Eur Biophys J (1994) 23:139–143

European Biophysics Journal

Determination of the accessible surface of globular proteins by means of tritium planigraphy

A. V. Volynskaya, E. A. Kasumov, E. N. Bogascheva, A. V. Shishkov, V. I. Goldanskii

N. N. Semenov Institute of Chemical Physics, Russian Academy of Science, Kosyginstr. 4, 117334 Moscow, Russia

Received: 8 November 1993 / Accepted: 21 December 1993

Abstract. Results are presented for proteins with known three-dimensional structure (lysozyme, myoglobin, ribonuclease), which show that the probability of label incorporation upon bombardment by "hot" tritium atoms may be quantitatively linked with the surface area of the protein accessible to water molecules. Possible deviations from simple linear dependency caused by particular mechanisms of label introduction are discussed. The data obtained in experiments with model systems were used to determine the accessible surface area of human serum albumin, for which structural data is not sufficiently accurate to allow estimation of accessible surface area. Experimental data correlate reasonably well with estimations based on conventional concepts of the relationship between accessible surface area and molecular weight for globular proteins.

Key words: Accessible surface – Tritium planigraphy

Introduction

The surface are of a globular protein accessible to the solvent is responsible for its main features; namely, the unique spatial architecture, the ability to interact with components of the surrounding medium, biological activity, and structural dynamics.

Until quite recently the only way of obtaining information about the accessible surface was through high resolution X-ray analysis. From these data, using an algorithm first suggested by Lee and Richards (1971), it was possible to calculate the accessibility of atoms constituting a macromolecule. All the other techniques allow only the localization of certain surface groups and, perhaps, the possibility of reacting some conclusions about their immediate surroundings; they give no information about the surface as a whole.

Correspondence to: A. V. Volynskaya

In this paper, we suggest a new approach for experimental measurements of the accessible surface area (ASA) of the protein by means of the tritium planigraphy method (tritium labelling). This technique is based on the ability of hot tritium atoms with energy ≥ 0.3 eV to substitute hydrogen atoms in the hydrocarbon groups (CH^-, CH_2^-, CH_3^-) of amino acid residues without destroying the polypeptide chain of the protein (Goldanskii et al. 1988).

"Hot" tritium atoms are produced by thermal dissociation of molecular tritium on a tungsten wire heated to 2000 K. The substitution of T for H proceeds only during the first collision of a tritium atom with the target (protein molecule), through which a stable radioactive label (CT^-, CT_2^-, CT_3^-) comes to be incorporated into the surface accessible to the direct transit of tritium atoms. The depth of penetration of the tritium atoms into the protein does not exceed 5Å (Volynskaya et al. 1982).

We have reported earlier that for the N-terminal fragment of myoglobin (Goldanskii et al. 1982) and from the lysozyme molecule (Volynskaya et al. 1985) a linear dependency between the value for the specific radioactivity of amino acid residues (A_i) and that for the area of their surface accessible to water molecules (S_i) does exist ($A_i = K \times S_i$, $K = 0.11 \pm 0.03$; linear correlation coefficient $r = 0.986$). This allows us to suggest that the tritium planigraphy method can be successfully applied in the quantitative estimation of the accessible surface of globular proteins.

The possibility that the accessibility to tritium atom is correlated with that for water molecules is illustrated in the present work by data obtained during the study of three globular proteins: hen egg lysozyme, sperm-whale myoglobin and pancreatic ribonuclease. Quantitative parameters obtained in experiments with these model systems are used for the description of the accessible surface area of human serum albumin.

Materials and methods

Sperm-wale myoglobin (Sigma), human serum albumin (Sigma) and hen egg lysozyme (Serva) were used without

additional purification. Bovine pancreatic ribonuclease (Reonal; Leningrad Medicine Factory) was purified by means of gel filtration, using a column with SP-sephadex G-25 in 0.1 M sodium phosphate buffer (pH 6.5). The protein fraction was concentrated by ultrafiltration on a UM-2 membrane (Amicon) and made salt-free by using a 20 × 300 mm column with AI S 01 × 8D.

The equipment for hot tritium bombardment experiments was described by Yusupov and Spirin (1988). The tritium label was introduced according to the procedure described in Volynskaya et al. (1985). Frozen (77 K) 1% aqueous solutions were subjected to bombardment by tritium atoms. The labeled solvent as well as the tritium in the exchangeable groups (OH, SH, NH, NH_2, COOH) of the amino acid residues were removed by repeated lyophilization and gel filtration. Proteins were hydrolysed to free amino acids under standard conditions and were analysed on a Hitachi-835 analyser, along with the detection of the radioactivity of the fractions. The specific radioactivity values of the amino acids were calculated by dividing the measured activity by the amount of the amino acid and then normalizing to the total radioactivity level of the initial preparation. The last value corresponds to the flux of the tritium atoms. In Tables 1 and 3 the average radioactivity values obtained from a set of 5 to 13 parallel experiments are given. The root-mean-square error ranges from 5 to 50%. We did not determine the label incorporation of Trp, Cys and Met residues, whose analysis is only possible by means of special techniques. As hydrolyzed amino acid amides are transformed into the corresponding acids, Tables 1 and 3 show values for (Asn + Asp) and for (Gln + Glu).

Results and discussion

The experimental values of the specific radioactivity of each type of amino acid residue (A_i) and the accessible surface areas (S_i) calculated according to the Lee and Richards method are given for the three model proteins in Tables 1 and 2. The calculations were carried out for water molecules by assuming them to be spheres with a radius of 1.4 Å and for tritium atoms by assuming them to be sphere with a radius of 0.8 Å. For the calculations, data from the data bank of three dimensional protein structures were used (Berstein et al. 1977). Only hydrocarbon groups of amino acid residues were taken into account. For these groups, the tritium atoms are not exchanged with the solvent (CT-bonds) and they are not removed during the purification, hydrolysis and analysis procedure. The Van der Waal's radii values of the atoms composing the protein molecule were taken from Lee and Richards (1971). The S_i values, given in Tables 1 and 2, are the average for each type of amino acid residue: $S_i = \sum_n S_i/n$, where n is the number of i-type residues. The probability of tritium incorporation was defined as the ratio specific of the radioactivity value to the accessible surface area: $K_i = A_i/S_i$.

The analysis of data given in Table 1 shows that the correlation between A_i and S_i revealed earlier is retained for the majority of amino acid residues in all three proteins. This is the case for 30 of the 45 residues (the starred ones), whose K_i coefficients coincide with an accuracy of 40%, this being close to the experimental error. $\bar{K}_i$ (average for 30 residues) = 0.10 ± 0.025; $r = 0.981$. However, as may be seen from Table 1, the results for ribonuclease are in good agreement for only 6 residues, and in the case of myoglobin for only 10 residues. The largest deviations from $\bar{K}_i$ are observed for Val and Leu in myoglobin and for Phe in ribonuclease. This casts some doubt upon the validity of the application of the tritium labelling method for calculating the accessible protein surface. Furthermore, it is difficult to decide what values of the coefficient K_i have to be used for this purpose, whether

Table 1. Specific radioactivity (A_i) accessible surface area for water (S_i) and coefficients $K_i = A_i/S_i$ for lysozyme (1), ribonuclease (2) and myoglobin (3)

Residue	A_i			S_i			K_i		
	1	2	3	1	2	3	1	2	3
D+N	1.3	0.6	1.8	12.3	12.2	12.9	0.10*	0.05*	0.14*
T	2.5	2.4	4.4	25.8	26.1	39.9	0.10*	0.09*	0.11*
S	1.4	1.0	2.9	13.7	25.2	18.7	0.10*	0.04	0.15*
E+Q	2.0	0.6	2.9	9.9	15.2	23.2	0.20	0.04	0.12*
P	8.7	5.7	10.5	79.1	56.3	58.5	0.11*	0.10*	0.18
G	1.5	2.3	1.4	16.0	27.2	17.5	0.09*	0.08*	0.08*
A	1.5	2.8	4.4	12.3	26.6	27.7	0.12*	0.10*	0.16*
V	3.1	2.1	14.1	27.3	18.3	29.9	0.11*	0.11*	0.47
I	1.5	2.1	3.6	15.6	6.3	15.2	0.10*	0.33*	0.23
L	1.3	5.3	7.2	16.2	24.4	14.7	0.08*	0.22	0.49
Y	1.8	1.2	2.8	24.9	36.2	20.6	0.07*	0.03	0.13*
F	2.1	2.3	1.8	23.5	4.0	19.1	0.09*	0.57	0.09*
K	4.2	2.2	3.1	35.2	58.6	46.0	0.12*	0.04	0.07*
H	4.1	5.1	3.6	25.7	17.8	35.0	0.16*	0.29	0.10*
R	3.6	8.9	6.6	37.4	36.1	22.9	0.10*	0.25	0.28

* The asterisks mark the 30 K_i values which agree well. The average value is 0.10 ± 0.025 (see text)

Table 2. Accessible surface area for tritium atoms (S_i), coefficients K_i and average values $\bar{K}_i$ for tritium ($\bar{K}_i^T$) and water ($\bar{K}_i^W$). Lysozyme (1), ribonuclease (2), myoglobin (3). Radius of tritium atom 0.8 Å

Residue	S_i			K_i			$\bar{K}_i^T \pm \sigma$	$\bar{K}_i^W \pm \sigma$
	1	2	3	1	2	3		
D+N	16.8	10.8	17.4	0.08	0.08	0.10	0.09 ± 0.01	0.10 ± 0.04
T	29.5	26.0	38.7	0.08	0.09	0.11	0.09 ± 0.01	0.10 ± 0.01
S	14.8	22.9	19.5	0.09	0.04	0.15	0.09 ± 0.05	0.10 ± 0.05
E+Q	20.8	19.6	18.5	0.10	0.05	0.16	0.10 ± 0.05	0.12 ± 0.08
P	74.0	52.2	58.2	0.12	0.11	0.18	0.14 ± 0.04	0.13 ± 0.04
G	16.3	23.6	16.1	0.09	0.10	0.09	0.09 ± 0.01	0.08 ± 0.01
A	17.6	29.2	29.8	0.09	0.10	0.15	0.11 ± 0.03	0.13 ± 0.03
V	36.3	34.2	41.6	0.08	0.06	0.34	0.16 ± 0.15	0.23 ± 0.20
I	37.4	22.4	40.7	0.04	0.09	0.09	0.07 ± 0.03	0.22 ± 0.11
L	35.3	33.4	40.3	0.04	0.16	0.18	0.13 ± 0.07	0.26 ± 0.21
Y	29.5	40.6	33.8	0.06	0.03	0.08	0.06 ± 0.02	0.08 ± 0.05
F	41.5	21.9	41.1	0.05	0.10	0.04	0.06 ± 0.03	0.25 ± 0.28
K	45.3	59.4	53.5	0.09	0.04	0.06	0.06 ± 0.02	0.08 ± 0.04
H	38.0	20.4	41.3	0.11	0.25	0.09	0.15 ± 0.09	0.18 ± 0.09
R	44.1	38.5	34.0	0.08	0.23	0.19	0.17 ± 0.07	0.21 ± 0.09

Note: $\bar{K}_i^W$ is calculated from the K_i values of Table 1

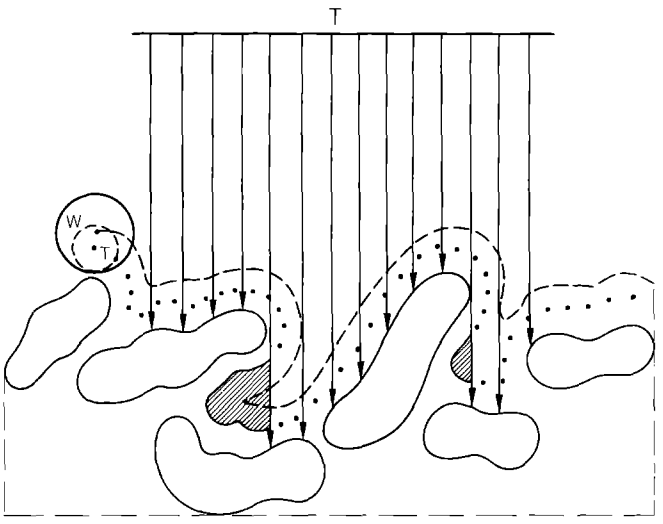

Fig. 1. Schematic representation of the accessible surface of a typical globular protein. The hatched regions indicate surface areas which, on the average, are experimentally less accessible to tritium incorporation than to the computed access of a water or tritium atom surface probe

the most probable $\bar{K}_i = 0.10$, or K_i values specific for each type of residue. To answer this question, some possible reasons for the large scattering of the K_i values have to be considered.

We are of the opinion that there are at least three reasons fore the above anomalies. The first one is connected with the difference in the radii of the "probes" from which the accessible surface area is calculated. The tritium atom radius is nearly half that of the water molecule. One would therefore expect that the accessibility of residues, especially those located in the interior of the protein, would be higher for tritium than for a water molecule. This is supported by data in Table 2 where surface areas accessible to tritium, together with the corresponding K_i coefficients, are presented for different amino acid residues. It can be seen that the accessible Val, Ile, Leu and Phe surfaces for all three proteins (with the exeption of the Val in lysozyme) are larger for the tritium atom than for the water molecule. For this reason, the dispersion of K_i magnitudes is smaller than in Table 1.

The second reason is connected with the different methods for ASA determination. In calculations of ASA values, according to the Lee and Richards method, the procedure of rolling out the surface by probes with different radii is applied. The tritium planigraphy method allows one to determine only that region of the surface which is accessible to the direct flight-path of the tritium atoms. In consequence, this surface may appear to be smaller than that found by rolling out. This is illustrated in Fig. 1, where the region of a protein inaccessible to the direct flight-path, is shown by hatching. It can be clearly seen that the residue surface accessible for rolling out by a water molecule, or by tritium, is substantially higher in this region than the surface which could incorporate the tritium label. This leads to the underestimated K_i values. Though the relative proportion of these residues is small (6–8 out of 45), their presence, nevertheless reduces the correlation between ASA and radioactivity.

Finally, the third reason for the scatter in K_i values may be associated with the influence of the chemical structure of the amino acid residues on their ability to interact with the tritium atoms. Experiments with free amino acids and short peptides (Baratova et al. 1982) have shown that tritium label incorporates into most residues with much the same probability. Exceptions are Val, Leu, and Ile, whose structures include a tertiary carbon atom. The energy of hydrogen separation from the latter is 1.5 times smaller than that for separation from a primary carbon atom (Berdle and Le Roy 1952) resulting in higher radioactivity values. In proteins, however, a tertiary carbon atom is generally inaccessible. Table 2 shows that this factor is most pronounced for the Val in myoglobin.

Thus, the analysis presented above shows that a good correlation between the radioactivity of residues and their surface accessible to water molecules is not, in principle, possible. For the tritium atoms, it can be seen (Table 2) that the contribution of factors which weaken the correlation is not too large. For 38 out of 45 residues, the K_i values coincide with an accuracy not exceeding the experimental error ($\bar{K}_i = 0.09 \pm 0.03$; $r = 0.92$) and for more than half of them (25) the correlation is very good ($\bar{K}_i = 0.1 \pm 0.01$; $r = 0.99$).

A comparison of accessibilities of residues for tritium atoms and water molecules reveals that the major differences are observed, as would be expected, for those of them found in the interior of a protein (Val, Leu, Ile, Phe). For the other residues $\bar{K}_i^T$ and $\bar{K}_i^W$ coefficients are very nearly equal. As can be seen from Table 2, it is these interior residues which will introduce the largest error into the estimation of the accessible surface area of proteins. It is also evident that for the estimation of the ASA the individual $\bar{K}_i$ value for each type of residue have to be used. The mean $\bar{K}_i^T$ value for all 45 residues is equal to 0.1 ± 0.06 and this is coincident with that of the most probable $\bar{K}_i^W$. It is clear that using the most probable $\bar{K}_i^W$ quantity for ASA calculations will give results which is appear to be less accurate than they in fact are.

Thus, using individual $\bar{K}_i^W$ values for each type of residue, we have estimated the ASA of the amino acid residues of human serum albumin and its total accessible surface area. It should be noted that this study was completed in 1985 (Volynskaya et al. 1985), when data on the X-ray analysis of this protein were still not available. The physical and chemical properties of serum albumin had been extensively studied, and models of its spatial structure had been proposed (Peters 1985; Brown 1978). These are the reasons for us choosing this protein to study. In 1989, Carter and colleagues published their X-ray investigation of the three-dimensional structure of SA (Carter et al. 1989). Unfortunately the 6 Å resolution is not sufficient for ASA calculations.

Table 3 represents some data on specific radioactivities, accessible surface areas of CH-groups of residues (calculated using the $\bar{K}_i^W$ value for each particular residue), and total surface areas of residues including all constituent atoms. Total ASA values have been calculated using the following approximation. According to Lee and Richards (1971), for each type of residue in lysozyme,

Table 3. Accessible surface areas of human serum albumin. A_i – specific radioactivity; S_i^W, S_i^T, S_{tot}^W, S_{tot}^T – hydrophobic part and total residue accessible surface areas for water and tritium atoms; γ_i – proportionality coefficient between S_i and S_{tot}. n – number of residues in interdomain connections (see text)

Residue	A_i	$S_i^W = A_i/\bar{K}_i^W$	$S_i^T = A_i/\bar{K}_i^T$	γ_i	S_{tot}^W	S_{tot}^T	n (%)
D+N	0.3	3.0	3.3	7.8	23.4	25.7	34
T	2.6	26.0	28.9	2.4	62.4	69.4	11
S	0.7	7.0	7.8	2.9	20.3	22.6	21
E+Q	0.7	5.8	7.0	6.6	38.3	46.2	12
P	4.0	30.8	28.6	1.1	33.9	31.5	37
G	1.1	13.7	12.2	2.2	30.1	26.8	8
A	4.4	33.8	39.6	1.4	47.3	55.4	18
V	5.6	24.3	35.0	1.3	31.6	45.5	24
I	11.0	50.0	157.1	1.1	55.0	172.8	37
L	10.0	38.5	76.9	1.3	50.0	99.9	21
Y	2.8	35.0	46.7	2.0	70.0	93.4	5
F	4.3	17.2	71.7	1.5	25.8	107.5	26
K	2.1	26.2	35.0	2.1	55.0	73.5	17
H	4.8	26.7	32.0	1.9	50.7	60.8	12
R	4.2	20.0	24.7	3.6	72.0	88.9	16

ribonuclease, myoglobin, and in model peptides with the sequence Ala-X-Ala, the relationship between the accessibility of the whole residue and that of its hydrocarbon part is much the same: $\gamma_i = S_{tot}/S_{CH}$. The γ_i coefficients given in Table 3 have been used for the calculation of the total accessible surface of serum albumin amino acid residues. Knowing the amino acid composition of serum albumin, it is easy to determine its ASA value. This value is equal to $\sim 24 \times 10^3$ Å^2 and $\sim 34 \times 10^3$ Å^2 for water molecules and trutium atoms, respectively. According to emperical estimations (Chothia 1975; Janin 1979; Sprang et al. 1979) the ASA value of a typical globular protein with a molecular mass 66.3 kDa should be equal to $(23-26) \times 10^3$ Å^2. Furthermore, the average ASA value for a residue, from our data, is equal to ~ 44 Å^2, which is very similar to the value (~ 48 Å^2) obtained for 46 globular proteins (Miller et al. 1987). Therefore we have shown that our results are rather realistic.

An analysis of data presented in Table 3 allows us to draw some additional conclusions about the details of the spatial structure of serum albumin. Considerable attention must be given to the fact that the surface of some serum albumin residues (Asx, Ser, Glx, Pro), accessible to tritium atoms, is far smaller than for these residues in the "reference" proteins, whereas this value is substantially higher for Ile, Leu and Phe (see Table 2).

As we have demonstrated above, the most significant discrepancy between the accessible surfaces determined by the Lee-Richards method and those found by tritium planigraphy is likely to be found for residues positioned in regions with a rough surface profile (Fig. 1). It is reasonable to suppose that serum albumin has a more uneven surface than the "reference" proteins, as shown by the presence of a large number of "anomalous", in terms accessible surface, residues. Furthermore, the higher accessibility of hydrophobic residues may point to the existence of some regions within the protein with a loose intramolecular packing (perhaps cavities and channels). Inasmuch as serum albumin consists of three domains, each consisting of two subdomains (Carter et al. 1989), these loosely packed regions are most likely to be the interfaces of domains and subdomains.

Carter and coworkers (1989) managed to identify regions of the serum albumin polypeptide chain which link together domains and subdomains. There are five such "binders", which involve 116 amino acid residues, i.e. 20% of the total number. The analysis of the amino acid composition of these "binders" reveals that "anomalous", according to our data, residues really have a tendency to be located at interfaces of domains and subdomains (see the last column in Table 3). The predominantly hydrophobic character of the interfaces makes them the potential sites for the binding of various hydrophobic ligands. Furthermore, the high number of Asp residues at interfaces, which are able to form salt bridges, provides a logical explanation for the observed reversible stretching-compression transition at pH ≤ 3.5 (N $\leftrightarrow$ F transition). According to our own data, the accessible surface of amino acid residues increases two- to three-fold owing to acid expansion. The accessibility of Asx residues increases upon acidification by a factor of five. These observations argue in favour of the N $\leftrightarrow$ F transition mechanism involving the separation of the protein's subunits.

Regardless of all the assumptions and reservations, in the case of serum albumin the method of tritium labelling evidently works. Regarding the possibility of quantitative determination of the accessible surface value, we believe that at the this stage we are only able to make approximate evaluations. To get more reliable results, it is necessary to extend the range of reference proteins with known structure and to increase the experimental accuracy.

Acknowledgements. The authors would like to thank Dr. L. A. Baratova for making the amino acid analysis and N. M. Stirikovich for help in translating the text into English.

References

Baratova LA, Goldanskii VI, Rumjantsev YuM, Unukovich MS, Shishkov AV (1982) Study of the three-dimensional structure of proteins by means of tritium labelling. 1. Free amino acids as a model of residues in an unfolded polypeptide chain. Mol Biol 16:117–122 (in Russian)

Berdle MR, Le Roy DJ (1952) The reaction of atomic hydrogen with ethane. J Chem Phys 20:200–201

Bernstein FC, Koetzle TF, Williams GYB, Meyer EF, Brice MD, Rodgers YR, Kennard O, Tasumi M, Shimanouchi T (1977) The protein data bank: a computer-based archival file. J Mol Biol 112:535–542

Brown JR (1978) Structure and evolution of serum albumin. In: Albumin structure, biosynthesis, function. Oxford, Pergamon Press, pp 1–10

Carter DC, Hiao-min H, Munson SH, Twigg FD, Gernert KM, Broom MB, Miller TY (1989) Three-dimensional structure of human serum albumin. Science 244:1195–1198

Chothia C (1975) Structural invariants in protein folding. Nature 254:304–308

Goldanskii VI, Rumyantsev YuM, Shishkov AV, Baratova LA, Beljakova LP (1982) Study of three-dimensional structure of proteins by means of tritium labelling. 2. Intramolecular distribution of tritium in the N-terminal part of myoglobin and the tertiary structure of protein. Mol Biol 16:528–534 (in Russian)

Goldanskii VI, Kashirin IA, Shishkov AV, Baratova LA, Grebenshcikov NI (1988) The use of thermally activated tritium atoms for structural-biological investigations: topography of the TVM protein accessible surface of the virus. J Mol Biol 201:567–574

Janin J (1979) Surface and inside volumes in globular proteins. Nature 277:491–492

Lee B, Richards FM (1971) The interpretation of protein structures: Estimation of static accessibility. J Mol Biol 55:379–400

Miller S, Janin J, Lesk AM, Cothia C (1987) Interior and surface of monomeric proteins. J Mol Biol 196:641–656

Peters T (1985) Serum albumin. Adv Prot Chem 37:161–245

Sprang S, Yang D, Fletteric RJ (1979) Solvent accessibility properties of complex proteins. Nature 280:333–335

Volynskaya AV, Scripkin AYu, Shishkov AV, Goldanskii VI (1982) The research of the structure of surface active substances adsorption layers by means tritium labelling. Dokl Acad Nauk USSR 266:871–874 (in Russian)

Volynskaya AV, Scripkin AYu, Jafarov ES, Rumyantsev YuM, Shishkov AV, Goldanskii VI (1985) Determination of the accessible surface of lysozyme and serum albumin by the total tritium labelling method. Mol Biol 19:1294–2300 (in Russian)

Yusupov MM, Spirin AS (1988) Hot tritium bombardment technique for ribosoms surface topography. Methods Enzymol 164:426–439