Accessibility and partner number of protein residues, their relationship and a webserver, ContPlot for their display

Arunay Pal1, Ranjit Prasad Bahadur1, Partha Sarathi Ray2 and Pinak Chakrabarti*1,2

Address: 1Department of Biochemistry, Bose Institute, P-1/12 CIT Scheme VIIIM, Kolkata 700 054, India and 2Bioinformatics Centre, Bose Institute, P-1/12 CIT Scheme VIIIM, Kolkata 700 054, India

Email: Arumay Pal - arumay@boseinst.ernet.in; Ranjit Prasad Bahadur - ranjtip_bahadur@yahoo.com; Partha Sarathi Ray - partharay18@gmail.com; Pinak Chakrabarti* - pinak@boseinst.ernet.in
* Corresponding author

Abstract

Background: Depending on chemical features residues have preferred locations – interior or exterior – in protein structures, which also determine how many other residues are found around them. The close packing of residues is the hallmark of protein interior and protein-protein interaction sites.

Results: The average values of accessible surface area (ASA) and partner number (PN, the number of other residues within a distance of 4.5 Å from any atom of a given residue) of different residues have been determined and a webserver, ContPlot has been designed to display these values (relative to the average values) along the protein sequence. This would be useful to visually identify residues that are densely packed, or those involved in protein-protein interactions. The skewness observed in the distribution of PNs is indicative of the hydrophobic or hydrophilic nature of the residue. The variation of ASA with PN can be analytically expressed in terms of a cubic equation. These equations (one for each residue) can be used to estimate the ASA of a polypeptide chain using the PNs of the individual residues in the structure.

Conclusion: The atom-based PNs (obtained by counting surrounding atoms) are highly correlated to the residue-based PN, indicating that the latter can adequately capture the atomic details of packing. The average values of ASA and PN associated with each residue should be useful in protein structure prediction or fold-recognition algorithm. ContPlot would provide a handy tool to assess the importance of a residue in the protein structure or interaction site.

Background

Hydrophobic interactions have long been hypothesized to play a dominant role in organizing and stabilizing the architecture of proteins [1,2]. This causes the non-polar side chains to segregate into the interior of globular proteins constituting the core, which is compact, densely packed and can be viewed as being solid-like [3]. The charged and polar side-chains remain exposed to the solvent. Thus the pattern of hydrophobic and polar residues may constitute the binary code responsible for the fold of a protein [4,5]. The dependence of the degree of burial of a residue on its hydrophobicity [6] would mean that its
solvent accessibility and the number of contacts with the surrounding residue would also show distinct pattern. Indeed, it has been shown that when the contact is defined in terms of the number of atoms (i.e., partner number) surrounding a given residue the above two parameters are also related [7]. Parameters such as these obtained from a database of known protein structures constitute the knowledge base for deriving scoring functions or statistical potentials used in protein fold recognition, usually by threading, concerned with protein-structure prediction for cases in which a target sequence does not have unambiguous sequence homology to any known structure [8-12]. These knowledge-based potential functions could be at the residue level or atom-based [13-15]. As the earlier work [7] defined the partner number (PN) in terms of atoms, the present study examines it in terms of residues and finds the relationship with the solvent accessible surface area (ASA). The analytical expression can be used to calculate the ASA corresponding to any observed PN, which can then be compared to the observed ASA to assess the importance of any residue at the binding site [16]. Finally, a webserver is described that scans a protein sequence in a structure to indicate the residues for which the two parameters deviated significantly from the corresponding average value for the same residue.

Results and discussion

Partner number of residues

The number of partner residues gives the quantitative measure of the packing of residues in the protein structures. Figure 1 and 2 show the distributions of the number of partner groups around the whole residue of a particular type or just its side chain, and follows the trend observed when this number was calculated counting the individual atoms around the central group [7]. For example, the largest residue Trp has the highest average partner number (12.6 and 8.8 for whole residue and side chain, respectively), whereas the smallest, Gly has the least (7.4). For residues other than Gly the difference in partner numbers, especially the ones involving the side chains. Considering Asp and Lys as examples for hydrophilic residues it can be seen that the distributions have positive skewness, with the peaks shifting towards lower partner numbers. In contrast, the distributions have negative skewness (peaks shifting towards higher partner numbers) for hydrophobic residues, such as Leu, Ile and Met.

Variation of the accessible surface area with the number of partner residues

Various contact measures, such as the one discussed in the previous section, have been used to evaluate the accuracy of three-dimensional protein models, as also the contact area between residues [21]. To relate the two we have plotted the variation of the mean values of the ASA of the residues (considering either the whole residue or only the side chain) with the partner numbers in Figure 3 for two representative cases (others are given in the Additional file 1, Figure S1). The dependence can be represented by a cubic equation (Table 1). Other curve fitting equations (exponential and quadratic) were also tried, but the R² values were poorer compared to those obtained from the cubic fitting. For a few residues and especially for the side chain, the fitted plots tend to indicate negative values of ASA at higher partner numbers and it may be better to level off the curves at ASA = 0 at such high values of PN. In the majority of the cases, extrapolation to x = 0 (i.e., no partner) leads to a value which is close to the value for the residue (X) obtained in the fragment Gly-X-Gly in an extended conformation. At this value of x, the calculated ASA is given by the value of b₀, which should be equivalent to the value of a₁ in the earlier work where an exponential equation was fitted to relate ASA to the atom-based PN [7]. Indeed, the correlation coefficients between the two sets of values are 0.90 and 0.98 considering the whole residue and the side chain, respectively. Considering the whole residue only three amino acids (Cys, Gln and Leu) have values of b₀ that differ from those provided by the Gly-X-Gly tripeptide by more than 15%. Considering only the side chain, the variation is in the range 10–15% for Ala, Cys and Thr. The deviation is the maximum for Cys and this is the only residue for which the fitted parameter b₁ is positive. The discrepancy must be due to the fact that Cys residues with free sulphydryl group, as well as those involved in disulfide bridges have been considered together.
Calculation of ASA of a residue (or the whole protein) using the equation involving the partner number

Advances have been made in the prediction of the ASA of amino acid residues in proteins [22]. Wodak and Janin [23] have developed an analytical approximation to ASA, expressed as a function of interatomic distances only. Using the equation derived here relating ASA and partner-number, it is possible to calculate the ASA ($<ASA>_{\text{calc}}$) corresponding to the average number of partners for the whole residue and the side chain. These estimates of average ASA compare very well with the means of the observed values ($<ASA>_{\text{obs}}$) (Table 2); the correlation coefficient between the observed and calculated values is 0.99 for both the sets involving the whole residue and the side chain.

Figure 1
Histogram showing the distribution of the number of partner residues in contact with a particular amino acid residue in proteins. The bars corresponding to the side chain (sc) and the whole residue (w) appear grouped towards the left and the right sides of the plot, respectively, and the average value (with standard deviation) of partner-residue in each case is shown. The residues are arranged according to their volume [33]. Only ten residues are shown.
Figure 2
Same as in Figure 1, but for the remaining ten residues.
chain. It may be mentioned that other studies also report quite similar values for the mean ASA (absolute value, or that relative to the standard state) of residues [24,25]. Instead of the mean, some studies have used median ASA [24]. As such, median partner numbers (PNm) are also provided in Table 2.

We also compared the observed ASA (ASA_{obs}) of a structure to ASA_{calc}, calculated by considering each residue in the structure in turn and calculating its PN, which was then employed to derive its ASA using the equation given in Table 1. Two metrics were computed: $R_A = \sum \text{ASA}_{i\text{(calc)}}/\sum \text{ASA}_{i\text{(obs)}}$ (summed over all the residues, $i$), and

$$R_D = \sum \left| \text{ASA}_{i\text{(calc)}} - \text{ASA}_{i\text{(obs)}} \right|/\sum \text{ASA}_{i\text{(obs)}}.$$ 

These values were also calculated when the partner number was atom-based (i.e., the number of surrounding atoms) and an exponential equation was used to relate PN to ASA_{calc} [7]. Results presented in Table 3 indicate that $R_A$ is closer to 1 and $R_D$ is smaller when the partner number is based on residues, suggesting a closer match with ASA_{obs} for the residue-based, rather than the atom-based calculation. This can also be seen from the smaller difference, $(\text{ASA}_{\text{calc}} - \text{ASA}_{\text{obs}})$, for the former (Additional file 1, Figure S2). For the majority of the structures ASA_{obs} is greater than ASA_{calc} (individual values are available in Table S1, Additional file 1). We have identified at least one feature that may be responsible for the greater discrepancy involving the atom-based calculation – structures that are less compact, being made up of smaller domains, Figure 4 showing an example, the residue-based ASA_{calc} is closer to ASA_{obs}.

The webserver, ContPlot

Having computed the average values of the ASA and PN for twenty residue types we felt that a software that would
Table 1: Parameters for fitting cubic equation \( y = b_6 + b_1 x + b_2 x^2 + b_3 x^3 \) to relate ASA \( y \) to the partner number, PN \( x \) for different residues.

| Residue | ASA (Å²) | Whole residue | Side chain | R² |
|---------|----------|---------------|------------|----|
| Gly 6245 | 83.9 | 84.4 | 8.3 | 0.07 | 0.991 |
| | | 4.40 | 2.31 | 0.33 | 0.01 |
| Ala 6688 | 116.4 (55.4) | 130.5 | 3.42 | 0.49 | 0.02 |
| | | 6.53 | 0.29 | 0.05 | 0.01 |
| Ser 4614 | 125.7 (69.1) | 140.0 | 4.79 | 0.55 | 0.02 |
| | | 3.89 | 2.04 | 0.29 | 0.05 |
| Cys 1367 | 141.5 (82.1) | 65.6 | 3.4 | -1.5 | 0.06 |
| | | 12.37 | 1.5 | 0.5 | 0.02 |
| Thr 4724 | 148.1 (88.6) | 156.7 | 2.06 | 0.28 | 0.01 |
| | | 4.17 | -0.9 | 0.06 | 0.01 |
| Asp 4463 | 155.4 (97.8) | 170.9 | 1.78 | 0.22 | 0.01 |
| | | 4.02 | -0.1 | 0.03 | 0.01 |
| Pro 3701 | 144.8 (106.4) | 147.2 | 2.06 | -1.3 | 0.08 |
| | | 4.16 | 0.28 | 0.01 | 0.01 |
| Asn 3537 | 168.9 (109.9) | 173.0 | 2.10 | -0.7 | 0.05 |
| | | 4.49 | 0.27 | 0.05 | 0.01 |
| Val 5619 | 162.2 (103.1) | 167.8 | 3.05 | -0.4 | 0.04 |
| | | 7.37 | 0.36 | 0.04 | 0.01 |
| Glu 3843 | 187.2 (123.5) | 190.8 | 1.77 | -1.0 | 0.05 |
| | | 4.47 | 0.20 | 0.01 | 0.01 |
| Gin 2645 | 189.2 (129.7) | 231.8 | 3.78 | 0.42 | 0.02 |
| | | 9.56 | -0.2 | 0.02 | 0.01 |
| His 1789 | 198.5 (141.3) | 209.9 | 1.47 | -0.4 | 0.04 |
| | | 3.31 | 0.18 | 0.04 | 0.01 |
| Leu 6575 | 198.0 (141.5) | 233.2 | 4.91 | 0.6 | 0.06 |
| | | 12.42 | 0.55 | 0.06 | 0.01 |
| Ile 4362 | 190.0 (130.7) | 214.5 | 2.17 | 0.26 | 0.03 |
| | | 5.24 | 0.26 | 0.03 | 0.01 |
| Met 1533 | 210.6 (150.4) | 223.0 | 2.61 | -0.8 | 0.05 |
| | | 5.88 | 0.31 | 0.05 | 0.01 |
| Lys 3861 | 207.5 (148.0) | 231.0 | 3.42 | -0.4 | 0.03 |
| | | 8.66 | 0.38 | 0.03 | 0.01 |
| Phe 3225 | 223.3 (164.2) | 210.3 | 3.20 | -1.0 | 0.06 |
| | | 7.59 | 0.37 | 0.06 | 0.01 |
| Tyr 2927 | 238.3 (180.0) | 220.1 | 3.40 | -1.2 | 0.06 |
| | | 8.06 | 0.39 | 0.06 | 0.01 |
| Arg 3024 | 249.3 (190.2) | 274.3 | 1.76 | 0.1 | 0.02 |
| | | 5.26 | 0.17 | 0.02 | 0.01 |
| Trp 1230 | 265.4 (209.6) | 265.8 | 3.25 | -0.1 | 0.02 |
| | | 10.60 | 0.29 | 0.02 | 0.01 |

Standard error for each parameter is given below in italics.

* The number of each residue in the non-redundant dataset of 432 polypeptide chains is provided.

* The values for the whole residue and the side chain (in parentheses) for the Gly-based tripeptide are taken from [7].
display the two parameters relative to the average values for all the residues in a polypeptide chain would be useful for the analysis of residue packing in relation to the secondary structures and the overall fold of the protein molecule. ContPlot is such a webserver that creates a plot based on either of the two parameters. For ASA, though the observed value is checked against average ASA value of a particular residue, what is plotted is the relative accessibility (i.e., what percentage of the surface area of residue (X) is accessible to solvent relative to its value in the standard state, taken as the ASA of X in the tripeptide, Ala-X-Ala (in the program NACCESS), so that all the residues are displayed in the range 0–100. The plot enables one to visually identify residues that are tightly packed, but the presence of hydrophobic residues that are less than optimally packed on the surface may indicate a binding site for another protein molecule. Indeed there is an option in the server to simultaneously display the results from another

| Residue | Whole Residue | Side Chain |
|---------|---------------|------------|
|     | PN<sub>4</sub> | <PN> | ASA<sub>obs</sub> | ASA<sub>calc</sub> | PN<sub>4</sub> | <PN> | ASA<sub>obs</sub> | ASA<sub>calc</sub> |
| Gly    | 7.5 | 7.4 (2.2) | 26.6 (24.5) | 25.9 |
| Ala    | 8.0 | 8.6 (2.5) | 28.1 (30.9) | 23.9 | 4.0 | 3.6 (1.9) | 18.2 (21.2) | 13.9 |
| Ser    | 8.0 | 7.9 (2.6) | 39.2 (33.2) | 34.7 | 5.0 | 3.6 (2.0) | 27.9 (24.2) | 23.6 |
| Cys    | 9.5 | 10.0 (2.3) | 17.1 (21.0) | 15.1 | 5.0 | 5.6 (1.7) | 10.3 (15.7) | 8.0 |
| Thr    | 8.5 | 8.5 (2.6) | 44.2 (36.0) | 39.9 | 6.0 | 4.5 (2.2) | 35.5 (30.0) | 31.1 |
| Asp    | 9.5 | 7.9 (2.5) | 58.1 (37.2) | 54.3 | 5.5 | 3.9 (2.1) | 48.0 (30.6) | 43.9 |
| Pro    | 8.5 | 7.7 (2.6) | 54.2 (39.5) | 51.4 | 6.0 | 4.9 (2.4) | 43.4 (32.8) | 40.5 |
| Asn    | 9.0 | 8.3 (2.7) | 57.9 (40.8) | 53.7 | 6.5 | 4.4 (2.3) | 48.0 (33.7) | 42.9 |
| Val    | 9.5 | 10.3 (2.6) | 24.1 (32.0) | 20.3 | 5.5 | 6.0 (2.1) | 19.0 (27.2) | 13.5 |
| Glu    | 9.5 | 8.4 (2.5) | 73.4 (41.9) | 70.7 | 5.5 | 4.0 (2.2) | 63.8 (36.3) | 59.9 |
| Gln    | 9.0 | 9.0 (2.7) | 68.6 (43.3) | 62.1 | 6.5 | 4.6 (2.4) | 59.6 (37.7) | 54.6 |
| His    | 9.5 | 9.7 (2.9) | 53.8 (44.6) | 48.2 | 6.0 | 5.7 (2.5) | 46.5 (38.7) | 40.1 |
| Leu    | 9.5 | 11.0 (2.7) | 28.8 (38.0) | 21.9 | 6.5 | 6.6 (2.3) | 23.3 (32.6) | 16.9 |
| Ile    | 8.5 | 11.0 (2.7) | 25.0 (35.2) | 18.9 | 6.5 | 6.9 (2.2) | 20.5 (30.6) | 15.2 |
| Met    | 9.5 | 11.2 (3.1) | 35.5 (45.8) | 26.5 | 6.5 | 6.7 (2.6) | 28.9 (38.1) | 21.1 |
| Lys    | 9.5 | 8.4 (2.4) | 95.8 (42.9) | 92.5 | 6.5 | 4.0 (2.1) | 85.4 (37.0) | 82.3 |
| Phe    | 10.0 | 11.9 (2.9) | 31.0 (39.8) | 24.5 | 7.0 | 7.9 (2.5) | 25.3 (35.1) | 18.5 |
| Tyr    | 10.0 | 11.5 (3.1) | 45.5 (45.0) | 38.2 | 7.5 | 7.5 (2.6) | 39.7 (40.2) | 33.8 |
| Arg    | 11.0 | 10.1 (3.1) | 85.5 (53.3) | 80.0 | 8.0 | 5.8 (2.7) | 77.3 (48.0) | 70.4 |
| Trp    | 11.0 | 12.6 (3.2) | 43.5 (47.6) | 36.9 | 8.0 | 8.8 (2.9) | 37.8 (43.7) | 29.3 |

The standard deviations are in parentheses. <ASA><sub>calc</sub> at a given <PN> is obtained using the equation provided in Table 1.
calculation including other chain(s) in a binary (or higher oligomeric) complex, thus providing a method to visually identify the residues in the interface [26]. Figure 5 shows the ContPlot (using ASA) of barstar alone and in complexation with barnase. Residues 27–46 constitute the longest segment of the polypeptide chain that forms the interface [27]. For this stretch there is a substantial decrease in ASA of residues on complexation. In the isolated state residue Tyr29 (red square) has an ASA value that is greater than 1.0σ times the average ASA of Tyr. There are quite a few residues of this type in the above stretch of the chain (Figure 6). When the calculation is done on the basis of PN, many of these residues would also be shown in red square, indicating a value of the PN that is smaller than 1.0σ times the average PN of the same residue type. In the complex state many of the residues in this stretch are shown in green (or blue) crosses, indicating the ASA values to be within (or less than) 1.0σ of the average value. Although 1.0σ level has been used here, the σ values are rather large (Table 2) and it may be easier to identify the deviant residues at a lower level. Many other issues are discussed in the help file associated with the server.

**Conclusion**

A parameter, PN has been defined to indicate how many residues are found with 4.5 Å of a given residue. The histogram of the variation of PN has a skewness that indicates the hydrophobic or hydrophilic nature of the residue (Figures 1 and 2). ASA of a residue decreases with the increase in PN and a cubic equation has been used to represent the dependence (Table 1, Figures 3 and S1, Additional file 1). The average values of ASA and PN associated with the residues have been used to develop a webserver, ContPlot that reads a PDB file to display the deviation of these values from the average values of the corresponding residues along the protein sequence (Figure 5). Extending the concept that each residue in a protein sequence would try to achieve a value of ASA or PN that is close to its average value we have devised an algorithm that can identify the native fold from a set of decoys representing near-native structures (Bahadur and Chakrabarti, unpublished). Furthermore, the presence of cluster of hydrophobic residues with unsatisfied (i.e., below-average) PNs (or above-average ASAs) (Figure 5) could indicate the existence of a protein-protein interaction site and this is being used to identify interface patches in protein structures (Pal and Chakrabarti, unpublished).

**Methods**

Atomic coordinates were obtained from the Protein Data Bank (PDB) [28]. The analysis was carried out using the same dataset of 432 polypeptide chains used in [7] and the same criteria on B-factor were used to select only the well-ordered residues. The solvent accessible surface area (ASA) was computed using the program NACCESS [29], which is an implementation of the algorithm by Lee and Richards [30], using the default probe radius of 1.4 Å. This ASA has been designated as the observed value (ASA<sub>obs</sub>) to distinguish it from the one calculated (ASA<sub>calc</sub>) using the equation derived in this paper. Calculations were restricted to the atoms of a particular subunit only. Two residues are in contact if any atom-to-atom distance between the pair are within 4.5 Å [31]. The partner number of a residue is the number of other residues within a distance of 4.5 Å from any atom of the residue under consideration; the flanking residues were not considered as partner if the interaction was only with the main-chain atoms. The secondary structures were determined using DSSP [32]. The webserver, ContPlot uses the

| Parameters | Partner atoms | Partner residues |
|------------|---------------|-----------------|
| $R_A$      | 0.93 (0.11)   | 0.98 (0.07)     |
| $R_D$      | 0.37 (0.07)   | 0.30 (0.06)     |

Average values based on 275 structures (considering only the well-ordered structures, as given in Methods, and excluding those with missing atoms/residues).
average values of ASA and PN (and the associated standard deviations) for all the residues given in Table 2 and is available, along with necessary documentation at: http://www.boseinst.ernet.in/resources/bioinfo/stag.html.

Authors' contributions
PC conceptualized the work that was carried out by AP, RPB and PSR. AP, RPB and PC participated in interpretation of the data and writing the manuscript. All the authors have read and accepted the final version of the manuscript.

Additional material

Additional File 1
Supplementary data. Table S1. Observed and calculated ASAs, and the match between them, in different protein structures. Figure S1. Variation of the mean accessible surface area (Å²) with partner number for residues other than Lys and Met. Figure S2. Plot of the difference, $(ASA_{calc} - ASA_{obs})$, for 275 individual PDB files.
Click here for file [http://www.biomedcentral.com/contentplementary/1471-2105-10-103-S1.doc]
Acknowledgements
The work was supported by grants from the Department of Biotechnology, and the Council of Scientific and Industrial Research, India.

References
1. Kauzmann W: Some factors in the interpretation of protein denaturation. Adv Protein Chem 1959, 14:1-63.
2. Dill KA: Dominant forces in protein folding. Biochemistry 1990, 29(31):7133-7155.
3. Richards FM: The interpretation of protein structures: total volume, group volume distributions and packing density. J Mol Biol 1974, 82(1):1-14.
4. Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH: Protein design by binary patterning of polar and nonpolar amino acids. Science 1993, 262(5140):1680-1685.
5. Huang ES, Subbiah S, Levitt M: Recognizing native folds by the arrangement of hydrophobic and polar residues. J Mol Biol 1995, 252(5):709-720.
6. Miller SJ, Janin J, Lesk AM, Chothia C: Interior and surface of monomeric proteins. J Mol Biol 1987, 196(3):641-656.
7. Samanta U, Bahadur RP, Chakrabarti P: Quantifying the accessible surface area of protein residues in their local environment. Protein Eng 2002, 15(8):659-667.
8. Wodak SJ, Rooman MJ: Generating and testing protein folds. Curr Opin Struct Biol 1993, 3:247-259.
9. Torda AE: Perspectives in protein-fold recognition. Curr Opin Struct Biol 1997, 7(2):200-205.
10. Moult J: Comparison of database potentials and molecular mechanics force fields. Curr Opin Struct Biol 1997, 7(2):194-199.
11. Finkelstein AV: Protein structure: what is it possible to predict now? Curr Opin Struct Biol 1997, 7(1):60-71.
12. Jones DT: Protein structure prediction in the postgenomic era. Curr Opin Struct Biol 2000, 10(3):371-379.
13. Delarue M, Koehl P: Atomic environment energies in proteins defined from statistics of accessible and contact surface areas. J Mol Biol 1995, 249(3):675-690.
14. Jernigan RL,Bahar I: Structure-derived potentials and protein simulations. Curr Opin Struct Biol 1996, 6(2):195-209.
15. Melo F, Feytmans E: Novel knowledge-based mean force potential at atomic level. J Mol Biol 1997, 267(1):207-222.
16. Samanta U, Chakrabarti P: Assessing the role of tryptophan residues in the binding site. Protein Eng 2001, 14(1):7-15.
17. Nishikawa K, Ooi T: Prediction of the surface-interior diagram of globular proteins by an empirical method. Int J Pept Protein Res 1980, 16(1):19-32.
18. Panjikar SK, Biswas M, Vishveshwara S: Determinants of backbone packing in globular proteins: an analysis of spatial neighbours. J Mol Biol 1997, 257(5):627-637.
19. Karlin S, Zhu YZ, Baud F: Atom density in protein structures. Proc Natl Acad Sci USA 1999, 96(22):12500-12505.
20. Zhang C, Kim SH: Environment-dependent residue contact energies for proteins. Proc Natl Acad Sci USA 2000, 97(24):12550-12555.
21. Abagyan RA, Totrov MM: Contact area difference (CAD): a robust measure to evaluate accuracy of protein models. J Mol Biol 1997, 268(3):678-685.
22. Wang JY, Lee HM, Ahmad S: SVM-Cabins: Prediction of solvent accessibility using accumulation cutoff set and support vector machine. Proteins 2007, 68(1):82-91.
23. Wodak SJ, Janin J: Analytical approximation to the accessible surface area of proteins. Proc Natl Acad Sci USA 1980, 77(4):1736-1740.
24. Lins L, Thomas A, Brassearer R: Analysis of accessible surface of residues in proteins. Protein Sci 2003, 12(7):1406-1417.
25. Wang JY, Ahmad S, Gromiha MM, Sari A: Look-up tables for protein solvent accessibility prediction and nearest neighbor effect analysis. Biopolymers 2004, 75(3):209-216.
26. Chakrabarti P, Janin J: Dissecting protein-protein recognition sites. Proteins 2002, 47(3):334-343.
27. Pal A, Chakrabarti P, Bahadur R, Rodier F, Janin J: Peptide segments in protein-protein interfaces. J Biochem 2007, 12(1):101-11.
28. 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(1):235-242.
29. Hubbard SJ: NACCESS: program for calculating accessibilities. Department of Biochemistry and Molecular Biology 1992 [http://www.biomedcentral.com/103]. University College of London.
30. Lee B, Richards FM: The interpretation of protein structures: estimation of static accessibility. J Mol Biol 1971, 55(3):379-400.
31. Chakrabarti P, Bhattacharyya R: Geometry of nonbonded interactions involving planar groups in proteins. Prog Biophys Mol Biol 2007, 95:83-137.
32. Kabsch W, Sander C: Dictionary of protein secondary structure. Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983, 22(12):2577-2637.
33. Chothia C: Structural invariants in protein folding. Nature 1975, 254(5498):304-308.
34. Buckle AM, Schreiber G, Ferstl AR: Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-A resolution. Biochemistry 1994, 33(30):8878-8889.

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