Vol. 81, No. 2, 1978 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
March 30, 1978 Pages 410-415

Solvent Accessibility Calculations for Sperm Whale Ferrimyoglobin based on Refined Crystallographic Data

James B. Matthew, George I. H. Hanania and Frank R. N. Gurd

Department of Chemistry, Indiana University, Bloomington, IN 47401

Received November 16,1977

SUMMARY: The calculation of solvent accessibility parameters from protein crystallographic data, by the method of Lee and Richards has been simplified to allow treatment of single atoms and their immediate environment. New accessibility values for all titratable amino acid residues in myoglobin have been computed from recent X-ray structure data of Takano. A number of prominent differences appear between these values and those from the older structure data of Watson. Differences are interpreted in terms of the proximity of neighboring residues.

INTRODUCTION: The structure of a protein prescribes its surface topology, and its reactivity depends on the extent of exposure of particular sites to solvent and to other reacting species in solution. Lee and Richards (1) first showed how X-ray crystallographic data can be effectively used to calculate and interpret the van der Waals surface properties of three proteins, ribonuclease-S, lysozyme and myoglobin. Lee and Richards introduced the concept of static solvent accessibility, and they defined the term "accessible surface area" as the area around an atom over which the center of a solvent (water) molecule can be placed, enabling it to maintain van der Waals contact with that atom without penetrating any other atom in the protein. They calculated accessibilities for all the protein atoms other than hydrogen, and they made the striking point that about half the surface area of a globular protein is occupied by non-polar atoms in the folded as well as the unfolded state.

The above approach was applied by Shrake and Rupley (2) to the computation of spatial environment and exposure to solvent of the atoms in lysozyme and insulin, as well as to changes in exposure and to the nature of contacts

0006-291X/78/0812-0410$01.00/0

that develop through folding of these two proteins. The idea was further developed by Chothia (3-6) who recently showed that a large and constant proportion (75%) of the polar surface is buried in the native structure of six monomeric proteins.

The concept of a static and accessible surface area has also been demonstrated to be a valuable parameter applicable to the investigation of electrostatic effects in myoglobin. Shire et al. (7-9) introduced the solvent accessibility of individual groups in myoglobin for the calculation, based on the Tanford-Kirkwood discrete charge model, of electrostatic free energies to estimate intramolecular as well as ionic strength effects on pK values for individual ionizing groups to obtain their effective charges, and consequently the net protein charge as a function of pH, the titration curve. Original calculations were based on the early X-ray crystallographic coordinates of Watson (10) and on Lee and Richards' accessibilities (1) derived from the same atomic coordinates. Takano (11) has determined a new set of refined crystallographic coordinates at 2.0Å resolution (11) which differs from the original ones in several respects, making it necessary to recalculate accessibilities and also electrostatic interactions in myoglobin.

In this report, we present a simple and efficient computer method for the solvent accessibility calculations and give detailed results for the titratable groups in sperm whale ferrimyoglobin. Some prominent differences from the previously published values are commented on.

PROCEDURE: The method of Lee and Richards (1) treats the protein structure as a set of interlocking spheres of appropriate van der Waals radii. This continuous structure is sliced by a series of parallel planes with predetermined spacing at 0.25Å. The intersection of each sphere with a given plane appears as a circle, and the outermost trace of these circles becomes the envelope of a van der Waals surface of the protein section in that plane. For calculation, a sphere is centered at each atomic site in the coordinate space and is assigned a radius equal to the sum of radii for the atom and the solvent (1.40Å for water). The surface section is the locus of the center of a solvent molecule as it rolls along the protein molecule, and any part of the atom's surface that lies on the envelope is regarded as accessible. The atom's solvent accessibility is thus proportional to the length of its exposed arcs, summed over each of the sections which pass through the atom.

The present computational method follows essentially that of Lee and Richards except that an initial scan of distances is first made to eliminate

all atoms beyond 6.40Å from the selected atom, this being the largest sum of two radii each including water. Radii were taken as follows: 3.10Å for main chain alpha carbon atoms, 2.92Å for main chain carbonyl oxygens, 2.95Å for main chain amide nitrogens, and 2.30Å for all other atoms (1, 12). At a 0.25Å spacing, 25 sections can be cut parallel to the X-Y plane through the principal atom. Within each section, atoms intersected by the plane give arcs of various radii. For the principal atom, the length of the arc in that plane is obtained by determining the fraction of the circumference that does not lie within the effective radius of a neighboring atom. Following Lee and Richards, the accessible surface area (A) is obtained by summation over the 25 slices:

$$A = \sum_{i} (0.25(R)(L_i) / (R^2 - Z_i^2)^{\frac{1}{2}}) \qquad (1)$$

where R is the effective radius, L_i is the length of an arc in a given section i, and Z_i is the perpendicular distance from the center of the atomic sphere to the planar section i, all values being in Å. The accessibility of the atom is then defined as the accessible surface area in $Å^2$, normalized and expressed as percentage:

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

A clearer appraisal of the influence of protein structure on the solvent accessibility of individual atoms or groups is obtained by normalization to the corresponding accessibility in a model tripeptide where, it may be assumed, accessibilities attain their maximum values. The "fractional static accessibility" (SA) is expressed as:

$$SA = (A \text{ in protein})/(A \text{ in tripeptide}) \qquad (3)$$

Accessible surface areas have been calculated for the two beta trans configurations of Ala-X-Ala peptides, where X is an amino acid residue (1). We have observed that accessibility values intermediate between those for the two peptide structures are obtained if one considers the tripeptide Ala-X-Ala coordinates as taken from helical regions.

In the present method of computation, run on a CDC-6600 computer system, the determination of accessibilities took about 5 seconds per atom, the only required input data being the atomic coordinates. The structure data of Takano (11) for sperm whale ferrimyoglobin were obtained from the Brookhaven National Laboratory. A computer program with full details is available on request.

RESULTS AND DISCUSSION: To test the validity of this simplified computational method, a comparison was first made between our results and those of Lee and Richards (1) using the same original set of atomic coordinates for sperm whale ferrimyoglobin (10). The test was applied to the titratable groups as well as several other types of exposed or buried residues of interest. All results were concordant within the limits of uncertainty stated by Lee and Richards.

Accessibility values were then calculated from the refined data of

Table I

Comparison of Accessibility Differences for Some Residues in Sperm Whale Myoglobin as Derived from Watson (10) and Takano (11) X-ray Coordinates

Residue	Helix	Atom	Ref.	X-Ray Coordinates X	Y	Z	Number of Atoms at Distance of ≤ 6.4Å	≤ 3.2Å	(Equation 2) Accessibility	Change in Accessibility (T-W)
1 Val	1NA	N	W	-2.90	17.60	15.50	24	4	24.8	+13.8
			T	-4.42	16.67	16.94	15	4	38.6	
12 His	10A	NE2	W	15.00	9.20	24.30	22	6	8.4	+ 7.4
			T	14.40	9.23	24.95	19	4	15.8	
16 Lys	14A	NZ	W	17.00	8.60	23.30	27	3	9.2	- 8.6
			T	17.25	8.70	22.18	38	2	0.6	
45 Arg	3CD	NH1	W	24.10	31.80	6.30	41	4	12.2	- 3.6
			T	24.28	31.80	6.03	44	4	8.6	
145 Lys	22H	NZ	W	-1.70	34.30	5.40	28	7	1.7	+26.0
			T	-2.17	28.01	7.78	31	3	27.7	
148 Glu	25H	OE1	W	-3.20	35.10	4.20	25	4	26.4	+15.1
			T	-3.67	35.05	3.91	15	3	41.5	

Takano (11). On the whole, the results were similar to those derived from Watson's data (10), but the magnitudes were significantly different for many residues, as illustrated with some examples shown in Table I. Changes in atomic coordinates from Watson (W) to Takano (T) are seen to be variable, and in some cases quite large. As a consequence, the numbers of neighboring atoms within 6.40Å of the principal atom differ considerably, and also those within a single radius of 3.20Å. Accessibility differences, listed in the last column of Table I, can therefore be traced to the repositioning of the principal atom or of its neighboring atoms.

The N-terminal atom of 1-Val is 14% more accessible to solvent as judged by the new data, the amino group being angled further into solvent, thus removing 137-Leu from its sphere of influence. For 12-His, the atom NE2, which holds the dissociable proton, was only 2.3Å away from the positively charged NZ atom of 16-Lys according to the old coordinates, and the calculated electrostatic effect was inconsistent with the experimentally determined acid dissociation pK value for 12-His as derived from proton NMR measurements (13). The new coordinates place the NZ atom more than 4Å away from 12-His, allowing a simple interpretation of observed pK data. Although 16-Lys is

Table II

Accessibilities for Some Selected Groups of Sperm Whale Ferrimyoglobin

Residue	Helix	Atom	Accessibility (Equation 2)	Residue	Helix	Atom	Accessibility (Equation 2)
1 Val	1NA	N	38.6	78 Lys	1EF	NZ	23.9
3 Ser	1A	OG	13.2	79 Lys	2EF	NZ	29.5
4 Glu	2A	OE1	38.4	81 His	4EF	NE2	30.9
6 Glu	4A	OE1	26.3	83 Glu	6EF	OE1	31.2
12 His	10A	NE2	15.8	85 Glu	8EF	OE2	12.2
16 Lys	14A	NZ	0.6	87 Lys	2F	NZ	53.2
18 Glu	16A	OE	25.7	96 Lys	1FG	NZ	54.7
20 Asp	1B	OD1	24.9	97 His	2FG	ND1	9.1
27 Asp	8B	OD2	13.0	98 Lys	3FG	NZ	45.0
31 Arg	12B	NH1	30.7	102 Lys	3G	NZ	34.1
34 Lys	15B	NZ	48.8	103 Tyr	4G	OH	21.3
35 Ser	16B	OG	20.6	105 Glu	6G	OE1	32.5
36 His	1C	NE2	9.9	109 Glu	10G	OE1	35.5
38 Glu	3C	OE1	21.8	113 His	14G	NE2	15.5
41 Glu	6C	OE2	22.5	116 His	17G	NE2	28.7
42 Lys	7C	NZ	13.6	118 Arg	19G	NH2	13.9
44 Asp	2CD	OD1	21.8	119 His	1GH	ND1	5.5
45 Arg	3CD	NH1	8.6	122 Asp	4GH	OD1	29.4
47 Lys	5CD	NZ	30.5	126 Asp	3H	OD1	35.3
48 His	6CD	NE2	16.8	133 Lys	10H	NZ	24.0
50 Lys	8CD	NZ	33.4	136 Glu	13H	OE1	26.4
52 Glu	2D	OE2	30.4	139 Arg	16H	NH1	3.1
54 Glu	4D	OE1	38.3	140 Lys	17H	NZ	31.4
56 Lys	6D	NZ	19.3	141 Asp	18H	OD1	9.7
59 Glu	2E	OE1	36.7	145 Lys	22H	NZ	27.7
60 Asp	3E	OD2	7.7	146 Tyr	23H	OH	0.0
62 Lys	5E	NZ	44.3	147 Lys	24H	NZ	52.7
63 Lys	6E	NZ	52.3	148 Glu	25H	OE1	41.5
64 His	7E	ND1	1.8	151 Tyr	2HC	OH	33.0
66 Val	9E	CG1	26.8	153 Gly	4HC	OXT	8.3
77 Lys	20E	NZ	30.3	200 Hem		FE	0.0
				200 Hem		01A	14.7
				200 Hem		01D	21.0

now further away from 12-His, its NZ atom has reduced accessibility because of its closer proximity to 119-His and to 122-Asp in the GH helical turn. 45-Arg shows only slight alteration. 145-Lys and 148-Glu may be considered together. Watson's coordinates had these latter two groups held in an intra-helical closed salt bridge. Takano's coordinates now place 145-Lys (NZ) beyond 6Å away from 148-Glu (OE1 and OE2 atoms) thus opening the salt bridge and allowing greater solvent accessibility for both residues. Several other important groups undergo similar spatial reorientation by the new data. It is clear that previous computations of surface areas and accessibilities in myoglobin need to be reconsidered in the light of the newly available data.

Table II lists solvent accessibilities (Equation 2) for all titratable

groups in sperm whale ferrimyoglobin. As the Table indicates, about half of the accessibility values differ by more than 20% from the previously published data, and in some cases the difference exceeds 50%. The use of accessibility parameters in electrostatic calculations is demonstrated in the following report (14).

Acknowledgments: We wish to thank T. M. Rothgeb, and W. D. Matthew for helpful discussions and D. Embry for typing of these manuscripts. This work was supported by United States Public Health Service Research Grant HL-05556 and Biochemistry Training Grant T01-GM-1046 (J.B.M). It is the 90th paper in a series dealing with coordination complexes and catalytic properties of proteins and related substances.

References

1. Lee, B. and Richards, F. M. (1971) J. Mol. Biol. 55, 379-400.
2. Shrake, A. and Rupley, J. A. (1973) J. Mol. Biol. 79, 351-371.
3. Chothia, C. (1974) Nature 248, 338-339.
4. Chothia, C. (1974) Nature 254, 304-308.
5. Chothia, C. (1976) J. Mol. Biol. 105, 1-14.
6. Chothia, C., Wodak, S. and Janin, J. (1976) Proc. Natl. Acad. Sci. USA, 73, 3793-3797.
7. Shire, S. J., Hanania, G. I. H. and Gurd, F. R. N. (1974) Biochemistry 13, 2967-2974.
8. Shire, S. J., Hanania, G. I. H. and Gurd, F. R. N. (1974) Biochemistry 13, 2974-2979.
9. Shire, S. J., Hanania, G. I. H. and Gurd, F. R. N. (1975) Biochemistry 14, 1352-1357.
10. Watson, H. C. (1969) Prog. Stereochemistry 4, 299-333.
11. Takano, T. (1977) J. Mol. Biol. 110, 537-568.
12. Bondi, A. (1964) J. Phys. Chem. 68, 441-451.
13. Botelho, L. H. (1975) Ph.D. Thesis, Indiana University.
14. Matthew, J. B., Friend, S. H., Botelho, L. H., Lehman, L. D., Hanania, G. I. H., and Gurd, F. R. N. (1978) Biochem. Biophys. Res. Commun., immediately following paper.