

Database

Open Access

## ASAView: Database and tool for solvent accessibility representation in proteins

Shandar Ahmad\*<sup>1</sup>, Michael Gromiha<sup>2</sup>, Hamed Fawareh<sup>3</sup> and Akinori Sarai<sup>1</sup>

Address: <sup>1</sup>Department of Biochemical Engineering and Science, Kyushu Institute of Technology, Iizuka 820 8502, Fukuoka-ken, Japan, <sup>2</sup>Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-41-6 Aomi, Koto-ku, Tokyo, Japan and <sup>3</sup>Computer Science Department, Zarka Private University, Zarka 13110, Jordan

Email: Shandar Ahmad\* - shandar@bse.kyutech.ac.jp; Michael Gromiha - michael-gromiha@aist.go.jp; Hamed Fawareh - fawareh@zpu.edu.jo; Akinori Sarai - sarai@bse.kyutech.ac.jp

\* Corresponding author

Published: 01 May 2004

Received: 20 November 2003

BMC Bioinformatics 2004, 5:51

Accepted: 01 May 2004

This article is available from: <http://www.biomedcentral.com/1471-2105/5/51>

© 2004 Ahmad et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

### Abstract

**Background:** Accessible surface area (ASA) or solvent accessibility of amino acids in a protein has important implications. Knowledge of surface residues helps in locating potential candidates of active sites. Therefore, a method to quickly see the surface residues in a two dimensional model would help to immediately understand the population of amino acid residues on the surface and in the inner core of the proteins.

**Results:** ASAView is an algorithm, an application and a database of schematic representations of solvent accessibility of amino acid residues within proteins. A characteristic two-dimensional spiral plot of solvent accessibility provides a convenient graphical view of residues in terms of their exposed surface areas. In addition, sequential plots in the form of bar charts are also provided. Online plots of the proteins included in the entire Protein Data Bank (PDB), are provided for the entire protein as well as their chains separately.

**Conclusions:** These graphical plots of solvent accessibility are likely to provide a quick view of the overall topological distribution of residues in proteins. Chain-wise computation of solvent accessibility is also provided.

### Background

Key functional properties of proteins and so-called active amino acid sites strongly correlate with amino acid solvent accessibility or accessible surface area (ASA) [1,2]. For example, DNA-binding probability of a residue is significantly higher for residues with higher solvent accessible area [2]. Recognizing the importance of ASA, several groups have developed methods for predicting it from amino acid sequence [3-7] similar to secondary structure prediction. We have recently developed a prediction server, which provides real-valued predictions of solvent accessibility rather than burial categories [8].

Although useful methods for representing secondary structures have been developed and are widely used, good tools for representing solvent accessibility have been conspicuously missing. As a case in point *PDBsum* carries plots of secondary structure [9] but gives no mention of accessibility, which may be even more important for the estimate of active sites [10]. We have therefore developed a method to provide quick visualization of solvent accessibility in terms of a compact spiral plot, which may reveal deep insights into protein structure along with secondary structure, composition and other summary information. We also developed a tool to generate postscript graphical

output of solvent accessibility from solvent accessibility data in different file formats such as DSSP and other programs. Further, the output obtained from the real-value prediction can also be used to display the ASA. Postscript graphics produced by our program have been converted to acrobat PDF and PNG formats using Latex2HTML tools [11].

### Implementation

This so-called *ASAView* algorithm involves carrying out the following steps:

1. Calculation of the solvent accessibility of each amino acid residue: If the complete three-dimensional structures are known, ASA values may be calculated using programs such as ACCESS [12], DSSP [13], ASC [14], NACCESS [15] and GETAREA [16]. The ASA values can also be obtained directly from the DSSP database, if the corresponding PDB code is known. GETAREA gives the ASA online and executable files are available for other programs. We have used DSSP for calculating ASA for all proteins contained in the February 2003 release of PDB. However, one can use the computer program to get these plots for any protein, which is freely available from the corresponding author. If ASA values are taken from a prediction, a real-value prediction of ASA is necessary, as category predictions (e.g., classification as buried or exposed) cannot be plotted. Further, the ASA values obtained from the real-value prediction algorithm [8] can also be used as the ASA inputs for *ASAView*.

2. Representation of each amino acid residue by a filled circle: Equivalent radii are calculated from the ASA values obtained in step 1; consequently, the size of each circle representing a residue is proportional to its relative solvent accessibility. If the available ASA values are not in relative scale (as is mostly the case), the absolute ASA values are changed to relative values using appropriate scaling factors [2], thus normalizing the view for relative exposed surfaces rather than absolute area. For the scaling the ASA of the extended states of Ala-X-Ala for every residue X are used (assuming that the absolute values include side chain and backbone surface area). These values are (in Å<sup>2</sup>) 110.2 (Ala), 144.1 (Asp), 140.4 (Cys), 174.7 (Glu), 200.7 (Phe), 78.7 (Gly), 181.9 (His), 185.0 (Ile), 205.7 (Lys), 183.1 (Leu), 200.1 (Met), 146.4 (Asn), 141.9 (Pro), 178.6 (Gln), 229.0 (Arg), 117.2 (Ser), 138.7 (Thr), 153.7 (Val), 240.5 (Trp), and 213.7 (Tyr) respectively.

3. Color-coding is assigned to the residues: In the online version, gray, red, blue and green are used to represent hydrophobic, negatively charged, positively charged and polar neutral residues, respectively. Cystein residues are shown in yellow color due to its unique properties.

4. A residue number, a residue name, and an equivalent radius now identify each residue. These residues are then sorted in the order of their equivalent radii, calculated in step (2).

5. A two-dimensional spiral plot in postscript language is then generated through appropriate placement of the circles representing amino acid residues. The residue with the smallest relative ASA is placed at the origin of the spiral, and residues with larger ASAs are successively placed on the spiral, whose radius is properly scaled.

6. The size of the spiral plot is forced to remain within one page and hence a protein with large number of residues will have a smaller size of circles for the same ASA. For the actual value of ASA, bar plots (see next point) or the textual data can be used as a reference.

7. Bar plots are also generated for the protein by retaining the order of residues as they occur in the original input file. This will show the ASA of residues for a protein sequence, similar to hydrophobicity plot [17,18].

*ASAView* software also provides several additional features for better visualization:

1. Input file formats: To generate images, *ASAView* can make use of ASA inputs in four different formats:

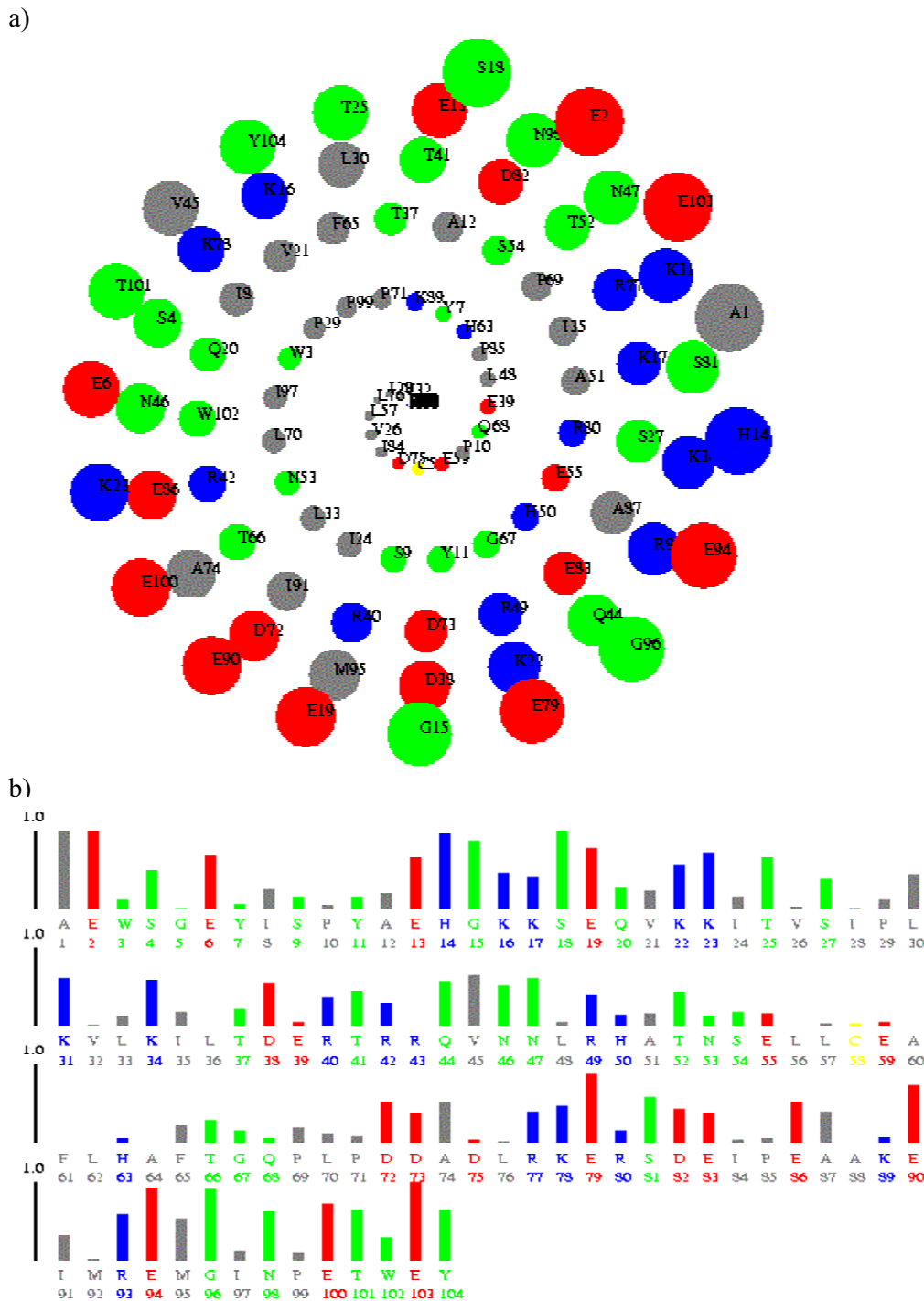
- (a) DSSP: Files from DSSP, the most popular database of secondary structure and solvent accessibility, may be directly input into *ASAView* in the form of PDB code.

- (b) RVP: Real-value prediction obtained from RVP-Net may also be directly input into *ASAView* [8].

- (c) Percentages: Solvent accessibility values obtained by any other methods (ASC, GETAREA, ACCESS, Naccess) may be used for plots, provided they are written in a two column format in which the first column contains a list of residues (single letter codes), and the second column contains the corresponding solvent accessibility values as percentages. This will help to compare the ASA from different methods, visually.

- (d) Relative ASA: Relative ASAs normalized to a value of 1 are the default input for this program.

2. Image rescaling: Although postscript is a vector graphic method of generating images, we also provide an "Image Shrinking" option to reduce the size of plotted images. This is especially desirable when the number of residues is large.



**Figure 1**

ASAView of a DNA binding protein (PDB code ICMA chain A). (a) The spiral view, which shows amino acid residues of ICMA, in the order of their solvent accessibility. Most accessible residues come on the outermost ring of this spiral. Blue, red, green, gray colors are used for positively charged, negatively charged, polar and non-polar residues respectively. Yellow color is used for Cystein residues. Radius of the solid circles representing these residues corresponds to the relative solvent accessibility (b) Solvent accessibility of residues, with residues arranged in the original order as in their PDB file. Length of the bar represents the ASA in units relative to extended state ASA of that residue.

3. A selected number of most exposed residues (those with the largest ASA values) may be plotted to avoid cluttering the view in a large protein.

#### Database design and update plan

ASA values for the entire protein databank, their post-script plots and PDF and PNG formatted image files are stored in compressed flat and image files. Upon receiving a query request these compressed files are expanded and served through links which are generated on the fly. New paths to the resulting image and textual data are also created in the final step. If a wrong PDB code is entered or if the database does not have a data corresponding to the submitted query, a message to this effect is displayed. A local mirror of Protein Data Bank is being maintained and updated as part of database included in Bioinfo Bank [19]. Updates of ASAView database are planned to be undertaken upon every update of this PDB mirror.

#### Results and discussion

Snapshots generated by ASAView are shown in Figure 1 (a and b). The plots for proteins and their chains are available online [20] and one can obtain a plot of these proteins by simply entering the PDB code for that protein [21]. On the other hand, we have also implemented a feature in the server by which coordinate files in PDB format can be uploaded and ASA calculations will be performed by the server and a graphical plot will be provided. Graphical plots of solvent accessibility have several applications in molecular biology. Especially, the spiral plot can be used to immediately provide an overall visual summary of the protein. For example, a plot with a large number of positively charged residues instantly tells that the given protein is charged as such. Similarly, concentration of gray circles suggests hydrophobic nature of proteins. This kind of information may not be quickly seen from the overall composition as more than one residue make for the hydrophobic or electrostatic charge property of the protein. Outward distribution of higher solvent accessible residues also provides the view of distribution of charged, hydrophobic or polar residues in different ranges of solvent accessibility. The information about the residues with similar ASA may be helpful for further analyzing the relative number and nature of contacts in protein structure.

Topological distribution of residues and packing density are qualitatively visible from the way residues are distributed in various ASA ranges. A tightly packed protein will have a large number of residues in the interior of the spiral plot and hence the ASAView spiral of such proteins will have a narrow thread of residues in its interior. A more loosely packed protein on the other hand will have few residues in the interior and relatively more residues with

higher solvent accessibility, which is visible from large number of circles having greater radii.

Possible active sites potentially lie in the higher accessibility region. Charged residues on the surface will fall on the outermost ring of the spiral and hence these plots automatically suggest potential binding sites of the protein.

With these applications of solvent accessibility plots, ASAView complements protein summary information such as *PDBbsum*. As solvent accessibility is an important property for predicting protein mutant stability [22-26], ASAView may be useful to gain insights about the mutant positions for the thermodynamic data available for proteins and mutants in ProTherm [27]. Thus ProTherm database has already been linked to ASAView, through automatically generated query hyperlinks.

#### Conclusions

A database and web server for graphical representation of solvent accessibility has been developed. This is expected to assist in structural analysis of the proteins, particularly for observing the topological distribution of residues in a nutshell.

#### Availability and requirements

The entire implementation of ASAView for all PDB proteins, as a whole or for an individual chain may be accessed at <http://www.netasa.org/asaview/>. Requirements for the use are simply the PDB code or the coordinate file.

#### Authors' contributions

Corresponding author (SA) conceived the project and implemented it with initial computational inputs from HF, under the project guidance of AS. MMG provided useful contributions in writing the manuscript, adding references and checked the errors in the website and the manuscript.

#### Acknowledgement

Corresponding author (S.A.) would like to acknowledge Advanced Technology Institute Inc., Tokyo for partially supporting this research.

#### References

1. Bartlett GJ, Porter CT, Borkakoti N, Thornton JM: **Analysis of catalytic residues in enzyme active sites.** *J Mol Biol* 2002, **324**:105-121.
2. Ahmad S, Gromiha MM, Sarai A: **Analysis and Prediction of DNA-binding proteins and their binding residues based on Composition, Sequence and Structural Information.** *Bioinformatics* 2004, **20**:477-486.
3. Rost B, Sander C: **Conservation and prediction of solvent accessibility in protein families.** *Proteins* 1994, **20**:216-226.
4. Cuff JA, Barton GJ: **Application of multiple sequence alignment profiles to improve protein secondary structure prediction.** *Proteins* 2000, **40**:502-511.
5. Pollastri G, Baldi P, Fariselli P, Casadio R: **Prediction of coordination number and relative solvent accessibility.** *Proteins* 2002, **47**:142-153.

6. Ahmad S, Gromiha MM: **NETASA: Neural network based prediction of solvent accessibility.** *Bioinformatics* 2002, **18**:819-824.
7. Ahmad S, Gromiha MM, Sarai A: **Real-value prediction of solvent accessibility from amino acid sequence.** *Proteins* 2003, **50**:629-635.
8. Ahmad S, Gromiha MM, Sarai A: **RVP-Net: online predictions of real-value accessible surface area of proteins from single sequences.** *Bioinformatics* 2003, **19**:1849-1851.
9. Lakowski RA: **PDBsum: summaries and analyses of PDB structures.** *Nucleic Acids Res* 2001, **29**:221-222.
10. Nielsen JE, Beier L, Otzen D, Borchert TV, Frantzen HB, Andersen KV, Svendsen A: **Electrostatics in the active site of an alpha-amylose.** *Eur J Biochem* 1999, **264**:816-824.
11. **Latex2html software** [<http://www.latex2html.org>]
12. Richmond TJ, Richards FM: **Packing of alpha-helices: geometrical constraints and contact areas.** *J Mol Biol* 1978, **119**:537-555.
13. Kabsch W, Sander C: **Dictionary of protein secondary structure: Pattern recognition of hydrogen-bond and geometrical features.** *Biopolymers* 1983, **22**:2577-2637.
14. Eisenhaber F, Argos P: **Improved strategy in analytical surface calculation for molecular system- handling of singularities and computational efficiency.** *J Comp Chem* 1993, **14**:1272-1280.
15. **NACCESS, Computer program, Department of Biochemistry and Molecular Biology** [<http://wolf.bi.umist.ac.uk/unix/naccess.html>]
16. Fraczekiewicz R, Braun W: **Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules.** *J Comp Chem* 1998, **19**:319-333.
17. Kyte J, Doolittle RF: **A simple method for displaying the hydrophobic character of a protein.** *J Mol Biol* 1982, **157**:105-132.
18. Ponnuswamy PK, Gromiha MM: **Prediction of transmembrane helices from hydrophobic characteristics of proteins.** *Int J Pept Protein Res* 1993, **42**:326-341.
19. **Bioinfo Bank, Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, Iizuka, Japan** [<http://gibk26.bse.kyutech.ac.jp/jouhou/>]
20. **ASAVIEW: Solvent accessibility graphics for proteins** [<http://www.netasa.org/asaview/>]
21. 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.
22. Gilis D, Rooman M: **Stability changes upon mutation of solvent-accessible residues in proteins evaluated by database-derived potentials.** *J Mol Biol* 1996, **257**:1112-1126.
23. Gilis D, Rooman M: **Predicting protein stability changes upon mutation using database-derived potentials: solvent accessibility determines the importance of local versus non-local interactions along the sequence.** *J Mol Biol* 1997, **272**:276-290.
24. Gromiha MM, Oobatake M, Kono H, Uedaira H, Sarai A: **Role of structural and sequence information in the prediction of protein stability changes: comparison between buried and partially buried mutations.** *Protein Engg* 1999, **12**:549-555.
25. Gromiha MM, Oobatake M, Kono H, Uedaira H, Sarai A: **Importance of surrounding residues for protein stability of partially buried mutations.** *J Biomol Struct Dyn* 2000, **18**:281-95.
26. Gromiha MM, Oobatake M, Kono H, Uedaira H, Sarai A: **Importance of mutant position in Ramachandran plot for predicting protein stability of surface mutations.** *Biopolymers* 2002, **64**:210-220.
27. Bava KA, Gromiha MM, Uedaira H, Kitajima K, Sarai A: **ProTherm, version 4.0: Thermodynamic Database for Proteins and Mutants.** *Nucleic Acids Res* 2004, **32**:D120-D121.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

