 amino acids) and the proteins are about 550 amino acids long. A perfect result in this case would be if a program first finds all proteins that contain all four domains and then those that contain three of them, then two of them and, finally, proteins with only one shared domain.

To investigate if and how program performance was affected by small variations in the query structures, we compared the results obtained with the average, energy-minimized NMR model of the glucocorticoid receptor's DNA-binding domain (PDB code 1gdc) and those obtained with five "raw" NMR models of the same domain (2gda). For each of the 24 models in 2gda, we used LSQMAN[38] to calculate the RMS value of the RMSD on Cα atoms to each of the other 23 models. Model 12 had the lowest RMS (RMSD) value (0.89 Å) and, hence, was considered to be the central model.[39] Conversely, model 20 had the highest value (1.48 Å) and was, therefore, the "most atypical" of the ensemble. Three additional models (2, 7, and 18) were selected at random.

Some structural biologists are hesitant to submit their newly determined structures over the Internet to a server whose level of security and confidentiality may not be known. A partial solution to this problem is to submit only the Cα coordinates (possibly with all residue types changed to alanine as well). We tested the ability of the various programs to deal with Cα-only models by submitting only the Cα atoms of cyclophilin structure 1a33 and comparing the results with those obtained for the intact structure.

One of the most frequently posed questions after a new structure has been determined is whether it has a new fold or not. Therefore, it is of interest to check if the programs, when confronted with a new fold, are able to recognize this fact by failing to find significant hits in their database. To investigate this, we selected a number of proteins classified as singletons in CATH (i.e., without any structural neighbors at the topology level). We selected four test cases, one from each of the main classes in the CATH classification: transcription factor Stat-4 (1bgf; CATH code 1.10.532.10; mainly α), phosphomannose isomerase (1pmi; CATH code 2.30.41.10; mainly β), colicin Ia (1cii; CATH code 3.30.305.10; mixed α–β), and pyruvyl-dependent histidine decarboxylase (1pya; CATH code 4.10.510.10; few SSEs).

RESULTS

Functionality Assessment

The functionality assessment was to some extent subjective and obviously not as important as the performance tests. Nevertheless, the available features, options, and databases can be important criteria when deciding which server to use. The result of the assessment is summarized

TABLE V. Results of the Cyclophilin Test Case for All Servers (see text)

Program	Found itself	Other true positives	Rank of first false positive	Elapsed time per query (min)
CE	9	9	either none found, or rank 4	(used precomputed results)
DALI	2	9	either none found, or rank 3	3–9
DEJAVU	9	9	none found	5–55
LOCK	1	8	2	8–15
MATRAS	3	9	3	10–45
PRIDE	6	8	1–8	<1
SSM	6	8	none found	2–3
TOP	7	9	either none found, or rank 6–8	7–10
TOPS	5	9	2–6	2–5
TOPSCAN	5	9	4–6	<1
VAST	2	9	either none found, or rank 3	5–25

TABLE VI. Details of the Fold Families Used to Assess the Performance of the Servers

Class	CAT[a]	Nr H[b]	PDB entries	Name
Mainly-α	1.10.40	2	1rlr 1yfm 1fur 1auw 1jsw 1hyl 1i0a	ribonucleotide reductase protein R1, domain 1
Mainly-α	1.10.164	3	1aq6 1c3u 1fez 1jud 1zrn	L-2-haloacid dehalogenase, domain 2
Mainly-α	1.25.30	3	1b3u 1bk6 1gcj 1ial 1ibr 1qbk 2bct	armadillo repeat
Mainly-β	2.30.110	2	1ci0 1dnl 1eje 1i0r	PNP oxidase
Mainly-β	2.40.100	1	1a33 1awg 1cyn 1dyw 1ihg 1lop 1qng 1qoi 2rmc	cyclophilin
Mainly-β	2.100.10	3	1c3k 1ciy 1jac 1jot 1dlc 1vmo	vitelline membrane outer layer protein I, subunit A
Mixed α-β	3.10.70	2	1bkf 1grj 1pbk 1rot 1yat	GreA transcript cleavage factor, domain 2
Mixed α-β	3.40.91	3	1bhm 1cfr 1d2i 1fok	restriction endonuclease
Mixed α-β	3.70.10	3	1axc 1b77 1czd 1dml 1ge8 1plq	proliferating cell nuclear antigen
Few SSEs	2.40.20	1	1b2i 1cea 1kdu 1kiv 1krn 1pk4 1pml 5hgp	plasminogen kringle 4

[a] Class-Architecture-Topology code according to CATH.
[b] Number of Homologous superfamilies (H level in CATH) of the Topology level.

TABLE VII. Results of the Functionality Assessment[†]

Program	Operation[a]	Presentation[a]	Implementation[a]
CE	−	0	+
DALI	+	0	0
DEJAVU	+	+	+
LOCK	0	+	−
MATRAS	+	0	0
PRIDE	+	0	−
SSM	+	0	+
TOP	+	0	+
TOPS	+	−	0
TOPSCAN	0	−	−
VAST	0	+	0

[†] Refer to Table III for the criteria considered for each of the three categories in this table.
[a] Results are scored as "−" (poor), "0" (average), and "+" (better than average).

in Table VII. The complete set of results is available on our Web site (http://xray.bmc.uu.se/~marian/servers/index.htm).

Representatives of Main Structural Classes

The performance of the programs was tested with folds from each of the four main structural classes in CATH (mainly-α, mainly-β, mixed α–β, and few SSEs), as described in Materials and Methods. The results are summarized in Table VIII, and the complete set of results is available on our Web site (http://xray.bmc.uu.se/~marian/servers/index.htm).

Difficult Cases

For the 11 difficult cases, none of the programs attained a 100% success rate. The best performance was obtained with CE and DALI (10 out of 11 successes), and LOCK, MATRAS, and VAST (9 successes). TOPSCAN (6), DEJAVU (6), TOP (5), and TOPS (5) were only successful in about half of the cases, whereas SSM (1) and PRIDE (0) failed in almost all cases.

Ten of the difficult cases were used a few years ago in a comparison of three of the programs (CE, DALI, and VAST).[11] A comparison of the results obtained then and now is not without surprises. The results for CE are similar in both investigations, but the DALI and the VAST results differ. DALI did find both previously missed similarities (2azaA/1paz and 1fxiA/1ubq), but now missed the similarity between 1bgeB and 2gmfA. VAST was still unable to find the pair 3hlaB/2rhe and now also failed for the pair 1cid/2rhe. On the other hand, it now did succeed in finding the similarity between 1fxiA and 1ubq.

In the case of ribosome factor IF6, eight of the programs found amidino-transferases as hits, including DALI, which had failed to do so originally. (Some of its parameters were changed since then; see Ref. 37.) Moreover, DALI, MATRAS, and VAST found another structure as a very good hit, namely, N^G,N^G-dimethylarginine dimethylaminohy-

TABLE VIII. Program Performance for Representatives From Each of the Four Main Structural Classes

Program	Mainly α (19)[a]	Mainly β (19)[a]	Mixed α-β (15)[a]	Few SSEs (8)[a]	Overall (%)[b]
CE	17	19	13	8	93
DALI	14	19	14	8	90
DEJAVU	14	19	9	4	75
LOCK	0	14	11	8	54
MATRAS	11	19	14	8	85
PRIDE	14	14	7	3	62
SSM	5	13	10	5	54
TOP	14	18	12	7	84
TOPS	2	15	14	7	62
TOPSCAN	15	12	9	7	70
VAST	12	17	15	7	87

[a] The number of cases for which at least one true positive (different from the query structure itself) was retrieved are listed. At most, this is equal to the number of members of the structural family (this number is listed in brackets in t he column headers).

[b] The overall success rate is defined as the percentage of cases of all four classes in which a true positive (apart from the query structure itself) was retrieved. For instance, for TOPS this is (100% * (2 + 15 + 14 + 7)/(19 + 19 + 15 + 8)) = 62%.

drolase (1h70),[40] which has been solved more recently and was not yet classified in CATH. Both visual inspection and use of the SSAP program (as used in CATH) confirmed that 1h70 has a fold similar to that of ribosome factor IF6.

We also tested if the algorithms are symmetric (i.e., if they produce identical values for the number of aligned residues and the RMSD after swapping the target structure and the query. We found that most programs yield symmetrical results, but not DEJAVU, PRIDE, TOP, and VAST.

Multidomain Proteins

Four of the programs (CE, DALI, MATRAS, and VAST) gave perfect results for both multidomain structures, as defined in Materials and Methods. LOCK did not find any structures that shared all four domains, but it did find the representative structure for each of the domains contained in the Src kinases. TOP and DEJAVU behaved the other way around: they could find the structures with more than one shared domain but were blind to similarities in individual domains, whereas SSM only found structures containing all four domains. TOPSCAN and PRIDE found only one structure with a kinase domain and then for only one of the structures. TOPS was unable to handle multidomain proteins at all and did not find any true positives.

Structural Variation

The sensitivity to small structural variations was assessed by using a set of NMR models of the DNA-binding domain of the glucocorticoid receptor derived from the same experimental data. The results for 1gdc, which is an averaged, energy-minimized NMR model, were compared with those obtained for five "raw" models (models 2, 7, 12, 18, and 20) of 2gda. CE, LOCK, and TOP gave identical results with all models, whereas for the other programs the results differed, depending on the choice of NMR model. DALI, PRIDE, MATRAS, and VAST retrieved different numbers of true and false positives for different models but always had a true positive as the top result. MATRAS retrieved one true positive for most of the models, but two with 1gdc and three with the central model (model 12) of 2gda. It is somewhat surprising that VAST retrieved one true positive for most models, but no fewer than four when using 1gdc or the most atypical model of the ensemble (model 20). TOPS and TOPSCAN were very sensitive to the chosen model, failing in two and three cases, respectively, to retrieve any true positives. DEJAVU failed in all cases, except for model 2 of 2gda for which it found a single hit. SSM succeeded in five of the cases but failed to process model 18 of 2gda because of lack of well-defined SSEs.

Our results suggest that, at least for some programs, one has to be careful with NMR-derived models, and one may want to repeat searches with several models rather than just one. In addition, energy-minimized average models appear better suited as query structures than individual members of an ensemble, possibly due to a better definition of their secondary structure.

Cα-Only Structures

MATRAS, LOCK, SSM, CE, and TOPS failed to work at all with Cα-only structures. DEJAVU, PRIDE, DALI, and VAST gave the same results with all-atom and Cα-only models. The results for TOP differed in the scores of the hits, but the ranking was almost the same. For TOPSCAN, the results obtained with the Cα-only model were substantially worse than those obtained with the intact model.

New Folds

The best detector of new folds proved to be VAST, which in two cases found only the query structure itself. In the other two cases, it did retrieve one other structure but with high P values (0.0286 and 0.002). CE, DALI, and MATRAS also demonstrated a potential to detect a new fold, although in some of the cases, they retrieved other structures with significant scores (albeit lower than for the query structure). Conversely, TOPS failed in all cases; it always found a lot of hits with supposedly significant scores.

TABLE IX. Results Obtained in All Performance Tests

Program	Average performance (%)[a]	Multidomain proteins[b]	Cα-only structures[b]	Structural variation[c]	New folds[b]	Overall judgment[d]	Typical processing time (min)
CE	93	+	−	+	+	+	10–20
DALI	90	+	+	+	+	+	5–15
DEJAVU	73	0	+	−	0	0	10–30
LOCK	58	0	−	+	0	−	5–15
MATRAS	84	+	−	0	+	+	10–25
PRIDE	52	−	+	+	−	−	<1
SSM	47	0	−	−	0	−	1
TOP	78	0	0	+	0	0	20–60
TOPS	60	−	−	+	−	−	1–5
TOPSCAN	68	−	0	−	−	−	<1
VAST	86	+	+	0	+	+	10–20

[a] The average performance is defined as the percentage of cases (using the 61 structures from the 4 main structural classes and the 11 difficult structures) in which a true positive (different from the query structure itself) was retrieved.
[b] Results for the special test cases are listed as "−" (failed), "0" (worked, but not without problems), and "+" (worked very well).
[c] Results are listed as "−" (success or failure depends on the choice of model), "0" (different numbers of true positives depending on the choice of model), and "+" (always the same set of true positives, irrespective of the choice of model).
[d] The overall (subjective) judgment is listed as "−" (poor), "0" (average), and "+" (better than average).

Summary

The results of the performance assessment are summarized in Table IX. This table also lists average processing times (real time elapsed between submission of the query and receipt of the results) as well as our overall (subjective) judgment of each server's performance. Detailed results for all tests are available on our Web site (http://xray.bmc.uu.se/~marian/servers/index.htm).

DISCUSSION AND CONCLUSIONS

The aim of this study was to assess the functionality and performance of several publicly accessible fold comparison programs with Web-based interfaces. Such programs are easy to use and do not require installation of any software or maintenance of any databases on in-house computers. Therefore, such programs can also be used by people who are not experts in structure comparison or Unix system administration.

The functionality of these programs and servers was assessed by using a list of criteria related to operation, presentation, and implementation. The results were not purely objective, because different users have different requirements and levels of experience. It is impossible to capture all the information about the functionality of the programs in one table, so Table VII only provides the briefest of summaries. The weak and strong sides of each server will be addressed below. In general, all the programs were user-friendly, usually worked fast, and differed mainly in the presentation of the results (especially in the presentation of structural alignments and links to visualization programs) and their databases.

The performance tests used the CATH classification system as the "standard-of-truth." CATH was chosen because it represents a middle way between automatic (FSSP) and manual classification (SCOP). However, CATH is not perfect either, because it does not classify many multidomain proteins, even if they were deposited years ago. CATH also contains some obvious mistakes[32] and suffers from a relatively low update frequency. For these reasons, the data from CATH were augmented with information obtained from SCOP, FSSP, and BLAST searches of the PDB.

The tests with representatives of the four main structural classes (Table VIII) showed that the programs could be divided into three groups based on their performance: CE, DALI, MATRAS, TOP, and VAST had >80% success rates in retrieving at least one true positive; DEJAVU and TOPSCAN had >70% success rates; and LOCK, PRIDE, SSM, and TOPS had <65%. Despite the relatively small size of the samples, it is worth noting that most programs performed poorest on mainly-α proteins and best on mainly-β proteins.

To test the ability of the programs to find nontrivial similarities, a set of test cases listed by Fischer et al.[34] was used, augmented with the more recently published structure of IF6.[35] The latter is an example of a structural similarity that was missed (at the time of publication) by DALI, and the structure was incorrectly described as having a new fold. It turned out later that its structure is similar to that of the amidino-transferases.[36,37] Although DALI now also detected the similarity to the amidino-transferases, this example provides a warning for users. True positives can usually be distinguished quite easily from false positives (mainly through visual inspection of the structural superposition and the structure-based sequence alignment), but the absence of any positives according to only one server should not be taken as proof that a protein has a new fold. Instead, other programs should be tried, and one can be reasonably confident that the fold is unique, only if none of the programs detect any structural similarity.

If the difficult structures are taken into account as well, the servers can be grouped as follows: CE, DALI, MATRAS, and VAST all have >80% success rates; DEJAVU, TOP, and TOPSCAN score between 60 and 80%; and LOCK, PRIDE, SSM, and TOPS score ≤60% . The strong and weak sides of each of the fold comparison servers are discussed below.

CE was the only server that did not allow uploading of a structure from a Windows computer and only accepted structures uploaded using Macintosh or Unix workstations. The processing times were very long in some cases (up to 20 h), but this may have been due to excessive loads or network or server problems, because the timings were not reproducible. To avoid long waiting periods, we mostly used the precomputed results available from the CE Web site, after verifying that they were indeed identical to those obtained from the server. In its list of results, only significant matches with the query structure were shown (Z score > 3.7). CE gave identical results for various NMR models and recognized all the domains in Src kinases. CE further proved to have the potential to identify a new fold. However, it did not work with a Cα-only structure.

DALI was one of the oldest of the programs compared here, which was reflected in its presentation of the results. It was very difficult to interpret the structural alignments, and graphic representations of the superimposed structures were not provided. The processing times were on the order of minutes, but the server was very busy occasionally. The list of results returned by DALI contained only structures with a Z score > 2.5. The results from the difficult cases indicated that DALI has undergone continuous development, because it found both structures that it had not been able to detect in a similar experiment 3 years earlier.[11] On the other hand, DALI mysteriously lost the ability to find one similarity, which it did identify then. The new fold tests in three of the four cases resulted in a few false positives (i.e., unrelated structures with a Z score > 2.5), albeit that the z score for the query structure was always substantially higher.

DEJAVU presented its results in a useful format, complete with structure-based sequence alignments and a graphic representation of superimposed structures. DEJAVU gave correct results with a Cα-only model, but its algorithm is not symmetric. DEJAVU found true positives for only one of the NMR models and performed fairly well with multidomain proteins but was sensitive to the size of the domains. DEJAVU also turned out to have difficulties with structures that contain few SSEs.

LOCK also provided a very useful visualization of superimposed structures. The processing times increased substantially with the length of the protein. Fortunately, the results were kept on the LOCK server longer than the announced 6 h (>4 days, in our experience). LOCK offered six databases, which differed in the level of redundancy and the lowest allowed resolution of the structures. The largest database contained only 582 structures, and it had not been updated since 1997. The poor databases substantially limited the performance of LOCK, although the results from the difficult cases were promising (all of these, except IF6, are older structures[34]). On the other hand, LOCK had difficulties with mainly-α structures. LOCK could not handle Cα-only models, because it used DSSP[41] to assign secondary structure, and this program requires calculation of main-chain hydrogen bonds.

The presentation of the results from MATRAS was very similar to that of DALI, with all the advantages and disadvantages of the DALI format (in particular, the poor presentation of structural alignments). The significance of the results was expressed as a Z score, and only results with a Z score > 4.5 were reported. MATRAS was somewhat sensitive to small variations in structures, producing different numbers of true positives for different NMR models. However, on the whole, it did very well in the performance tests.

PRIDE was extremely fast, providing results within seconds, but its performance was rather poor (in particular for mixed α–β structures and those with few SSEs). PRIDE did not provide structural alignments or superimposed structures. The server failed in all the difficult cases: none of the expected similarities were found among the first 100 hits. The average PRIDE score for proteins with the same CATH topology classification is ~ 0.35,[17] which could explain the failure with the difficult cases, because hits with such scores appeared only near the bottom of the list of hits after many false positives. Because the method is based on a comparison of Cα-distance distributions, it worked with Cα-only structures. It handled NMR models well but failed with multidomain proteins.

SSM is a very intuitive, user-friendly, and fast server. The longest processing time was 5 min for multidomain proteins, but on average it ran in under 1 min. The list of hits was transparent and informative. The performance was rather poor (especially with mainly-α structures): SSM succeeded in about half of the cases. SSM failed to work with Cα-only models and also failed for one of the NMR models. For multidomain proteins, it found only hits that contained all the domains.

TOP used the SCOP database instead of the entire PDB. In some cases, this server took several hours to process queries. TOP could handle Cα-only structures but with worse results than when the whole structure was used, probably because it uses the Cβ atoms to optimize the structure superposition. TOP did not recognize all four domains of the Src kinases but only found the two larger domains. Therefore, TOP was sensitive to the size of domains. TOP did not have a symmetrical algorithm.

TOPS was a very fast program. Structures with a score < 10 (a threshold given by the authors without explanation) were supposed to display significant similarity to the query structure, although sometimes true positives with a score > 10 were found. TOPS was not able to handle Cα-only models or multidomain proteins, and it failed in all four cases to identify a new fold. The program displayed sensitivity to small structural variations, sometimes failing to find any true positives, depending on the choice of NMR model. Despite all this, TOPS may sometimes give valuable information rapidly, as in the case of ribosome antiassociation factor IF6. The server performed very poorly with mainly-α structures but did quite well with mixed α–β structures.

TOPSCAN was constructed as a very simple and fast prescreen to more sophisticated programs. It proved to be the fastest program, never taking more than a minute to produce its results. Given the simplicity of the method, it performed surprisingly well with the representatives of the four main structural classes, but it failed with multidomain proteins and new folds. TOPSCAN worked with Cα-only models, but the results were completely different from those obtained with the intact structures. When used with NMR models, its performance showed critical dependence on the choice of model.

VAST probably had the best presentation of results, including a significance score, RMSD value, a graphic presentation of secondary structure content, and visualization of superimposed structures. On the other hand, VAST had difficulties parsing some of the uploaded PDB files. The processing times depended on the length of the protein and reached up to 1 h. VAST has undergone continuous development as was shown by its success with the difficult cases. One new similarity was found, and one was somehow lost compared to the results of Shindyalov and Bourne.[11] The VAST algorithm is not symmetrical, and the values it produced for the RMSD and the number of aligned residues were worse than for the other programs. It worked properly with the Cα-only model, but the NMR models retrieved different numbers of true positives with VAST. On the other hand, VAST showed the best performance in recognizing a new fold.

The combined results of all tests and assessments did not reveal any overall best server. Instead, they suggested that it is a prudent strategy to take advantage of several different programs and to use more than one of them to confirm the results. In particular, if no hits are found by one server (suggesting that a fold might be new), one does well to err on the side of caution and to repeat the search using one or more of the other servers.

Finally, we expect that the authors of the programs tested here (and of future programs) will benefit from our findings. Besides providing a wide variety of realistic test cases and results that can be used as benchmarks, this work also reveals the strong and weak sides of the various programs and servers and, hence, which aspects of each server have potential for improvement.

ACKNOWLEDGMENTS

The authors thank Prof. T. Alwyn Jones (Uppsala University) for a critical reading of the manuscript and Terese Bergfors (Uppsala University) for indispensable linguistic comments and corrections. MN thanks Dr. Petr Folk (Charles University, Prague) for his interest in this work, and Dr. Eugene Krissinel (European Bioinformatics Institute, Hinxton, UK) for information about the SSM server. We are also grateful to the authors and administrators of the various servers for the services they provide to the community and, in a number of cases, for answering our queries. Finally, we thank the two anonymous referees for their comments and suggestions. MN was supported through a grant from the Ministry of Education, Youth and Sports of the Czech Republic to Dr. Folk, as well as a student mobility grant from the Socrates II/Erasmus Programme of the European Union. GJK is a Royal Swedish Academy of Sciences (KVA) Research Fellow, supported through a grant from the Knut and Alice Wallenberg Foundation. He is supported by KVA, Uppsala University, the Linnaeus Centre for Bioinformatics, and the Swedish Structural Biology Network (SBNet).

REFERENCES

1. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
2. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
3. Holm L, Sander C. Searching protein structure databases has come of age. Proteins 1994;19:165–173.
4. Brown NP, Orengo CA, Taylor WR. A protein structure comparison methodology. Comput Chem 1996;20:359–380.
5. Eidhammer I, Jonassen I, Taylor WR. Structure comparison and structure patterns. J Comput Biol 2000;7:685–716.
6. Koehl P. Protein structure similarities. Curr Opin Struct Biol 2001;11:348–353.
7. Gibrat JF, Madej T, Bryant SH. Surprising similarities in structure comparison. Curr Opin Struct Biol 1996;6:377–385.
8. Kleywegt GJ, Jones TA. Detecting folding motifs and similarities in protein structures. Methods Enzymol 1997;277:525–545.
9. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. J Mol Biol 1993;233:123–138.
10. Jung J, Lee B. Protein structure alignment using environmental profiles. Protein Eng 2000;13:535–543.
11. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng 1998;11:739–747.
12. Madsen D, Kleywegt GJ. Interactive motif and fold recognition in protein structures. J Appl Crystallogr 2002;35:137–139.
13. Levitt M, Gerstein M. A unified statistical framework for sequence comparison and structure comparison. Proc Natl Acad Sci USA 1998;95:5913–5920.
14. Singh AP, Brutlag DL. Hierarchical protein structure superposition using both secondary structure and atomic representations. Proc Intell Syst Mol Biol 1997;97:284–293.
15. Horn BKP. Closed form solution of absolute orientation using unit quaternions. J Opt Soc Am 1987;4:629–642.
16. Kawabata T, Nishikawa K. Protein structure comparison using the Markov transition model of evolution. Proteins 2000;41:108–122.
17. Carugo O, Pongor S. Protein fold similarity estimated by a probabilistic approach based on Cα-Cα distance comparison. J Mol Biol 2002;315:887–898.
18. Vlahovicek K, Carugo O, Pongor S. The PRIDE server for protein three-dimensional similarity. J Appl Crystallogr 2002;35:648–649.
19. Lu G. TOP: a new method for protein structure comparisons and similarity searches. J Appl Crystallogr 2000;33:176–183.
20. Gilbert D, Westhead D, Nagano N, Thornton J. Motif-based searching in TOPS protein topology databases. Bioinformatics 1999;15:317–326.
21. Westhead DR, Hatton DC, Thornton JM. An atlas of protein topology cartoons available on the World-Wide Web. Trends Biochem Sci 1998;23:35–36.
22. Martin ACR. The ups and downs of protein topology; rapid comparison of protein structure. Protein Eng 2000;13:829–837.
23. Orengo CA, Taylor WR. SSAP: sequential structure alignment program for protein structure comparison. Methods Enzymol 1996;266:617–635.
24. Madej T, Gibrat JF, Bryant SH. Threading a database of protein cores. Proteins 1995;23:356–369.
25. Harrison A, Pearl F, Mott R, Thornton J, Orengo C. Quantifying the similarities within fold space. J Mol Biol 2002;323:909–926.
26. 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;247:536–540.
27. Brenner SE, Chothia C, Hubbard TJ, Murzin AG. Understanding protein structure: using scop for fold interpretation. Methods Enzymol 1996;266:635–643.
28. Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB,

Thornton JM. CATH—a hierarchic classification of protein domain structures. Structure 1997;5:1093–1108.

29. Holm L, Sander C. The FSSP database of structurally aligned protein fold families. Nucleic Acids Res 1994;22:3600–3609.
30. Holm L, Sander C. Mapping the protein universe. Science 1996;273: 595–602.
31. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–410.
32. Hadley C, Jones DT. A systematic comparison of protein structure classifications: SCOP, CATH and FSSP. Structure 1999;7:1099–1112.
33. McGuffin LJ, Bryson K, Jones DT. What are the baselines for protein fold recognition? Bioinformatics 2001;17:63–72.
34. Fischer D, Elofsson A, Rice D, Eisenberg D. Assessing the performance of fold recognition methods by means of a comprehensive benchmark. Proc Pacific Symp Biocomput 1996;96:300–318.
35. Groft CM, Beckmann R, Sali A, Burley SK. Crystal structures of ribosome anti-association factor IF6. Nat Struct Biol 2000;7:1156–1164.
36. Paoli M. An elusive propeller-like fold. Nat Struct Biol 2001;9:744.
37. Groft CM, Beckmann R, Sali A, Burley SK. Response to Paoli. Nat Struct Biol 2001;9:745.
38. Kleywegt GJ. Use of non-crystallographic symmetry in protein structure refinement. Acta Crystallogr 1996;D52:842–857.
39. van Aalten DMF, Milne KG, Zou JY, Kleywegt GJ, Bergfors T, Ferguson MAJ, Knudsen J, Jones TA. Binding site differences revealed by crystal structures of *Plasmodium falciparum* and bovine acyl-CoA binding protein. J Mol Biol 2001;309:181–192.
40. Murray-Rust J, Leiper J, McAlister M, Phelan J, Tilley S, Santa Maria J, Vallance P, McDonald N. Structural insights into the hydrolysis of cellular nitric oxide synthase inhibitors by dimethylarginine dimethylaminohydrolase. Nat Struct Biol 2001;8:679–683.
41. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
42. Holm L, Sander C. Alignment of three-dimensional protein structures: network server for database searching. Methods Enzymol 1996;266:653–662.
43. Kleywegt GJ, Jones TA. Halloween…masks and bones. In: Bailey S, Hubbard R, Waller DA, editors. From first map to final model. Daresbury: SERC Daresbury Laboratory; 1994. p 59–66.
44. Gilbert D, Westhead D, Viksna J, Thornton J. A computer system to perform structure comparison using TOPS representations of protein structure. Comput Chem 2001;26:23–30.
45. Pearl FM, Lee D, Bray JE, Sillitoe I, Todd AE, Harrison AP, Thornton JM, Orengo CA. Assigning genomic sequences to CATH. Nucleic Acids Res 2000;28:278–282.
46. Subbiah S, Laurents DV, Levitt M. Structural similarity of DNA binding domains of bacteriophage repressors and the globin core. Curr Biol 1993;3:141–148.