W50–W55 *Nucleic Acids Research, 2005, Vol. 33, Web Server issue*
doi:10.1093/nar/gki416

CAMPO, SCR_FIND and CHC_FIND: a suite of web tools for computational structural biology

Alessandro Paiardini, Francesco Bossa and Stefano Pascarella[1,*]

Dipartimento di Scienze Biochimiche 'A. Rossi Fanelli' and [1]Centro Interdipartimentale di Ricerca per la Analisi dei Modelli e dell'Informazione nei Sistemi Biomedici (CISB), Università La Sapienza, Piazzale Aldo Moro 5, 00185 Roma, Italy

Received February 8, 2005; Revised and Accepted March 21, 2005

ABSTRACT

The identification of evolutionarily conserved features of protein structures can provide insights into their functional and structural properties. Three methods have been developed and implemented as WWW tools, CAMPO, SCR_FIND and CHC_FIND, to analyze evolutionarily conserved residues (ECRs), structurally conserved regions (SCRs) and conserved hydrophobic contacts (CHCs) in protein families and superfamilies, on the basis of their 3D structures and the homologous sequences available. The programs identify protein segments that conserve a similar main-chain conformation, compute residue-to-residue hydrophobic contacts involving only apolar atoms common to all the 3D structures analyzed and allow the identification of conserved amino-acid sites among protein structures and their homologous sequences. The programs also allow the visualization of SCRs, CHCs and ECRs directly on the superposed structures and their multiple structural and sequence alignments. Tools and tutorials explaining their usage are available at http://schubert.bio.uniroma1.it/SCR_FIND, http://schubert.bio.uniroma1.it/CHC_FIND and http://schubert.bio.uniroma1.it/CAMPO.

INTRODUCTION

The results obtained from the concurrent detection of structurally conserved interaction patterns and the analysis of sequence conservation in a protein family can be of great value for deciphering complex biological phenomena, such as protein folding or the evolutionary emergence of distinct catalytic properties from a common scaffold, and for planning protein design and engineering experiments (1,2). In this sense, the rapid increase in the number of sequences and structures owing to structural and genomic projects represents a major challenge, i.e. how to best exploit this information in order to extract biologically relevant features. Here, we present a suite of publicly available web services we implemented for the identification of evolutionarily conserved regions and contacts in protein families and superfamilies: CAMPO, SCR_FIND and CHC_FIND.

CAMPO is a fully automated web tool that enables the assessment of the evolutionary conservation grade of protein residues. Usually, the evolutionary conservation grade is a useful measure of the importance of a residue: for example, the catalytic center of an enzyme is highly conserved since, if a mutation occurs at that site, the catalytic activity of the enzyme is likely to be lost, leading to a decreased fitness of the organism (3). The evolutionary conservation grade can be determined by the variability of residues in the columns of a multiple sequence alignment of homologous proteins (4). The algorithm implemented in CAMPO assigns a score to each column of a multiple sequence alignment throughout the application of a user-defined mutational matrix and incorporates a weight based on the percentage of sequence identity between proteins being compared. The results obtained can be mapped onto a reference protein structure to allow the identification of functionally important residues and surface regions. Optionally, CAMPO also allows to measure the evolutionary conservation of a spatial region of arbitrary radius, centered on every atom of the 3D structure. Once identified, most of the evolutionarily conserved residues (ECRs) of homologous proteins can be further analyzed to assess if their conservation reflects a functional role (i.e. substrate binding, catalysis, interaction with other macromolecules) or a structural role [i.e. residues interacting through hydrophobic contacts, which are thought to be necessary for the proper fold and structural stability of proteins (5,6)].

SCR_FIND and CHC_FIND are able to identify structurally conserved regions (SCRs), similar 3D patterns of protein segments that conserve the same main-chain conformation, and their conserved hydrophobic contacts (CHCs), in members belonging to a family or superfamily of proteins (7). SCRs and

*To whom correspondence should be addressed. Tel: +39 06 49917574; Fax: +39 06 49917566; Email: stefano.pascarella@uniroma1.it

CHCs are presumably subjected to similar constraints during the divergent evolution of a family or superfamily of proteins from a common ancestor; therefore, they possibly contain most of the determinants necessary to maintain the fold (8). Although many public domain tools and WWW servers are able to analyze structural and sequence relationships (9,10), a server devised to the extraction of SCRs and CHCs from aligned protein structures at different similarity thresholds is, to our knowledge, not yet available.

Finally, an interface to the CE–MC multiple protein structure alignment algorithm was made available (11), modified so that its output is suitable for the SCR_FIND and CHC_FIND tools (http://schubert.bio.uniroma1.it/CEMC).

METHODS

CAMPO, SCR_FIND and CHC_FIND are coded in C, PERL-CGI and JavaScript and run on a Digital Alpha station, under UNIX operating system.

CAMPO makes use of a procedure similar to that adopted by ConSurf to obtain a fully automated multiple sequence alignment starting from a sequence probe (12). CAMPO utilizes the standalone version of BLAST, with an *E*-value threshold set by the user to accept or reject the sequences (13). Identified sequences are filtered out (see below) and then aligned with ClustalW (14); in addition, CAMPO allows the choice of the following options: (i) protein database used to retrieve homologous sequences [currently nrdb (15) and SwissProt (16) are incorporated, and other databases can be readily added]; (ii) minimum and maximum percentages of sequence identity to the probe, and minimum percentage of residues aligned to the probe to accept the sequences found. Furthermore, CAMPO allows the user choose the most appropriate mutational matrix (PAM and BLOSUM series are implemented) to align and assign a conservation score to the filtered sequences (Supplementary Material 1). Since in extensive tests of sequence alignments the BLOSUM and PAM matrix series on average gave superior results compared with matrices based on physicochemical properties (17), it seemed appropriate to adopt these mutational matrices to assign a score for the amino acid exchanges (18). To measure the sequence conservation, CAMPO assigns to each position of the multiple sequence alignment the following score, which is formally similar to the one proposed by Karlin and Brocchieri (19):

$$O_k = \frac{1}{[n(n-1)/2]} \times \left[\frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left[\frac{\mathrm{Bscore}_{kij}}{1/2\left(|\mathrm{Bscore}_{kii}| + |\mathrm{Bscore}_{kjj}|\right)} \cdot \left(1 - \frac{\mathrm{nid}_{ij}}{\mathrm{nal}_{ij}}\right) \right]}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(1 - \frac{\mathrm{nid}_{ij}}{\mathrm{nal}_{ij}}\right)} \right], \quad \mathbf{1}$$

where O_k is the score assigned for every position k of the multiple sequence alignment, n is the number of sequences included in the alignment, i and j refers to the ith and the jth sequence, respectively, Bscore_{kij}, Bscore_{kii} and Bscore_{kjj} are the scores assigned to the residue exchange in position k between the ith and the jth sequence according to the BLOSUM or PAM mutational matrix, nid_{ij} is the number of identical residues and nal_{ij} is the number of aligned residues between the ith and the jth sequence, respectively. For every possible exchange at a particular position of the multiple alignment, a normalized conservation index is computed, based on the score of a mutational matrix. Since the matrix scores for matching the same amino acids vary for different residues, conservation indices for invariant positions of the multiple sequence alignment would depend on residue type; normalization is used to avoid different conservation scores for invariant positions. Indels are assigned a fixed gap penalty score, according to the mutational matrix chosen by the user. At variance with the method of Karlin and Brocchieri (19), a weighting scheme is incorporated in which every residue exchange is corrected by the inverse of the sequence similarity between the proteins being compared, measured as their percentage identity. Thus, sequence weighting attempts to normalize against redundancy in the alignment. More sophisticated tree-based weighting schemes were adopted by Altschul *et al.* (13) and Armon *et al.* (20). However, as stated by Valdar (21), tree-based weighting schemes require more assumptions than those based directly on the sequence alignments and can introduce additional uncertainty to the final score.

The mean $\bar{O}$ and the standard deviation σ for the distribution of O_k values are then determined; the significance R of every conservation index of the alignment is then calculated by dividing the difference between O_k and $\bar{O}$ by σ. The scores computed on every position of the multiple alignment can be optionally utilized as weights to compute the evolutionary conservation of a spatial region of arbitrary radius D, centered on every atom of the 3D structure, applying a percolation theory-inspired technique (22):

$$R_{i\mathrm{new}} = R_i + \frac{\sum_{j=1}^{n} \begin{cases} \frac{R_j}{d_{ij}}, & \text{if } d_{ij} \leqslant D, \\ 0, & \text{if } d_{ij} > D, \end{cases}}{\sum_{j=1}^{n} \begin{cases} \frac{1}{d_{ij}}, & \text{if } d_{ij} \leqslant D, \\ 0, & \text{if } d_{ij} > D, \end{cases}} \quad \mathbf{2}$$

where n is the number of atoms of the molecule, $R_{i\mathrm{new}}$ represents the recalculated score for atom i, R_i and R_j are the initial scores assigned to atom i and j, respectively, corresponding to the conservation score computed for the residues to which the atoms belong, d_{ij} is the distance between the centers of mass of the residues to which atoms i and j belong, and D is the user-defined arbitrary radius. This approach enables the user to achieve a better resolution of the boundaries of conserved patches on the 3D protein structure, possibly interacting with small ligands and macromolecules.

In order to assess the reliability of the conservation indexes computed on the ECRs of the described sample cases (see below), the following null hypothesis is tested: the average evolutionary conservation of the ECRs for a given sample is no higher than that obtained by randomly resampling the original dataset of n sequences with m sites each, and realigning them. Two m sites are randomly drawn from each sequence a million times (random sequences) and fifty times (pseudoreplicate sequences), and their position interchanged. The new sequences are realigned and conservation scores recomputed on every position of the multiple alignment. Five hundred multiple sequence alignments are generated for each sample case. The normalized distribution of the obtained conservation

values for each sample case, compared with the average evolutionary conservation of the ECRs, is shown in Supplementary Material 2. The results show that the ECRs of all proteins analyzed are statistically ($P < 0.05$) more conserved than expected for a distribution of scores derived from random and pseudoreplicate sequences.

SCR_FIND is a tool for the identification of SCRs, starting from a multiple structure alignment and the corresponding superposed 3D coordinates. These two files can be (i) obtained by using the interface to the CE–MC multiple protein structure alignment algorithm (11), at http://schubert.bio.uniroma1.it/CEMC/index.html; (ii) downloaded directly from the CE site (http://cl.sdsc.edu/ce.html); (iii) manually edited, following the CE standard file format. For every structurally equivalent position i of the multiple structural alignment, SCR_FIND computes a score SC_i based on the root-mean-square deviation (RMSD) from the center of mass of the structurally equivalent Cα atoms and an arbitrary gap penalty GP, which is added for every gap found (N_{gaps}):

$$SC_i = \sqrt{\frac{\sum_{j=1}^{N}\left(x_{ji}-\bar{x}_i\right)^2+\left(y_{ji}-\bar{y}_i\right)^2+\left(z_{ji}-\bar{z}_i\right)^2}{N}} + GP \cdot N_{gaps}, \quad \mathbf{3}$$

where x_{ji}, y_{ji} and z_{ji} are the Cartesian coordinates of the jth Cα atom at position i of the alignment and $\bar{x}_i$, $\bar{y}_i$ and $\bar{z}_i$ are the coordinates of the center of mass computed over the N atoms found at position i. A window of arbitrary size w is then scrolled through the alignment. Each time three or more consecutive positions with a mean score below a user-given score threshold value are found, w is increased iteratively by 1 position until the mean score does not raise above the threshold value, or until the window reaches the end of the alignment. The scores computed for every position of the alignment and details concerning the SCRs found (residues constituting the SCRs, positional RMSD, mean positional RMSD for each SCR and score for each position), constitute the output of the program, along with the 3D coordinates of the SCRs.

The output of SCR_FIND can be used as input for CHC_FIND. This program exploits the algorithm by Drabløs (6), which computes pairwise atom contact areas between non-polar atoms from a standard PDB coordinate file, to calculate the pairwise residue contact areas for every possible pair of residues belonging to the SCRs of the structures analyzed. CHCs are then classified on the basis of their location (intra-SCR and inter-SCR CHCs), the number of structures in which the hydrophobic contact is conserved and the mean apolar contact area of the structurally equivalent residues of each structure. If two positions of the multiple structural alignment, x and y, have residues in hydrophobic contact in at least two of the structures, then a candidate CHC is detected. CHCs are then classified on the basis of their strength s_{xy}, defined as:

$$s_{xy} = \frac{\sum_{i=1}^{N} A_i(x,y)}{N}, \quad \mathbf{4}$$

where A_i is the apolar contact area of the ith structure between residues at absolute positions x and y of the structural alignment, and N is the number of superposed structures.

Finally, ECRs, SCRs and CHCs can be mapped onto the sequences and structures with a color code reflecting the mean and standard deviation of the values found, through a CHIME/Rasmol (23) interface (Supplementary Material 3).

RESULTS

Two sample cases are presented here and will be utilized to demonstrate how these tools may be used and what kind of information they are expected to present. The first example explains the usage of CAMPO in a well-studied case, the potassium channel from *Streptomyces lividans* (Kcsa, PDB code 1BL8), and its performance compared with ConSurf (20), one of the most commonly used servers for the identification of functionally important regions and ECRs. In the second example, it is shown how SCR_FIND and CHC_FIND can be used to predict and explain experimental data, through the analysis of the well-studied trypsin inhibitor protein fold. For additional examples see Supplementary Materials 4–6.

The potassium channel from *S.lividans*

To demonstrate the ability of CAMPO to detect evolutionarily conserved patches that are likely to be required for protein activity and stability, we report as example the analysis carried out on the potassium channel from *S.lividans* (Kcsa, PDB code 1BL8), a well-studied protein for which suitable sequence and structural information is known, and regions of functional importance have already been determined. The potassium channel from *S.lividans* is an integral membrane protein with sequence similarity to all known K^+ channels, particularly in the pore region. It has been observed that sequence conservation among K^+ channels is strongest for the amino acids corresponding to the pore region (residues 61–85) and the inner helix (residues 86–119), whereas the N-terminal, outer helix (residues 23–60) is less conserved (24). CAMPO's results for 1BL8, chain A, are available at http://schubert.bio.uniroma1.it/transitoCAMPO/kcsa/ (see also Figure 1). CAMPO identified 68 homologous sequences using default parameters (E-value threshold of BLAST, 0.001; minimum and maximum percentages of identity to accept a sequence for further analysis, 20 and 80%, respectively; minimum percentage of residues aligned to the probe to filter the sequences found, 80%). A BLOSUM62 mutational matrix was chosen to align the sequences and assign the conservation score. CAMPO was able to detect the most conserved residues facing the inner face of the channel (Phe 114, Leu 110, Val 106, Gly 104, Leu 105, Gly 99, Ile 100, Thr 74, Thr 75, Trp 68 and Pro 83) and interacting with the other subunits that constitute the tetrameric structure (Trp 67, Tyr 78 and Asp 80). In particular, residues Gly 77, Tyr 78 and Gly 79, which are known to interact with the K^+ ion and to be absolutely required for K^+ selectivity, were highlighted as the most conserved ones in the inner protein core. The difference between the inner and outer surface of the channel was even more evident when the initial scores were clustered into spatial regions of increasing radii, allowing a 'percolation' of the evolutionary conservation to detect the most conserved patches (Figure 1A and B). At 5 Å radius, when the ratio between interacting and not-interacting atoms enclosed by the sphere is maximum, the differences between the mean conservation values obtained for the inner

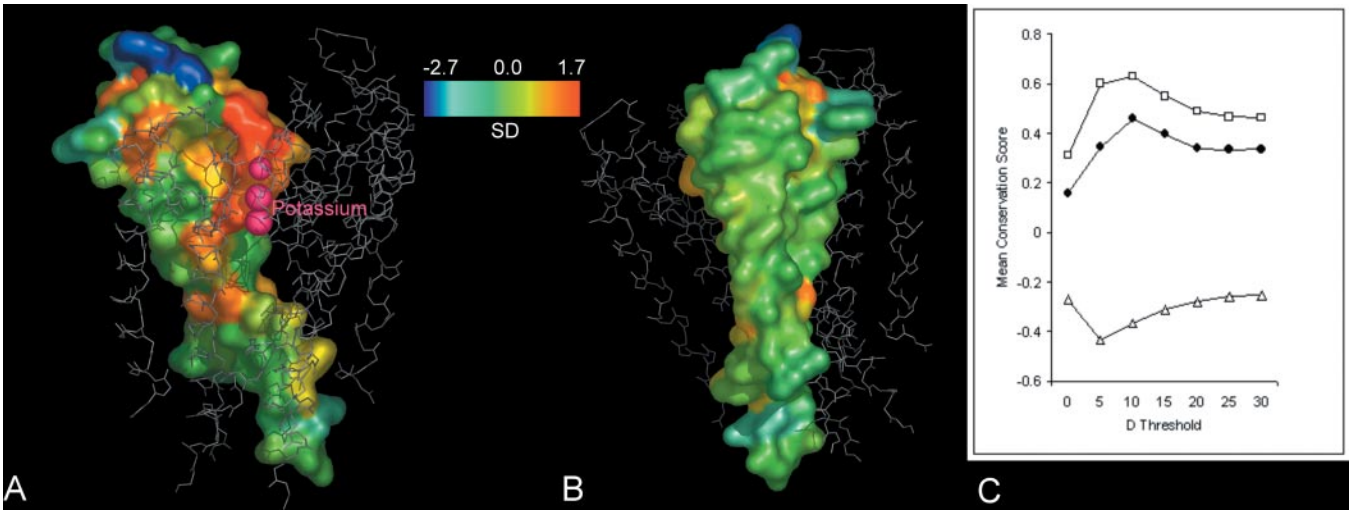

Figure 1. Mapping of evolutionary conservation on the (**A**) inner and (**B**) outer surface of the potassium channel protein from *S.lividans*, as scored by CAMPO, and (**C**) plot of the mean conservation scores against different radius (*D*) thresholds for the outer helix (open triangle), inner helix (filled circles) and the pore region (open squares) of the potassium channel. The results obtained are expressed in units of standard deviation from the mean conservation value. According to CAMPO's color scheme, dark blue corresponds to maximal variability, red to maximal conservation. Potassium ions are displayed as pink CPKs.

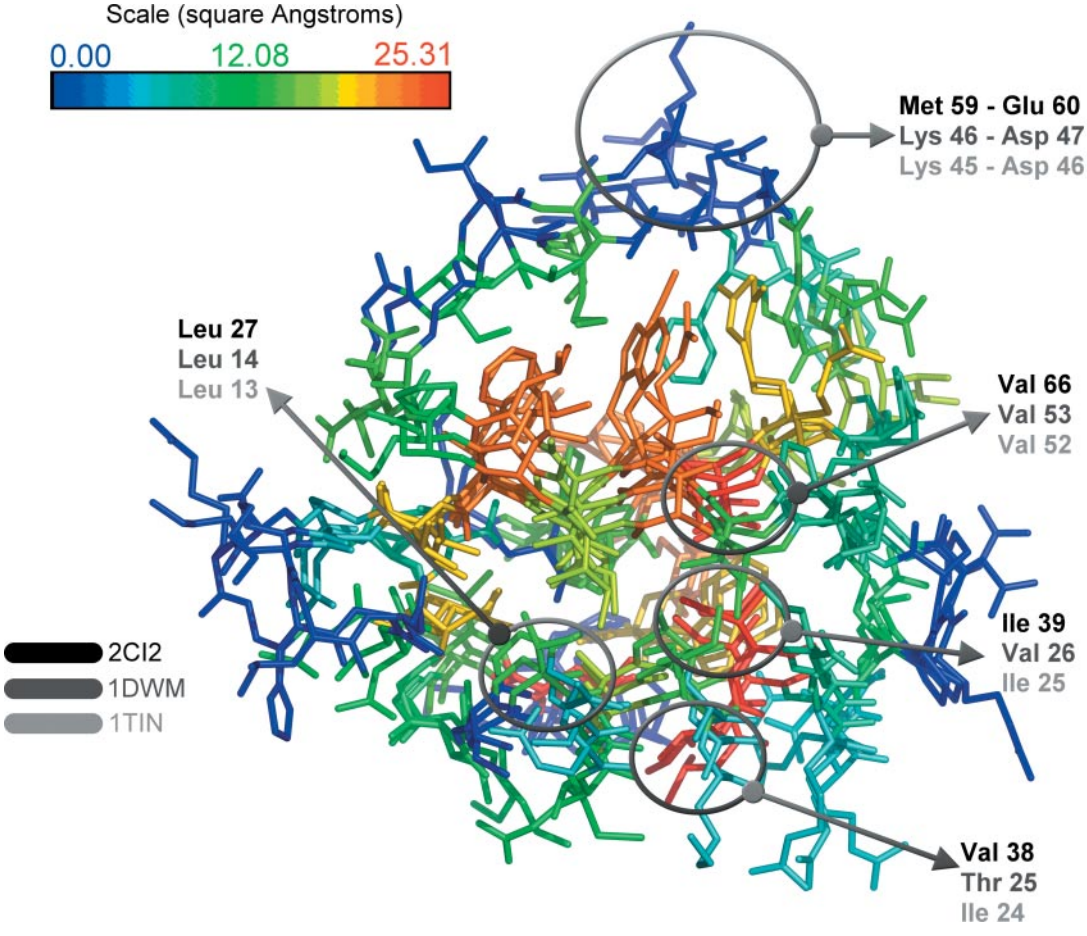

Figure 2. An adapted sample output of SCR_FIND and CHC_FIND, showing the CHCs found. Chymotrypsin inhibitor 2 (PDB code 2CI2), *L.usitatissimum* trypsin inhibitor (PDB code 1DWM) and *C.maxima* trypsin inhibitor (PDB code 1TIN). The 3D structures are colored according to the strongest value of mean apolar contact surface in which they are involved. Residues involved in the strongest hydrophobic contacts and the protease-binding loop are also highlighted. More information is available at http://schubert.bio.uniroma1.it/SCR_FIND and http://schubert.bio.uniroma1.it/CHC_FIND.

helix, the outer helix and the pore region are the most evident (Figure 1C). This observation is in agreement with the results previously obtained by Doyle *et al.* (24). The analysis performed by ConSurf on the same protein and the comparison of the results obtained by the two servers are presented as Supplementary Material 4, as well as the analysis carried out by CAMPO on three conserved hypothetical proteins, 1VKI, 1VKB and 1VK4 (Supplementary Material 5).

The trypsin inhibitor fold

The folding process of many small globular proteins is often a spontaneous event *in vitro* that takes place in an apparent two-state reaction mechanism (25). These reactions are characterized by the presence of a single, rate-limiting transition state separating the unfolded and the folded states, with no other apparent observable intermediates (26). It is suggested that the interactions in the transition state ensemble are mostly native-like, with the residues involved forming a nucleation hydrophobic core (8). So far, site-directed mutagenesis approaches have been applied to obtain insights into the folding mechanism of a variety of small, globular proteins [P22 Arc repressor (27); CI2 (28); CheY (29)]. A well-studied case, the trypsin inhibitor fold, will be discussed to demonstrate a possible usage of SCR_FIND and CHC_FIND to help the user highlight possible targets of site-directed mutagenesis experiments.

Hordeum vulgare chymotrypsin inhibitor 2 (CI2) [PDB code 2CI2 (30)], *Linum usitatissimum* trypsin inhibitor [PDB code 1DWM (31)] and *Cucurbita maxima* trypsin inhibitor [PDB code 1TIN (32)] are small single domain proteins that share a similar fold (Figure 2). An extended nucleus of interactions is identified using SCR_FIND and CHC_FIND, structured around the N-terminal α-helix and the C-terminal β-sheet of the proteins. Most conserved hydrophobic interactions are engaged in by (numbering refers to 2CI2): Leu 27 with Val 38 (mean apolar contact surface: 23.9 $Å^2$) and Ile 39 with Val 66 (mean apolar contact surface: 25.3 $Å^2$). Other hydrophobic residues display well-conserved patterns of interactions: Trp 24 and Ala 35 with Leu 27 (19.8 and 17.0 $Å^2$, respectively), Val 50 with Leu 68 (17.7 $Å^2$), Val 70 with Ile 76 (17.7 $Å^2$), Leu 68 with Ile 76 (17.0 $Å^2$), and Leu 51 with Phe 69 (22.7 $Å^2$). Some of these contacts have been previously identified in folding intermediates, using engineering approaches (33). It has been demonstrated that complementation of peptide fragments to gain a native-like structure occurs only when the cleavage is located in the protease-binding loop at position Met 59-Glu 60 (34). Accordingly, this region is not involved in any CHC (Figure 2). The results for the trypsin inhibitor are available at http://schubert.bio.uniroma1.it/transitoCHC/ci2/.

Another example, the acyl CoA binding protein fold, is discussed in the Supplementary Material 6.

CONCLUSIONS

We presented a suite of web services for structural analysis, CAMPO, SCR_FIND and CHC_FIND, along with several examples to explain their usage and show their capabilities. We suggest that the use of these tools, along with others already available such as ConSurf, can shed light into the evolutionary history and functional properties of protein families for which suitable structural information is available.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Prof. Donatella Barra for support. Thanks to Daniele Tronelli for helpful advices. The authors are grateful to Dr Chittibabu Guda for having provided us with the source code of the CE–MC algorithm. This work was supported in part by the Italian 'Ministero dell'Istruzione, dell' Università e della Ricerca' (MIUR). Funding to pay the Open Access publication charges for this article was provided by MIUR.

Conflict of interest statement. None declared.

REFERENCES

1. Prasad,T., Prathima,M.N. and Chandra,N. (2003) Detection of hydrogen-bond signature patterns in protein families. *Bioinformatics*, **19**, 167–168.
2. Kuhlman,B., Dantas,G., Ireton,G.C., Varani,G., Stoddard,B.L. and Baker,D. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science*, **302**, 1364–1368.
3. Pils,B. and Schultz,J. (2004) Inactive enzyme-homologues find new function in regulatory processes. *J. Mol. Biol.*, **340**, 399–404.
4. Lesk,A. and Chothia,C. (1980) How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J. Mol. Biol.*, **136**, 225–270.
5. Gallet,X., Charloteaux,B., Thomas,A. and Brasseur,R. (2000) A fast method to predict protein interaction sites from sequences. *J. Mol. Biol.*, **302**, 917–926.
6. Drabløs,F. (1999) Clustering of non-polar contacts in proteins. *Bioinformatics*, **15**, 501–509.
7. Murzin,A.G., Brenner,S.E., Hubbard,T. and Chothia,C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **247**, 536–540.
8. Shakhnovich,E., Abkevich,V. and Ptitsyn,O. (1996) Conserved residues and the mechanism of protein folding. *Nature*, **379**, 96–98.
9. Sobolev,V., Sorokine,A., Prilusky,J., Abola,E.E. and Edelman,M. (1999) Automated analysis of interatomic contacts in proteins. *Bioinformatics*, **15**, 327–332.
10. Shindyalov,I.N. and Bourne,P.E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.*, **11**, 739–747.
11. Guda,C., Lu,S., Scheeff,E.D., Bourne,P.E. and Shindyalov,I.N. (2004) CE-MC: a multiple protein structure alignment server. *Nucleic Acids Res.*, **32**, 100–103.
12. Glaser,F., Pupko,T., Paz,I., Bell,R.E., Bechor-Shental,D., Martz,E. and Ben-Tal,N. (2003) ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*, **19**, 163–164.
13. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
14. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
15. Holm,L. and Sander,C. (1998) Removing near-neighbour redundancy from large protein sequence collections. *Bioinformatics*, **14**, 423–429.
16. Junker,V., Contrino,S., Fleischmann,W., Hermjakob,H., Lang,F., Magrane,M., Martin,M.J., Mitaritonna,N., ÒDonovan,C. and Apweiler,R. (2000) The role SWISS-PROT and TrEMBL play in the genome research environment. *J. Biotechnol.*, **78**, 221–234.

17. Miyata,T., Miyazawa,S. and Yashunaga,T. (1979) Two types of amino acid substitutions in protein evolution. *J. Mol. Evol*., **12**, 219–236.
18. Vogt,G., Etzold,T. and Argos,P. (1995) An assessment of amino acid exchange matrices in aligning protein sequences: the twilight zone revisited. *J. Mol. Biol*., **249**, 816–831.
19. Karlin,S. and Brocchieri,L. (1996) Evolutionary conservation of RecA genes in relation to protein structure and function. *J. Bacteriol*., **178**, 1881–1894.
20. Armon,A., Graur,D. and Ben-Tal,N. (2001) ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. *J. Mol. Biol*., **307**, 447–463.
21. Valdar,W.S. (2002) Scoring residue conservation. *Proteins*, **48**, 227–241.
22. Harrison,P. (2001) Percolation theory. In Purich,D.L. (ed.), *Computational Methods in Physics Chemistry and Biology: An Introduction*. John Wiley & Sons Inc., NY, pp. 209–224.
23. Sayle,R.A. and Milner-White,E.J. (1995) RASMOL: biomolecular graphics for all. *Trends Biochem. Sci*., **20**, 374.
24. Doyle,D.A., Morais Cabral,J., Pfuetzner,R.A., Kuo,A., Gulbis,J.M., Cohen,S.L., Chait,B.T. and MacKinnon,R. (1998) The structure of the potassium channel: molecular basis of K conduction and selectivity. *Science*, **280**, 69–77.
25. Kragelund,B.B., Osmark,P., Neergaard,T.B., Schiodt,J., Kristiansen,K., Knudsen,J. and Poulsen,F.M. (1999) The formation of a native-like structure containing eight conserved hydrophobic residues is rate limiting in two-state protein folding of ACBP. *Nature Struct. Biol*., **6**, 594–601.
26. Ptitsyn,O.B. (1991) How does protein synthesis give rise to the 3D-structure? *FEBS Lett*., **285**, 176–181.
27. Milla,M.E., Brown,B.M., Waldburger,C.D. and Sauer,R.T. (1997) P22 Arc repressor: transition state properties inferred from mutational effects on the rates of protein unfolding and refolding. *Biochemistry*, **34**, 13914–13919.
28. Jackson,S.E., elMasry,N. and Fersht,A.R. (1993) Structure of the hydrophobic core in the transition state for folding of chymotrypsin inhibitor 2: a critical test of the protein engineering method of analysis. *Biochemistry*, **32**, 11270–11278.
29. Lopez-Hernandez,E. and Serrano,L. (1996) Structure of the transition state for folding of the 129 aa protein CheY resembles that of a smaller protein CI-2. *Fold Des*., **1**, 43–45.
30. McPhalen,C.A. and James,M.N. (1987) Crystal and molecular structure of the serine proteinase inhibitor CI-2 from barley seeds. *Biochemistry*, **26**, 261–269.
31. Cierpicki,T. and Otlewski,J. (2000) Determination of a high precision structure of a novel protein (*Linum usitatissimum* trypsin inhibitor LUTI) using computer-aided assignment of NOESY cross-peaks. *J. Mol. Biol*., **302**, 1179–1192.
32. Krishnamoorthi,R., Gong,Y.X. and Richardson,M. (1990) A new protein inhibitor of trypsin and activated Hageman factor from pumpkin (*Cucurbita maxima*) seeds. *FEBS Lett*., **273**, 163–167.
33. Fersht,A.R., Itzhaki,L.S., elMasry,N.F., Matthews,J.M. and Otzen,d.E. (1994) Single versus parallel pathways of protein folding and fractional formation of structure in the transition state. *Proc. Natl Acad. Sci. USA*, **91**, 10426–10429.
34. Otzen,D.E., Itzhaki,L.S., elMasry,N.F., Jackson,S.E. and Fersht,A.R. (1994) Structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding. *Proc. Natl Acad. Sci. USA*, **91**, 10422–10425.