A Review of Methods Available to Estimate Solvent-Accessible Surface Areas of Soluble Proteins in the Folded and Unfolded States

Article *in* Current Protein and Peptide Science · March 2014

DOI: 10.2174/1389203715666140327114232 · Source: PubMed

Md Imtaiyaz Hassan
Jamia Millia Islamia
536 PUBLICATIONS **10,110** CITATIONS
SEE PROFILE

Asimul Islam
Jamia Millia Islamia
275 PUBLICATIONS **4,650** CITATIONS
SEE PROFILE

Faizan Ahmad
Jamia Millia Islamia
375 PUBLICATIONS **9,199** CITATIONS
SEE PROFILE

A Review of Methods Available to Estimate Solvent-Accessible Surface Areas of Soluble Proteins in the Folded and Unfolded States

Syed Ausaf Ali, Md. Imtaiyaz Hassan, Asimul Islam and Faizan Ahmad*

Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, Jamia Nagar, New Delhi 10025, India

Abstract: Solvent accessible surface area (SASA) of proteins has always been considered as a decisive factor in protein folding and stability studies. It is defined as the surface characterized around a protein by a hypothetical centre of a solvent sphere with the van der Waals contact surface of the molecule. Based on SASA values, amino acid residues of a protein can be classified as buried or exposed. There are various types of SASAs starting from relative solvent accessibility to absolute surface areas. Direct estimation of accurate SASAs of folded proteins experimentally at the atomic level is not possible. However, the SASA of a native protein can be estimated computationally from the atomic coordinates. Similarly, various simulation methods are available to compute the SASA of a protein in its unfolded state. In efforts to estimate the changes in SASA related to the protein folding, a number of the unfolded state models have been proposed. In this review, we have summarized different algorithms and computational tools for SASA estimations. Furthermore, online resources for SASA calculations and representations have also been discussed in detail. This review will be useful for protein chemists and biologists for the accurate measurements of SASA and its subsequent applications for the calculation of various biophysical and thermodynamic properties of proteins.

Keywords: Crystal structure, protein folding, protein stability, solvent accessible surface area, thermodynamic properties, unfolded state.

1. INTRODUCTION

The concept of solvent accessibility of residues in globular proteins was first introduced by Lee and Richards in 1971 [1]. They defined the term "solvent accessible surface area" (SASA) as the extent to which atoms on the surface of a protein can form contacts with solvent. The solvent in biological systems is taken as water with a radius of 1.4 Å. It represents the locus of the center of the solvent molecule as it rolls along the protein, making the maximum permitted van der Waals contacts without penetrating any other atom. It was also described as the "static accessible surface area" because potential flexibility was not included [1].

For calculation purposes, a sphere is focused at each atomic position in the co-ordinate list as shown in (Fig. **1**). It is assigned a radius equal to the sum of the radius of the atom and that of the solvent molecule. According to the first algorithm proposed by Lee and Richards [1], SASA can be calculated using following relations:

$$\mathrm{SASA} = \Sigma\ [R\ /\ (R^2 - Z_\mathrm{i}^2]\ L_\mathrm{i}.\ D \qquad (1)$$

$$\mathrm{D} = \Delta Z/2 + \Delta'\ Z$$

where R is the radius of the sphere R, L_i is the length of the arc drawn on a given section i, $Z_{\mathrm{i\ is}}$ the perpendicular distance from the centre of the sphere to the section i, ΔZ is the spacing between the sections, and Δ' is $\Delta Z/2$ or R-Z_i, whichever is smaller. Summation is all over the arcs drawn for the given atom.

The accessibility to the solvent can be calculated as [1]:

$$\text{Accessibility} = 100.\ \mathrm{SASA}\ /\ 4\pi R^2 \qquad (2)$$

The relative ASA (RSA) is the percentage of real ASA with respect to ASA of amino acids in the Ala-X-Ala extended state [2]. SASA is generally measured in $Å^2$ [3].

The early computer programs were developed to compute the SASA of a number of model compounds prior to their application to proteins. The first model compounds were tripeptides of the form Gly-X-Gly and Ala-X-Ala, where 'X' is the residue whose SASA was to be computed [1]. From his studies on shape and surface area of folded proteins, Gates [4] arrived at several important conclusions. (i) The SASA of a protein has been stated to be proportional to the two-third of the real volume. (ii) It was also concluded that the shape is not important in estimating the solvent accessibility, and spheres are good models for protein subunits. (iii) When the volume of the ordered figure used to calculate the SASA is calculated geometrically, the accessible area becomes proportional to the 0.77 power of that volume. SASA is a quantity of particular interest in protein folding and functional studies. It plays an important role to understand the structure-function relationship of proteins and their residues [5]. It is well known fact that burial of hydrophobic residues is a key factor in protein folding. Naturally, the exposure of these residues to the solvent and the hydrophobic core is

*Address correspondence by this authors at the Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, Jamia Nagar, New Delhi 10025, India; Tel: +91-11-2632-1733; Fax: +91-2698-3409; E-mail: fahmad@jmi.ac.in

1389-2037/14 $58.00+.00 © 2014 Bentham Science Publishers

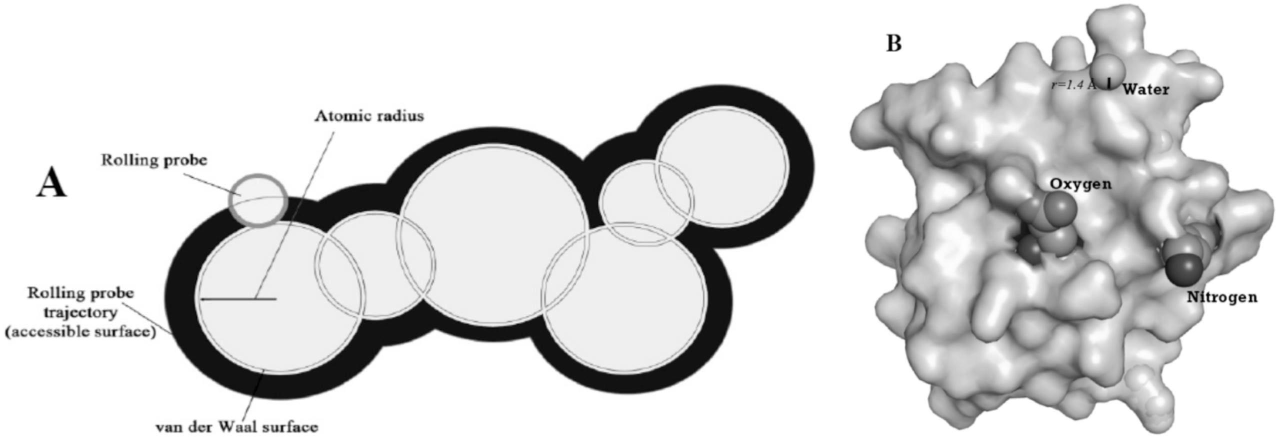

Fig. (1). (A). A cross-section of a part of a macromolecule in space rolling probe, van der Waals and accessible surface areas are indicated in the figure. **(B).** Three-dimensional representation of SASA of a protein on which water molecule is rolling (Green). The relative size of water molecule can be compared with nitrogen (blue) and oxygen (red).

directly related to the stability of the protein. Moret and Zebende [6] analyzed the disparity of the SASA of amino acids in small fragments of proteins to investigate the hydrophobic effect of the residues. The estimation of the loss of SASA proved to be a measure of the probability of an amino acid to have a non-polar or polar side chain [6]. Likewise, the environment free energy of amino acids depends on an accurate and rapid estimation of solvent accessible surface area [7]. The stability of proteins is also governed by the heat capacity, enthalpy and entropy changes, which are strongly associated to the change in SASA experienced by the polypeptide chain [8].

A precise prediction of tertiary structure of proteins has been one of the most challenging problems for biologists [9, 10]. The accuracy of predictions depends on a number of factors. It has been proposed that the relative solvent accessibility of residues might be an effective factor for increasing the accuracy of protein secondary and tertiary structure prediction [11]. Roknabadi *et al.* [12] investigated the effect of the alteration of the solvent accessibility threshold on the accuracy of protein structure prediction. Kurt and Cavagnero [13] investigated the influence of chain truncation on the disclosure of nonpolar solvent-accessible surface area (NSASA) for (a) unfolded state, represented as a fully extended chain, and (b) native-like folded state [14]. Changes in protein stability upon point mutations have been analyzed in terms of Relative Solvent Accessibility (RSA) as classifiers for potentials [2]. RSA can also be used to describe the physical and evolutionary properties of a protein [15]. The concept of SASA has also been used to understand the protein-protein and protein-nucleic acid interactions. Protein-DNA recognition is significant for different cellular processes and the binding free energy of protein-DNA complex is related with the change in interface accessible surface area upon binding [12]. Protein-protein interactions play very important role nearly in all biological processes. Bahadur *et al.* [16] analyzed the SASA of protein-protein interactions and established a method for predicting the interface surface area. Mandell *et al.* [17] carried out kinetic experiments to estimate the relation between SASA and bindings specificity of thrombin-thrombomodulin interface. Hughes *et al.* [18] analyzed the role of SASA to understand the structure-function relationship of protein-protein networks. Geometric representations of proteins and ligands including SASA values can be applied to characterize interactions between and within proteins, ligands and the solvent [19]. These studies have revealed the potential use of SASA in the field of drug design and discovery [5]. The SASA analysis can also be used as a probe to study the protein-DNA binding interface and protein mobility [20]. Ochoa *et al.* [21] used information on predicted SASA to analyze co-evolution based protein interactions that are quantified in the form of phylogenetic trees.

In view of the immense significance of SASA in the field of protein structure, function and stability [2, 22, 23], here our aim is to review various available methods for calculation of SASA. In this review, we have focused on various types of SASA and its application in the field of structural biology and protein biochemistry. Furthermore, we have described computational algorithms and prediction methods for calculating/predicting SASA from the crystal structure of protein [24]. Since proteins cannot crystallize in the unfolded conformation such an estimation of SASA in this conformation is not possible experimentally [8]. Therefore, various theoretical models for predicting SASA in the unfolded state have also been emphasized. We have also compared different values of SASA for a particular residue of the protein using different SASA prediction methods.

SASA can be broadly classified as (i) unfolded state SASA, which is represented as X_{ASA}, and (ii) folded state SASA, that is represented as F_{ASA}. [25]. Both types of SASA can be applied to the individual residues, domains, subdomains, and even to the whole protein molecule [8, 26]. The absolute or numerical SASA for a residue/protein is estimated using atomic coordinates of the crystal structure (for that residue/protein) in Protein Data Bank (PDB) [27] by different computational algorithms [5]. The relative solvent accessibility for the i^{th} amino acid (RSA_i) is defined as the ratio of the absolute SASA of that residue observed in a given structure, denoted as SA_i, and the maximum attainable value of the solvent-exposed surface area for this residue, denoted as MSA_i. Thus, RSA_i can have a value between 0 and 100%, with 0% corresponding to a fully buried and 100% to a fully accessible residue, respectively [28].

$$RSA_i\,(\%) = 100.\ (SA_i / MSA_i) \quad (3)$$

Analytical SASA is estimated computationally using analytical equations and their first and second derivatives [29]. One of the widely used programs to estimate the analytical SASA is GETAREA [30]. The SASA of a particular residue can be further classified into hydrophobic and hydrophilic surface areas. On the basis of hydrophobic to hydrophilic ratio of the surface area, amino acids can be clustered into three groups, namely (i) hydrophobic amino acids (G <A < V,C,P < I,L < F,M), (ii) hydrophilic amino acids (D,N < E,Q <R), and (iii) miscellaneous group (S, T, H, K,Y,W) [25].

The SASA of a residue protein can also be estimated as mean or average SASA and median SASA. Mean SASA of a residue is the average of all the SASA values for that residue in different proteins in either folded or unfolded conformation. In median SASA the middle value is selected of all the SASA values for a particular amino acids [25]. The above mentioned SASA types have been summarized in (Fig. **2**).

2. ALGORITHMS TO CALCULATE SASA IN THE FOLDED STATE

2.1. Z-layer Integration Method

This is another version of Lee and Richards' method to calculate the SASA in the folded state. In this method, a series of two-dimensional sections is taken at a number of equally placed separations throughout the molecule. For each atom cutting that particular plane, the circle of intersection is sampled to determine the proportion masked by other atoms. By replacing each of these circles with a cylinder of the same length as the separation between planes, it is possible to find a fairly accurate value for the exposed area of atoms. Evidently, as the number of planes increases and their separation decreases, the value of derived SASA improves [31]. This algorithm is most widely used and is implemented by many computational programs like SERF [31], ACCESS [32], NACCESS [33], etc., to calculate SASA from the atomic coordinates of protein crystal structures.

2.2. Intersection Method

This method was initially proposed by Gibson and Scheraga [34] to calculate the SASA and volumes of atomic coordinates of a protein. This approach is articulated in terms of the surface area lost when spheres intersect each other. The surface area of sphere *i* is calculated as the surface area of the free sphere S_A minus the area of each intersection with spheres *B, C,* and *D* [31]:

$$S_i = S_A - S_A \cap S_B + S_A \cap S_B \cap S_C - S_A \cap S_B \cap S_C \cap S_D \quad (4)$$

The problem with this method is mainly in the identification of the number and type of two-, three-, or four-fold intersection. Dodd and Theodorou [35] proposed a more general approach to avoid shortcomings in this formulation. An analytical formula was derived by Richmond [36] for the calculation of the surface area exterior to an arbitrary number of overlapping spheres. This process was based on minimization procedures used with molecular docking algorithms and energy calculations. An analytical formula for the calculation of the excluded volume (enclosed within the SASA) was also proposed.

Wodak and Janin [37] proposed an analytical and statistical method for the calculation of SASA of a protein. This was implemented in computational programs like SERF [31]. It was not an appropriate method for the calculation of individual atomic areas but could be suitable when approximating whole residues as large single spheres. Hasel *et al.* [38] further modified this approach to better cope with atomic accessibility. They optimized the equations proposed by Wodak and Janin [36] for different atom types and applied empirically derived correction factors by standardizing their method against independently calculated SASAs.

2.3. The Shrake and Rupley Algorithm

The algorithm proposed by Shrake and Rupley [39] was one of the simplest and oldest methods to calculate SASA of a protein molecule. The calculation is based on the Monte Carlo numerical integration of the atomic coordinates of the protein involving the assumptions like those of Lee and Richards [1]. It was initially applied to insulin and lysozyme

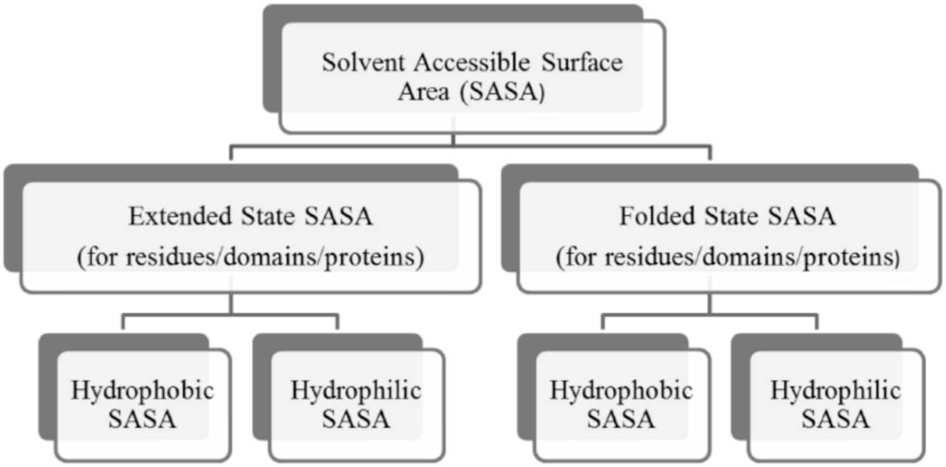
Solvent Accessible Surface Area (SASA)
Extended State SASA (for residues/domains/proteins)
Folded State SASA (for residues/domains/proteins)
Hydrophobic SASA
Hydrophilic SASA
Hydrophobic SASA
Hydrophilic SASA

Fig. (2). The classification of solvent accessible surface area.

[39]. In this method, the spherical surface of each atom is shielded with 92 equally placed points. The points that lie within other expanded atoms are also estimated. The number of exposed atoms determines the proportion of the total surface area that is accessible to the solvent. This approach is general, robust and straight forward. The computational implementation of this approach suggested many generalizations, for example, glycine is among those residues whose exposure is least affected by protein folding [39, 40].

2.4. Linear Combinations of Pairwise Overlaps (LCPO) Method

Weiser *et al.* [41] proposed a methodology to estimate the SASA, which is based on few analytical equations. This method involves linear combinations of terms composed from pairwise overlaps of hard spheres. For better performance, neighbor-list reduction is applied SASA preprocessing step. Initially LCPO method was used to compute the SASA of 18 different compounds including proteins. This method, when implemented computationally, gave the SASA and first derivatives of penicillopepsin (having 2366 atomic coordinates) in just 0.87 seconds on an SGI R10000/194 MHz processor [41]. Weiser *et al.* [29] further extended their work to derive a fast analytical equation based on tetrahedrally directed neighbour densities for the calculation of molecular SASA. They used a Gaussian function to calculate the neighbor density in four tetrahedral directions in three-dimensional space. SASA and the first derivatives of penicillopepsin could be computed in just 0.13 seconds on an SGI R10000/194 MHz processor [29].

2.5. Power Diagram Method

During the last decade, bio-macromolecular simulation has made remarkable inroads to help explicate biological processes. This progress was possible due to the massive increase of computer power in addition to the improvement of the models used in the simulations. In these simulations, an atom is modelled as a ball bounded by a sphere and a molecule as the union of a finite collection of such balls. This combination is called space filling diagram of the molecule. [42]. The power diagram contains power cells having polygonal faces and vertices. The dual of the power diagram is called Weighted or Delaunay triangulation. Implicit solvent models usually analyze the interaction of water with non-polar atoms of a molecule as a weighted sum of SASA of each atom. Bryant *et al.* [42] developed a new version of the alpha-shape software, ALPHAVOL [43] that implements area and weighted area derivatives to compute SASA of peptides and proteins. It is indeed fast, accurate and robust method that is distributed as an open source program (http://biogeometry.duke.edu/software/proshape/) [40].

3. MODELS FOR ESTIMATING SASA IN THE UNFOLDED STATE

If we define protein stability as the Gibbs free energy change associated with the equilibrium, Folded state ↔ Unfolded state, the unfolded state of proteins is as significant as the native state in determining protein stability [44]. However, most of the folding studies are focused on the native state because of the increased facility of studying folded proteins. Fitzkee and Rose [45] demonstrated that the random-coil model does predict the experimentally determined coil dimensions of denatured proteins successfully and they are biased toward specific conformations. The surface area lost upon folding is taken as the difference in SASA between the native and unfolded states. The calculation is simple and straightforward for a defined native structure by applying Lee and Richards algorithm [1]. The estimation of the SASA for the unfolded state of protein is problematic and indirect [3]. It is based on simulation and modeling of individual residues in combination with other residues in a polypeptide (the standard state) [46]. Here, we describe various models used to estimate the SASA of residues for the unfolded state of protein. A comparison of residue SASA in the extended state by different approaches has been shown in Table **1**.

3.1. Some Early Models

The SASA in the unfolded state was not well defined till early 1970s. Therefore, the solvent accessibilities of the amino acids in the denatured polypeptide chain were approximated as standard states [47]. These standard states included a single blocked residue proposed by Sneddon and Tobias [48]. Shrake and Rupley [38] used a stochastic standard state defined as average solvent accessibility of X in Gly-X-Gly tripeptides with dihedral angles reflecting the observed distribution in the protein structures data base. In 1985, Rose *et al.* [49] correlated the residue hydrophobicity with the average area buried on folding. Thus the average exposed area on unfolding could be roughly estimated. In 1987, Miller *et al.* [50] defined the standard state accessibility of residue X to be the accessible surface area of X in an extended Gly-X-Gly tripeptide. Both the extended state and stochastic approaches tend to overestimate the standard state solvent accessibilities, for in a real polypeptide chain the nearest neighbours of a residue are bulkier than glycine [47]. Zielenkiewicz and Saenger [47] used a new approach for calculating the standard state residue solvent accessibilities. It is based on averaging the surface area of a central residue in Ala-X-Ala tripeptides from a molecular dynamics simulation at a temperature of 368 K. Livingstone *et al.* [51] concluded that the large negative heat capacity changes observed in protein folding provide a quantitative measure of the reduction in the solvent-accessible nonpolar surface area. Although, all these models were valid standard state models for comparison of a buried surface with an exposed surface, none was a true model of an unfolded state.

3.2. The Upper and Lower bound Model

It was proposed by Creamer *et al.* [46] in 1995. According to them, the tripeptide was not a good descriptor of an unfolded state. That is why they proposed two reference states to bracket the expected behavior of the unfolded chain between the reliable extremes. For upper bound values, flexible peptides were simulated using hard-sphere Monte Carlo approximations. On the other hand, the peptide fragments excised from high resolution proteins structures were used to approximate the lower bound values. They proposed that the actual SASA of the side chain of a residue would lie between these two extreme limits under attainable folding conditions. They showed that this improved model could be

Table 1. Extended state value predicted for each residue by different algorithms.

S.No.	**Residue**	**SASA Value Proposed for Each Residues in Extended State (in Å^2)**					
		Ala-X-Ala [148]	**Gly-X-Gly** [148]	**Lins** ***et al.*** [25]	**ASC (ECEPP/2)** [182]	**Miller** [47]	**Zielenkiewicz and Saenger** [47]
1	A	102.68	116.40	111	110.20	113	111.60
2	C	127.72	141.48	157	140.40	140	136.90
3	D	141.61	155.37	160	144.10	151	154.70
4	E	173.46	187.16	187	174.70	183	179.90
5	F	209.64	223.29	208	200.70	218	210.60
6	G	70.27	83.91	86	78.70	85	75.60
7	H	184.79	198.51	191	181.90	194	187.20
8	I	176.22	189.95	173	185.00	182	188.40
9	K	193.73	207.49	212	205.70	211	209.90
10	L	184.33	197.99	179	183.10	180	192.20
11	M	196.87	210.55	201	200.10	204	196.60
12	N	155.22	168.87	166	146.40	158	151.20
13	P	126.78	144.80	135	141.90	143	146.20
14	Q	175.42	189.17	194	178.60	189	183.20
15	R	235.45	249.26	250	229.00	241	231.40
16	S	111.97	125.68	125	117.20	122	123.20
17	T	134.28	148.06	144	138.70	146	154.80
18	V	148.51	162.24	149	153.70	160	164.80
19	W	251.78	265.42	249	240.50	259	242.10
20	Y	224.68	238.30	227	213.70	229	218.00

used to evaluate the driving force for helix formations in proteins [46] They repeated the same exercise for the backbone surface of a residue in the unfolded state [3]. A data set of 43 proteins was selected. Then excised segments of a particular length were modeled to approximate the lower bound values. For upper bound values, the proteins were modelled in an extended conformation with backbone dihedrals set to ϕ = -120^o and Ψ = +120^o and side chain torsions set to 180^o. They concluded that most of the backbone surface was buried within the local structures [3]. As a practical measure, Schellman [52] took the mean of these two extreme values of a particular protein group, and Auton and Bolen [53] followed suit.

3.3. Model Based on Sequence Specificity

Bernado *et al.* [44] used a data set of 19 non-homologous proteins containing from 98 to 579 residues and reported average SASAs for all types of residues. These values were considerably lower than those for tripeptides and close to the lower limit reported by Creamer *et al.* [3]. They observed very high sequence dependence for the solvent accessibilities of all residues. Very small influences of both protein size and protein amino acid composition were observed in the averaged residue SASA for individual proteins. It has been implemented in an online web application known as ProtSA [7].

3.4. Statistical Model

Goldenberg [54] developed a model based on ensembles of polypeptide conformations of specific selected proteins. A simple Monte Carlo approach was used for this purpose. The SASA was estimated for each of the conformations generated in the simulations with excluded volume. The average values for the ensembles representing each of the proteins were plotted as a function of molecular weight. The tripeptide values were found to be lower than the average values from the ensembles [54].

3.5. Model Based on Solvent Quality

This physically rigorous model was developed by Gong and Rose [55] to estimate SASA values of amino acid residues in unfolded proteins. This model was based on the fact that SASA values would be larger in a good solvent where solute–solvent interactions dominate and promote chain extension. Likewise, they would be smaller in a poor solvent where solute–solute interactions dominate and promote chain collapse. These solvent-dependent effects were modeled by Boltzmann-weighting of a simulated ensemble for poor or good solvent quality. Solvent quality was dependent on intramolecular hydrogen bond strength. For the backbone, these midpoint SASA values were found to be in agreement with the earlier estimate of unfolded state SASA given by the mean of Creamer's upper and lower bound values [3].

3.6. Critical Analysis of Various Algorithms

All the tools and programs calculate SASA of the folded soluble proteins using their X-ray coordinates of their crystal structures. Atomic coordinates have been used to figure out the flexibility of the native proteins relative to SASA to simulate the unfolding of the these proteins [56]. Because of the lack of available coordinates, there is no direct empirical procedure to evaluate the SASA of the residues of unfolded proteins accurately. As we have mentioned and shown in section 3, researchers have used various models to simulate SASA of residues in the unfolded state. In the absence of the experimental data, the accuracy of these methods cannot be determined easily. The answer of the question, "Which model is the best?", is not straightforward, simple and easy. As evident from Table **1**, there is a wide range in the output values predicted by different models for the same set of residues. All proposed models have their own pros and cons. The problem with tri-peptide models is that the central residue does not always mimic the exact biological neighbourhood in the native structure. Thus these models tend to overestimate the SASA of the unfolded residues. Nearly all sequence based methods that provide absolute SASA values give drastically different outputs for the same residue. To cope up with the problem the 'extreme value models' were proposed wherein the SASA value of the residue lies between the two extreme values (see Table **2**). The shortcoming of these models is that they are based on simulation studies of some of the proteins that may not be generalized universally. To avoid such problems Singh *et al.* [57] performed the normalization of SASA values by taking into consideration the *context dependent* Highest Observed SASA (HOA) instead of using *context free* Extended State SASA (ESA). They studied the statistics of HOA of residues in all possible 400 combinations of each type of residue. Even this approach is not foolproof. The reason for saying this is that most of the hydrophobic residues are buried inside the native protein [58]. Therefore, their HOAs do not necessarily correspond to their fully extended conformation. Thus we conclude here that further research is required to provide more insights into the evaluation, validation, normalization and visualization of SASA in the unfolded state of proteins. It is hoped that we shall be able to answer this difficult question in the years to come with the help of the new advancements in structural sciences and technology.

4. PROGRAMS AND TOOLS FOR SASA PREDICTION

Since estimating SASA has far ranging biological applications, it, therefore, became a driving force for the prediction of SASA from amino acid sequences in computational biology [59]. A number of tools have been put forward to ascribe the location of residues as buried, partially buried and exposed based on specific thresholds of SASA [8]. Furthermore, some recent programs have been developed for predicting the real value of SASA using the sequence information [60]. Here we discuss some such useful and commonly used tools and programs.

4.1. Artificial Neural Network (ANN)

Rost and Sander [61] analyzed the three-dimensional structure of a protein from its amino acid sequence using their relative solvent accessibility. They developed an ANN system for predicting the SASA of each residue using evolutionary profiles of amino acid substitutions based on multiple sequence alignments. The predicted accuracy depends on the extent to which the residues are buried within the 3D structure. They have also developed a web server, PHDacc, for assigning the location of residues based on solvent accessibility. Cuff and Barton [62] used neural networks along with PSI-BLAST and HMMER profiles to improve the accuracy of solvent accessibility prediction. Ahmad and Gromiha [8] optimized ANN with many new features in its architecture and training method to develop a web server NETASA to accurately predict the SASA of amino acid residues. Adamczak *et al.* [28] developed a regression based neural network algorithm to accurately predict the relative SASA of residues in proteins. They also developed a novel method for secondary structure prediction that uses predicted relative SASA in addition to the attributes derived from evolutionary profiles [10]. Bravo *et al.* [63] optimized a method for analyzing codification schema for amino acids in an arbitrary space using neural networks. It was applied to optimize the amino acid codifications for the prediction of the SASA values of the proteins using feed-forward neural networks. Bondugula and Xu [64] proposed a novel method for SASA prediction based on known structure and sequence information. First the relative SASA of the query protein is estimated using fuzzy mean operator from the known structure fragments. The estimated SASA is then integrated with the position specific scoring matrix of the query protein using ANN. This approach has also been integrated into a web server MUPRED [64] that is available at http://digbio.missouri.edu/mupred. Recently, Faraggi and co-workers [65] developed a multi-step neural network algorithm called SPINE X to predict secondary structure, SASA and backbone torsion angles in an iterative manner. This method was applied to a dataset of 2640 proteins to achieve 82.0% accuracy on a 10-fold cross validation. Dor and Zhou [66] implemented an integrated system of neural networks (called Real-SPINE) for real-value prediction of physical parameters of proteins. They applied this method to predict residue-SASA and backbone dihedrals of proteins based on information derived from sequences only. An ensemble of artificial neural networks was trained by Petersen *et al.* [67] on a set of experimentally solved protein structures to predict the relative exposure of the amino acids. The method provides a reliability score to each surface accessibility prediction as an integral part of the training process. A Pearson's correlation coefficient of 0.72 is comparable to the performance of one of the most accurate available methods, Real-SPINE [67]. Both methods associate a reliability score with the individual predictions.

Table 2. The extreme SASA values in Å^2 proposed for each residue in the unfolded protein.

S.No.	Residue	Creamer *et al.* [3]		Gong and Rose values [55]		Bernado *et al.* [44]	
		Lower Bound	Upper Bound	Poor Solvent (Minimum)	Good Solvent (Maximum)	Minimum	Maximum
1	A	66.40	99.50	93.80	101.90	58.10	83.60
2	C	81.10	117.5	122.00	131.00	88.30	76.00
3	D	97.30	128.7	126.50	137.10	83.00	117.60
4	E	120.70	157.4	156.80	166.30	108.40	145.50
5	F	134.00	173.1	188.60	198.10	131.30	160.9
6	G	54.60	75.7	67.90	74.90	36.20	65.50
7	H	118.80	228.2	167.00	175.80	109.30	140.2
8	I	115.30	158.8	158.10	167.10	106.40	136.20
9	K	160.80	192.6	187.00	202.60	130.90	167.30
10	L	116.10	148.4	164.10	173.60	108.80	146.10
11	M	122.00	144.65	173.80	183.00	121.50	148.60
12	N	102.1	173.3	113.10	124.60	91.60	123.40
13	P	102.40	116.6	125.80	131.00	81.00	121.90
14	Q	122.20	162.1	138.70	152.00	107.10	140.40
15	R	174.00	196.15	209.90	230.30	154.80	193.70
16	S	83.50	108.3	101.40	111.60	59.20	89.90
17	T	95.90	120.7	121.80	132.50	78.10	109.20
18	V	97.70	135.8	134.70	143.50	84.00	116.10
19	W	169.80	190.4	226.10	239.20	160.40	185.50
20	Y	148.70	185.8	205.70	213.80	141.50	172.20

4.2. Support Vector Machine (SVM)

During the last decade various researchers have used the support vector machine to predict the SASA of proteins from their primary structures. For instance, Yuan and Huang [68] used this model to predict the relative SASAs of residues with correlation coefficient of 0.66. Wang *et al.* [69] established a new method of generating numerical real values of SASA, using accumulation cut-off set and SVM. It was called SVM cabins method. It first estimates discrete states of SASA of amino acid residues from their evolutionary profile and then maps the predicted states on to a real valued linear space by simple algebraic methods. They claimed this rigorous method to be comparable with best methods available so far. Liu *et al.* [70] used SASA among few other features to generate SVM classifiers to identify interaction sites of the proteins. Folkman *et al.* [23] analysed various features like solvent accessibility for the prediction of protein stability. They used SVM based library to implement their model.

4.3. Markov Chain Model

Markov Chain Model (MCM) approach was adopted by Wang *et al.* [71] for statistical modeling of relative SASA of protein residues. Prediction results for two different data sets and different cut-off thresholds were evaluated and compared with other existing methods, like as ANN, information theory and SVM. The best prediction accuracies achieved by the MCM method were 78.9% for the two-state prediction problem and 67.7% for the three-state prediction problem, respectively. The prediction accuracy and the correlative coefficient of the MCM method are better than or comparable to those obtained by the other prediction methods. Another advantage of this method is the lower computation complexity and better time-consuming performance.

4.4. Other Methods

During the last two decades researchers have used different methods to predict the SASA of proteins. For example, Bayesian statistics was applied by Thompson and Goldstein [72] for this purpose. A method called PredAcc was proposed by Mucchielli-Giorgi *et al.* [73] for predicting SASA in four states with machine learning based on an improved logistic function. Pascarella *et al.* [74] developed a simple method based on amino acid exchange and compositional preference matrices for each of the three accessible states:

buried, exposed, and intermediate. The prediction method proposed by Carugo [75] was based on the comparison of the observed and the average values of the SASA for a set of 338 monomeric, non-homologous and high-resolution protein crystal structures. A jackknife procedure was applied to each entry. McConkey *et al.* [19] described computational methods based on the Voronoi procedure that provided rapid and exact solutions to SASAs, volumes, and atom contacts within the proteins. Garg *et al.* [76] developed a multiple sequence alignment based method for predicting the real value of SASA from the sequence using evolutionary information. A multiple linear regression method was applied by Wang *et al.* [77] to predict real values of SASA from the sequence and evolutionary information to obtain the coefficients of regression and correlation between different parameters. A simple quadratic programming method was employed by Xu *et al.* [78] based on the parameter set of amino acids for tendency to become buried (or the "buriability"). This method, called QBES (Quadratic programming and Buriability Energy function for Solvent accessibility prediction) is the first method based on energy optimization. Likewise, multiple linear regression method was applied by Li and Pan [79] to predict the SASA of protein from its amino acid sequence. A novel method was introduced by Naderi-Manesh *et al.* [80] based on information theory to predict the SASA of amino acid residues in various states defined by their thresholds. Gianese *et al.* [81] described a new approach for predicting solvent accessibility from single sequence. This method was based on probability profiles calculated on an amino acid sequence centred on the residue whose accessibility has to be predicted. Gianese and Pascarella [82] proposed a consensus prediction method that combines the results of three different methods. Two methods, JPRED and ACCpro were based on neural networks while the remaining system, PP was based on probability profiles. PSAIA (Protein Structure and Interaction Analyzer) was developed by Mihel *et al.* [83] to compute geometric parameters including SASA and RSA for large sets of protein structures to investigate protein-protein interaction sites. This method was based on random triplet sampling and averaging of parameters. In order to predict metabolism sites for cytochrome P450, Rydberg *et al.* [84] constructed 2DSASA, a method for the calculation of the atomic SASA, that is independent of three-dimensional structure. The method was implemented in the SMARTCyp site of metabolism prediction models and improved the results by up to 4% for nine cytochrome P450 isoforms. Zhan *et al.* [85] carried out a comparison of the native structures predicted from various SASA models for the peptide Met-enkephalin (Tyr-Gly-Gly-Phe-Met). Both ECEPP/2 and ECEPP/3 force fields in conjunction with ten different sets of SASA parameterization were used for this purpose.

Richardson and Barlow [86] devised a simple method to predict residue SASA in proteins with the intention that it should be used as a baseline by which more sophisticated approaches could be judged. Guvenchand Brooks [87] introduced a fast and approximate all-atom SASA method parameterized using a set of folded and heat-denatured conformations of globular proteins. The parameters were shown to be transferable to folded and heat-denatured conformations for another set of proteins. For a 4644 atom protein, the calculation of the approximate SASA requires only 1/11th the CPU time required for estimation of the non-bonded interactions. Yu *et al.* [88] proposed a modified HMM for the prediction of RSA and backbone torsion angles (BTA) of local protein motifs.

Noval *et al.* [89] introduced a method that applies top-down Fourier transform mass spectrometry (FTMS) for the rapid profiling of amino acid side-chain reactivity. Many techniques were used to monitor SASA such as Edman degradation, NMR, etc. The reactivity of side-chain groups can be used to infer residue-specific SASA and can also be used to probe protein structure and interactions. Wang *et al.* [90] designed dictionaries of two-, three-, and five-residue patterns in proteins and computed the SASA of the central residues in their native proteins. These dictionaries serve as a look-up table for making predictions of SASA of amino acids. They concluded that the predictions made in this way are comparable to those made using more sophisticated methods of SASA prediction. They also assessed the effect of immediate neighbours on residue SASA. This method can accessed at the URL, www.netasa.org/look-up/.

Protein solvation energies are often taken to be proportional to their SASAs. Street and Mayo [91] originated a pair-residue approximation for SASAs to express all energy terms as single residue and pair residue terms. The main source of error in this methodology is the overhanging burial of side-chain surfaces in the protein core. Zhang *et al.* [92] gave a new pair-residue approximation which greatly reduces this overlap error by the use of optimized generic side-chains. They tested the generic-side-chain method for the ten proteins studied by Street and Mayo [91] and for 377 single-domain proteins from the CATH database [93]. The new method clearly reduces error for total areas as well as residue-by-residue areas by more than a factor of two.

Chen and Zhou [94] studied the prediction of SASA and sites of deleterious mutations from protein sequences. They tried to refine a set of five methods over a large dataset and described a meta-method based on an ensemble average, leading to a two-state classification of residue burial (buried or exposed) with an accuracy of 80%. Residues that were most positively classified as buried were proposed as sites of deleterious mutations. Out of the total of 130 residues predicted as sites of deleterious mutations, 104 were correct.

Fleming and co-workers [95] reported a novel method in which hydrophobic SASA is estimated after positioning water oxygen in hydrogen-bonded orientations proximate to all accessible peptide/protein backbone N and O atoms. This conditional hydrophobic accessible surface area was termed CHASA. The CHASA method was validated by predicting the polyproline-II and beta-strand conformational preferences of non-proline residues in the coil library. Further, the method successfully rationalized the previously unexplained solvation energies in polyalanyl peptides.

Gong *et al.* [96] developed a large-scale protein domain interaction interface database called InterPare. It contains both inter- and intra-chain interfaces. InterPare uses three methods to detect interfaces: (i) the geometric distance method for checking the distance between atoms that belong to different domains, (ii) estimation of SASA to detect the buried region of a protein, and (iii) the Voronoi diagram, a computational geometry method that uses a mathematical definition of interface regions. InterPare includes visualization tools to display protein interior, surface, and interaction interfaces. The atom coordinates belonging to interface, surface, and interior regions can be downloaded from the website http://interpare.net.

Hayryan *et al.* [97] suggested a new analytical method to compute protein SASA and its gradient. They considered the transformation that maps the spherical circles formed by intersection of the atomic surfaces in 3D space onto the circles on a two-dimensional plane. The algorithm is suitable for parallelization. It is comparable to other analogous algorithms. Mandell *et al.* [98] proposed methods to get the SASA of protein-ligand and protein-protein interfaces. The kinetics of SASA at the protein-protein interface is monitored by amide hydrogen/deuterium (H/2H) exchange detected by matrix assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (MS). Methods were also described for the measurement of tightly bound complexes of large interactions such as antibody-antigen complexes.

Kim and Choi [99] introduced a simple analytical method for the determination of the SASA of protein, consisting of linear functions of distances between unified atoms. Their formulation was much simpler than previously developed methods. It showed equipollent performance with the mean relative error on total SASA of 0.49 and 2.16% for 89 denatured and native protein structures, respectively. Chang *et al.* [10] studied the position specific scoring matrix (PSSM) based features for real value SASA prediction by considering the physicochemical properties and solvent propensities of residue types. They described a systematic method for identifying residue groups with respect to protein SASA. The amino acid columns in the PSSM profile belonging to a particular residue type are merged to generate novel features. Finally, support vector regression (SVR) was applied to construct a real value SASA predictor. Experimental results show that the proposed method is one of the best among several of existing packages for performing SASA prediction.

Leaver-Fay *et al.* [100] proposed a method to maintain accurate SASA during a Monte Carlo search of sequence and rotamer space for a fixed protein backbone. Le Grand and Merz [101] algorithm was extended that discretizes the SASA for each atom by placing dots on a sphere and combines Boolean masks to determine the exposed dots. They also optimized a SASA-based measure of protein packing for the complete redesign of a large set of proteins and protein-protein interactions. Pollastri *et al.* [102] developed a high-throughput machine learning systems for the prediction of protein secondary structure and SASA based on homology to proteins of known structure. These systems were compared to their *ab initio* counterparts in which these parameters were extracted directly from the templates. They concluded that the structural information from templates greatly improves secondary structure and SASA prediction quality. The predictive system is available at http://distill.ucd.ie.

Half-sphere exposure (HSE) is a newly developed measure of 2D solvent exposure. Song *et al.* [103] proposed a novel approach to predict the HSE measures based on a well-prepared non-homologous protein structure dataset. SVR was applied to assess the relationship between HSE measures and protein sequences and evaluate its prediction performance. The successful application of SVR approach suggested that it should be more useful in quantifying the protein sequence-structure relationship and predicting the structural property profiles from protein sequences. The prediction webserver is accessible at http://sunflower.kuicr.kyoto-u.ac.jp/~sjn/hse/.

Craig *et al.* [104] tried to analyze the protein SASA using a new experimental approach based on the reaction of the photochemical reagent diazirine with the polypeptide chain. By virtue of its size, diazirine is a reasonable analogue of aqueous solvent. They structurally characterized denatured states of the protein alpha-lactalbumin. Covalent tagging resulting from unspecific methylene reaction leads to a global determination of SASA and to map out solvent accessibility along the protein sequence. The virtual absence of a defined long-range organization brings about a feature less labeling pattern for the unfolded state.

Dynerman *et al.* [105] derived an analytical formula for estimation of protein SASA by deriving a simple model of desolvation energy as a differentiable function of atomic positions. They developed CUSA and CUDE, codes that calculate SASA and desolvation using the CUDA programming language. They showed that this algorithm was very well suited for hardware acceleration on graphics processing units (GPUs), outperforming the CPU by up to two orders of magnitude. They explored the scaling of this desolvation algorithm and presented implementation details applied to general pairwise algorithms.

Exact sequence context, the folding state of the residues, and the actual environment of a folded protein together enforce some additional constraints on the highest observed values (HOA) of SASA. Singh and Ahmad [57] analyzed the statistics of these constraints and assessed the normalization of absolute ASA values using context-dependent HOA instead of using context-free extended state SASA (ESA) of residues. They anthologized the statistics of HOA of residues in their different contexts and examined their distribution in all 400 possible combinations for each residue type. They concluded that many tripetides are more exposed than ESA. They also showed that HOA residues are often found in turn, coil and bend conformations.

The generalized Born/SASA implicit solvent model (GB/SA) stands to benefit from hybrid GPU/CPU computers, using the GPU for the GB calculation and the CPU for the SASA calculation. Tanner *et al.* [106] examined the computational challenges facing GB/SA calculations on hybrid GPU/CPU computers and analyzed the utilization of GPUs and CPUs by NAMD, a parallel molecular dynamics program, for fast GB/SA simulations. The hybrid computation principles thus explained, are generally applicable to parallel applications employing hybrid GPU/CPU calculations.

5. ONLINE RESOURCES TO CALCULATE SASA OF PROTEINS/RESIDUES

Several tools and softwares to estimate SASA specially in the folded state of proteins and their residues using the atomic coordinates [5], are available online. Very few online resources are mentioned below which are useful, widely used, reliable and user-friendly. All such tools and programs are having their own idiosyncrasies. Commonly used online resources are summarized in Table **3**.

Most of the online tools and servers that calculate SASA of a folded soluble protein are based on the Lee and Richards' fundamental definition [1]. Hence the overall SASA value of a folded protein, predicted by these methods, does not drastically deviate from one another. However, as far as individual residue SASA is concerned, it may vary depending on the method used (see Table **4**). We can conclude that more or less there is no superiority of one tool over the other. Nonetheless, DSSP remains the most widely used program for the calculation of SASA of folded protein residues, perhaps because of its cross linking with PDB [5].

5.1. PDBePISA

PDBePISA [107] is an interactive tool for the exploration interfaces, structures, assemblies of protein. Krissinel and Henrick [108] developed this method based on chemical thermodynamics for the detection of macromolecular assemblies [109]. The server provides a single-button analysis of X-ray structures, including the assessment of multimeric state, symmetry number, protein-protein interfaces, accessible and buried surface areas of the residues and proteins, etc. It was available on European Bioinformatics Institute (EBI) web server that is a part of European Molecular Biology Lab (EMBL), United Kingdom. PISA has claimed to reproduce 90% of complex structures verified by independent (noncrystallographic) experimental studies [110]. Its home page is shown in supplementary Fig. **S1**.

5.2. Collaborative Computational Project, Number 4 (CCP4)

CCP4 [111] software suite is an integrated cluster of about 200 autonomous programs and software libraries used for estimating a wide range of physico-chemical and structural parameters based on X-ray crystallographic data of the proteins. CCP4 is a community-based resource that supports all inclusive researcher community and academicians. This software has been regularly updated [112]. The current release of CCP4 program suite v.6.3.0 can be freely downloaded to a local machine or a server before being run and has graphical user interface. Its online version are also available [113]. To calculate SASA, there is a separate program called AreaMol that takes .pdb file as input and gives the SASA of the individual residues and the protein as well. A snapshot of the software has been shown in supplementary Fig. **S2**.

5.3. ProtSA

ProtSA [8] is a novel web application that calculates sequence-specific SASA in the unfolded state. If the atomic coordinate file is provided, it can estimate the change in the accessible surface after protein folding. Its architecture mainly consists of three parts: (i) the web browser of the user, (ii) a Common Gateway Interface application running on a web server, and (iii) the server part that calculates SASA of the unfolded-state protein ensemble. The server utilizes three external software tools to perform the calculations: (i) Flexible-Meccano for backbone conformation generation [44], SCCOMP [114] for side chain building, and ALPHASURF [43] for SASA calculations of each conformation of the unfolded protein ensemble. The program is very simple, user-friendly but cannot be downloaded on the local machine. The output file sent to the user by e-mail includes chain wise description of average residue SASA, main chain SASA, side chain SASA, polar and non-polar SASA all in folded and unfolded states with their difference [8]. Supplementary Figs. **S3** and **S4** depict the submission page and the output file of the program.

5.4. GETAREA

This method was developed by Fraczkiewicz and Braun [30] in late 1990s. It is considered as one of the fastest and exact analytical method available today to find SASA and its gradient for proteins [115]. It finds solvent-exposed vertices of intersecting atoms, and avoids calculating buried vertices that are not required to estimate the SASA by the Gauss-Bonnet theorem [116]. The surface routine was implemented in FANTOM [117], a program for energy minimization and Monte Carlo simulation [116]. The CPU time for the accurate determination of the SASA has been reduced by a factors of 2.2 for Met-enkephalin as compared with their previous approach [116]. The efficiency of this method is almost similar to the approximate methods like MSEED [118] and SASAD [119]. The input is the Cartesian coordinates of the protein stored in PDB format in local disk and the output is SASA different of formats. This is a very straightforward and user friendly program. The input interface and the result page has been shown in supplementary Figs. **S5** and **S6**.

5.5. Define Secondary Structure of Protein (DSSP)

DSSP [120] is one of the oldest and most widely accepted tool to estimate the SASA of individual residues of the protein. The program was originally developed in standard Pascal by Kabsch and Sander [120, 121] to standardize the secondary structure of the protein. A database of SASA was set up for secondary structure assignments for all protein entries deposited in PDB. This tool describes the secondary structure of a protein based on its three-dimensional structure. DSSP is the de facto standard in the field of secondary structure determination [122]. Recently, Hekkelman [122, 123] developed a new software that gives same output as the original DSSP, but is much faster and easier to maintain. This new software is called DSSP and the original software is referred to as DSSPold. The home page and PDB cross-linked data file have been depicted in supplementary Figs. **S7** and **S8**.

5.6. Other Online Resources

There are numerous tools available for estimating SASA in the folded ensemble of the proteins using atomic coordinates [124, 125]. The programs NACCESS developed by Hubbard and Thornton [33], ACCESS introduced by Richmond and Richards [32], POPS (Parameter Optimized Surfaces) developed by Cavallo *et al.* [126], Alpha Shapes software developed by the research group of Edelsbrunner [127], ASAP (Accessible Surface Area Predictor) developed by

Table 3. List of computational programs available for SASA calculation.

S.No.	Program	Specification	URL	Developer	Refreence
1	PDBePISA	Gives residue SASA for folded state of protein	http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html	EBI (EMBL), U.K.	[107]
2	CCP4	Complete structure analysis suite	http://www.ccp4.ac.uk/	RcaH, STFC Rutherford A. Labs, U.K.	[111, 112]
3	ProtSA	Sequence specific average SASA of unfolded ensemble	http://oldwebapps.bifi.es/protSA/	BIFI, Spain	[8, 44]
4	GETAREA	Analytical calculation of SASA based on MC simulation	http://curie.utmb.edu/getarea.html	SCSB, University of Texas, U.S.A.	[30]
5	DSSP	Database of secondary structure assignments	http://swift.cmbi.ru.nl/gv/dssp/	CMBI, Nijmegen, Netherlands	[120, 121]
6	NACCESS	Calculates atomic and residue ASA for proteins and nucleic acids	http://wolf.bms.umist.ac.uk/naccess/	University of Manchester, U.K.	[33]
7	ACCESS	Gives SASA of backbone and side chain atom of each residue	http://www.csb.yale.edu/	Yale University, U.S.A	[32]
8	POPS-R	Fast and analytical method, residue based approach for large assemblies like ribosomes	http://mathbio.nimr.mrc.ac.uk/~ffranca/POPS	NIMR, London, U.K.	[126]
9	SERF	Facilitates the use of SASA in in structure analysis like changes during binding and complexation	guitar.rockeller.edu/pub/jpo/serf.tar.	DPMS, U.K.	[31]
10	ASAP	SVM based tool for calculating SASA of transmembrane residues	http://ccb.imb.uq.edu.au/ASAP/	University of Queensland, Australia	[128]
11	SABLE	Linear regression based method for RSA prediction	http://sable.cchmc.org	CHRF, Cincinnati, U.S.A	[28, 129]

Yaun *et al.* [128] for soluble and transmembrane residues, sequence based method SABLE [28, 129] for proteins of unknown structures, so on and so forth. The available online resources for SASA calculations have been summarized in Table **3**. A comparison of Folded state SASA values for the first 30 residues of Chymotrypsin Inhibitor 2 (PDBID 2CI2) using different available resources is given in Table **4**.

6. PICTORIAL ILLUSTRATION OF SASA

The pictorial representation of SASA values is very useful for conception of the location of each residue in the structure of the protein. It depicts the quantitative burial level of a particular residue within the three-dimensional structure of the protein [5, 124].

6.1. ASAView

ASAView is an algorithm of schematic illustration of SASA of amino acid residues within proteins. Its application and database was developed by Ahmad *et al.* [130]. An emblematic 2D spiral plot of SASA caters to a convenient graphical view of residues in terms of their exposed surface areas. Several additional features are provided for better visualization. It takes input in different formats which are outputs of programs like DSSP, RVP, GETAREA, Naccess, etc. This is useful in structural analysis of the proteins, specifically for monitoring the topological distribution of residues in a nutshell. The user interface and the resulting SASA representation have been shown in supplementary Figs. **S9** and **S10**. The URL is http://www.abren.net/asaview/.

6.2. POLYVIEW-2D

This user friendly visualization tool developed by Porollo *et al.* [131] for structure and function based annotation of proteins. The POLYVIEW server can be used to generate protein sequence annotations that include relative SASA of proteins. 2D graphical illustration in a user friendly format may be generated for both known and unknown protein structures. In this tool, the relative solvent accessibility is represented with numerical values starting from 0 to 9, with 0 corresponding to fully buried (0-9% RSA) and 9 corresponding to fully exposed residue (90-100% RSA), respectively [5, 124]. The input page and the resulting graphical presentation have been depicted in supplementary Figs. **S11** and **S12**.

Table 4. Folded state SASA of the first 30 residues of chymotrypsin inhibitor 2 (PDB ID-2CI2) using different tools.

S.No	Residue	DSSP	PISA	ProtSA	CCP4	GETAREA
1	N	208.00	191.79	193.00	200.10	191.07
2	L	119.00	129.91	128.11	126.00	127.97
3	K	81.00	82.05	82.08	76.30	84.79
4	T	57.00	55.35	55.65	55.70	53.41
5	E	84.00	72.37	73.25	79.00	72.18
6	W	0.00	0.00	0.00	0.00	0.00
7	P	64.00	69.84	70.06	68.80	71.18
8	E	91.00	86.92	87.35	87.00	88.18
9	L	2.00	1.68	1.45	2.10	0.85
10	V	67.00	69.52	69.30	67.90	70.64
11	G	47.00	48.19	48.83	49.60	49.94
12	K	82.00	81.47	81.38	77.40	84.20
13	S	19.00	23.99	23.20	20.30	23.99
14	V	10.00	12.55	12.50	10.90	12.61
15	E	100.00	99.02	100.09	101.30	99.83
16	E	105.00	90.57	91.23	98.20	88.67
17	A	0.00	0.00	0.00	0.00	0.00
18	K	95.00	90.74	90.57	93.70	89.46
19	K	136.00	135.11	136.54	135.40	138.61
20	V	50.00	52.80	52.52	51.00	52.39
21	I	0.00	0.00	0.00	0.00	0.00
22	L	77.00	78.58	79.15	77.90	79.65
23	Q	154.00	144.08	143.79	147.10	138.24
24	D	72.00	64.57	63.40	69.00	60.41
25	K	9.00	10.11	10.78	7.60	9.05
26	P	100.00	105.71	106.11	104.00	109.52
27	E	98.00	92.69	94.19	94.90	95.65
28	A	8.00	9.02	8.68	10.40	8.52
29	Q	118.00	105.68	106.40	108.90	104.37
30	I	40.00	37.12	37.71	40.50	33.63

7. THE CORRELATION OF SASA WITH OTHER PHYSICAL PARAMETERS OF PROTEINS AND ITS APPLICATIONS

Hydrophobic force has been considered as one of the major driving force for protein folding [132]. Chothia [133] has tried to establish a relationship between SASA and hydrophobicity of different residues that constitute a protein. He concluded that there exists a linear relationship between SASA and hydrophobicity of non polar side chains and side chains with hydroxyl groups (Ser, Thr, Tyr) with correlation coefficient of 0.998. Residues having similar SASA with polar groups were found to be 1 kcal less hydrophobic than those with non polar groups.

Chothia [134] then analyzed a set of 15 proteins to find an association between the loss of SASA on protein folding and hydrophobicity. He deduced that loss of SASA by monomeric proteins on folding was a function of their molecular mass. He also calculated the SASA and the proportion of the polar groups forming intramolecular hydrogen bonds in all of the 15 proteins including insulin, lysozyme, subtilisin, etc. He further analyzed the volume occupied by amino acid residues in the interior of the 9 proteins and concluded that they occupy the same volume as they do in amino acid crystals. In the experiments carried out by Rose *et al.* [49], proteins of known structure were used to measure the average area that each residue buries upon folding. This characteristic quantity, average area buried, was correlated with residue hydrophobicity. The work of Moret and Zebenda [6] indicates that the variation of the SASA describes an alternative hydrophobicity scale.

Teller [135] further examined the empirical equations of Chothia [134] for SASA of 12 monomeric proteins and demonstrated that SASA of a monomeric protein varies 2/3 power of its molecular weight. The total SASA of a polypeptide chain in fully extended state was found to be linear function of its molecular weight which is contributed by each residue. Islam and Weaver [136] simply estimated the SASA of 58 unique proteins (39 monomeric and 19 dimeric ones) in crystalline state and simply related that to their molecular weight. They deduced that there was a direct relationship between SASA and molecular weight in both types of proteins.

Stellwagen and Wilgus [137] examined the relationship between thermal stability and SASA of protein. For globular proteins, as surface to volume ratio decreases with chain length, they related observed T_{m} (midpoint of thermal denaturation) values to the fractional SASA of the protein polypeptide. They obtained the results for 30 proteins and established a linear relationship between these parameters.

Eisenberg and McLachlan [138] devised a method for calculating the protein stability in water from its atomic coordinates. The contribution of each protein atom to the solvation free energy was estimated as the product of its SASA and atomic solvation parameter for transfer from the interior of a protein to aqueous solution. Their method effectively combined the calculation of SASA and estimating energies from transfer-free energies by weighting the effect of each atom by its polar or apolar nature.

Miller *et al.* [139] studied relationship between SASA and the stability of the oligomeric proteins. They showed that the SASA of oligomeric proteins and buried surface area are directly related to relative molecular mass. It has important implications for these surface areas in the stability and activity of oligomeric proteins. The structural features of oligomeric proteins could also be related to their thermodynamic and physical properties. Ooi *et al.* [140] described a method to use SASA as a measure of thermodynamic parameters of hydration of peptides. The free energy of hydration comprises of a cumulative contributions of various functional groups. The hydration of each group was assumed to be proportional to SASA of that group. Proportionality constants depicting the free energy of hydration per unit SASA, were evaluated for 7 classes of groups (present in peptides) using least-squares fitting of experimental free energies of solution. The same method was also applied to the modeling of the enthalpy and heat capacity of hydration, each of which was calculated from the SASA of the peptides.

Delarue and Koehl [141] derived atomic contact potentials by statistical analysis of atomic contact areas with respect to atom type in a set of non-homologous protein structures. The atomic environment is characterized by SASA and the surface of contacts with polar and non-polar atoms. These atomic potentials clearly discriminate misfolded from correct structural models. It was suggested that these potentials reflect the atomic short range non-local interactions in proteins. For characterization of atomic solvation alone, similar potentials were derived as a function of the percentage of SASA alone.

Masuda *et al.* [142] used a SASA-scaled atomistic method to estimate molecular hydrophobicity (log P) by taking into account the proximity effect of substituent groups as well as the importance of solute-solvent interaction in the partition phenomena. This novel method reassigns atomic parameters when the molecule is fully exposed to surrounding solvent. Each atom in a molecule contributes to the log P by an amount of its atomic parameter multiplied by the degree of exposure to the surrounding solvent, which is dependent on its SASA value.

Krishnan and Cosman [143] derived an empirical relationship between rotational correlation time and SASA of 75 proteins with known structures. The theoretical correlation between SASA and correlation time was also considered. SASA was determined from the structure and correlation time was computed from diffusion tensor calculations.

Bustamante *et al.* [144] explored five enzymes for which polymorphic sequence variation within *Escherichia coli* and/or *Salmonella enterica* was available. Single and multivariate logistic regression models were developed that could evaluate physicochemical properties like SASA as predictors of polymorphism. The proposed model predicts an increase in the probability of amino acid polymorphism with increasing SASA for each protein regardless of physicochemical properties, secondary structure element, or size of the amino acid. The results showed a strong decrease in purifying selection with increasing SASA of the proteins.

Courtenay *et al.* [145, 146] carried out thermodynamic analysis of interactions between denaturants and protein surface exposed on unfolding for interpreting urea and guanidinium chloride m-values and their correlation with changes in SASA using preferential interaction coefficients and the local-bulk domain model. They did quantify the interactions of urea and guanidinium chloride with native bovine serum albumin surface using vapour pressure osmometry.

Luise *et al.* [147] found that the main factor influencing water residence time in proximity to a specific protein site was its SASA. The protein atom with low accessible surface in an intraprotein hydrogen bond modulates the length of the water residence time. All atomic surfaces having high SASA, independently of their character, are surrounded by water molecules which rapidly exchange with the bulk solvent.

Samata *et al.* [148] worked on the quantification of the packing of residues in proteins in relation to their SASAs. The number of atoms in contact (within a distance of 4.5 Å) can be used to describe the local environment of a residue. As this number increases, the SASA of the residue decreases exponentially. This exponential equation provides a method to estimate the SASA of a protein molecule. They also concluded that the average RSA of different residues is inversely correlated with their hydrophobicity values.

Wohlfahrt *et al.* [149] described the significance of SASA and secondary structure elements in positioning of anchor groups in protein loop prediction. For 550 insertions and 544 deletions they tested all possible positions for anchor groups with an inserted loop of a length between 3 and 12 residues. They concluded that amino acids with lower SASA are better anchor group.

Efimov and Brazhnikov [150] showed that intramolecular hydrogen bonding in proteins depends on the SASA of donor and acceptor groups. The frequency of occurrence of H-bonded side chains in proteins is inversely proportional to the SASA of their donor and acceptor groups. They concluded that intramolecular H-bonding interactions of buried and half-buried donors and acceptors can contribute favorably to the stability of a protein, while those of solvent-exposed polar atoms become less favorable.

Hou and Xu [151] devised a novel method for the calculations of 1-octanol/water partition coefficient (log *P*) of organic molecules. This method, SLOGP v1.0 estimates the log *P* values by summing the contribution of atom-weighted SASA and other correction factors.

Zou and Zou [152] studied the average contribution of individual residue to folding stability and its dependence on buried SASA. They showed that the contribution of a residue has a significant correlation with buried SASA and the regression slopes (buriability parameter) of all 20 amino acid residues are all positive. The buriability parameter is a quantitative measure of the driving force for the burial of a residue. The large buriability gap observed between hydrophobic and hydrophilic residues is the reason for the burial of hydrophobic residues in globular proteins.

Bogatyreva and Ivankov [153] studied the relationship between protein SASA and the number of native contacts in its structure. They showed that a decrease in SASA caused by the change in protein conformation during its folding is accompanied by the corresponding increase in the number of native contacts. This correlation can be used for an accurate and rapid calculation of the protein SASA from the number of native contacts.

Goodarzi *et al.* [154] have tried to use SASA as an ingredient of a robust amino acid substitution matrix along with residue charge and volume. This so called SCV matrix (SASA charge volume matrix) supports the uncontaminated nature of this matrix regarding the genetic code. SCV matrix which has some similarities with a number of previously available cost measure matrices, results in a more significant optimality in the error-buffering capacity of the genetic code.

Liu *et al.* [155] developed a method called SP(4) for fold recognition that is based on protein SASA and other parameters like, residue-depth, structure-derived sequence profiles, etc. SP(4) was found to be improved over the previous method SP(3) in the sensitivity of fold recognition. The SP(4) server and its local usage package are available on http://sparks.informatics.iupui.edu/SP4.

Negi *et al.* [156] designed a web server called InterProSurf that predicts amino acid residues in proteins that are supposed to interact with other proteins. This method is based on SASA of residues in the isolated subunits, a propensity scale for interface residues and a clustering algorithm to identify surface regions with residues of high interface propensities.

am Busch *et al.* [157] tested the Coulomb/Accessible Surface Area (CASA) solvent model for protein stability, ligand binding, and protein design. They carried out a new optimization using a set of experimental stability changes for single point mutations of various proteins and peptides as a target. The optimization procedure is general in nature and could be used with other force fields.

Pearlman *et al.* [158] described a new approach FURSMASA (function for rapid scoring using an MD-averaged grid and the accessible surface area) which can be used for rapidly ranking the binding of ligands to proteins as well as for estimating relative aqueous molecular solubilities. One of its novel features is the inclusion of a term that depends on the change in the SASA on an atomic basis.

Pal *et al.* [159] deduced a relationship between average SASA values of amino acid residues and their partner number (PN). PN is the number of other residues within a distance of 4.5 Å from any atom of a given residue. A web server, ContPlot has been developed to display these values (relative to the average values) along the protein sequence. This is useful to visually identify residues that are closely packed, or that involved in protein-protein interactions.

A study was performed by Shaytan *et al.* [160] to analyze the distribution of residues based on their SASA in more than 8000 protein structures. Using extensive statistical sampling, they evaluated residue apparent free energies of transfer between protein interior and surface. The correlation of these statistical energies with several experimental hydrophobicity scales was presented. They proposed three types of statistical apparent transfer-free energy scales and showed that each of these scales is in better correlation with one of the experimental hydrophobicity scales (water/vapour, water/cyclohexane, and water/octanol transfer scales).

A new method was developed by Tuncbag *et al.* [161] to predict computational hot spots based on SASA, conservation and statistical pair-wise residue potentials of the interface residues. This empirical method is a simple approach in hot spot prediction with high accuracy and efficacy. The list of training and test sets are available as supplementary data at http://prism.ccbb.ku.edu.tr/hotpoint/supplement.doc.

Vranken and Rieping [162] presented a relationship between the chemical shift values obtained from NMR experiments and per-atom SASA as per their 3D coordinates. The repository for chemical shift data is the BioMagResBank that provides NMR-STAR files with this type of information. Atoms with zero per atom SASA have significantly larger chemical shift dispersion and generally have a different chemical shift distribution. With higher per atom ASA, the chemical shift values also tend towards random coil values. The data are available online on http://www.ebi.ac.uk/pdbe/docs/NMR/shiftAnalysis/index.html.

Zhan *et al.* [163] investigated the relationship between the residue flexibility (B-factor) and its relative solvent accessibility (RSA) in the context of local neighborhood and related concepts like residue depth. They observed that the flexibility of a given residue is strongly influenced by the SASA of the adjacent neighbours. Correlation between the local RSA and B-factor is shown to be stronger than the correlation that considers local distance- or volume-based residue depth. They found that the correlation coefficients between B-factor and RSA, called flexibility-exposure correlation index, are strongly correlated with the stability scale that signifies the average contributions of each amino acid to the folding stability.

Chang *et al.* [164] presented a novel sequence-based method for predicting protein-protein interactions based on the assumption that protein-protein interactions are more related to amino acids at the surface than those at the core. The proposed method maintains the advantage of relying on only sequence data by including a SASA predictor. The prediction performance achieved by using the SASA predictor is close to that using the surface obtained from protein structures. This proposed method of surface identification improved the prediction performance with an F-measure of 5.1%.

Recently, Zellner *et al.* [165] also implemented the program PresCont which predicts amino acid residues constituting protein-protein interfaces (PPIs). The core of PresCont is an SVM that assesses properties like SASA, conservation, hydrophobicity and the local environment of each residue on the protein surface. For PPIs of permanent complexes, SASA and hydrophobicity contribute most to classification quality whereas for PPIs of transient complexes, the assessment of the local environment is most significant. PresCont is available as a web server at http://www-bioinf.uni-regensburg.de/.

Quite recently, Basse *et al.* [166] have developed a database called 2P2Idb, a hand-curated structural database dedicated to protein-protein interactions with known orthosteric modulators. It includes all interactions for which both the protein-protein and protein-ligand complexes have been structurally characterized. A large range of descriptors are computed including buried accessible surface area, gap volume, non-bonded contacts, hydrogen-bonds, atom and residue composition, number of segments and secondary structure contribution. All together the 2P2I database represents a structural source of information for scientists from academic institutions or pharmaceutical industries, which is freely accessible at http://2p2idb.cnrs-mrs.fr.

A novel approach was given by Gao *et al.* [167] for accurate prediction of protein folding rates from residue flexibility and SASA. Protein folding rates depend on the topology of the fold and composition of the sequence. The proposed sequence based predictor, PFR-AF, includes three linear regressions for proteins with two-state, multistate and mixed-state folding kinetics. They showed that increased flexibility of coils facilitates faster folding, and the proteins with larger SASA values may fold at a slower pace.

Nunez *et al.* [168] evaluated a novel scoring method based on SASA descriptors for its database enrichment potential against the virtual screening for ligand-receptor interactions. Several proteins such as adenosine deaminase and estrogen receptor alpha were used for the evaluation purpose. These SASA descriptors display an outstanding robustness to generate satisfactory early enrichments for a large variety of target classes. These novel topological descriptors comprise a valuable *in silico* tool in hit finding practices.

Xia *et al.* [169] introduced an efficient tool called APIS (A combined model based on Protrusion Index and Solvent accessibility) for accurate prediction of hot spots in protein interfaces by combining protrusion index with SASA. It uses support vector machine to predict hot spot residues in protein interfaces. This proposed method yields significantly better prediction accuracy than those previously reported in the literature. They demonstrated the predictive power of this method by modelling two protein complexes: the calmodulin/myosin light chain kinase complex and the heat shock locus gene products U and V complex. The data and source code are available on web site http://home.ustc.edu.cn/~jfxia/hotspot.html.

Alcaro *et al.* [170] carried out computational study to investigate the conformational properties of telomerase and the relationships between the target affinity and SASA of the ligands. Their aim was to rationalize the different experimental activities of known telomerase inhibitors because the telomerase had been regarded as one of the most attractive targets in cancer treatment. It provided useful preliminary information to discriminate end-stacking ligand binding affinities, revealing itself as a helpful predictive tool in drug design and lead optimization processes.

Lu *et al.* [171] designed a computational method to identify the carboxylation sites of proteins using SASA as one of the predictors. This prediction method was implemented as Carboxylator (http://csb.cse.yzu.edu.tw/Carboxylator/), a web-based tool for the identification of carboxylated proteins. The protein carboxylation is very important due to its involvement in biological processes such as blood clotting cascade and bone growth. The experimental identification of carboxylation sites via mass spectrometry-based methods is very expensive, time-consuming and labour-intensive. They intended to investigate the protein carboxylation by considering the composition of amino acids that surround modification sites. Cross-validation using the combined features of SASA, amino acid sequence and composition yields the highest accuracy levels.

Marsh and Teichmann [172] have observed the relationships between protein structures and the conformational changes they undergo upon binding. They used the RSA to predict the magnitude of binding-induced conformational changes from the structures of monomeric proteins. They observed considerable enrichment of intrinsically disordered sequences in proteins predicted to undergo large conformational changes. They demonstrated that the RSA of monomeric proteins can be used as a simple measure for protein flexibility.

Chen *et al.* [173] introduced a novel method based on neural networks for the prediction of protein interaction sites. It was based on the extraction on a many features like SASA, entropy, sequence profiles, entropy, relative entropy, conservation weight etc. The proposed method is claimed to perform better than the other related methods such as SVMs.

Wang and Hou [174] devised a fast approach to estimate the conformational entropy based on SASA calculations. According to their approach, the conformational entropy of a molecule can be obtained by summing up the contributions of all buried or exposed atoms. To parameterize entropy model, entropies were calculated for a large set of small molecules taking the solvent effect into account. The weighted solvent accessible surface area (WSAS) model was thoroughly evaluated in three tests. This model could find its applications in the fields like molecular docking, high throughput screening and rational protein design.

Duann *et al.* [175] analyzed the effect of hydrophobic force on 5 protein complexes in which helical chains are bound together, using Molecular Dynamics (MD) simulations. The simulation study employed three different methods to treat hydrophobic effect, the standard Generalized Born (GB) method that lacks explicit hydrophobic force, the LCPO method in which the explicit hydrophobic force is based on SASA, and a proposed packing enforced GB (PEGB) method inclusive of explicit hydrophobic force based on the radius of gyration of the protein complex. They proposed that the PEGB method proved to be quite promising for MD simulation of large, multi-domain packed proteins in implicit solvent mode.

Ghattyvenkatakrishna *et al.* [176] described an MD simulation study of the effect of trehalose concentration on the structure and dynamics of hen egg-white lysozyme. They found that the changes in trehalose concentration do not alter the global structural characteristics of the protein as measured by standard quantities like SASA, mean square deviation, radius of gyration, etc.

Li *et al.* [177] presented an overview of experimentally validated de novo designed proteins and a comparison of few available programs such as RosettaDesign, EGAD, Liang-Grishin and RosettaDesign-SR, by analyzing designed sequences computationally. Computational analysis includes the total SASA, the recovery of native sequences, the calculation of sizes of hydrophobic patches, etc. It can be inferred from this computational assessment that the next-generation protein-design scoring function will arise from the right balance of complementary interaction term.

Wei and co-workers [178] studied the significance of the SASA within biologically relevant oligomeric assemblies. They found that the SASA from biological assemblies causes statistically significant improvement in prediction over the SASA of monomers from protein crystal structures. This information with sequence-based features in an SVM leads to 82% accuracy on a balanced dataset of 50% disease-associated mutations from SwissVar and 50% neutral mutations from primate sequence differences in orthologous proteins.

7.1. Applications of Various Models of the Denatured State for the Estimation of a Stability Parameter

Auton and Bolen [53] showed that the solvent dependent cooperative folding-unfolding free energy changes can be predicted using transfer free-energies of protein groups from water to osmolyte solution and the fractional increase in the exposure of the surface area of these groups on unfolding of the protein. The *m*-value is a measure of the osmolytic cooperativity of the transition between N and D states. It measures the efficacy of the osmolyte in favoring the folding or unfolding of the protein. It can be calculated from the following relation [53]:

$$m_{\text{calc}} = \Delta G^{1M}_{tr,D} - \Delta G^{1M}_{tr,N} = \sum n_i \ \Delta g_{tr,i} \ \Delta\alpha_i \qquad (5)$$

where n_i is the number of residues of type i whose transfer-free energy from water to 1 M osmolyte solution is given by $\Delta g_{tr,i}$ and $\Delta\alpha_i$ is the fractional change in SASA of ith type of residue while going from water to osmolyte solution. $\Delta\alpha_i$ is determined from the relation [179, 180]:

$$\Delta\alpha_i = \Sigma\ (X_{ASA,i} - F_{ASA,i}) / X_{ASA,i} \qquad (6)$$

where $X_{ASA,i}$ and $F_{ASA,i}$ are SASA values of ith type of residue in the denatured and native states, respectively. These equations (5 and 6) were used to estimate the *m*-values of a protein, barstar (PDBID 1BTA) in 1 M sarcosine solution using different models for denatured states given in Tables **1** and **2**. We calculated $F_{ASA,i}$ value for each type of residue in the native structure of barstar using DSSP server. For the estimation of $X_{ASA,i}$, we used each model for the denatured state given in Tables **1** and **2**. Transfer-free energy values for all 20 residues were taken from the work of Auton and Bolen [53]. We obtained *m*-values of 2.79, 2.90, 2.88, 2.83, 2.87, 2.87, 2.22, 2.68, 2.63, 2.73, 1.89, and 2.51 kcalmol^{-1}M^{-1} using models Ala-X-Ala, Gly-X-Gly, Lins *et al.*, ASC, Miller, Zielenkiewicz and Saenger, Creamer (LB), Creamer (UB), Gong and Rose (poor solvent), Gong and Rose (good solvent), Bernado (minimum) and Bernado (maximum), respectively. The experimental *m*-value of barstar is 2.33 kcalmol^{-1}M^{-1} [181]. A comparison of this value with *m*-values predicted by different models suggests that the deviation of predicted values from the experimental one ranges from 19% to 24%. The accepted error in the determination of *m*-values is within 10%. Thus, significantly different *m*-values from different models tend to erode our confidence in these models for estimating SASA of residues in unfolded proteins. As far as the estimation of this thermodynamic parameter is concerned, only two models, Creamer (LB) and Bernado (maximum) give the most reasonable *m*-values.

CONCLUSIONS

SASA has been considered as one of the most important physical parameters in protein science for the prediction of various physico-chemical and thermodynamic properties. In this review, we have defined the various types of SASA and their significance. Early and recent models to compute SASA of protein groups in the folded state have been explained in detail. It has been observed that the prediction of values of SASA of folded protein groups using various prediction methods is very accurate and model independent. We have also described the various algorithms used to estimate the SASA of residues in unfolded states of proteins. Contrary to the folded state SASA, there does not exist any exact, accurate and error free method for SASA calculations of the unfolded proteins. Various online resources available for estimating SASA have also been listed. The applications, uses and correlations of SASA with various physico-chemical and thermodynamic properties have been thoroughly reviewed. Some of these properties include molecular weight, hydrophobicity, radius of gyration, free energy changes during protein folding, transfer-free energies, intermolecular hydrogen bonding, partition coefficients, etc. Furthermore, concept of solvent accessibility is very essential for finding protein-protein interfaces, analyzing the effects of mutations on protein stability, thermodynamic studies of proteins, discovery of *de novo* drugs, *in silico* molecular modelling and protein engineering. We hope that this review will be an impetus for protein researchers and scientific community at large.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

SAA, MIH and FA are thankful to the Department of Science and Technology, Government of India for financial assistance.

SUPPLEMENTARY MATERIALS

Supplementary material is available on the publisher's web site along with the published article.

REFERENCES

[1] Lee, B.; Richards, F.M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.*, **1971**, *55,* 379-400.

[2] Parthiban, V.; Gromiha, M.M.; Hoppe, C.; Schomburg, D. Structural analysis and prediction of protein mutant stability using distance and torsion potentials: role of secondary structure and solvent accessibility. *Proteins*, **2007**, *66,* 41-52.

[3] Creamer, T.P.; Srinivasan, R.; Rose, G.D. Modeling unfolded states of proteins and peptides. II. Backbone solvent accessibility. *Biochemistry*, **1997**, *36,* 2832-2835.

[4] Gates, R.E. Shape and accessible surface area of globular proteins. *J. Mol. Biol.,* **1979**, *127,* 345-351

[5] Gromiha, M.; Ahmad, S. Role of solvent accessibility in structure based drug design. *Curr. Comput. Aided Drug Des.*, **2005**, *1,* 65-72.

[6] Moret, M.A.; Zebende, G.F. Amino acid hydrophobicity and accessible surface area. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, **2007**, *75,* 011920.

[7] Durham, E.; Dorr, B.; Woetzel, N.; Staritzbichler, R.; Meiler, J. Solvent accessible surface area approximations for rapid and accurate protein structure prediction. *J. Mol. Model*, **2009**, *15,* 1093-1098.

[8] Estrada, J.; Bernado, P.; Blackledge, M.; Sancho, J. ProtSA: a web application for calculating sequence specific protein solvent accessibilities in the unfolded ensemble. *BMC Bioinformatics*, **2009**, *10,* 104.

[9] Ahmad, S.; Gromiha, M.M. NETASA: neural network based prediction of solvent accessibility. *Bioinformatics*, **2002**, *18,* 819-824.

[10] Chang, D.T.; Huang, H.Y.; Syu, Y.T.; Wu, C.P. Real value prediction of protein solvent accessibility using enhanced PSSM features. *BMC Bioinformatics*, **2008**, *9 Suppl 12,* S12.

[11] Adamczak, R.; Porollo, A.; Meller, J. Combining prediction of secondary structure and solvent accessibility in proteins. *Proteins*, **2005**, *59,* 467-475.

[12] Momen-Roknabadi, A.; Sadeghi, M.; Pezeshk, H.; Marashi, S.A. Impact of residue accessible surface area on the prediction of protein secondary structures. *BMC Bioinformatics*, **2008**, *9,* 357.

[13] Kurt, N.; Cavagnero, S. The burial of solvent-accessible surface area is a predictor of polypeptide folding and misfolding as a function of chain elongation. *J. Am. Chem. Soc.*, **2005**, *127,* 15690-15691.

[14] Spolar, R.S.; Record, M.T.; Jr. Coupling of local folding to site-specific binding of proteins to DNA. *Science*, **1994**, *263,* 777-784.

[15] Tien, M.Z.; Meyer, A.G.; Sydykova, D.K.; Spielman, S.J.; Wilke, C.O. Maximum allowed solvent accessibilites of residues in proteins. *PLoS One*, **2013**, *8,* e80635.

[16] Bahadur, R.P.; Chakrabarti, P.; Rodier, F.; Janin, J. A dissection of specific and non-specific protein-protein interfaces. *J. Mol. Biol.*, **2004**, *336,* 943-955.

[17] Mandell, J.G.; Baerga-Ortiz, A.; Akashi, S.; Takio, K.; Komives, E.A. Solvent accessibility of the thrombin-thrombomodulin interface. *J. Mol. Biol.*, **2001**, *306,* 575-589.

[18] Hughes, R.E.; Rice, P.A.; Steitz, T.A.; Grindley, N.D. Protein-protein interactions directing resolvase site-specific recombination: a structure-function analysis. *EMBO J.*, **1993**, *12,* 1447-1458.

[19] McConkey, B.J.; Sobolev, V.; Edelman, M. Quantification of protein surfaces, volumes and atom-atom contacts using a constrained Voronoi procedure. *Bioinformatics*, **2002**, *18,* 1365-1373.

[20] Qin, Y.; Yang, Y.; Zhang, L.; Fowler, J.D.; Qiu, W.; Wang, L.; Suo, Z.; Zhong, D. Direct probing of solvent accessibility and mobility at the binding interface of polymerase (Dpo4)-DNA complex. *J. Phys. Chem. A*, **2013**, *117,* 13926-13934.

[21] Ochoa, D.; Garcia-Gutierrez, P.; Juan, D.; Valencia, A.; Pazos, F. Incorporating information on predicted solvent accessibility to the co-evolution-based study of protein interactions. *Mol. Biosyst.*, **2013**, *9,* 70-76.

[22] Mirabello, C.; Pollastri, G. Porter, PaleAle 4.0: high-accuracy prediction of protein secondary structure and relative solvent accessibility. *Bioinformatics*, **2013**, *29,* 2056-2058.

[23] Folkman, L.; Stantic, B.; Sattar, A. Sequence-only evolutionary and predicted structural features for the prediction of stability changes in protein mutants. *BMC Bioinformatics*, **2013**, *14 Suppl 2,* S6.

[24] Kouranov, A.; Xie, L.; de la Cruz, J.; Chen, L.; Westbrook, J.; Bourne, P.E.; Berman, H.M. The RCSB PDB information portal for structural genomics. *Nucleic Acids Res.*, **2006**, *34,* D302-D305.

[25] Lins, L.; Thomas, A.; Brasseur, R. Analysis of accessible surface of residues in proteins. *Protein Sci.*, **2003**, *12,* 1406-1417.

[26] Wang, C.; Xi, L.; Li, S.; Liu, H.; Yao, X. A sequence-based computational model for the prediction of the solvent accessible surface area for alpha-helix and beta-barrel transmembrane residues. *J. Comput. Chem.*, **2012, *33,*** 11-17.

[27] Berman, H.M.; Henrick, K.; Nakamura, H.; Markley, J.; Bourne, P.E.; Westbrook, J. Realism about PDB. *Nat. Biotechnol.*, **2007**, *25,* 845-846; author reply 6.

[28] Adamczak, R.; Porollo, A.; Meller, J. Accurate prediction of solvent accessibility using neural networks-based regression. *Proteins*, **2004**, *56,* 753-767.

[29] Weiser, J.; Shenkin, P.S.; Still, W.C. Approximate solvent-accessible surface areas from tetrahedrally directed neighbor densities. *Biopolymers*, **1999**, *50,* 373-380.

[30] Fraczkiewicz, R.; Braun, W. Exact and Efficient Analytical Calculation of the Accessible Surface Areas and Their Gradients for Macromolecules. *J. Comput. Chem.*, **1998**, *19,* 319-333.

[31] Flower, D.R. SERF: a program for accessible surface area calculations. *J. Mol. Graph Model*, **1997**, *15,* 238-244.

[32] Richmond, T.J.; Richards, F.M. Packing of alpha-helices: geometrical constraints and contact areas. *J. Mol. Biol.*, **1978**, *119,* 537-555.

[33] Hubbard, S.J.; Thornton, J.M. NACCESS. London: Department of Biochemistry and Molecular Biology, University College London; 1993.

[34] Gibson, K.D.; Scheraga, H.A. Exact calculation of the volume and surface-area of fused hard-sphere molecules with unequal atomic radii. *Mol. Phys.*, **1987**, *62,* 1247-1265.

[35] Dodd, L.R.; Theodorou, D.N. Analytical treatment of the volume and surface area of molecules. *Mol. Phys.*, **1991**, *72,* 1313-1345.

[36] Richmond, T.J. Solvent accessible surface area and excluded volume in proteins. Analytical equations for overlapping spheres and implications for the hydrophobic effect. *J. Mol. Biol.*, **1984**, *178,* 63-89.

[37] Wodak, S.J.; Janin, J. Analytical approximation to the accessible surface area of proteins. *Proc. Natl. Acad. Sci. U. S. A.*, **1980**, *77,* 1736-1740.

[38] Hasel, W.; Hendrickson, T.F.; Still, W.C. A rapid approximation to the solvent accessible surface areas of atoms. *Tetrahedron Comput. Methods*, **1988**, *1,* 103-116.

[39] Shrake, A.; Rupley, J.A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.*, **1973**, *79,* 351-371.

[40] Edelsbrunner, H.; Koehl, P. The geometry of biomolecular solvation. *Discrete Comput. Geom.*, **2005**, *52,* 241-273.

[41] Weiser, J.; Shenkin, P.S.; Still, W.C. Approximate Atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comput. Chem.,* **1999**, *20,* 217-230.

[42] Bryant, R.; Edelsbrunner, H.; Koehl, P.; Levitt, M. The area derivative of a space-filling diagram. *Discrete Comput. Geom.*, **2004**, *32,* 293-308.

[43] Edelsbrunner, H.; Koehl, P. The weighted-volume derivative of a space-filling diagram. *Proc. Natl. Acad. Sci. U. S. A.*, **2003**, *100,* 2203-2208.

[44] Bernado, P.; Blackledge, M.; Sancho, J. Sequence-specific solvent accessibilities of protein residues in unfolded protein ensembles. *Biophys. J.*, **2006**, *91,* 4536-4543.

[45] Fitzkee, N.C.; Rose, G.D. Reassessing random-coil statistics in unfolded proteins. *Proc. Natl. Acad. Sci. U. S. A.*, **2004**, *101,* 12497-12502.

[46] Creamer, T.P.; Srinivasan, R.; Rose, G.D. Modeling unfolded states of peptides and proteins. *Biochemistry*, **1995**, *34,* 16245-16250.

[47] Zielenkiewicz, P.; Saenger, W. Residue solvent accessibilities in the unfolded polypeptide chain. *Biophys. J.*, **1992**, *63,* 1483-1486.

[48] Sneddon, S.F.; Tobias, D.J. The role of packing interactions in stabilizing folded proteins. *Biochemistry*, **1992**, *31,* 2842-2846.

[49] Rose, G.D.; Geselowitz, A.R.; Lesser, G.J.; Lee, R.H.; Zehfus, M.H. Hydrophobicity of amino acid residues in globular proteins. *Science*, **1985**, *229,* 834-838.

[50] Miller, S.; Janin, J.; Lesk, A.M.; Chothia, C. Interior and surface of monomeric proteins. *J. Mol. Biol.*, **1987**, *196,* 641-656.

[51] Livingstone, J.R.; Spolar, R.S.; Record, M.T., Jr. Contribution to the thermodynamics of protein folding from the reduction in water-accessible nonpolar surface area. *Biochemistry*, **1991**, *30,* 4237-4244.

[52] Schellman, J.A. Protein stability in mixed solvents: a balance of contact interaction and excluded volume. *Biophys. J.*, **2003**, *85,* 108-125.

[53] Auton, M.; Bolen, D.W. Predicting the energetics of osmolyte-induced protein folding/unfolding. *Proc. Natl. Acad. Sci. U. S. A.*, **2005**, *102,* 15065-15068.

[54] Goldenberg, D.P. Computational simulation of the statistical properties of unfolded proteins. *J. Mol. Biol.*, **2003**, *326,* 1615-1633.

[55] Gong, H.; Rose, G.D. Assessing the solvent-dependent surface area of unfolded proteins using an ensemble model. *Proc. Natl. Acad. Sci. U. S. A.*, **2008**, *105,* 3321-3326.

[56] Benson, N.C.; Daggett, V. Dynameomics: large-scale assessment of native protein flexibility. *Protein Sci.*, **2008**, *17,* 2038-2050.

[57] Singh, H.; Ahmad, S. Context dependent reference states of solvent accessibility derived from native protein structures and assessed by predictability analysis. *BMC Struct. Biol.*, **2009**, *9,* 25.

[58] Betts, S.; Haase-Pettingell, C.; Cook, K.; King, J. Buried hydrophobic side-chains essential for the folding of the parallel beta-helix domains of the P22 tailspike. *Protein Sci.*, **2004**, *13,* 2291-2303.

[59] Singh, Y.H.; Gromiha, M.M.; Sarai, A.; Ahmad, S. Atom-wise statistics and prediction of solvent accessibility in proteins. *Biophys. Chem.*, **2006**, *124,* 145-154.

[60] Ahmad, S.; Gromiha, M.M.; Sarai, A. RVP-net: online prediction of real valued accessible surface area of proteins from single sequences. *Bioinformatics*, **2003**, *19,* 1849-1851.

[61] Rost, B.; Sander, C. Conservation and prediction of solvent accessibility in protein families. *Proteins*, **1994**, *20,* 216-226.

[62] Cuff, J.A.; Barton, G.J. Application of multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins*, **2000**, *40,* 502-511.

[63] Arauzo-Bravo, M.J.; Ahmad, S.; Sarai, A. Dimensionality of amino acid space and solvent accessibility prediction with neural networks. *Comput. Biol. Chem.*, **2006**, *30,* 160-168.

[64] Bondugula, R.; Xu, D. Combining sequence and structural profiles for protein solvent accessibility prediction. *Comput. Syst. Bioinformatics Conf.*, **2008**, *7,* 195-202.

[65] Faraggi, E.; Zhang, T.; Yang, Y.; Kurgan, L.; Zhou, Y. SPINE X: improving protein secondary structure prediction by multistep learning coupled with prediction of solvent accessible surface area and backbone torsion angles. *J. Comput. Chem.*, **2012**, *33,* 259-267.

[66] Dor, O.; Zhou, Y. Real-SPINE: an integrated system of neural networks for real-value prediction of protein structural properties. *Proteins*, **2007**, *68,* 76-81.

[67] Petersen, B.; Petersen, T.N.; Andersen, P.; Nielsen, M.; Lundegaard, C. A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Struct. Biol.*, **2009**, *9,* 51.

[68] Yuan, Z.; Huang, B. Prediction of protein accessible surface areas by support vector regression. *Proteins*, **2004**, *57,* 558-564.

[69] Wang, J.Y.; Lee, H.M.; Ahmad, S. SVM-Cabins: prediction of solvent accessibility using accumulation cutoff set and support vector machine. *Proteins*, **2007**, *68,* 82-91.

[70] Liu, R.; Jiang, W.; Zhou, Y. Identifying protein-protein interaction sites in transient complexes with temperature factor, sequence profile and accessible surface area. *Amino Acids*, **2010**, *38,* 263-270.

[71] Wang, M.; Li, A.; Wang, X.; Feng, H. Prediction of protein solvent accessibility with Markov chain model. *Sheng. Wu. Yi. Xue. Gong. Cheng. Xue. Za. Zhi.*, **2006**, *23,* 1109-1113.

[72] Thompson, M.J.; Goldstein, R.A. Predicting solvent accessibility: higher accuracy using Bayesian statistics and optimized residue substitution classes. *Proteins*, **1996**, *25,* 38-47.

[73] Mucchielli-Giorgi, M.H.; Hazout, S.; Tuffery, P. PredAcc: prediction of solvent accessibility. *Bioinformatics*, **1999**, *15,* 176-177.

[74] Pascarella, S.; De Persio, R.; Bossa, F.; Argos, P. Easy method to predict solvent accessibility from multiple protein sequence alignments. *Proteins*, **1998**, *32,* 190-199.

[75] Carugo, O. Predicting residue solvent accessibility from protein sequence by considering the sequence environment. *Protein Eng.*, **2000**, *13,* 607-609.

[76] Garg, A.; Kaur, H.; Raghava, G.P. Real value prediction of solvent accessibility in proteins using multiple sequence alignment and secondary structure. *Proteins*, **2005**, *61,* 318-324.

[77] Wang, J.Y.; Lee, H.M.; Ahmad, S. Prediction and evolutionary information analysis of protein solvent accessibility using multiple linear regression. *Proteins*, **2005**, *61,* 481-491.

[78] Xu, Z.; Zhang, C.; Liu, S.; Zhou, Y. QBES: predicting real values of solvent accessibility from sequences by efficient, constrained energy optimization. *Proteins*, **2006**, *63,* 961-966.

[79] Li, X.; Pan, X.M. New method for accurate prediction of solvent accessibility from protein sequence. *Proteins*, **2001**, *42,* 1-5.

[80] Naderi-Manesh, H.; Sadeghi, M.; Arab, S.; Moosavi Movahedi, A.A. Prediction of protein surface accessibility with information theory. *Proteins*, **2001**, *42,* 452-459.

[81] Gianese, G.; Bossa, F.; Pascarella, S. Improvement in prediction of solvent accessibility by probability profiles. *Protein Eng.*, **2003**, *16,* 987-992.

[82] Gianese, G.; Pascarella, S. A consensus procedure improving solvent accessibility prediction. *J. Comput. Chem.*, **2006**, *27,* 621-626.

[83] Mihel, J.; Sikic, M.; Tomic, S.; Jeren, B.; Vlahovicek, K. PSAIA - protein structure and interaction analyzer. *BMC Struct. Biol.*, **2008**, *8,* 21.

[84] Rydberg, P.; Rostkowski, M.; Gloriam, D.E.; Olsen, L. The contribution of atom accessibility to site of metabolism models for cytochromes P450. *Mol. Pharm.*, **2013**, *10,* 1216-1223.

[85] Zhan, L.; Chen, J.Z.; Liu, W.K. Comparison of predicted native structures of Met-enkephalin based on various accessible-surface-area solvent models. *J. Comput. Chem.*, **2009**, *30,* 1051-1058.

[86] Richardson, C.J.; Barlow, D.J. The bottom line for prediction of residue solvent accessibility. *Protein Eng.*, **1999**, *12,* 1051-1054.

[87] Guvench, O.; Brooks, C.L., 3rd. Efficient approximate all-atom solvent accessible surface area method parameterized for folded and denatured protein conformations. *J. Comput. Chem.*, **2004**, *25,* 1005-1014.

[88] Yu, J.; Xiang, L.; Hong, J.; Zhang, W. HMM-based prediction for protein structural motifs' two local properties: solvent accessibility and backbone torsion angles. *Protein Pept. Lett.*, **2013**, *20,* 156-164.

[89] Novak, P.; Kruppa, G.H.; Young, M.M.; Schoeniger, J. A top-down method for the determination of residue-specific solvent accessibility in proteins. *J. Mass Spectrom.*, **2004**, *39,* 322-328.

[90] Wang, J.Y.; Ahmad, S.; Gromiha, M.M.; Sarai, A. Look-up tables for protein solvent accessibility prediction and nearest neighbor effect analysis. *Biopolymers*, **2004**, *75,* 209-216.

[91] Street, A.G.; Mayo, S.L. Pairwise calculation of protein solvent-accessible surface areas. *Fold Des.*, **1998**, *3,* 253-258.

[92] Zhang, N.; Zeng, C.; Wingreen, N.S. Fast accurate evaluation of protein solvent exposure. *Proteins*, **2004**, *57,* 565-576.

[93] Orengo, C.A.; Michie, A.D.; Jones, S.; Jones, D.T.; Swindells, M.B.; Thornton, J.M. CATH--a hierarchic classification of protein domain structures. *Structure*, **1997**, *5,* 1093-1108.

[94] Chen, H.; Zhou, H.X. Prediction of solvent accessibility and sites of deleterious mutations from protein sequence. *Nucleic Acids Res.*, **2005**, *33,* 3193-3199.

[95] Fleming, P.J.; Fitzkee, N.C.; Mezei, M.; Srinivasan, R.; Rose, G.D. A novel method reveals that solvent water favors polyproline II over beta-strand conformation in peptides and unfolded proteins: conditional hydrophobic accessible surface area (CHASA). *Protein Sci.*, **2005**, *14,* 111-118.

[96] Gong, S.; Park, C.; Choi, H.; Ko, J.; Jang, I.; Lee, J.; Bolser, D.M.; Oh, D.; Kim, D.S.; Bhak, J. A protein domain interaction interface database: InterPare. *BMC Bioinformatics*, **2005**, *6,* 207.

[97] Hayryan, S.; Hu, C.K.; Skrivanek, J.; Hayryane, E.; Pokorny, I. A new analytical method for computing solvent-accessible surface area of macromolecules and its gradients. *J. Comput. Chem.*, **2005**, *26,* 334-343.

[98] Mandell, J.G.; Baerga-Ortiz, A.; Falick, A.M.; Komives, E.A. Measurement of solvent accessibility at protein-protein interfaces. *Methods Mol. Biol.*, **2005**, *305,* 65-80.

[99] Kim, R.G.; Choi, C.Y. A linear function for the approximation of accessible surface area of proteins. *Protein Pept. Lett.*, **2006**, *13,* 549-553.

[100] Leaver-Fay, A.; Butterfoss, G.L.; Snoeyink, J.; Kuhlman, B. Maintaining solvent accessible surface area under rotamer substitution for protein design. *J. Comput. Chem.*, **2007**, *28,* 1336-1341.

[101] Le Grand, S.M.; Merz, K.M. Rapid approximation to molecular surface area via the use of Boolean logic and look-up tables. *J. Comput. Chem.*, **1993**, *14,* 349-352.

[102] Pollastri, G.; Martin, A.J.; Mooney, C,.; Vullo, A. Accurate prediction of protein secondary structure and solvent accessibility by consensus combiners of sequence and structure information. *BMC Bioinformatics*, **2007**, *8,* 201.

[103] Song, J.; Tan, H.; Takemoto, K.; Akutsu, T. HSEpred: predict half-sphere exposure from protein sequences. *Bioinformatics*, **2008**, *24,* 1489-1497.

[104] Craig, P.O.; Gomez, G.E.; Ureta, D.B.; Caramelo, J.J.; Delfino, J.M. Experimentally approaching the solvent-accessible surface area of a protein: insights into the acid molten globule of bovine alpha-lactalbumin. *J. Mol. Biol.*, **2009**, *394,* 982-993.

[105] Dynerman, D.; Butzlaff, E.; Mitchell, J.C. CUSA and CUDE: GPU-accelerated methods for estimating solvent accessible surface area and desolvation. *J. Comput. Biol.*, **2009**, *16,* 523-537.

[106] Tanner, D.E.; Phillips, J.C.; Schulten, K. GPU/CPU Algorithm for Generalized Born/Solvent-Accessible Surface Area Implicit Solvent Calculations. *J. Chem. Theory Comput.*, **2012**, *8,* 2521-2530.

[107] Krissinel, E.; Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **2007**, *372,* 774-797.

[108] Krissinel, E.; Henrick, K. Detection of protein assemblies in crystals. *Lecture Notes Comput. Sci.*, **2005**, *3695,* 163-174.

[109] Bhat, T.N.; Bourne, P.; Feng, Z.; Gilliland, G.; Jain, S.; Ravichandran, V.; Schneider, B.; Schneider, K.; Thanki, N.; Weissig, H.; Westbrook, J.; Berman, H.M. The PDB data uniformity project. *Nucleic Acids Res.*, **2001**, *29,* 214-218.

[110] Krissinel, E. Crystal contacts as nature's docking solutions. *J. Comput. Chem.*, **2010**, *31,* 133-143.

[111] Winn, M.D.; Ballard, C.C.; Cowtan, K.D.; Dodson, E.J.; Emsley, P.; Evans, P.R.; Keegan, R.M.; Krissinel, E.B.; Leslie, A.G.; McCoy, A.; McNicholas, S.J.; Murshudov, G.N.; Pannu, N.S.; Potterton, E.A.; Powell, H.R.; Read, R.J.; Vagin, A.; Wilson, K.S. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.*, **2011**, *67,* 235-242.

[112] Potterton, L.; McNicholas, S.; Krissinel, E.; Gruber, J.; Cowtan, K.; Emsley, P.; Murshudov, G.N.; Cohen, S.; Perrakis, A.; Noble, M. Developments in the CCP4 molecular-graphics project. *Acta. Crystallogr. D Biol. Crystallogr.*, **2004**, *60,* 2288-2294.

[113] Potterton, E.; McNicholas, S.; Krissinel, E.; Cowtan, K.; Noble, M. The CCP4 molecular-graphics project. *Acta. Crystallogr. D Biol. Crystallogr.*, **2002**, *58,* 1955-1957.

[114] Eyal, E.; Najmanovich, R.; McConkey, B.J.; Edelman, M.; Sobolev, V. Importance of solvent accessibility and contact surfaces in modeling side-chain conformations in proteins. *J. Comput. Chem.*, **2004**, *25,* 712-724.

[115] Rychkov, G.; Petukhov, M. Joint neighbors approximation of macromolecular solvent accessible surface area. *J. Comput. Chem.*, **2007**, *28,* 1974-1989.

[116] Mumenthaler, C.; Braun, W. Folding of Globular Proteins by Energy Minimization and Monte Carlo Simulations with Hydrophobic Surface Area Potentials. *J. Mol. Model*, **1995**, *1,* 1-10.

[117] Schaumann, T.; Braun, W.; Wuthrich, K. The Program FANTOM for Energy Refinement of Polypetides and Proteins Using a Newton-Raphson Minimizer in Torsion Angle Space. *Biopolymers*, **1990**, *29,* 679-693.

[118] Perrot, G.; Cheng, B.; Gibson, K.D.; Vila, J.; Palmer, K.A.; Nayeem, A.; Maigret, B.; Scheraga, H.A. MSEED: A program for the rapid analytical determination of accessible surface areas and their derivatives. *J. Comput. Chem.*, **1992**, *13,* 1-11.

[119] Sridharan, S.; Nicholls, A.; Sharp, K.A. A rapid method for calculating derivatives of solvent accessible surface areas of molecules. *J. Comput. Chem.*, **1995**, *16,* 1038-1044.

[120] Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **1983**, *22,* 2577-2637.

[121] Kabsch, W.; Sander, C. How good are predictions of protein secondary structure? *FEBS Lett.*, **1983**, *155,* 179-182.

[122] Joosten, R.P.; te Beek, T.A.; Krieger, E.; Hekkelman, M.L.; Hooft, R.W.; Schneider, R.; Sander, C.; Vriend, G. A series of PDB related databases for everyday needs. *Nucleic Acids Res.*, **2010**, *39,* D411-D419.

[123] Hekkelman, M.L.; Te Beek, T.A.; Pettifer, S.R.; Thorne, D.; Attwood, T.K.; Vriend, G. WIWS: a protein structure bioinformatics Web service collection. *Nucleic Acids Res.*, **2010**, *38,* W719-W723.

[124] Gromiha, M. Protein Bioinformatics: from sequence to function. First ed: Academic Press, Elsevier; 2010.

[125] Gibas, C.; Jambeck, P. Developing Bioinformatics Computer Skills. 4th Indian Reprint ed. LeJeune L, editor. Mumbai: O'Reilly & Associates, Inc.; 2003.

[126] Cavallo, L.; Kleinjung, J.; Fraternali, F. POPS: A fast algorithm for solvent accessible surface areas at atomic and residue level. *Nucleic Acids Res.*, **2003**, *31,* 3364-3366.

[127] Liang, J.; Edelsbrunner, H.; Fu, P.; Sudhakar, P.V.; Subramaniam, S. Analytical shape computation of macromolecules: I. Molecular area and volume through alpha shape. *Proteins*, **1998**, *33,* 1-17.

[128] Yuan, Z.; Zhang, F.; Davis, M.J.; Boden, M.; Teasdale, R.D. Predicting the solvent accessibility of transmembrane residues from protein sequence. *J. Proteome Res.*, **2006**, *5,* 1063-1070.

[129] Wagner, M.; Adamczak, R.; Porollo, A.; Meller, J. Linear regression models for solvent accessibility prediction in proteins. *J. Comput. Biol.*, **2005**, *12,* 355-369.

[130] Ahmad, S.; Gromiha, M.; Fawareh, H.; Sarai, A. ASAView: database and tool for solvent accessibility representation in proteins. *BMC Bioinformatics*, **2004**, *5,* 51.

[131] Porollo, A.A.; Adamczak, R.; Meller, J. POLYVIEW: a flexible visualization tool for structural and functional annotations of proteins. *Bioinformatics*, **2004**, *20,* 2460-2462.

[132] Kauzmann, W. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.*, **1959**, *14,* 1-63.

[133] Chothia, C. Hydrophobic bonding and accessible surface area in proteins. *Nature*, **1974**, *248,* 338-339.

[134] Chothia, C. Structural invariants in protein folding. *Nature*, **1975**, *254,* 304-308.

[135] Teller, D.C. Accessible area, packing volumes and interaction surfaces of globular proteins. *Nature*, **1976**, *260,* 729-731.

[136] Islam, S.A.; Weaver, D.L. Molecular interactions in protein crystals: solvent accessible surface and stability. *Proteins*, **1990**, *8,* 1-5.

[137] Stellwagen, E.; Wilgus, H. Relationship of protein thermostability to accessible surface area. *Nature*, **1978**, *275,* 342-343.

[138] Eisenberg, D.; McLachlan, A.D. Solvation energy in protein folding and binding. *Nature*, **1986**, *319,* 199-203.

[139] Miller, S.; Lesk, A.M.; Janin, J.; Chothia, C. The accessible surface area and stability of oligomeric proteins. *Nature*, **1987**, *328,* 834-836.

[140] Ooi, T.; Oobatake, M.; Nemethy, G.; Scheraga, H.A. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. *Proc. Natl. Acad. Sci. U. S. A.*, **1987**, *84,* 3086-3090.

[141] Delarue, M.; Koehl, P. Atomic environment energies in proteins defined from statistics of accessible and contact surface areas. *J. Mol. Biol.*, **1995**, *249,* 675-690.

[142] Masuda, T.; Jikihara, T.; Nakamura, K.; Kimura, A.; Takagi, T.; Fujiwara, H. Introduction of solvent-accessible surface area in the calculation of the hydrophobicity parameter log P from an atomistic approach. *J. Pharm. Sci.*, **1997**, *86,* 57-63.

[143] Krishnan, V.V.; Cosman, M. An empirical relationship between rotational correlation time and solvent accessible surface area. *J. Biomol. NMR*, **1998**, *12,* 177-182.

[144] Bustamante, C.D.; Townsend, J.P.; Hartl, D.L. Solvent accessibility and purifying selection within proteins of Escherichia coli and Salmonella enterica. *Mol. Biol. Evol.*, **2000**, *17,* 301-308.

[145] Courtenay, E.S.; Capp, M.W.; Saecker, R.M.; Record, M.T., Jr. Thermodynamic analysis of interactions between denaturants and protein surface exposed on unfolding: interpretation of urea and guanidinium chloride m-values and their correlation with changes in accessible surface area (ASA) using preferential interaction coefficients and the local-bulk domain model. *Proteins*, **2000**, *Suppl 4,* 72-85.

[146] Courtenay, E.S.; Capp, M.W.; Record, M.T., Jr. Thermodynamics of interactions of urea and guanidinium salts with protein surface: relationship between solute effects on protein processes and changes in water-accessible surface area. *Protein Sci.*, **2001**, *10,* 2485-2497.

[147] Luise, A.; Falconi, M.; Desideri, A. Molecular dynamics simulation of solvated azurin: correlation between surface solvent accessibility and water residence times. *Proteins*, **2000**, *39,* 56-67.

[148] Samanta, U.; Bahadur, R.P.; Chakrabarti, P. Quantifying the accessible surface area of protein residues in their local environment. *Protein Eng.*, **2002**, *15,* 659-667.

[149] Wohlfahrt, G.; Hangoc, V.; Schomburg, D. Positioning of anchor groups in protein loop prediction: the importance of solvent accessibility and secondary structure elements. *Proteins*, **2002**, *47,* 370-378.

[150] Efimov, A.V.; Brazhnikov, E.V. Relationship between intramolecular hydrogen bonding and solvent accessibility of side-chain donors and acceptors in proteins. *FEBS Lett.*, **2003**, *554,* 389-393.

[151] Hou, T.J.; Xu, X.J. ADME evaluation in drug discovery. 2. Prediction of partition coefficient by atom-additive approach based on atom-weighted solvent accessible surface areas. *J. Chem. Inf. Comput. Sci.*, **2003**, *43,* 1058-1067.

[152] Zhou, H.; Zhou, Y. Quantifying the effect of burial of amino acid residues on protein stability. *Proteins*, **2004**, *54,* 315-322.

[153] Bogatyreva, N.S.; Ivankov, D.N. The relationship between protein accessible surface area and number of native contacts in its structure. *Mol. Biol. (Mosk)*, **2008**, *42,* 1048-1055.

[154] Goodarzi, H.; Katanforoush, A.; Torabi, N.; Najafabadi, H.S. Solvent accessibility, residue charge and residue volume, the three ingredients of a robust amino acid substitution matrix. *J. Theor. Biol.*, **2007**, *245,* 715-725.

[155] Liu, S.; Zhang, C.; Liang, S.; Zhou, Y. Fold recognition by concurrent use of solvent accessibility and residue depth. *Proteins*, **2007**, *68,* 636-645.

[156] Negi, S.S.; Schein, C.H.; Oezguen, N.; Power, T.D.; Braun, W. InterProSurf: a web server for predicting interacting sites on protein surfaces. *Bioinformatics*, **2007**, *23,* 3397-3399.

[157] am Busch, M.S.; Lopes, A.; Amara, N.; Bathelt, C.; Simonson, T. Testing the Coulomb/Accessible Surface Area solvent model for protein stability, ligand binding, and protein design. *BMC Bioinformatics*, **2008**, *9,* 148.

[158] Pearlman, D.A.; Rao, B.G.; Charifson, P. FURSMASA: a new approach to rapid scoring functions that uses a MD-averaged potential energy grid and a solvent-accessible surface area term with parameters GA fit to experimental data. *Proteins*, **2008**, *71,* 1519-1538.

[159] Pal, A.; Bahadur, R.P.; Ray, P.S.; Chakrabarti, P. Accessibility and partner number of protein residues, their relationship and a webserver, ContPlot for their display. *BMC Bioinformatics*, **2009**, *10,* 103.

[160] Shaytan, A.K.; Shaitan, K.V.; Khokhlov, A.R. Solvent accessible surface area of amino acid residues in globular proteins: correlation of apparent transfer free energies with experimental hydrophobicity scales. *Biomacromolecules*, **2009**, *10,* 1224-1237.

[161] Tuncbag, N.; Gursoy, A.; Keskin, O. Identification of computational hot spots in protein interfaces: combining solvent accessibility and inter-residue potentials improves the accuracy. *Bioinformatics*, **2009**, *25,* 1513-1520.

[162] Vranken, W.F.; Rieping, W. Relationship between chemical shift value and accessible surface area for all amino acid atoms. *BMC Struct. Biol.*, **2009**, *9,* 20.

[163] Zhang, H.; Zhang, T.; Chen, K.; Shen, S.; Ruan, J.; Kurgan, L. On the relation between residue flexibility and local solvent accessibility in proteins. *Proteins*, **2009**, *76,* 617-636.

[164] Chang, D.T.; Syu, Y.T.; Lin, P.C. Predicting the protein-protein interactions using primary structures with predicted protein surface. *BMC Bioinformatics*, **2010**, *11 Suppl 1,* S3.

[165] Zellner, H.; Staudigel, M.; Trenner, T.; Bittkowski, M.; Wolowski, V.; Icking, C.; Merkl, R. PresCont: predicting protein-protein interfaces utilizing four residue properties. *Proteins*, **2012**, *80,* 154-168.

[166] Basse, M.J.; Betzi, S.; Bourgeas, R.; Bouzidi, S.; Chetrit, B.; Hamon, V.; Morelli, X.; Roche, P. 2P2Idb: a structural database dedicated to orthosteric modulation of protein-protein interactions. *Nucleic Acids Res.*, **2013**, *41,* D824-D827.

[167] Gao, J.; Zhang, T,.; Zhang, H.; Shen, S.; Ruan, J.; Kurgan, L. Accurate prediction of protein folding rates from sequence and sequence-derived residue flexibility and solvent accessibility. *Proteins*, **2010**, *78,* 2114-2130.

[168] Nunez, S.; Venhorst, J,.; Kruse, C.G. Assessment of a novel scoring method based on solvent accessible surface area descriptors. *J. Chem. Inf. Model*, **2010**, *50,* 480-486.

[169] Xia, J.F.; Zhao, X.M.; Song, J.; Huang, D.S. APIS: accurate prediction of hot spots in protein interfaces by combining protrusion index with solvent accessibility. *BMC Bioinformatics*, **2010**, *11,* 174.

[170] Alcaro, S.; Artese, A.; Costa, G.; Distinto, S.; Ortuso, F.; Parrotta L. Conformational studies and solvent-accessible surface area analysis of known selective DNA G-Quadruplex binders. *Biochimie*, **2011**, *93,* 1267-1274.

[171] Lu, C.T.; Chen, S.A.; Bretana, N.A.; Cheng, T.H.; Lee, T.Y. Carboxylator: incorporating solvent-accessible surface area for identifying protein carboxylation sites. *J. Comput. Aided Mol. Des.*, **2011**, *25,* 987-995.

[172] Marsh, J.A.; Teichmann, S.A. Relative solvent accessible surface area predicts protein conformational changes upon binding. *Structure*, **2011**, *19,* 859-867.

[173] Chen, Y.; Xu, J.; Yang, B.; Zhao, Y.; He, W. A novel method for prediction of protein interaction sites based on integrated RBF neural networks. *Comput. Biol. Med.*, **2012**, *42,* 402-407.

[174] Wang, J.; Hou, T. Develop and test a solvent accessible surface area-based model in conformational entropy calculations. *J. Chem. Inf. Model*, **2012**, *52,* 1199-1212.

[175] Duan, L.L.; Zhu, T.; Mei, Y.; Zhang, Q.G; Tang, B.; Zhang, J.Z. An implementation of hydrophobic force in implicit solvent molecular dynamics simulation for packed proteins. *J. Mol. Model*, **2013**, *19,* 2605-2612.

[176] Ghattyvenkatakrishna, P.K.; Carri, G.A. The effect of complex solvents on the structure and dynamics of protein solutions: The case of Lysozyme in trehalose/water mixtures. *Eur. Phys. J. E. Soft Matter*, **2013**, *36,* 9828.

[177] Li, Z.; Yang, Y.; Zhan, J.; Dai, L.; Zhou, Y. Energy Functions in De Novo Protein Design: Current Challenges and Future Prospects. *Annu. Rev. Biophys.*, **2013**, *42,* 315-335.

[178] Wei, Q.; Xu, Q.; Dunbrack, R.L.; Jr. Prediction of phenotypes of missense mutations in human proteins from biological assemblies. *Proteins*, **2013**, *81,* 199-213.

[179] Ahmad, F.; Bigelow, C.C. Thermodynamics of Solvation of Proteins in Guanidine Hydrochloride. *Biopolymers*, **1990**, *29,* 1593-1598.

[180] Chothia, C. Principles that determine the structure of proteins. *Annu. Rev. Biochem.*, **1984**, *53,* 537-572.

[181] Pradeep, L.; Udgaonkar, J.B. Osmolytes induce structure in an early intermediate on the folding pathway of barstar. *J. Biol. Chem.*, **2004**, *279,* 40303-40313.

[182] Ahmad, S.; Gromiha, M.M.; Sarai, A. Real value prediction of solvent accessibility from amino acid sequence. *Proteins*, **2003**, *50,* 629-635.

Received: August 19, 2013 Revised: March 24, 2014 Accepted: March 25, 2014