

# The Diamond STING server

Goran Neshich\*, Luiz C. Borro, Roberto H. Higa, Paula R. Kuser,  
Michel E. B. Yamagishi, Eduardo H. Franco, Joao N. Krauchenco, Renato Fileto,  
André A. Ribeiro, George B. P. Bezerra, Thiago M. Velludo, Tomás S. Jimenez,  
Noboru Furukawa<sup>2</sup>, Hirofumi Teshima<sup>2</sup>, Koji Kitajima<sup>2</sup>, Abdulla Bava<sup>2</sup>, Akinori Sarai<sup>2</sup>,  
Roberto C. Togawa<sup>1</sup> and Aduino L. Mancini

Núcleo de Bioinformática Estrutural, Embrapa/Informática Agropecuária, Campinas, Brazil, <sup>1</sup>Laboratório de Bioinformática, Embrapa/Recursos Genéticos e Biotecnologia, Brasília, Brazil and <sup>2</sup>Department of Bioscience and Bioinformatics, Kyushu Institute of Technology (KIT), Iizuka, Japan

Received March 11, 2005; Revised and Accepted March 14, 2005

## ABSTRACT

**Diamond STING is a new version of the STING suite of programs for a comprehensive analysis of a relationship between protein sequence, structure, function and stability. We have added a number of new functionalities by both providing more structure parameters to the STING Database and by improving/expanding the interface for enhanced data handling. The integration among the STING components has also been improved. A new key feature is the ability of the STING server to handle local files containing protein structures (either modeled or not yet deposited to the Protein Data Bank) so that they can be used by the principal STING components: <sup>Java</sup>Protein Dossier (<sup>PD</sup>) and STING Report. The current capabilities of the new STING version and a couple of biologically relevant applications are described here. We have provided an example where Diamond STING identifies the active site amino acids and folding essential amino acids (both previously determined by experiments) by filtering out all but those residues by selecting the numerical values/ranges for a set of corresponding parameters. This is the fundamental step toward a more interesting endeavor—the prediction of such residues. Diamond STING is freely accessible at <http://sms.cbi.cnptia.embrapa.br> and <http://trantor.bioc.columbia.edu/SMS>.**

## INTRODUCTION

The first appearance of the STING on the internet was in 1998. STING was hosted by the Protein Data Bank (PDB) server,

located at that time at the Brookhaven National Laboratory. The initial goal of the STING was to provide a simple graphical environment for the analysis of the sequence and structure by coupling the information from those two domains. To our surprise, such a simple idea quickly attracted the attention of many users. The addition of a few extra STING components further triggered the interest of the scientific community although STING use was limited to the public PDB (1) files and only a few categories of the structural analysis involving description of the interfaces between protein chains, hydrogen bond network description and simple statistics on local residue environment. Gradually, STING was expanded by including many more components for a complex analysis (2–8) of the structure, function, stability and binding as well as for the relationship between them. During this period, STING also got several mirror sites which conveniently covered different geographical regions of the users around the world.

We describe in this paper some of the new features available in Diamond STING Suite, the four new per-residue parameter classes and a whole new per-chain descriptor class, as well as the higher level of integration among existing and improved STING components. We also make a mention of our effort toward the integration of structural information and thermodynamic data of proteins, which is crucial for the understanding of the mechanism of protein stability.

Also, we emphasize here that the main difference that continues to work in favor of the STING and distinguishes it from other servers in this category, which are also available on the Internet (9), is the capability to present the largest number of descriptors for the sequence, structure, function, stability and binding in a concise and visually compelling manner, as well as the ability to select/focus for/on those residues that satisfy a user defined parameter/descriptor values. As we proceed to further explain in this paper, this feature presents answers to some very interesting and

\*To whom correspondence should be addressed. Tel: +55 19 3789 5774; Fax: +55 19 3289 9594; Email: neshich@cnptia.embrapa.br

important questions, such as: is there a set of parameters (protein structure descriptors) that can define UNIQUELY an amino acid ensemble coinciding with the active site of a given protein or coinciding with amino acids identified experimentally as crucial for the folding/stability.

## DIAMOND STING INTRINSICS

The Diamond STING components have achieved a higher level of integration among themselves when compared with the Gold STING version. As before, the main entry components in the STING Suite are the Sting Millennium (2,3), which is basically the molecular sequence and structure viewer, the <sup>Java</sup>Protein Dossier (<sup>J</sup>PD) (4) with the Java display of the STING\_DB parameters and the STING Report (8) which compiles into a single HTML document a report now containing 23 images showing parameters in context designed plots together with 17 tables. STING Report displays focused information in per-residue fashion. However, STING Report is an exhaustive, yet static presentation of the structure/function descriptors for a given residue; so a user may be able to handle all the parameters in a dynamic way by selecting the residues, e.g. those described by the parameters/descriptors within a certain user defined value range. In such a case, the STING suite entry component should be the <sup>J</sup>PD.

All the other STING components can be invoked from the 'Modules' and 'Images' menu at the sequence frame of the STING Millennium. However, a user can opt to go directly to the specific STING component (say, Graphical Contacts) and focus only on the type of information that this particular component offers.

The STING component integration is elevated by allowing a user to always receive focalized data about selected residues from any of the STING components if a residue in one of them has already been selected and/or clicked on.

## STING\_DB UPDATE AND RENEWAL

The STING\_DB is regularly updated in synchrony with the PDB updates. This happens once a week. All the parameters for the new PDB entries are calculated by the STING server. However some missing items, namely the HSSP (10–12) related parameters are usually not available until a couple of weeks (or even a couple of months) after the new structure has been published at the PDB. They will be added as soon as the STING\_DB generator identifies that the new HSSP release contains the data for this particular PDB ID.

In addition, with any new HSSP version and new SwissProt (13) version, all the data on conservation have to be re-calculated for the entire STING\_DB. Since this puts an enormous load on our CPUs, the STING\_DB renewal is done every two months.

## NAVIGATING DIAMOND STING

Some of the Diamond STING options require a basic knowledge of the nature of the parameter chosen for inspection/calculation. In addition, some STING components require a user to be familiar with the file input options and output

interpretation. In order to minimize the time required for a successful STING use, we made four different entry points a user may consult for a quicker comprehension on how to interpret and proceed with the data. These four options are: the 'Hypervisual' java guide with easy to use and quick to access information about any STING component and/or parameter calculated and stored in the STING\_DB, the extensive HELP pages which are content sensitive, the frequently asked questions (FAQ) with which many of the users are familiar and finally the <sup>J</sup>PD and STING Report Legends containing a graphic description of the parameters, which when clicked on would display a corresponding parameter help page.

We found that the new users appreciated the combination of three entry HELP points but always started with the 'Hypervisual' java guide and then continued with the extensive use of the content sensitive help pages.

## DIAMOND STING HANDLING OF PUBLIC AND LOCAL PDB FORMATTED FILES

Generally, STING operates with both PDB public files and local files in PDB format. However, in order to properly handle STING\_DB parameters for a local file, STING needs to pre-calculate those. Similarly, the STING <sup>J</sup>PD can handle both public and local files and this can be done for a single structure as well as for two structurally aligned proteins.

### Interactive and 'batch' job modes

In Diamond STING, a user can initiate STING Millennium by using one of the three possible input file routes: a public PDB formatted file available at the PDB (a user can inspect any or all STING parameters as they are pre-calculated), a local PDB-formatted file (this will initiate STING relatively quickly (in ~30 s at most) but a user will only be able to inspect a reduced number of STING parameters, and finally a TGZ file containing all STING parameters pre-calculated by the STING server upon request. The request to generate the TGZ file is a part of the separate form at STING Millennium entry page and it is one of the best in the Diamond STING offerings.

### Advanced input options for local files in <sup>J</sup>PD

- (i) For the single structure, a user may submit a job to the STING Server and obtain in his/her e-mail box a message with attached TGZ file containing all parameters that <sup>J</sup>PD can display and analyze.
- (ii) For two structurally aligned files the <sup>J</sup>PD can handle both two publicly available (PDB deposited) files or two local (non public) PDB formatted files or a combination of a public and a local file. If there are two local files, a user needs to submit to the STING server two separate requests and obtain two TGZ files before doing a structural alignment.

### Queuing job requests

Currently the STING server has three major tasks to accomplish: the STING\_DB updating (a weekly activity), the STING\_DB renewal (a bi-monthly activity) and a job

**Table 1.** Complete list of STING\_DB parameters reported in 'per-residue' fashion) including newly added ones (11,29–32) accessible in Diamond STING by means of <sup>1</sup>PD and STING Report

- 
1. ResBoxes
  2. Intra-chain atomic contacts [ITC] and
  3. The inter-chain atomic contacts [IFC]
  4. ITC contacts energy and
  5. IFC contacts energy
  6. Interface area [IF]
  7. Water contacting [WC]
  8. Ligand pocket forming [LP] and
  9. Surface forming [SF] residues
  10. Prosite
  11. ProTherm (14)
  12. Secondary structure indicator [PDB]
  13. Secondary structure indicator [DSSP](18)
  14. Secondary structure [STRIDE](19)
  15. Multiple occupancy
  16. Temperature factor
  17. Hot spots
  18. Sequence conservation [HSSP]
  19. Sequence conservation [SH<sub>2</sub>Q<sup>2</sup>]
  20. Solvent accessibility
  21. Dihedral angles
  22. Pockets/cavities
  23. Electrostatic potential
  24. Hydrophobicity
  25. Curvature
  26. Distance from the N-/C-terminal
  27. Density
  28. Sponge
  29. Order of cross presence
  30. Order of cross link
  31. Rotamers
  32. Space clash
- 

processing that contains a request for calculating all the STING parameters for the local user files.

In order to optimize/maximize the CPU usage at our lab, we have now implemented the STING\_Queue procedure that basically saves all job requests in the list, which is then executed by CRON at indicated hours [normally identified (by constant monitoring) as the lower CPU activity period]. Currently, we are dedicating four CPUs for this activity, a capacity that should provide enough processing power for generating all STING parameters for ~300 structures per day (each structure of ~200 AA).

### ADDITIONAL STRUCTURE/FUNCTION PARAMETERS IN DIAMOND STING

We have added some extra parameters into the STING database (8).

In Table 1, we show a complete list including the newly added structure/function parameters (in their order of appearance in the actual <sup>1</sup>PD window). All these parameters are reported in per-residue fashion. In Table 2, per-chain parameters that are introduced in Diamond STING are listed.

The total number of available numerical values for structure/function parameters being reported by Diamond STING is brought to 306. Although we have a total of 98 STING\_DB parameters belonging to 30 distinct parameter classes, some of them are calculated with a variety of default conditions (the variable volume size for a probing sphere, the atom at

**Table 2.** List of newly added parameters reported in 'per-chain' fashion, also accessible in Diamond STING by means of <sup>1</sup>PD and STING Report

- 
- Total volume of the chain
  - Total surface area of the chain
  - Ratio of volume to surface
  - Ratio of total surface to surface at interface
  - Total number of
    - H-bonds
    - salt bridges
    - hydrophobic interactions
    - cystein bridges
    - aromatic stacking interactions
  - Total energy of established contacts for the chain
  - Energy of the contacts divided by the volume
  - Total surface area occupied by hot spots
  - Hot spot area divided by total surface area of molecule
  - Total pocket surface area
  - Total pocket volume area
  - Total cavity volume
- 

which the center of the probing sphere is placed or the variable size for a sliding window).

One of the new features in Diamond STING is the integration between structural parameters and thermodynamic parameters in ProTherm (14), which compiles a large amount of experimental thermodynamic data on protein stability and the effects of amino acid mutations. Thermodynamic data are crucial for the understanding of the mechanism of protein stability and the function of proteins such as enzymatic reaction and ligand binding. We have created links from per-residue structural parameters to the corresponding thermodynamic data for the mutations of that residue, as well as the opposite link from particular thermodynamic data in ProTherm to STING Report. This will enable users to interpret the structural characteristics around that residue in terms of thermodynamic effects of the mutation, as well as to interpret the thermodynamic data in terms of structural characteristics (see for an example the <sup>1</sup>PD description of the 1stn.pdb and corresponding ProTherm information).

### Selecting parameter range in Diamond STING

The select feature continues to be one of the major pillars of the STING Suite. In <sup>1</sup>PD, residue selection can be performed according to multiple criteria. The selection of amino acids can be made with any combination of conditions, permitting powerful identification of functionally/structurally important regions or sites. Consequently in <sup>1</sup>PD a user can make decisions, backed by evidence from the set of protein descriptors, about the possible role of specific amino acids in defining the function of the protein. At the same time, a user may also see the candidate residues which contribute critically to the protein stability or binding or even judiciously evaluate the type of effect one mutation might have on function/stability/binding.

### EXAMPLES OF THE DIAMOND STING APPLICATION

In order to emphasize the Diamond STING capabilities and its potential impact on the biological side of a problem under study in the following exercises, we present two interesting examples.

In the first example, we asked a very simple question: is there a set of parameters (protein structure descriptors) that can define UNIQUELY an amino acid ensemble coinciding with the active site of a given protein? An *in silico* experiment done by <sup>J</sup>PD and STING\_DB allowed us to test this hypothesis and to confirm the existence of such a set of parameters on HIV-integrase.

The HIV-1 integrase is an essential enzyme in the life cycle of the virus responsible for catalyzing the insertion of the viral genome into the host cell chromosome; it provides an attractive target for an antiviral drug design. The evidence was obtained from site-directed mutagenesis experiments that even the most conservative substitutions of any of the three absolutely

conserved carboxylate residues, D64, D116 and E152 (the so-called DDE motif), abolished catalytic activity (15).

By combining selected ranges for a number of structure parameters, we have obtained at the end of a Select procedure (see Figure 1) an ensemble of amino acids which coincides with the three critical residues identified experimentally as the active site of the HIV integrase. The Select procedure used the following parameters and their numerical values/regions:

- (i) Conservation: SH<sub>2</sub>Q<sup>s</sup>, relative entropy <30
- (ii) Physical-chemical: electrostatic potential, average <-20 kT/J/mol
- (iii) Geometric: pocket/cavity in complex, volume >0.



**Figure 1.** The HIV-integrase (1biu.pdb) active site residues: Asp<sub>64</sub> (Cyan), Asp<sub>116</sub> (Magenta) and Glu<sub>152</sub> (purple). Those three residues were selected by means of <sup>J</sup>PD and its select feature, using a set of parameters and the range for corresponding values: (i) conservation: SH<sub>2</sub>Q<sup>s</sup>: relative entropy <30; (ii) physical-chemical: electrostatic potential: average <-20 kT/J/mol; (iii) geometric: pocket/cavity in complex: volume >0.



**Figure 2.** The structure of acylphosphatase, AcP (2acy.pdb) and the key residues Tyr<sub>11</sub> (green), Pro<sub>54</sub> (cyan) and Phe<sub>94</sub> (magenta) found from the transition-state analysis to be crucial for the folding of this protein. While all three residues show cross presence  $\geq 1$ , only the Tyr<sub>11</sub> and the Pro<sub>54</sub> have long-range contacts. Those three residues were selected by means of <sup>1</sup>PD and its select feature, using a set of parameters: (i) total unused contact energy; (ii) density; (iii) sponge; (iv) cross presence order; (v) secondary structure element; (vi) conservation Sh<sub>2</sub>Q<sup>3</sup>/evolutionary pressure; (vii) electrostatic potential @ Ca (viii) electrostatic potential @ surface. For the range of corresponding values see Table 3.

In the second exercise, we asked a similar question: is there a set of parameters (protein structure descriptors) that can define UNIQUELY an amino acid ensemble coinciding with the residues identified as essential in the folding process of a given protein?

We used the protein 2acy (see Figure 2) and again we were able to identify the very same amino acid triplet that Vendruscolo (16) and colleagues have identified in their experiment as crucial for the folding process. An important detail here is that one of the parameters used was the order of cross link. The cross links are defined in STING as the contacts [any type from a possible five classes: (i) hydrophobic interaction, (ii) hydrogen bonding, (iii) aromatic stacking, (iv) salt bridging and (v) cystein-bridging] established among residues that are far apart in the protein primary sequence, but are close in its 3D fold. The order of cross link is identified as a number of such cross links established among independent stretches of sequence (in STING we used three sizes for a sequence stretch: 15, 20 or 30). Only a single occurrence is counted for the order of cross link even though several could be registered (a central amino acid can make more than one contact and each one of them can be established with a different amino acid belonging

to the same stretch of probing sequence size). The higher the order, the greater the importance that residue may have for the protein folding/stability. This specific STING parameter is calculated by varying three input parameters: (i) the size of the sequence stretch separating the residues in contact (15, 20 or 30); (ii) the size of the radius of the probing sphere within which the contacts are counted (3.5, 5 and 8.5 Å); (iii) the center of the probing sphere [C<sub>α</sub>, C<sub>β</sub> or last heavy atom (LHA) in a side chain].

In Table 3 we show the set of parameters and values used to obtain a positive result in searching for the amino acid triplet in 2acy.pdb structure.

## CONCLUDING REMARKS

Diamond STING is described here as an improved interactive suite of tools for browsing the structure related STING\_DB. The STING\_DB has been expanded and the new parameters introduced contribute significantly to the understanding of phenomena such as protein stability. They also describe more fully the protein function. Furthermore, the integration of structural information with the thermodynamic data of

**Table 3.** List of parameters and their numerical values/ranges used to filter out all amino acids from the 2acy.pdb except for the active site ones

Parameter name	Parameter value (range) or description
Total unused contact energy	>72 Kcal/mol
Density	>1.08 for probing sphere 3A centered at Ca
Sponge	>0.75 for probing sphere 3A centered at Ca
Cross presence order	≥1,1,2 (at C <sub>α</sub> , C <sub>β</sub> and LHA)
Secondary structure element	β sheet
Conservation Sh <sub>2</sub> Q <sup>2</sup> /evolutionary pressure	<28
Electrostatic potential at Ca	≥7
Electrostatic potential at surface	≤0

proteins and mutants enables us to establish the relationship between structure and thermodynamics and also helps us decipher the function of proteins.

There is an increased number of problems that STING tools can help tackle as they are maturing. With this maturation process we are closer to the general goal of the STING project: capture and describe in full detail nuances among amino acids flocking in regions with specific importance to the protein stability and function. Diamond STING is presented here as an intermediary but becomes a mature and powerful tool toward the final stage of its own development, having already the ability to identify with precision the flocks of residues with a very unique role in a protein: the active site amino acids and the folding essential amino acids. This step is the fundamental one toward the more interesting quest: predicting such residues. The discussion on the details of how general is the finding that the specific parameter set in a determined range of values can be used for the whole family of proteins in order to filter out all other amino acids but the ones belonging to the active site or are essential for the folding; and what this finding can achieve in terms of learning and predicting such important sites in the proteins is beyond the scope of this publication and will be discussed in a separate one.

It is clear that the STING Suite of programs may be used both in research and teaching. Diamond STING plots offer visual but powerful tools for detecting the constellations of amino acids with peculiar characteristics within a protein structure or in the structural alignment of two proteins. Such ability is crucially dependent on the Select feature in <sup>3</sup>PD and consequently is recommended as the filter for accepting and/or rejecting a hypothesis described in an experiment or in the literature.

Although we are adding many new parameters and features to the STING over the years, our initial goal of having most parameters at the same place represented in an easy to see and easy to interpret fashion has not yet been challenged. On the contrary, the more the number of features/parameters added, the greater the grasp of the capability of the STING in composing with ever higher resolution.

## FUTURE DEVELOPMENTS

Diamond STING is already being expanded and the STING\_DB being ported into relational database.

The new version—STAR STING—will be able to accept users' local files in a much larger quantity (we expect to double the CPU capacity dedicated exclusively to local file requests). By porting the STING\_DB into relational database we will be able to ask questions that are now only answered within a single PDB file. Those questions can be formulated in a way similar to this, for example: which are the protein structures available in the PDB that do possess a cavity of the determined volume, and the residues forming the cavity are conserved to a determined value of the relative entropy/evolutionary pressure, and the cavity forming residues have the electrostatic potential at the surface which is in a negative range and those proteins also do have at least two residues in their structure that are identified as having the Order of Cross Links of at least {1, 1, 1}. What the above question has just defined is the search across the PDB for the binding site that can accommodate a ligand of a determined size/volume, formed by amino acids which are very conserved, with the negative EP at the pocket surface (providing for the adequate activity of the protein) and that protein probably has a pocket with the mouth too tight for the ligand penetration. Consequently, it has to open for the ligand binding by some gating mechanism that involves an amino acid which can make the cross links across the pocket mouth. Upon a ligand binding, that residue will be responsible for keeping the pocket closed until the chemical action is completed. This residue is definitely making at least one Cross Link of Order equal or higher than {1, 1, 1} (C<sub>α</sub>, C<sub>β</sub>, LHA).

STAR STING should also present to the user an option for the graphical molecular viewer. STING is currently dependent on Chime. However, we would like to make it more OS independent and for that reason we plan to integrate STING with the JMOL (<http://jmol.sourceforge.net>). By making available JMOL as the STING molecular viewer, we also hope to get STING functional on Mac stations, an important achievement in light of widespread use of this platform.

With respect to new parameters that will be added to the STAR STING, we are working in complementing the current classes of the contacts with the important missing one: Protein–ligand contacts. In addition, we should be adding the moving protein parts identification (by identifying the residues involved in protein domain movements), as well as the preferred local environment reported in per-residue manner.

We have moved toward the integration between structural information and thermodynamic data of protein stability, by linking between STING and ProTherm. We plan to integrate thermodynamic data for molecular interactions, such as protein–nucleic acid and protein–ligand interactions, along this direction.

Finally, we are working to add more convenience to the STING Report in the STAR STING by adding a capability of creating the Reports in PDF format with a page break feature that allows for the figures being always integral and presented in a single page, as well as adding an important visual description [such as MolScript (17)] of the residue being explored by the STING Report.

## ACKNOWLEDGEMENTS

This work was supported in part by the following grants: FAPESP 01/08895-0, FINEP 1945/01 and CNPq 401695/

2003–2004. The authors wish to express special thanks to Phil Bourne and Wolfgang Bluhm from RCSB/PDB-San Diego, Jose Valverde from CNB-Madrid-Spain, Oscar Grau from LaPlata University in Argentina and Megan Restuccia from Columbia University at New York City for their collaboration in testing and maintaining STING at the respective STING mirror sites. Funding to pay the Open Access publication charges for this article was provided by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo), FINEP (Financiadora de Estudos e Projetos) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico).

*Conflict of interest statement.* None declared.

## REFERENCES

- Berman, H.M., Bourne, P.E. and Westbrook, J. (2004) The Protein Data Bank: a case study in management of community data. *Current Proteomics*, **1**, 49–57.
- Neshich, G., Togawa, R.C., Mancini, A.L., Kuser, P.R., Yamagishi, M.E.B., Pappas, G., Jr, Torres, W.V., Campos, T.F., Ferreira, L.L., Luna, F.M. *et al.* (2003) STING Millennium: a web based suite of programs for comprehensive and simultaneous analysis of protein structure and sequence. *Nucleic Acids Res.*, **31**, 3386–3392.
- Higa, R.H., Togawa, R.C., Montagner, A.J., Palandrani, J.C., Okimoto, I.K., Kuser, P.R., Yamagishi, M.E., Mancini, A.L. and Neshich, G. (2004) STING Millennium Suite: integrated software for extensive analyses of 3d structures of proteins and their complexes. *BMC Bioinformatics*, **5**, 107–115.
- Neshich, G., Rocchia, W., Mancini, A.L., Yamagishi, M.E., Kuser, P.R., Fileto, R., Baudet, C., Pinto, I.P., Montagner, A.J., Palandrani, J.F. *et al.* (2004) JavaProtein Dossier: a novel web-based data visualization tool for comprehensive analysis of protein structure. *Nucleic Acids Res.*, **32**, W595–W601.
- Higa, R.H., Montagner, A.J., Togawa, R.C., Kuser, P.R., Yamagishi, M.E., Mancini, A.L., Pappas, G., Jr, Miura, R.T., Horita, L.G. and Neshich, G. (2004) ConSSeq: a web-based application for analysis of amino acid conservation based on HSSP database and within context of structure. *Bioinformatics*, **20**, 1983–1985.
- Higa, R.H., Oliveira, A.G., Horita, L.G., Miura, R.T., Inoue, M.K., Kuser, P.R., Mancini, A.L., Yamagishi, M.E., Togawa, R.C. and Neshich, G. (2004) Defining 3D residue environment in protein structures using SCORPION and FORMIGA. *Bioinformatics*, **20**, 1989–1991.
- Mancini, A.L., Higa, R.H., Oliveira, A., Dominiquini, F., Kuser, P.R., Yamagishi, M.E., Togawa, R.C. and Neshich, G. (2004) STING Contacts: a web-based application for identification and analysis of amino acid contacts within protein structure and across protein interfaces. *Bioinformatics*, **20**, 2145–2147.
- Neshich, G., Mancini, A.L., Yamagishi, M.E., Kuser, P.R., Fileto, R., Pinto, I.P., Palandrani, J.F., Krauchenco, J.N., Baudet, C., Montagner, A.J. *et al.* (2005) STING Report: convenient web-based application for graphic and tabular presentations of protein sequence, structure and function descriptors from the STING database. *Nucleic Acids Res.*, **33**, D269–D274.
- Galperin, M.Y. (2005) The molecular biology database collection: 2005 update. *Nucleic Acids Res.*, **33**, D5–D24.
- Sander, C. and Schneider, R. (1991) Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins*, **9**, 56–68.
- Schneider, R., de Daruvar, A. and Sander, C. (1997) The HSSP database of protein structure–sequence alignments. *Nucleic Acids Res.*, **25**, 226–230.
- Schneider, R. and Sander, C. (1996) The HSSP database of protein structure–sequence alignments. *Nucleic Acids Res.*, **24**, 201–205.
- Gasteiger, E., Jung, E. and Bairoch, A. (2001) Swiss-Prot: connecting biological knowledge via a protein database. *Curr. Issues Mol. Biol.*, **3**, 47–55.
- Bava, K.A., Gromiha, M.M., Uedaira, H., Kitajima, K. and Sarai, A. (2004) ProTherm, version 4.0: thermodynamic database for proteins and mutants. *Nucleic Acids Res.*, **32**, D120–D121.
- Goldgur, Y., Dyda, F., Hickman, A.B., Jenkins, T.M., Craigie, R. and Davies, D.R. (1998) Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. *Proc. Natl Acad. Sci. USA*, **95**, 9150–9154.
- Vendruscolo, M., Paci, E., Dobson, C.M. and Karplus, M. (2001) Three key residues form a critical contact network in a protein folding transition state. *Nature*, **409**, 641–645.
- Kraulis, P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.*, **24**, 946–950.
- Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
- Heinig, M. and Frishman, D. (2004) STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.*, **32**, W500–W502.