J. Mol. Biol. (1973) **79**, 351–371

Environment and Exposure to Solvent of Protein Atoms. Lysozyme and Insulin

A. Shrake and J. A. Rupley†

Department of Chemistry
The University of Arizona
Tucson, Arizona 85721, U.S.A.

(*Received 12 March 1973, and in revised form 6 June 1973*)

A computer program is described for calculating the environment and the exposure to solvent of atoms of a protein. The computation is based on the atomic co-ordinates of the protein and on assumptions like those of Lee & Richards (1971). Results for lysozyme and insulin are presented. Changes in exposure to solvent and in the nature of contacts that develop through folding, association reactions and crystallization are described numerically. The computations suggest several generalizations. (a) Lattice contacts within the protein crystal are characterized by a significantly smaller involvement of non-polar side chains and a proportionately greater involvement of ionizable side chains than is found for protein folding or for protein association reactions important for biological function. (b) In helical regions the carbonyl oxygen of the first residue in the helix has high probability of being shielded from solvent. (c) Glycine is among the residues having exposure least affected by folding; this accords with the expectation that it lies at bends of the peptide chain on the surface of the molecule.

1. Introduction

Atomic co-ordinates derived from high resolution crystallographic analyses are available for more than 30 proteins (Dickerson, 1972). Some method of describing these structures in a way that allows simple and objective comparisons among them seems necessary. Particular importance is attached to descriptions of the molecular surface and the environments of reactive groups because these features should most closely relate to chemical properties. Lee & Richards (1971) have made an effective approach to this problem through developing a computer program for calculating the exposure of protein atoms to solvent. The present report extends this method by focusing attention on the nature of contacts between atoms. Results are given for native lysozyme and insulin and for changes in their surfaces that occur during folding and several association reactions, including crystallization.

2. Methods

Like Lee & Richards (1971), we describe the protein by a set of solvated van der Waals' spheres. The surface of a sphere is represented by a set of 92 test points that are nearly uniformly distributed. Each atom of the protein is considered separately as a *central atom* that is checked for overlap with all other atoms of the molecule—the *test atoms*. The latter

† To whom inquiries concerning this paper should be addressed.

TABLE 1. *Average areas exposed to solvent in Gly-X-Gly models for the unfolded state*

ALA (14)...		
N	2	1
CA	19	3
C	0	0
O	28	3
CB	75	2
BB	49	4
SC	75	2
ARG (13)...		
N	2	2
CA	13	3
C	0	0
O	27	3
CB	27	4
CG	26	7
CD	36	4
NE	11	2
CZ	4	2
NH1	46	6
NH2	52	5
BB	42	3
SC	202	7
ASN (18)...		
N	3	2
CA	9	4
C	0	0
O	28	3
CB	38	5
CG	1	1
NOD1	37	5
NOD2	45	6
BB	40	5
SC	121	6
ASP (7)...		
N	2	2
CA	12	4
C	0	0
O	25	5
CB	41	3
CG	1	1
OD1	32	5
OD2	41	5
BB	39	6
SC	115	5
CYS (20)...		
N	2	2
CA	5	5
C	0	0
O	25	4
CB	30	7
SG	32	5
BB	32	5
SC	62	10
GLU (10)...		
N	1	2
CA	13	5
C	1	1
O	23	2
CB	29	2
CG	38	6
CD	3	2
OE1	37	5
OE2	42	3
BB	38	5
SC	149	5
GLN (9)...		
N	2	2
CA	12	5
C	1	1
O	26	3
CB	25	5
CG	35	3
CD	2	1
NOE1	40	4
NOE2	49	3
BB	40	7
SC	150	5
GLY (18)...		
N	5	2
CA	53	5
C	1	1
O	30	6
BB	89	6
HIS (5)...		
N	2	2
CA	10	3
C	1	1
O	26	2
CB	33	2
CG	4	1
ND1	13	2
CE1	60	3
NE2	15	1
CD2	37	10
BB	39	2
SC	162	9
ILE (10)...		
N	1	2
CA	9	4
C	0	1
O	23	4
CB	6	3
CG1	27	3
CD1	74	3
CG2	53	4
BB	34	5
SC	160	4
LEU (19)...		
N	2	1
CA	8	4
C	0	1
O	23	5
CB	23	4
CG	10	4
CD1	65	7
CD2	66	7
BB	34	6
SC	164	5
LYS (7)...		
N	2	2
CA	11	2
C	0	1
O	26	3
CB	26	6
CG	25	5
CD	32	6
CE	51	3
NZ	41	5
BB	40	5
SC	174	6
MET (2)...		
N	1	0
CA	13	2
C	0	0
O	28	2
CB	20	1
CG	32	4
SD	43	3
CE	79	2
BB	42	4
SC	173	3
PHE (7)...		
N	2	2
CA	8	5
C	0	0
O	26	6
CB	29	5
CG	5	1
CD1	18	4
CE1	38	3
CZ	40	2
CE2	36	3
CD2	19	8
BB	37	5
SC	184	14
PRO (4)...		
N	0	0
CA	14	2
C	1	1
O	22	5
CB	37	5
CG	45	2
CD	30	3
BB	37	6
SC	113	5
SER (16)...		
N	2	2
CA	15	4
C	1	1
O	26	4
CB	51	6
OG	31	4
BB	44	6
SC	82	5
THR (11)...		
N	2	3
CA	10	4
C	0	1
O	25	4
CB	19	3
OG1	28	2
CG2	67	6
BB	38	5
SC	114	8
TRP (6)...		
N	1	1
CA	6	5
C	1	1
O	28	5
CB	30	6
CG	2	1
CD1	37	4
NE	17	1
CE1	9	4
CZ1	37	1
CH	37	1
CZ2	36	5
CE2	19	4
CD2	4	1
BB	36	6
SC	229	9
TYR (11)...		
N	2	2
CA	7	4
C	0	1
O	25	4
CB	34	4
CG	2	1
CD1	24	4
CE1	35	2
CZ	11	1
OH	40	1
CE2	35	3
CD2	22	6
BB	34	4
SC	202	10
VAL (14)...		
N	2	2
CA	10	4
C	0	0
O	24	4
CB	10	2
CG1	63	7
CG2	60	6
BB	36	8
SC	133	10
TERMINAL RESIDUES		
ALA (2)...		
N	2	1
CA	11	0
C	1	0
OEND	39	3
OEND'	41	3
CB	75	1
BB	94	1
SC	75	1
ASN (2)...		
N	2	0
CA	11	3
C	2	1
OEND	35	3
OEND'	42	4
CB	39	5
CG	2	1
NOD1	41	1
NOD2	38	2
BB	93	7
SC	120	2
GLY (2)...		
NEND	44	1
CA	64	4
C	2	1
O	32	1
BB	142	6
LEU (1)...		
N	2	
CA	13	
C	5	
OEND	41	
OEND'	39	
CB	25	
CG	3	
CD1	52	
CD2	71	
BB	99	
SC	152	
LYS (1)...		
NEND	37	
CA	24	
C	0	
O	33	
CB	32	
CG	13	
CD	33	
CE	51	
NZ	44	
BB	94	
SC	172	
PHE (2)...		
NEND	35	2
CA	13	2
C	1	1
O	32	1
CB	30	0
CG	4	1
CD1	27	4
CE1	39	1
CZ	38	1
CE2	38	1
CD2	25	2
BB	80	5
SC	202	0

The numbers of residues of each type included in the averaging (a total of 127 non-terminal residues in lysozyme and 94 in insulin dimer) are given in parentheses following the residue type. The atom designations are those used in Imoto *et al.* (1972). BB and SC stand for the sums over all backbone and over all side chain atoms of the residue, respectively. The first number column gives the average area (Å^2) exposed to solvent in the peptide model; the second gives the root-mean-square deviation. Mean values for terminal residues (Gly-X or X-Gly models) are given at the end of the list.

are divided into two categories. *Near test atoms* are those of a Gly-X-Gly tripeptide model in which residue X contains the central atom. The tripeptide model for a half-cystine residue of a disulfide includes the SG, CB and CA atoms of the partner half-cystine as near atoms. *Long test atoms* are all other test atoms.

The exposure of a particular central atom to solvent is the area of the solvated sphere that contains test points not occluded by test atoms. Each test point on the surface of a central atom is considered separately with respect to all the test atoms. The test atom given credit for occluding a test point is determined by the greatest value for the ratio of the solvated radius of the test atom to the distance from the test point to the center of the test atom. The list of interacting test atoms and the corresponding areas occluded on a particular central atom describes the environment of the central atom. This list, which is the basic output of the computation, is stored on magnetic tape for subsequent use in summations and comparisons.

The Gly-X-Gly tripeptide serves as a model for the environment of a central atom in the unfolded protein. This model assumes that side chains of adjacent residues in an unfolded chain on the average do not contact the central residue. The conformation used for the tripeptide is that for the corresponding atoms of the native protein. The area exposed to solvent for a particular type of atom (or residue) therefore varies for this model according to its location in the folded molecule. Table 1 gives averages for areas exposed to solvent in the unfolded state and the corresponding root-mean-square deviations (a) for each type of atom and (b) for the sums over all backbone atoms and over all side chain atoms for each type of residue. The small values of the root-mean-square deviations show that use of the native conformation for the tripeptide model introduces no systematic error and is likely a better representation of the random coil than that obtained using a single conformation for the tripeptide.

In calculations for unfolded proteins, the test atoms are near atoms only. In computations for folded molecules, the test atoms include both near and long atoms. Because near and long test atoms are considered on equal terms in determining which test atom is given credit for occluding a test point, the area assigned to near atoms in calculations for a folded protein is in general less than that occluded by the same near atoms in the unfolded model.

TABLE 2

Van der Waals' radii

	Radius† (Å)	Area of solvated sphere (Å²)
All nitrogen: —N<, —NH<, —NH_2, —NH_3^+	1·5	106
All oxygen: =O, —O—, —OH	1·4	99
All sulfur: —S—, —SH	1·85	133
Non-aromatic carbon: —CH<, —CH_2, —CH_3	2·0	145
Aromatic carbon: =CH, =C<	1·85	133
Carbonyl and all other carbon	1·5	106
Zn^{2+}	0·74	58
Solvent (water)	1·4	—

† From Pauling (1960).

The van der Waals' radii used in these computations and the surface areas of the solvated van der Waals' spheres are given in Table 2. The radius of the solvated sphere is the sum of the van der Waals' radius of the atom and that of the solvent (1·4 Å). The values of Table 2, which are those of Pauling (1960), differ somewhat from those of Lee & Richards (1971), who used the van der Waals' radii of Bondi (1964). Like Lee & Richards, we do not explicitly consider hydrogen atoms, which are incorporated into the van der Waals' radii for groups.

Atoms are classed as polar or non-polar. Hetero-atoms and carbon atoms bonded to two or more hetero-atoms are considered polar and all other atoms non-polar. Charged atoms are polar atoms that are part of an ionizable group, e.g. CD, OE1 and OE2 in glutamyl residues and ND1, CE1 and NE2 in histidyl residues.

Ninety-two points represent the surface of the solvated sphere with sufficient accuracy for this type of calculation. Decreasing the size of the set from approx. 400 to 92 changes the percentage of total surface area exposed to solvent by less than 2% (with respect to the total surface of the solvated van der Waals' sphere) for 90% of the atoms of lysozyme. The set of 92 surface points has 5-fold symmetry about an axis parallel to the z-axis of the Cartesian co-ordinate frame. Table 3 gives the atom contacts for the two Zn^{2+} ions of the insulin hexamer. The three dimer units of the hexamer are related by a 3-fold symmetry axis coincident with the z-axis. Thus, a rotation of 120° about this axis results in no significant change in contacts (the differences correspond to ± 2 surface points at most).

TABLE 3

Contacts of the two zinc atoms within the insulin hexamer

A. Zn^{2+} (0·0, 0·0, +8·1)†: Total area exposed to solvent in hexamer = 0·6 Å²

	Extents of contact‡		
	Dimer I	Dimer II	Dimer III
	(Å²)	(Å²)	(Å²)
B'10 His CE1	5·6	5·0	5·6
B'10 His NE2	6·3	7·5	6·3
B'10 His CD2	6·9	6·9	6·9

B Zn^{2+} (0·0, 0·0, −8·0)†: Total area exposed to solvent in hexamer = 8·2 Å²

	Extents of contact‡		
	Dimer I	Dimer II	Dimer III
	(Å²)	(Å²)	(Å²)
B10 His CE1	3·8	2·5	3·8
B10 His NE2	10·6	10·6	11·3
B10 His CD2	2·5	2·5	1·9

† Cartesian co-ordinates given by Guy Dodson (personal communication).
‡ Area of zinc atom occluded by protein atoms.

The input for a computation consists of the Cartesian co-ordinates of all atoms heavier than hydrogen and for each such atom the residue number and a designator of the atom type†. The co-ordinates for the tetragonal lysozyme crystal structure were obtained from D. C. Phillips and colleagues (Blake *et al.*, 1967; Blake, 1967) and those for the rhombohedral 2 zinc insulin crystal structure‡ from D. C. Hodgkin and colleagues (Blundell *et al.*, 1971). Both co-ordinate sets are those that were current in 1970 and both had been refined by the method of Diamond (1966). We emphasize that the co-ordinates used in

† The atoms are defined as by Imoto *et al.* (1972).
‡ The two zinc atoms were deleted in all calculations except in that giving the data in Table 3.

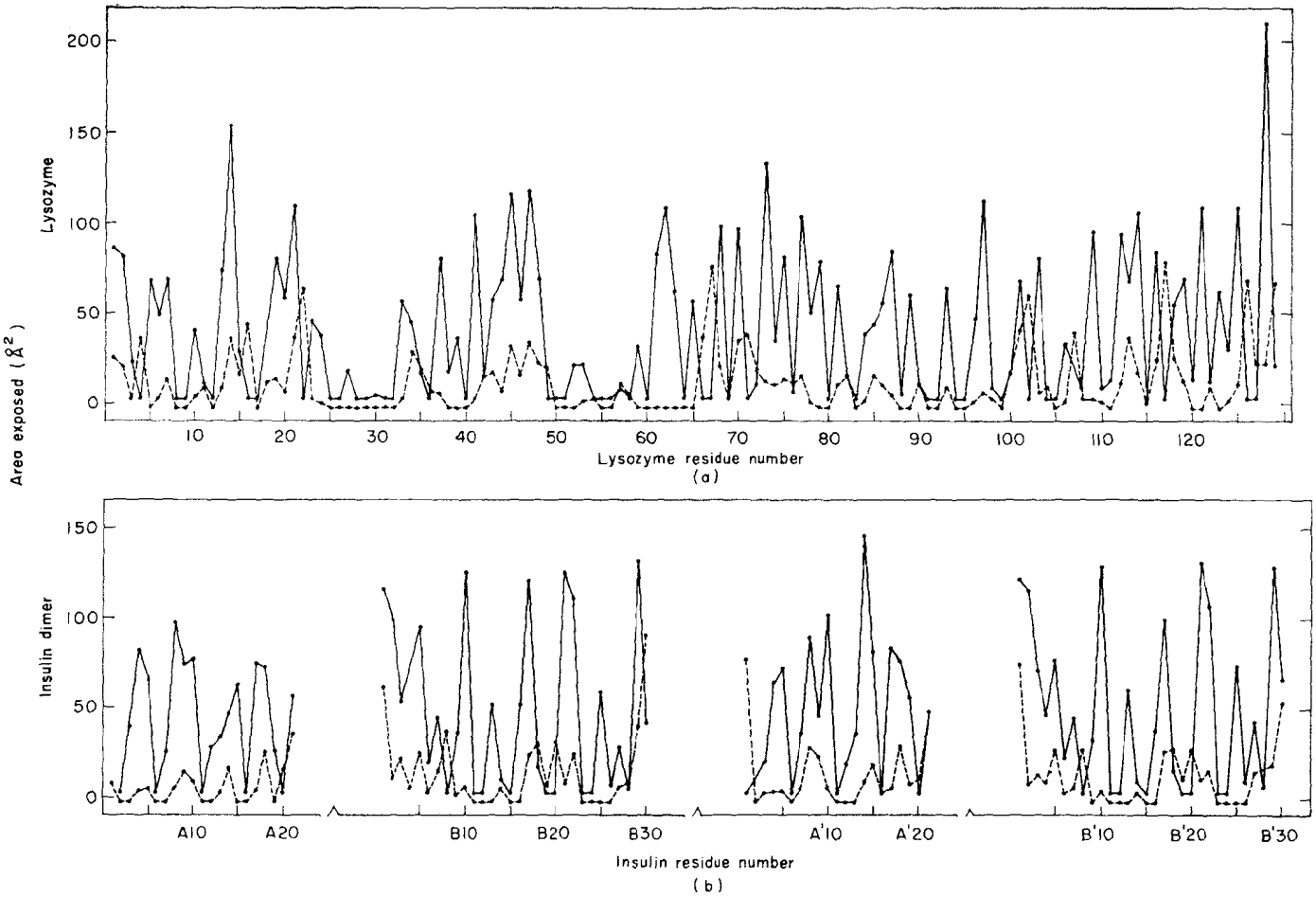

FIG. 1. Area exposed to solvent for backbone atoms (--●--●--) and side chain atoms (—●—●—). (a) Results for lysozyme; (b) results for insulin dimer.

TABLE 4

Exposure and environment results for lysozyme

1 LYS	111	200	216		SD	0	17	76		N	0	23	7		CG	0	28	28		CB	27	22	5
CA	21	30	25		CE	0	11	93		C	0	22	20		CD1	0	36	33		CG	2	38	5
NEND	5	38	16							O	0	32	18		CE1	6	32	43		CD	24	38	0
C	0	22	23	13 LYS	67	208	237		CB	0	35	36		CZ	22	17	36		NE	2	29	10
O	0	32	27		CA	0	33	39		CG	0	26	29		CE2	13	27	26		CZ	2	23	8
CB	8	17	28		N	0	31	15		CD1	4	22	29		CD2	6	26	26		NH2	32	25	0
CG	11	5	27		C	0	22	3		CE1	14	10	33							NH1	28	22	23
CD	8	17	38		O	9	13	16		CZ	6	10	40	35 GLU	33	151	319						
CE	35	21	19		CB	2	22	35		CE2	6	23	43		CA	2	30	32	46 ASN	73	210	61
NZ	24	18	13		CG	2	17	44		CD2	0	26	42		N	0	30	17		CA	13	30	8
					CD	5	17	49		OH	16	12	29		C	0	15	15		N	0	15	15
2 VAL	101	97	143		CE	35	30	21							O	15	4	20		C	1	21	9
CA	0	16	32		NZ	16	22	15	24 SER	38	166	136			CB	0	33	44		O	2	26	17
N	1	7	23							CA	0	52	14		CG	0	27	62		CB	28	32	2
C	0	9	22	14 ARG	189	153	58			N	1	26	16		CD	1	1	44		CG	0	22	0
O	18	3	16		CA	11	27	13		C	0	21	21		OE2	1	11	50		NOD2	26	18	7
CB	9	16	14		N	0	22	16		O	0	21	29		OE1	14	0	35		NOD1	2	47	3
CG2	38	17	22		C	0	15	7		CB	35	32	22										
CG1	35	28	14		O	25	7	0		OG	2	14	34	36 SER	7	206	195						
					CB	0	30	9							CA	6	44	36						
3 PHE	15	161	500		CG	32	13	9	25 LEU	3	178	377			N	0	26	18						
CA	0	22	39		CD	41	0	0		CA	0	32	38		C	0	22	24						
N	0	16	22		NE	7	7	0		N	0	29	20		O	1	17	41						
C	0	5	21		CZ	1	10	1		C	0	16	22		CB	0	60	46						
O	2	4	40		NH2	37	20	1		O	0	16	32		OG	0	36	30						
CB	0	14	51		NH1	36	2	1		CB	2	24	41											
CG	0	16	57							CG	0	24	71	37 ASN	86	119	137						
CD1	0	19	45	15 HIS	43	240	249			CD2	2	22	81		CA	0	32	28	48 ASP	81	79	114
CE1	4	22	50		CA	9	27	16		CD1	0	16	73		N	0	31	10		CA	13	8	9
CZ	9	19	55		N	0	28	11							C	4	15	30		N	4	10	15
CE2	0	14	63		C	0	22	10	26 GLY	0	103	182			O	6	4	37		C	0	9	2
CD2	0	10	58		O	5	20	22		CA	0	35	73		CB	24	14	9		O	11	11	13
					CB	0	24	38		N	0	21	33		CG	0	5	0		CB	38	9	9
4 GLY	35	79	132		CG	0	13	36		C	0	24	36		NOD2	49	0	0		CG	1	6	11
CA	35	27	36		ND1	0	16	26		O	0	20	41		NOD1	6	18	21		OD2	12	13	28
N	0	20	28		CE1	10	49	33												OD1	6	13	26
C	0	17	30		NE2	3	24	16	27 ASN	18	163	332		38 PHE	17	215	504						
O	0	15	39		CD2	14	19	39		CA	0	32	52		CA	0	27	46	49 GLY	20	116	60
										N	0	22	20		N	0	24	16		CA	19	38	27
5 ARG	71	204	270	16 GLY	44	98	87			C	0	20	31		C	0	17	38		N	1	24	5
CA	0	2	44		CA	36	36	21		O	0	16	37		O	0	20	45		C	0	18	15
N	1	8	17		N	0	32	7		CB	0	30	60		CB	0	24	35		O	0	35	14
C	0	10	15		C	0	16	28		CG	0	21	40		CG	0	22	38					
O	0	15	32		O	7	14	32		NOD2	11	15	44		CD1	0	16	50	50 SER	3	232	112
CB	16	3	39							NOD1	6	9	48		CE1	0	14	68		CA	0	52	22
CG	0	13	51	17 LEU	0	120	434								CZ	0	13	78		N	0	23	15
CD	16	16	43		CA	0	19	33	28 TRP	0	299	680			CE2	6	17	58		C	0	22	16
NE	10	13	2		N	0	16	15		CA	0	28	41		CD2	12	20	33		O	0	27	19
CZ	0	40	6		C	0	15	24		N	4	28	18							CB	8	66	33
NH2	28	36	7		O	0	21	30		C	0	23	22	39 ASN	37	181	270			OG	1	42	6
NH1	0	49	14		CB	0	19	60		O	0	19	37		CA	0	36	47					
					CG	0	13	84		CB	0	22	51		N	0	24	26	51 THR	3	275	194
6 CYS	52	97	103		CD2	0	11	96		CG	0	23	48		C	0	25	23		CA	0	49	32
CA	0	14	22		CD1	0	6	92		CD1	0	42	46		O	0	27	29		N	0	36	10
N	1	9	23							NE	0	33	29		CB	5	27	49		C	0	33	20
C	0	15	17	18 ASP	47	114	255			CE1	0	25	48		CG	0	11	29		O	0	34	27
O	1	16	26		CA	0	24	47		CZ1	0	19	69		NOD2	30	5	37		CB	0	54	32
CB	36	3	11		N	0	11	31		CH	0	10	84		ND1	2	26	39		CG2	3	36	55
SG	13	39	4		C	0	16	10		CZ2	0	10	81							OG1	0	33	19
					O	11	14	21		CE2	0	6	69	40 THR	0	257	234						
7 GLU	82	131	233		CB	11	6	62		CD2	0	12	38		CA	0	47	32	52 ASP	22	276	209
CA	11	24	32		CG	0	15	24							N	0	30	13		CA	0	49	36
N	1	17	22		OD2	19	5	36	29 VAL	0	165	307			C	5	26	25		N	0	17	25
C	0	20	22		OD1	6	22	22		CA	0	33	36		O	0	28	30		C	0	28	37
O	1	14	30							N	0	32	11		CB	0	58	39		O	0	27	39
CB	14	16	24	19 ASN	94	175	99			C	0	17	23		CG2	0	35	74		CB	0	49	30
CG	21	9	27		CA	0	39	17		O	0	12	35		OG1	0	32	20		CG	0	39	15
CD	2	8	25		N	0	36	10		CB	0	21	55							OD2	21	32	13
OE2	15	16	25		C	0	22	8		CG2	0	19	77	41 GLN	107	147	153			OD1	1	35	14
OE1	17	7	28		O	14	16	12		CG1	0	32	68		CA	0	28	33						
					CB	24	21	19							N	0	30	16	53 TYR	23	232	506
8 LEU	0	164	421		CG	1	10	8	30 CYS	3	127	205			C	0	5	28		CA	0	25	44
CA	0	28	43		ND2	32	20	3		CA	0	28	44		O	2	2	43		N	0	24	15
N	0	29	24		ND1	24	12	21		N	0	25	21		CB	11	36	14		C	0	21	24
C	0	18	22							C	0	18	16		CG	8	35	16		O	1	21	29
O	0	17	26							O	0	14	27		CD	1	1	0		CB	0	22	47
CB	0	21	57							CB	0	24	52		NOE2	48	0	3		CG	0	17	45
CG	0	19	74							SG	3	17	45		NOE1	36	10	0		CD1	0	9	58
CD2	0	14	82																	CE1	16	3	62
CD1	0	17	93						31 ALA	0	143	186		42 ALA	30	139	126			CZ	0	12	58
										CA	0	38	44		CA	16	32	25		CE2	0	33	46
9 ALA	0	119	235							N	0	41	10		N	0	28	22		CD2	0	30	35
CA	0	33	58							C	0	20	24		C	0	17	21		OH	6	15	44
N	0	28	23							O	0	16	34		O	0	20	30					
C	0	16	33							CB	0	28	73		CB	14	51	28	54 GLY	2	143	126
O	0	11	43																	CA	2	57	49
CB	0	32	77						32 ALA	0	167	199		43 THR	75	119	135			N	0	28	29
										CA	0	44	52		CA	0	27	32		C	0	32	23
10 ALA	44	111	147							N	0	24	28		N	1	16	11		O	4	27	26
CA	3	22	43	21 ARG	142	124	216			C	0	33	20		C	0	24	8					
N	0	24	26		CA	14	11	19		O	0	32	26		O	16	20	6	55 ILE	0	150	398
C	0	20	20		N	2	10	20		CB	0	33	74		CB	11	16	28		CA	0	27	27
O	0	18	29		C	0	9	8							CG2	19	16	49		N	0	23	17
CB	41	27	30		O	17	2	17	33 LYS	58	190	349			OG1	28	0	0		C	0	11	31
					CB	2	11	44		CA	0	43	39							O	0	16	39
11 ALA	19	150	158		CG	0	22	41		N	0	31	24	44 ASN	76	138	118			CB	0	16	54
CA	8	41	32		CD	30	14	14		C	0	22	16		CA	6	17	30			CG2	0	6	77
N	0	28	17		NE	8	8	10		O	2	25	16		N	0	24	21			CG1	0	16	76
C	0	29	24		CZ	1	9	14		CB	0	17	66		C	0	13	24			CD1	0	35	77
O	0	26	25		NH2	33	16	17		CG	0	16	62		O	0	28	26					
CB	11	27	60		NH1	34	10	11		CD	0	11	66		CB	11	30	9	56 LEU	0	98	435
										CE	16	16	49		CG	0	7	0		CA	0	17	54
12 MET	0	158	428	22 GLY	63	78	60			NZ	40	10	10		NOD2	21	20	8		N	0	13	28
CA	0	30	52		CA	33	27	21							NOD1	37	0	0		C	0	17	30
N	0	25	20		N	1	25	9	34 PHE	74	260	292								O	0	12	40
C	0	18	24		C	0	15	20		CA	3	28	21							CB	0	11	55
O	0	22	30		O	29	11	11		N	0	20	20	45 ARG	149	245	102			CG	0	6	62
CB	0	19	62							C	0	11	7		CA	0	24	21			CD2	0	14	76
															N	5	5	20						

57 GLN 19 268 223
CA 0 32 21
N 0 21 21
C 0 16 15
O 11 14 17
CB 3 36 30
CG 3 46 38
CD 0 33 26
NOE2 0 39 30
NOE1 2 28 26

58 ILE 5 177 323
CA 5 24 22
N 0 26 7
C 0 10 14
O 0 25 25
CB 0 21 43
CG2 0 13 66
CG1 0 25 69
CD1 0 33 77

59 ASN 33 207 189
CA 0 32 32
N 1 17 23
C 0 26 21
O 0 27 27
CB 21 28 25
CG 0 24 13
ND2 10 37 11
OD1 1 16 37

60 SER 0 255 142
CA 0 63 22
N 0 36 9
C 0 30 15
O 0 36 19
CB 0 60 46
OG 0 30 31

61 ARG 84 176 322
CA 0 24 47
N 0 21 20
C 0 8 25
O 0 19 31
CB 2 28 47
CG 0 25 44
CD 8 25 41
NE 0 8 22
CZ 1 0 17
NH2 28 14 13
NH1 46 3 15

62 TRP 109 191 350
CA 0 28 39
N 0 23 16
C 0 22 34
O 0 20 37
CB 0 9 49
CG 3 6 22
CD1 10 16 32
NE 13 6 20
CE1 4 1 20
CZ1 25 10 16
CH 27 23 7
CZ2 14 20 16
CE2 10 6 29
CD2 3 0 13

63 TRP 43 240 559
CA 0 33 35
N 0 14 16
C 0 23 26
O 0 35 26
CB 0 39 36
CG 0 20 39
CD1 3 12 63
NE 8 0 45
CE1 0 0 46
CZ1 17 1 50
CH 10 14 50
CZ2 4 19 50
CE2 0 20 38
CD2 0 9 38

64 CYS 0 142 177
CA 0 36 28
N 0 17 23
C 0 23 16
O 0 20 16
CB 0 25 43
SG 0 20 50

65 ASN 57 162 171
CA 0 39 33
N 0 30 15
C 0 24 21
O 0 17 33
CB 8 27 32
CG 1 8 5
ND2 37 3 13
ND1 11 13 20

66 ASP 35 181 171
CA 13 30 17
N 0 24 9
C 1 11 3
O 18 5 12
CB 3 27 39
CG 0 30 25
OD2 0 31 40
OD1 0 22 25

67 GLY 76 68 31
CA 39 32 6
N 0 29 3
C 0 8 14
O 36 0 7

68 ARG 120 262 91
CA 13 24 9
N 0 16 0
C 0 13 15
O 7 10 27
CB 27 17 0
CG 3 43 3
CD 19 33 3
NE 6 21 3
CZ 3 25 5
NH2 25 23 21
NH1 16 36 5

69 THR 3 209 221
CA 0 25 33
N 0 14 18
C 0 17 25
O 1 24 21
CB 0 49 38
CG2 2 49 58
OG1 0 31 27

70 PRO 131 54 61
CA 13 3 9
N 0 5 9
C 0 8 6
O 22 10 9
CB 39 0 0
CD 11 17 25
CG 46 11 3

71 GLY 39 72 92
CA 28 25 32
N 2 8 28
C 0 16 14
O 9 22 19

72 SER 32 159 102
CA 0 36 17
N 0 23 14
C 0 7 13
O 21 7 11
CB 11 49 27
OG 0 36 20

73 ARG 147 126 182
CA 13 14 27
N 0 18 13
C 0 9 26
O 0 26 26
CB 9 13 30
CG 25 14 11
CD 14 14 24
NE 10 6 0
CZ 2 3 0
NH2 55 0 0
NH1 17 8 24

74 ASN 47 165 179
CA 0 22 27
N 2 3 16
C 0 17 13
O 10 20 9
CB 24 13 36
CG 0 23 25
ND2 11 29 33
OD1 0 37 20

75 LEU 94 107 168
CA 3 19 14
N 0 26 6
C 0 13 14
O 11 14 20
CB 2 11 33
CG 8 8 28
CD2 65 0 2
CD1 6 16 51

76 CYS 20 113 188
CA 2 25 36
N 0 28 20
C 0 21 23
O 12 3 31
CB 6 16 38
SG 0 20 40

77 ASN 121 145 58
CA 0 41 13
N 0 39 9
C 0 11 10
O 17 1 19
CB 24 30 2
CG 1 9 2
ND2 47 8 2
ND1 32 5 0

78 ILE 52 132 287
CA 2 24 25
N 0 26 11
C 0 18 30
O 0 29 25
CB 2 11 36
CG2 0 24 58
CG1 0 0 54
CD1 49 0 47

79 PRO 79 119 181
CA 0 28 30
N 0 7 21
C 0 24 28
O 0 18 30
CB 25 27 11
CD 13 8 39
CG 41 6 22

80 CYS 0 113 245
CA 0 24 47
N 0 17 13
C 0 10 31
O 0 13 40
CB 0 21 60
SG 0 29 55

81 SER 75 72 135
CA 2 16 36
N 2 13 24
C 0 16 16
O 6 14 27
CB 49 8 19
OG 16 5 13

82 ALA 33 134 137
CA 16 35 28
N 0 34 8
C 0 18 17
O 1 26 20
CB 16 21 63

83 LEU 0 248 329
CA 0 36 35
N 0 22 18
C 0 25 17
O 0 32 24
CB 0 33 39
CG 0 35 60
CD2 0 38 68
CD1 0 27 68

84 LEU 42 206 260
CA 0 28 25
N 0 26 10
C 0 15 14
O 2 21 22
CB 9 28 22
CG 0 33 49
CD2 30 16 58
CD1 0 38 58

85 SER 61 147 76
CA 17 28 9
N 0 24 8
C 0 15 7
O 0 19 22
CB 39 27 17
OG 4 33 12

86 SER 69 50 164
CA 2 16 41
N 1 11 18
C 0 2 25
O 10 2 31
CB 35 8 33
OG 21 11 15

87 ASP 91 132 98
CA 5 17 19
N 0 18 17
C 0 22 11
O 1 33 14
CB 32 13 16
CG 2 8 7
OD2 43 0 0
OD1 9 20 14

88 ILE 5 122 315
CA 0 27 21
N 0 16 3
C 0 16 21
O 0 16 30
CB 0 2 41
CG2 0 11 71
CG1 5 17 44
CD1 0 17 84

89 THR 60 201 124
CA 0 35 35
N 0 24 17
C 0 20 14
O 0 27 22
CB 6 38 17
CG2 46 27 11
OG1 7 31 7

90 ALA 23 82 198
CA 9 17 46
N 0 10 30
C 0 17 25
O 2 12 39
CB 11 25 56

91 SER 0 103 313
CA 0 22 62
N 0 18 31
C 0 21 32
O 0 12 44
CB 0 17 88
OG 0 13 56

92 VAL 0 151 357
CA 0 25 47
N 0 26 24
C 0 22 23
O 0 16 36
CB 0 16 66
CG2 0 33 62
CG1 0 13 98

93 ASN 74 147 162
CA 2 35 25
N 0 28 22
C 0 20 11
O 9 10 28
CB 13 24 22
CG 1 9 14
ND2 32 6 25
ND1 18 17 14

94 CYS 2 139 199
CA 2 35 32
N 0 32 11
C 0 22 22
O 0 15 32
CB 0 27 39
SG 0 9 62

95 ALA 0 114 247
CA 0 32 62
N 0 26 24
C 0 20 29
O 0 14 44
CB 0 22 88

96 LYS 47 252 340
CA 0 35 47
N 0 28 16
C 0 25 28
O 2 22 30
CB 6 33 39
CG 0 27 54
CD 13 24 49
CE 5 32 60
NZ 21 26 17

97 LYS 119 169 146
CA 3 24 30
N 0 25 15
C 0 20 21
O 4 11 33
CB 5 24 22
CG 11 27 11
CD 8 25 14
CE 52 5 0
NZ 36 9 0

98 ILE 12 180 320
CA 0 25 33
N 0 23 15
C 0 17 13
O 4 26 15
CB 0 21 41
CG2 6 14 62
CG1 0 25 62
CD1 2 28 81

99 VAL 3 157 316
CA 0 33 32
N 0 20 23
C 0 15 15
O 3 18 29
CB 0 25 66
CG2 0 28 65
CG1 0 17 87

100 SER 34 202 102
CA 0 49 22
N 0 24 18
C 0 23 9
O 14 24 4
CB 19 51 24
OG 1 32 25

101 ASP 108 80 46
CA 11 22 6
N 0 23 8
C 0 0 0
O 29 0 0
CB 6 25 17
CG 0 0 6
OD2 35 0 7
OD1 27 10 1

102 GLY 61 104 21
CA 30 43 8
N 0 29 9
C 0 18 3
O 31 14 0

103 ASN 87 69 68
CA 6 6 13
N 0 5 0
C 0 10 9
O 0 30 13
CB 8 9 30
CG 1 0 0
ND2 33 8 3
ND1 39 0 0

104 GLY 12 91 126
CA 11 36 43
N 1 20 14
C 0 18 31
O 0 17 39

105 MET 0 89 526
CA 0 17 63
N 0 11 30
C 0 16 41
O 0 17 44
CB 0 3 79
CG 0 5 85
SD 0 12 81
CE 0 8 103

106 ASN 35 259 174
CA 0 46 27
N 0 21 22
C 0 34 11
O 2 41 10
CB 8 38 32
CG 1 30 16
ND2 23 26 18
ND1 1 24 39

107 ALA 62 91 123
CA 11 27 28
N 0 23 21
C 0 10 14
O 30 9 6
CB 21 22 54

108 TRP 13 259 518
CA 5 28 16
N 0 20 13
C 0 28 18
O 0 31 25
CB 0 30 30
CG 0 16 20
CD1 7 30 25
NE 0 14 24
CE1 1 9 48
CZ1 0 10 74
CH 0 10 71
CZ2 0 17 69
CE2 0 13 49
CD2 0 3 38

109 VAL 99 135 74
CA 3 25 16
N 1 20 8
C 0 14 17
O 0 28 21
CB 13 6 0
CG2 36 17 9
CG1 46 25 2

110 ALA 11 130 184
CA 3 32 49
N 0 17 29
C 0 23 21
O 0 20 30
CB 8 38 55

111 TRP 12 374 546
CA 0 39 38
N 0 18 25
C 0 24 16
O 0 30 19
CB 0 27 54
CG 0 26 36
CD1 0 36 48
NE 0 22 36
CE1 0 25 46
CZ1 9 17 56
CH 3 32 53
CZ2 0 35 48
CE2 0 25 42
CD2 0 19 29

112 ARG 102 267 237
CA 0 35 38
N 0 24 28
C 0 20 14
O 9 18 17
CB 0 36 33
CG 11 32 25
CD 27 36 11
NE 1 20 15
CZ 1 18 15
NH2 15 22 30
NH1 36 7 11

113 ASN 106 98 140
CA 6 21 21
N 0 24 17
C 0 11 13
O 32 0 7
CB 21 16 27
CG 1 10 9
ND2 30 5 20

	NOD1	15	11	27	117 GLY	79	23	62		CG1	11	39	39		CE2	1	1	50	127 CYS	26	113	137
					CA	43	8	19							CD2	1	1	36	CA	2	17	24
114 ARG	126	131	269		N	3	11	16	121 GLN	108	110	203							N	6	20	0
CA	9	30	13		C	0	3	15	CA	0	30	41	124 ILE	33	164	307			C	0	7	20
N	8	28	11		O	33	0	12	N	10	14	22	CA	0	33	25			O	22	11	15
C	0	20	15						C	0	21	26	N	6	23	16			CB	2	24	43
O	11	12	28	118 THR	81	122	77		O	0	19	37	C	0	25	14			SG	0	35	36
CB	6	19	28	CA	14	16	8		CB	8	19	33	O	3	25	18						
CG	8	9	44	N	0	13	0		CG	11	5	36	CB	6	15	46	128 ARG	233	31	77
CD	14	9	41	C	6	9	10		CD	2	0	1	CG2	0	14	79	CA	9	5	11
NE	9	0	17	O	13	10	15		NOE2	91	2	6	CG1	5	14	43	N	0	8	18
CZ	3	6	22	CB	17	21	13		NOE1	45	0	0	CD1	25	14	66	C	0	0	18
NH2	31	0	31	CG2	33	38	9										O	15	0	29
NH1	33	5	18	OG1	3	16	21	122 ALA	33	189	132	125 ARG	121	296	158	CB	32	0	0
								CA	0	49	38	CA	11	30	8	CG	24	13	0
115 CYS	0	140	205	119 ASP	81	128	109	N	0	33	25	N	0	26	10	CD	35	5	0
CA	0	22	38	CA	9	8	22	C	0	21	22	C	0	21	0	NE	11	1	0
N	0	20	20	N	0	6	9	O	11	22	21	O	2	42	0	CZ	5	0	0
C	0	24	23	C	6	11	17	CB	22	63	25	CB	28	24	5	NH2	55	0	0
O	8	26	27	O	3	15	21					CG	16	27	17	NH1	47	0	0
CB	0	25	43	CB	36	19	9	123 TRP	62	185	510	CD	22	33	16					
SG	0	23	55	CG	1	20	6	CA	0	25	35	NE	6	15	23	129 LEU	91	115	252
				OD2	28	10	10	N	0	22	16	CZ	1	23	24	CA	11	13	14
116 LYS	105	146	226	OD1	3	32	14	C	0	23	29	NH2	14	32	29	N	1	11	7
CA	2	21	36					O	0	21	37	NH1	21	23	26	C	5	10	2
N	0	15	28	120 VAL	13	206	208	CB	0	22	51					QEND	41	0	0
C	0	5	6	CA	0	24	33	CG	0	16	32	126 GLY	70	64	32	CB	14	14	24
O	19	10	9	N	0	16	8	CD1	1	23	42	CA	47	24	6	CG	0	13	47
CB	0	22	38	C	0	11	26	NE	10	13	15	N	6	10	10	CD2	8	17	76
CG	16	8	35	O	0	17	36	CE1	4	7	30	C	0	16	2	CD1	0	19	65
CD	0	19	46	CB	0	38	24	CZ1	14	7	38	O	17	14	13	QEND	12	17	17
CE	49	21	21	CG2	2	60	41	CH	16	1	43								
NZ	20	26	9					CZ2	13	1	56								

The values given for each atom are from left to right (columns 2 to 4): the area (Å^2) exposed to solvent, the area occluded by polar long atoms and the area occluded by non-polar long atoms, respectively. The corresponding sums of these areas over all atoms of each residue are given on the first line of each block.

this paper are preliminary and are being refined in the crystallographic laboratories. The effect of uncertainty in the co-ordinates is discussed in the Discussion (section (f)).

All computations were carried out on the CDC6400 computer of the University of Arizona Computer Center except for preliminary work, which was done on the Argus system of the Laboratory of Molecular Biophysics of Oxford University†. Run times on the CDC6400 were approx. 5·6 and 4·2 min for lysozyme and insulin dimer, respectively.

3. Results

(a) *Exposure and environment of atoms*

Individual atom data for native lysozyme and insulin dimer are presented in Tables 4 and 5. Comparison of the values of column 3 (area in Å^2 occluded by polar long atoms) with those of column 4 (area occluded by non-polar long atoms) gives a measure of the polarity of the environment within the folded protein. The values of column 2 (area exposed to solvent) can be compared with the areas exposed in the Gly-X-Gly models for the unfolded state that are given in Table 1 in order to see the effect of folding on exposure of an atom or residue.

(b) *Exposure and change in exposure of backbone and side chain elements*

Figure 1 shows the area exposed to solvent for the backbone and side chain of each residue of native lysozyme and insulin dimer. The values plotted are summations over the appropriate atoms of the results given in Tables 4 and 5. Graphs of this kind are a convenient way to present the changes in exposed surface area that follow from association reactions. Figure 2 shows the changes developed through binding of the

† During tenure of a Special Fellowship from the National Institutes of Health held by J. A. Rupley.

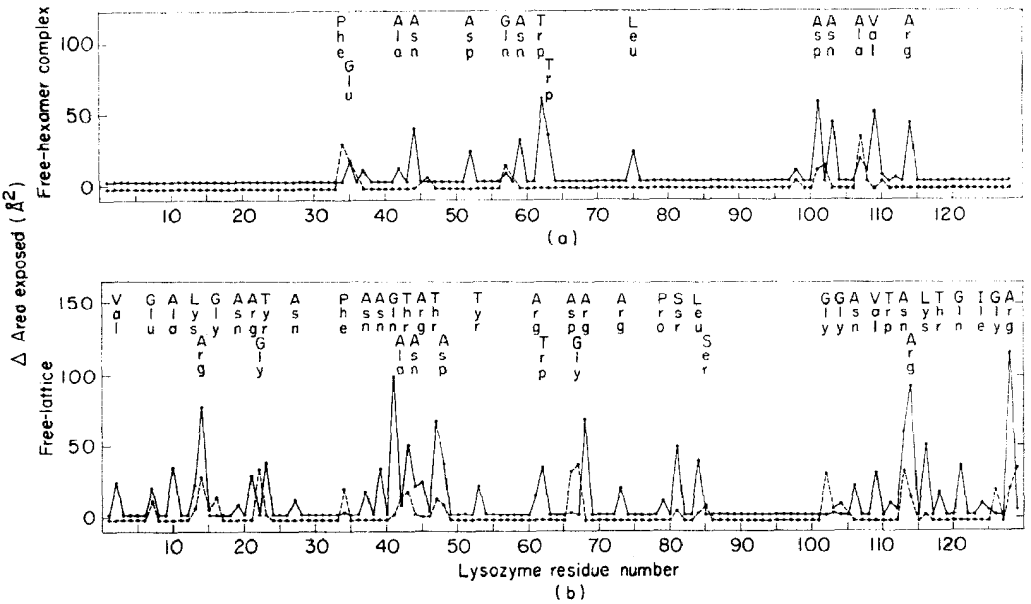

FIG. 2. Changes in area exposed to solvent for backbone atoms (—●—●—) and side chain atoms (—●—●—) for (a) binding of *N*-acetylglucosamine hexasaccharide to lysozyme; and (b) incorporation into the crystal lattice.

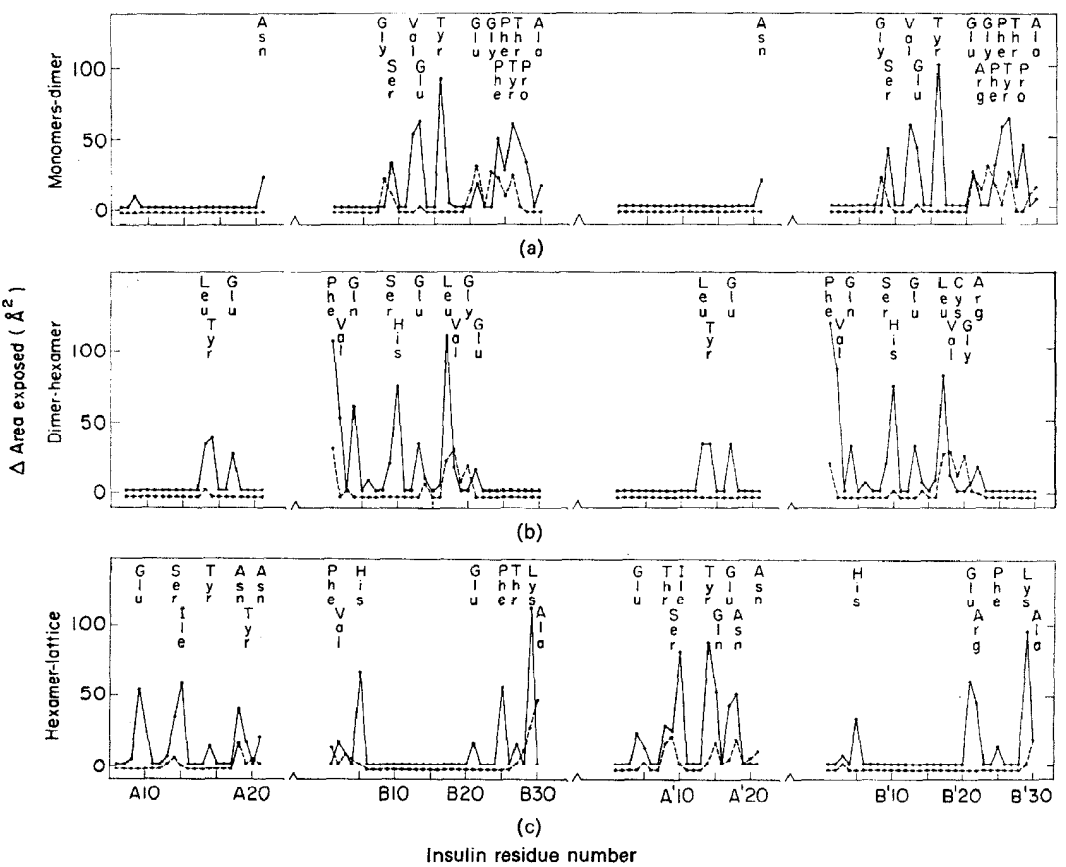

FIG. 3. Changes in area exposed to solvent for backbone atoms (—●—●—) and side chain atoms (—●—●—) for (a) association of insulin monomers to the dimer; (b) incorporation of dimer into the hexamer; and (c) incorporation of hexamer into the crystal lattice.

TABLE 5

Exposure and environment results for insulin dimer

A 1 GLY	8	129	172
CA	5	41	73
NENO	0	29	46
C	0	26	30
O	3	33	24
A 2 ILE	0	95	440
CA	0	21	41
N	0	8	22
C	0	11	25
O	0	21	28
CB	0	6	58
CG2	0	9	77
CG1	0	9	84
CD1	0	8	104
A 3 VAL	41	120	236
CA	0	16	43
N	0	13	28
C	0	20	18
O	0	21	28
CB	5	9	33
CG1	19	13	57
CG2	17	28	30
A 4 GLU	85	183	170
CA	2	28	25
N	0	20	20
C	1	18	18
O	1	26	20
CB	17	13	21
CG	41	14	6
CD	0	23	13
OE1	22	11	25
OE2	0	30	22
A 5 GLN	68	293	129
CA	5	49	22
N	0	20	15
C	0	41	11
O	0	43	11
CB	2	49	14
CG	14	39	21
CD	0	20	5
NOE1	5	24	26
NOE2	43	9	5
A 6 CYS	0	125	242
CA	0	28	41
N	0	25	9
C	0	22	28
O	0	20	26
CB	0	16	71
SG	0	13	68
A 7 CYS	25	158	113
CA	0	39	22
N	0	29	13
C	0	28	11
O	0	24	19
CB	0	30	32
SG	25	9	16
A 8 THR	101	163	66
CA	0	33	14
N	0	30	9
C	0	23	10
O	5	29	2
CB	16	17	17
OG1	15	16	15
CG2	65	14	0
A 9 SER	89	72	89
CA	5	21	21
N	0	20	11
C	0	10	21
O	10	7	24
CB	38	14	13
OG	36	0	0
A10 ILE	84	232	118
CA	0	44	11
N	0	30	5
C	0	25	9
O	9	24	3
CB	0	28	13
CG2	28	21	14
CG1	0	33	33
CD1	47	27	30
A11 CYS	0	180	136
CA	0	39	19
N	0	29	11
C	0	24	13
O	0	32	20
CB	0	28	35
SG	0	27	38
A12 SER	28	190	149
CA	0	47	27
N	0	34	10
C	0	18	32
O	0	28	29
CB	25	49	22
OG	3	15	29
A13 LEU	35	108	317
CA	0	16	36
N	0	11	20
C	0	16	18
O	2	19	30
CB	3	0	58
CG	0	5	51
CD1	27	3	66
CD2	3	38	38
A14 TYR	169	108	204
CA	0	22	24
N	0	17	13
C	0	16	16
O	17	12	16
CB	21	21	5
CG	3	6	10
CD1	19	1	16
CE1	36	1	12
CZ	6	0	23
OE2	22	0	29
CD2	17	12	20
OH	29	0	21
A15 GLN	64	263	122
CA	2	43	8
N	0	24	10
C	0	29	17
O	0	29	20
CB	0	46	13
CG	33	30	9
CD	2	13	10
NOE1	10	22	21
NOE2	17	26	13
A16 LEU	0	165	420
CA	0	28	46
N	0	20	18
C	0	13	26
O	0	14	41
CB	0	28	41
CG	0	25	76
CD1	0	14	90
CD2	0	22	82
A17 GLU	79	206	153
CA	0	44	32
N	0	28	23
C	0	29	7
O	5	20	15
CB	2	33	32
CG	17	28	25
CD	1	2	3
OE1	11	21	13
OE2	43	0	3
A18 ASN	97	142	117
CA	0	27	16
N	0	26	13
C	0	16	15
O	17	14	16
CB	17	21	21
CG	0	13	7
NOD1	22	13	24
NOD2	32	13	7
A19 TYR	24	235	468
CA	0	19	38
N	0	15	16
C	0	21	25
O	0	18	34
CB	0	21	58
CG	0	14	45
CD1	0	10	55
CE1	10	10	53
CZ	6	23	38
CE2	1	36	45
CD2	0	16	42
OH	6	32	19
A20 CYS	16	151	163
CA	0	30	35
N	0	21	20
C	0	11	20
O	13	9	24
CB	3	44	24
SG	0	35	42
A21 ASN	90	171	149
CA	9	25	19
N	0	15	24
C	0	25	14
OENO	25	13	7
OEND'	0	31	27
CB	9	24	25
CG	0	12	3
NOD1	6	20	29
NOD2	40	2	0
B 1 PHE	176	25	242
CA	13	0	17
NENO	23	0	24
C	0	0	20
O	26	0	19
CB	9	2	33
CG	0	0	24
CD1	23	0	4
CE1	40	0	0
CZ	32	0	14
CE2	4	10	43
CD2	6	4	43
B 2 VAL	108	54	90
CA	9	9	16
N	0	0	21
C	0	20	10
O	1	25	29
CB	6	0	0
CG1	39	0	14
CG2	52	0	0
B 3 ASN	75	53	233
CA	2	8	43
N	0	2	25
C	0	3	33
O	19	0	34
CB	21	11	35
CG	1	8	15
NOD1	28	2	22
NOD2	5	18	25
B 4 GLN	77	118	223
CA	5	19	24
N	0	21	21
C	0	16	24
O	0	22	32
CB	0	22	24
CG	22	11	27
CD	0	0	20
NOE1	3	6	40
NOE2	47	0	13
B 5 HIS	118	139	192
CA	0	22	32
N	2	1	25
C	0	13	18
O	21	3	16
CB	25	21	11
CG	3	19	14
ND1	1	29	10
CE1	25	32	29
NE2	11	0	16
CD2	29	0	20
B 6 LEU	22	145	329
CA	3	22	32
N	0	25	16
C	0	9	24
O	0	10	30
CB	0	22	44
CG	0	21	57
CD1	19	25	38
CD2	0	11	88
B 7 CYS	57	31	73
CA	2	0	22
N	0	9	13
C	0	3	10
O	12	1	19
CB	24	0	0
SG	20	17	9
B 8 GLY	36	66	122
CA	32	25	41
N	1	8	24
C	0	18	23
O	3	14	34
B 9 SER	36	64	196
CA	0	14	49
N	0	8	17
C	0	11	30
O	0	15	34
CB	19	3	58
OG	17	12	7
B10 HIS	129	193	121
CA	3	35	22
N	0	16	17
C	0	17	25
O	2	19	26
CB	16	33	16
CG	4	17	4
ND1	8	15	0
CE1	58	12	0
NE2	16	7	0
CD2	22	22	10
B11 LEU	0	178	341
CA	0	35	35
N	0	28	21
C	0	17	21
O	0	15	30
CB	0	32	27
CG	0	16	58
CD1	0	16	76
CD2	0	21	74
B12 VAL	0	138	388
CA	0	22	49
N	0	18	21
C	0	21	32
O	0	17	36
CB	0	11	73
CG1	0	0	103
CG2	0	19	74
B13 GLU	51	172	254
CA	0	25	43
N	0	24	23
C	0	20	28
O	0	12	36
CB	11	19	35
CG	0	24	51
CD	2	11	14
OE1	14	24	15
OE2	24	14	11
B14 ALA	14	138	208
CA	5	25	49
N	0	24	16
C	0	20	36
O	0	18	34
CB	9	21	73
B15 LEU	0	192	404
CA	0	33	51
N	0	25	23
C	0	24	24
O	0	19	24
CB	0	19	55
CG	0	24	62
CD1	0	35	71
CD2	0	13	95
B16 TYR	58	198	392
CA	0	28	44
N	0	28	22
C	0	16	5
O	0	20	15
CB	8	19	57
CG	0	7	42
CD1	0	12	59
CE1	3	25	46
CZ	4	12	33
CE2	27	12	20
CD2	0	13	27
OH	24	7	21
B17 LEU	142	130	142
CA	5	24	25
N	0	21	21
C	0	11	10
O	17	6	6
CB	0	19	32
CG	8	9	19
CD1	47	9	21
CD2	65	0	8
B18 VAL	44	173	171
CA	8	36	9
N	0	24	13
C	0	24	1
O	20	7	4
CB	0	27	38
CG1	6	38	43
CG2	9	16	63
B19 CYS	8	114	183
CA	0	21	32
N	0	20	17
C	0	10	8
O	6	12	24
CB	2	27	49
SG	0	25	53
B20 GLY	30	84	96
CA	28	36	28
N	0	30	11
C	1	9	18
O	0	9	37
B21 GLU	131	62	86
CA	5	16	21
N	2	9	1
C	0	8	16
O	0	13	28
CB	22	8	11
CG	25	3	5
CD	5	0	0
OE1	28	5	4
OE2	44	0	0
B22 ARG	134	252	133
CA	11	22	14
N	0	17	16
C	0	15	3
O	13	20	9
CB	19	32	8
CG	0	35	24
CD	19	36	24
NE	0	23	14
CZ	2	17	8
NH1	33	21	14
NH2	37	14	0
B23 GLY	0	93	165
CA	0	36	63
N	0	22	31
C	0	18	31
O	0	16	40
B24 PHE	1	182	578
CA	0	25	44
N	0	13	26
C	0	16	28
O	0	19	36
CB	0	28	46
CG	0	24	49
CD1	1	7	59
CE1	0	9	75
CZ	0	12	87
CE2	0	14	74
CD2	0	12	55
B25 PHE	58	171	395
CA	0	24	46
N	0	15	24
C	0	13	23
O	0	11	28
CB	0	14	57
CG	2	13	47
CD1	0	9	42
CE1	7	17	42
CZ	25	7	36
CE2	17	25	22
CD2	7	25	29
B26 TYR	6	87	647
CA	0	14	49
N	0	14	30
C	0	13	24
O	0	18	34
CB	0	9	58
CG	0	3	46
CD1	0	7	59
CE1	0	6	71
CZ	0	6	72
CE2	0	0	76
CD2	0	6	71
OH	6	3	56
B27 THR	33	84	259
CA	0	13	54
N	0	9	39
C	0	0	28
O	6	0	25
CB	0	25	47
OG1	13	16	16
CG2	14	21	51
B28 PRO	12	107	265
CA	0	17	41
N	0	6	11
C	0	13	18
O	7	19	11
CB	5	13	60
CG	0	27	52
CD	0	13	71
B29 LYS	166	99	125
CA	9	11	21
N	0	13	28
C	2	0	11
O	25	0	10
CB	9	16	11
CG	28	9	17
CD	19	21	17
CE	52	14	5
NZ	21	15	6
B30 ALA	131	58	31
CA	11	11	3
N	0	16	7
C	1	2	0
OENO	34	5	0
OEND'	144	0	0
CB	41	24	21
A* 1 GLY	76	60	93
CA	36	27	28
NENO	23	13	18
C	0	13	22
O	17	7	25
A* 2 ILE	11	166	313
CA	0	22	24
N	0	9	15
C	0	24	21
O	0	27	25
CB	0	14	35
CG2	2	13	81
CG1	0	24	49
CD1	9	33	65
A* 3 VAL	21	135	264
CA	2	14	38
N	0	9	26
C	0	11	15
O	0	13	26
CB	0	24	47
CG1	0	32	68
CG2	19	32	44
A* 4 GLU	67	136	275
CA	3	14	39
N	0	13	18
C	0	14	22
O	0	17	27
CB	21	11	21
CG	35	3	33
CD	8	8	32
OE1	0	14	43
OE2	9	12	40
A* 5 GLN	75	206	161
CA	3	35	25
N	0	28	14
C	0	29	20
O	0	25	22
CB	24	24	14

CG	11	28	22
CD	6	5	14
NOE1	21	12	11
NOE2	16	22	20
A* 6 CYS	6	179	175
CA	7	49	19
N	6	28	10
C	6	29	17
O	0	26	28
CB	0	33	41
SG	6	14	59
A* 7 CYS	38	106	101
CA	6	54	19
N	0	39	14
C	0	30	15
O	4	26	15
CB	6	38	27
SG	27	10	12
A* 8 THR	116	98	60
CA	2	21	14
N	0	26	11
C	0	11	6
O	26	6	2
CB	9	16	14
OG1	27	0	0
CG2	52	17	13
A* 9 SER	65	139	103
CA	14	21	13
N	0	23	9
C	1	8	29
O	5	20	29
CB	16	25	19
OG	29	12	4
A*10 ILE	106	122	183
CA	0	24	36
N	0	9	17
C	0	17	20
O	5	12	28
CB	3	17	19
CG2	35	24	16
CG1	6	13	30
CD1	55	6	17
A*11 CYS	3	146	135
CA	3	39	16
N	0	18	15
C	0	22	21
O	0	17	25
CB	0	39	16
SG	0	16	43
A*12 SER	20	171	160
CA	0	39	35
N	0	29	14
C	0	13	26
O	0	19	30
CB	17	51	32
OG	2	20	24
A*13 LEU	36	93	324
CA	0	13	39
N	0	13	22
C	0	11	22
O	0	14	34
CB	2	11	43
CG	0	5	43
CD1	16	27	51
CD2	19	0	71
A*14 TYR	155	86	220
CA	2	19	22
N	0	13	17
C	2	10	11
O	5	17	18
CB	24	9	13
CG	3	0	9
CD1	19	7	6
OE1	33	4	14
OZ	7	0	26
CE2	17	6	27
CD2	14	0	33
OH	28	0	22
A*15 GLN	85	205	176
CA	16	41	14
N	0	22	14
C	0	25	13
O	2	21	25
CB	6	25	30
CG	13	19	38
CD	0	17	15
NOE1	7	21	26
NOE2	40	13	2
A*16 LEU	0	193	309
CA	0	28	39
N	0	26	14
C	0	22	18
O	0	25	21
CB	0	16	49
CG	0	27	36
CD1	0	28	69
CD2	0	21	62
A*17 GLU	87	202	132
CA	0	47	21
N	0	28	15
C	0	30	5
O	4	32	5
CB	5	39	27
CG	9	22.	24
CD	2	0	13
OE1	27	2	17
OE2	46	1	6
A*18 ASN	104	104	100
CA	9	11	16
N	0	18	14
C	0	13	15
O	18	20	3
CB	24	6	27
CG	0	11	5
ND1	20	16	15
ND2	32	8	6
A*19 TYR	62	182	359
CA	6	22	25
N	0	14	11
C	0	15	21
O	1	19	28
CB	0	28	36
CG	0	13	26
CD1	6	23	23
CE1	19	14	22
CZ	3	3	39
CE2	3	10	56
CD2	0	12	48
OH	25	9	20
A*20 CYS	11	232	114
CA	0	51	22
N	0	25	18
C	0	34	7
O	11	39	1
CB	0	46	27
SG	0	38	39
A*21 ASN	95	168	162
CA	8	28	13
N	0	24	18
C	2	16	13
OEND	10	17	25
OEND'26		25	2
CB	32	22	8
CG	1	10	23
ND01	1	17	39
ND02	16	8	22
B* 1 PHE	196	30	219
CA	11	0	19
NEND	31	0	6
C	1	0	11
O	31	1	5
CB	2	2	44
CG	0	5	38
CD1	0	13	42
CE1	22	10	26
CZ	38	0	9
CE2	39	0	6
CD2	22	0	13
B* 2 VAL	122	38	73
CA	5	3	9
N	1	0	13
C	0	6	10
O	2	12	34
CB	11	2	0
CG1	47	16	2
CG2	55	0	5
B* 3 ASN	82	44	227
CA	0	8	39
N	1	2	23
C	6	10	23
O	12	10	26
CB	25	5	39
CG	3	3	17
ND01	17	2	24
ND02	23	3	36
B* 4 GLN	55	122	273
CA	9	13	30
N	0	16	23
C	6	14	30
O	6	25	32
CB	2	19	32
CG	0	27	47
CD	0	5	23
NOE1	25	1	14
NOE2	20	3	43
B* 5 HIS	151	232	167
CA	0	17	33
N	2	1	26
C	0	8	20
O	24	2	11
CB	17	13	28
CG	0	26	20
ND1	15	6	6
CE1	43	39	1
NE2	0	43	7
CD2	0	48	14
B* 6 LEU	25	148	341
CA	3	17	41
N	0	11	21
C	0	18	25
O	0	21	36
CB	0	17	41
CG	0	14	52
CD1	21	21	51
CD2	2	27	74
B* 7 CYS	50	50	99
CA	2	5	19
N	0	2	28
C	0	6	14
O	4	14	25
CB	32	0	8
SG	13	23	6
B* 8 GLY	27	75	119
CA	22	35	38
N	5	3	18
C	0	20	28
O	0	17	35
B* 9 SER	32	107	189
CA	0	19	49
N	0	11	24
C	0	30	18
O	0	26	28
CB	6	17	57
OG	26	3	13
B*10 HIS	131	174	133
CA	2	39	28
N	0	32	9
C	0	16	32
O	2	11	39
CB	9	28	22
CG	3	16	0
ND1	8	7	0
CE1	62	4	0
NE2	15	0	0
CD2	30	20	3
B*11 LEU	3	169	401
CA	0	32	47
N	0	29	22
C	0	24	25
O	0	21	29
CB	3	28	41
CG	0	13	62
CD1	0	11	87
CD2	0	11	88
B*12 VAL	0	103	397
CA	0	17	47
N	0	20	29
C	0	17	28
O	0	14	36
CB	0	9	84
CG1	0	8	98
CG2	0	17	76
B*13 GLU	60	216	193
CA	0	35	36
N	0	29	17
C	0	26	18
O	0	24	27
CB	9	25	28
CG	9	17	41
CD	3	13	6
OE1	7	30	17
OE2	38	17	2
B*14 ALA	9	107	200
CA	2	27	44
N	0	26	24
C	0	17	23
O	0	16	31
CB	8	21	77
B*15 LEU	0	182	422
CA	0	30	49
N	0	26	20
C	0	23	24
O	0	27	25
CB	0	24	54
CG	0	19	76
CD1	0	11	96
CD2	0	22	79
B*16 TYR	48	188	387
CA	0	30	35
N	0	21	24
C	0	15	15
O	1	14	29
CB	0	14	54
CG	0	1	39
CD1	1	4	39
CE1	17	12	30
CZ	3	22	20
CE2	0	29	38
CD2	1	10	53
OH	24	16	12
B*17 LEU	126	127	120
CA	5	25	14
N	0	25	15
C	1	14	8
O	20	4	5
CB	24	9	9
CG	0	21	8
CD1	46	16	25
CD2	30	13	35
B*18 VAL	39	147	159
CA	9	21	6
N	0	17	15
C	0	14	2
O	17	5	7
CB	0	32	27
CG1	3	43	35
CG2	9	16	66
B*19 CYS	11	96	179
CA	2	19	33
N	0	15	20
C	0	2	11
O	10	6	24
CB	0	22	43
SG	0	32	49
B*20 GLY	27	73	102
CA	27	27	33
N	0	22	20
C	0	11	17
O	0	13	32
B*21 GLU	141	24	79
CA	0	6	21
N	3	6	3
C	0	6	13
O	7	3	10
CB	28	0	0
CG	16	3	21
CD	3	0	0
OE1	43	0	3
OE2	40	0	0
B*22 ARG	117	275	134
CA	6	25	8
N	2	17	9
C	0	11	6
O	6	17	21
CB	25	30	5
CG	2	38	19
CD	22	36	14
NE	0	23	16
CZ	2	21	17
NH1	34	20	0
NH2	16	37	18
B*23 GLY	0	103	178
CA	0	46	63
N	0	24	25
C	0	18	44
O	0	15	46
B*24 PHE	0	205	532
CA	0	39	30
N	0	25	17
C	0	29	20
O	0	21	27
CB	0	28	44
CG	0	14	43
CD1	0	16	58
CE1	0	13	78
CZ	0	7	91
CE2	0	4	79
CD2	0	7	46
B*25 PHE	74	157	402
CA	0	19	44
N	0	16	22
C	0	15	31
O	1	12	50
CB	8	22	35
CG	5	13	35
CD1	1	26	19
CE1	17	14	36
CZ	23	0	40
CE2	17	9	48
CD2	1	12	42
B*26 TYR	9	109	662
CA	0	19	51
N	0	17	32
C	0	14	26
O	0	20	35
CB	0	16	58
CG	0	3	53
CD1	0	1	63
CE1	0	1	72
CZ	0	1	75
CE2	1	7	75
CD2	0	7	62
OH	7	1	58
B*27 THR	56	64	224
CA	0	11	41
N	1	11	30
C	0	0	34
O	13	0	25
CB	14	9	30
OG1	11	3	25
CG2	17	28	39
B*28 PRO	24	118	216
CA	2	9	33
N	0	6	20
C	2	7	7
O	14	7	10
CB	5	21	51
CD	2	30	47
CG	0	38	49
B*29 LYS	147	97	101
CA	8	19	16
N	0	17	18
C	0	3	16
O	11	12	22
CB	2	28	11
CG	27	16	9
CD	21	2	6
CE	46	0	2
NZ	33	0	0
B*30 ALA	118	61	71
CA	2	16	19
N	0	6	22
C	1	3	10
OEND	15	13	18
OEND'	34	5	0
CB	66	17	2

The insulin monomer consists of two polypeptide chains, A and B. The polypeptide chains of the second monomer unit of the dimer are distinguished by asterisks. See Table 4 for additional description.

Table 6

Contact information for lysozyme

1 LYS...	127 CYS 13	13 LYS 76	23 TYR 51	33 LYS 140	84 LEU 12	53 TYR...	57 GLN 17
3 PHE 110	10 ALA 6	19 ASN 74	105 MET 42	30 CYS 74	40 THR 11	51 THR 138	60 SER 10
86 SER 86	123 TRP 2	20 TYR 35	31 ALA 28	123 TRP 43	53 TYR 10	80 CYS 119	46 ASN 8
40 THR 71		16 GLY 29	28 TRP 24	110 ALA 29	44 ASN 7	84 LEU 74	53 TYR 4
7 GLU 47	10 ALA...	17 LEU 19	30 CYS 24	31 ALA 26	52 ASP 3	66 ASP 68	
39 ASN 32	13 LYS 52	23 TYR 12	25 LEU 20	115 CYS 12		57 GLN 67	60 SER...
2 VAL 31	129 LEU 49	24 SER 9	29 VAL 11	32 ALA 9	43 THR...	50 ILE 66	69 THR 63
41 GLN 27	7 GLU 40	26 TRP 5	115 CYS 11	36 SER 7	44 ASN 57	83 LEU 58	64 CYS 63
38 PHE 9	14 ARG 33	12 MET 1	116 LYS 4	35 GLU 3	53 TYR 47	43 THR 56	51 THR 60
84 LEU 3	6 CYS 21			111 TRP 2	52 ASP 31	60 SER 23	59 ASN 42
	8 LEU 19	19 ASN...	28 TRP...	37 ASN 2	84 LEU 30	42 ALA 17	66 ASP 31
2 VAL...	9 ALA 17	22 GLY 67	17 LEU 198		41 GLN 29	64 CYS 16	58 SER 23
39 ASN 68	12 MET 12	18 ASP 63	105 MET 140	35 GLU...	51 THR 22	81 SER 16	53 TYR 23
38 PHE 63	11 ALA 8	23 TYR 62	12 MET 99	57 GLN 96	42 ALA 22	52 ASP 10	72 SER 23
1 LYS 59	126 ARG 6	24 SER 44	25 LEU 80	110 ALA 82	57 GLN 16	68 ARG 9	80 CYS 21
3 PHE 16	25 LEU 2	17 LEU 18	23 TYR 80	31 ALA 60		56 LEU 2	62 TRP 21
37 ASN 11		21 ARG 12	20 TYR 57	108 TRP 50	44 ASN...		63 TRP 7
4 GLY 4	11 ALA...	25 LEU 5	99 VAL 53	32 ALA 48	62 ASP 91	54 GLY...	65 ASN 6
	15 HIS 55	20 TYR 2	31 ALA 41	56 LEU 28	51 THR 48	57 GLN 66	58 ILE 5
3 PHE...	8 LEU 51	28 TRP 1	27 ASN 35	109 VAL 23	57 GLN 43	84 LEU 32	61 ARG 5
38 PHE 129	14 ARG 44		56 LEU 35	36 SER 21	46 ASN 28	42 ALA 31	73 ARG 4
8 LEU 126	80 ILE 36	20 TYR...	24 SER 31	33 LYS 21	43 THR 24	40 THR 30	
1 LYS 115	7 GLU 32	96 LYS 172	108 TRP 30	34 PHE 18	45 ARG 9	83 LEU 25	61 ARG...
7 GLU 88	10 ALA 26	21 ARG 82	95 ALA 26	55 ILE 12	42 ALA 7	56 LEU 21	62 TRP 161
88 ILE 45	3 PHE 21	17 LEU 67	32 ALA 24	37 ASN 7	35 GLU 5	36 SER 20	71 GLY 94
40 THR 36	13 LYS 16	23 TYR 60	29 VAL 19	44 ASN 3	50 SER 1	53 TYR 14	59 ASN 50
11 ALA 33	9 ALA 14	28 TRP 54	26 GLY 15	30 CYS 1		39 ASN 11	69 THR 46
59 ILE 30	12 MET 12	99 VAL 51	30 CYS 8		45 ARG...	55 ILE 10	72 SER 39
2 VAL 28		16 GLY 43	18 ASP 4	36 SER...	68 ARG 124	41 GLN 5	48 ASP 36
86 SER 27	12 MET...	18 ASP 30	96 LYS 3	55 ILE 75	51 THR 78	91 SER 3	73 ARG 30
5 ARG 2	17 LEU 127	100 SER 26	21 ARG 1	39 ASN 72	49 GLY 59		50 SER 27
	8 LEU 92	22 GLY 24		32 ALA 52	58 SER 30	55 ILE...	60 SER 14
4 GLY...	28 TRP 70	19 ASN 6	29 VAL...	37 ASN 40	46 ASN 30	44 THR 92	63 TRP 1
7 GLU 77	88 ILE 47		123 TRP 98	42 ALA 38	44 ASN 27	36 SER 84	
8 LEU 36	15 HIS 44	21 ARG...	25 LEU 50	57 GLN 31		32 ALA 67	62 TRP...
38 PHE 35	9 ALA 38	100 SER 120	26 GLY 49	38 PHE 31	46 ASN...	56 LEU 62	61 ARG 154
3 PHE 30	25 LEU 32	20 TYR 73	38 PHE 42	35 GLU 24	50 SER 72	8 LEU 49	73 ARG 135
6 CYS 19	92 VAL 26	23 TYR 63	8 LEU 38	33 LYS 19	45 ARG 44	39 ASN 44	63 TRP 103
2 VAL 12	29 VAL 22	99 VAL 60	9 ALA 35	54 GLY 14	48 ASP 43	88 ILE 39	75 LEU 81
5 ARG 2	56 LEU 22	19 ASN 22	124 ILE 33	34 PHE 7	52 ASP 39	38 PHE 20	59 ASN 56
	10 ALA 20	132 GLY 2	32 ALA 32		44 ASN 30	3 PHE 21	60 SER 23
5 ARG...	11 ALA 19		12 MET 31	37 ASN...	47 THR 20	91 SER 20	64 CYS 9
123 TRP 109	32 ALA 7	22 GLY...	33 LYS 29	33 LYS 106	49 GLY 10	57 GLN 16	74 ASN 7
38 PHE 83	14 ARG 7	19 ASN 71	5 ARG 27	38 PHE 47	51 THR 8	35 GLU 12	76 CYS 2
29 VAL 52	16 GLY 4	23 TYR 34	28 TRP 20	39 ASN 47	59 ASN 6	83 LEU 10	71 GLY 1
125 ARG 48	13 LYS 2	20 TYR 19	31 ALA 15	36 SER 19		92 VAL 2	72 SER 1
6 CYS 46	18 ASP 2	24 SER 11	27 ASN 11	32 ALA 17	47 THR...	12 MET 2	
122 ALA 30	55 ILE 1	21 ARG 3	30 CYS 2	2 VAL 15	46 ASN 24	53 TYR 1	63 TRP...
8 LEU 29				35 GLU 5	48 ASP 24		98 ILE 187
9 ALA 25	13 LYS...	23 TYR...	30 CYS...		49 GLY 24	56 LEU...	62 TRP 122
7 GLU 17	129 LEU 101	105 MET 129	123 TRP 88	38 PHE...		108 TRP 101	76 CYS 95
124 ILE 15	18 ASP 80	21 ARG 78	34 PHE 50	3 PHE 133	48 ASP...	55 ILE 88	59 ASN 62
127 CYS 14	25 LEU 67	20 TYR 66	33 LYS 47	33 LYS 125	50 SER 87	58 ILE 67	75 LEU 73
3 PHE 3	10 ALA 54	28 TRP 61	27 ASN 35	5 ARG 83	61 ARG 41	28 TRP 43	58 ILE 65
126 GLY 2	9 ALA 31	19 ASN 57	26 GLY 31	8 LEU 70	46 ASN 31	35 GLU 37	64 CYS 37
	16 GLY 30	27 ASN 56	29 VAL 22	2 VAL 64	47 THR 27	95 ALA 35	94 CYS 36
6 CYS...	15 HIS 26	111 TRP 55	120 VAL 20	55 ILE 61	69 THR 4	54 GLY 26	97 LYS 33
5 ARG 66	11 ALA 19	104 GLY 52	32 ALA 13	29 VAL 42	70 PRO 3	91 SER 25	74 ASN 25
9 ALA 47	12 MET 18	99 VAL 41	114 ARG 10	32 ALA 32		32 ALA 25	101 ASP 17
4 GLY 23	17 LEU 13	25 LEU 5	28 TRP 10	36 SER 32	49 GLY...	12 MET 22	78 ILE 10
10 ALA 22	14 ARG 4	106 ASN 4	31 ALA 5	4 GLY 31	45 ARG 47	57 GLN 19	60 SER 7
126 GLY 13		17 LEU 4	118 THR 1	123 TRP 16	69 THR 36	92 VAL 13	73 ARG 5
8 LEU 10	14 ARG...	24 SER 3		40 THR 10	46 ASN 22	31 ALA 11	77 ASN 2
127 CYS 6	15 HIS 66	18 ASP 3	31 ALA...	39 ASN 8	47 THR 16	53 TYR 9	83 LEU 2
125 ARG 6	11 ALA 53	103 ASN 1	35 GLU 55	1 LYS 8	48 ASP 15	17 LEU 5	
7 GLU 3	10 ALA 41		111 TRP 43	37 ASN 6	51 THR 14	36 SER 4	64 CYS...
129 LEU 2	13 LYS 24	24 SER...	28 TRP 37		68 ARG 11	105 MET 3	60 SER 63
123 TRP 2	16 GLY 16	27 ASN 101	27 ASN 33	39 ASN...	70 PRO 9		74 ASN 48
	12 MET 10	19 ASN 47	115 CYS 27	2 VAL 108	50 SER 6	57 GLN...	78 ILE 48
7 GLU...		26 GLY 38	34 PHE 23	42 ALA 75		52 ASP 109	83 LEU 35
4 GLY 90	15 HIS...	28 TRP 30	56 LEU 23	36 SER 71	50 SER...	35 GLU 86	63 TRP 30
3 PHE 87	14 ARG 85	120 VAL 26	105 MET 17	41 GLN 60	48 ASP 78	53 TYR 68	59 ASN 23
1 LYS 49	92 VAL 75	18 ASP 16	33 LYS 16	55 ILE 43	46 ASN 68	54 GLY 57	80 CYS 19
10 ALA 44	89 THR 69	22 GLY 12	29 VAL 15	37 ASN 35	69 THR 54	44 ASN 52	53 TYR 13
11 ALA 20	11 ALA 60	121 GLN 12	30 CYS 14	1 LYS 34	59 ASN 51	42 ALA 26	79 PRO 11
6 CYS 27	88 ILE 55	23 TYR 9	119 ALA 11	40 THR 14	60 SER 28	36 SER 25	58 ILE 9
5 ARG 28	12 MET 53	124 ILE 6	108 TRP 8	54 GLY 5	51 THR 25	58 ILE 14	66 ASP 7
9 ALA 14	13 LYS 31	25 LEU 6	32 ALA 6	38 PHE 5	45 ARG 20	108 TRP 12	62 TRP 5
8 LEU 7	96 LYS 27		26 GLY 1		61 ARG 19	43 THR 12	73 ARG 4
	17 LEU 23	25 LEU...		40 THR...	52 ASP 1	59 ASN 12	65 ASN 3
8 LEU...	87 ASP 7	18 ASP 113	32 ALA...	55 ILE 87		55 ILE 9	94 CYS 2
3 PHE 120	93 ASN 3	124 ILE 85	55 ILE 64	1 LYS 84	51 THR...	56 LEU 7	72 SER 1
12 MET 89		28 TRP 85	56 LEU 44	84 LEU 79	53 TYR 108		
11 ALA 55	16 GLY...	9 ALA 51	35 GLU 40	54 GLY 34	45 ARG 62	58 ILE...	65 ASN...
55 ILE 47	18 ASP 39	29 VAL 49	36 SER 38	39 ASN 32	60 SER 57	83 LEU 87	74 ASN 64
38 PHE 47	20 TYR 32	13 LYS 47	38 PHE 36	3 PHE 27	44 ASN 44	53 TYR 70	79 PRO 59
29 VAL 46	13 LYS 30	24 SER 36	29 VAL 32	86 SER 26	60 ARG 43	98 ILE 62	80 CYS 59
4 GLY 41	96 LYS 30	129 LEU 28	28 TRP 30	85 SER 26	50 SER 27	56 LEU 60	67 GLY 53
5 ARG 37	14 ARG 19	12 MET 22	31 ALA 22	88 ILE 24	49 GLY 27	63 TRP 55	72 SER 34
32 ALA 25	12 MET 16	27 ASN 14	8 LEU 17	83 LEU 22	59 ASN 25	94 CYS 47	66 ASP 26
7 GLU 22	17 LEU 13	23 TYR 10	30 CYS 16	42 ALA 21	52 ASP 20	91 SER 36	78 ILE 22
88 ILE 19	15 HIS 7	17 LEU 9	37 ASN 11	41 GLN 16	46 ASN 17	52 ASP 22	64 CYS 12
6 CYS 16		19 ASN 4	34 PHE 7	38 PHE 13	66 ASP 16	64 CYS 18	60 SER 5
10 ALA 14	17 LEU...		12 MET 6		43 THR 11	95 ALA 16	
9 ALA 3	28 TRP 147	26 GLY...	33 LYS 3	41 GLN...	69 THR 8	108 TRP 9	66 ASP...
28 TRP 2	12 MET 135	120 VAL 63		84 LEU 126	58 ILE 4	59 ASN 8	68 ARG 88
	20 TYR 77	24 SER 41	33 LYS...	39 ASN 66		51 THR 7	53 TYR 59
9 ALA...	92 VAL 75	29 VAL 38	38 PHE 115	1 LYS 35	52 ASP...	57 GLN 2	80 CYS 54
25 LEU 55	96 LYS 38	124 ILE 34	37 ASN 109	43 THR 31	57 GLN 120	60 SER 2	69 THR 47
29 VAL 41	15 HIS 21	30 CYS 31	123 TRP 82	40 THR 24	44 ASN 107		60 SER 46
129 LEU 39	13 LYS 12	123 TRP 25	34 PHE 80	42 ALA 10	46 ASN 58	59 ASN...	65 ASN 38
6 CYS 37	56 LEU 11	25 LEU 23	30 CYS 58	54 GLY 9	59 ASN 51	83 TRP 67	51 THR 11
5 ARG 31	95 ALA 11	28 TRP 17	29 VAL 32		53 TYR 46	52 ASP 56	64 CYS 18
12 MET 30	19 ASN 9	121 GLN 6	35 GLU 20	42 ALA...	51 THR 40	50 SER 47	72 SER 7
124 ILE 24	25 LEU 8	27 ASN 6	32 ALA 17	39 ASN 68	58 ILE 27	62 TRP 45	79 PRO 2
13 LYS 21	23 TYR 8		31 ALA 15	57 GLN 38	43 THR 23	61 ARG 39	
8 LEU 20	18 ASP 2	27 ASN...	36 SER 12	43 THR 38	42 ALA 3	58 ILE 39	67 GLY...
7 GLU 19		120 VAL 108		54 GLY 36	50 SER 1	51 THR 38	65 ASN 38
11 ALA 14	18 ASP...	111 TRP 86	34 PHE...	41 GLN 22		64 CYS 26	68 ARG 28
	25 LEU 108	24 SER 85	114 ARG 205	36 SER 21			66 ASP 25

```
69 THR   9

68 ARG...
45 ARG 120
66 ASP  81
70 PRO  45
69 THR  43
51 THR  40
49 GLY  13
53 TYR  11

69 THR...
72 SER  72
66 ASP  72
70 PRO  71
60 SER  43
50 SER  42
49 GLY  36
68 ARG  33
61 ARG  25
71 GLY  20
51 THR   7
48 ASP   5
67 GLY   3

70 PRO...
72 SER  43
68 ARG  40
69 THR  19
49 GLY  11
48 ASP   2

71 GLY...
61 ARG  69
70 PRO  36
69 THR  28
73 ARG  27
72 SER   5

72 SER...
69 THR  64
61 ARG  48
70 PRO  47
65 ASN  36
60 SER  30
73 ARG  28
74 ASN   4
66 ASP   4

73 ARG...
62 TRP  86
75 LEU  68
71 GLY  55
61 ARG  43
72 SER  28
74 ASN  21
60 SER   3
64 CYS   3

74 ASN...
77 ASN 102
65 ASN  70
78 ILE  38
64 CYS  33
75 LEU  24
76 CYS  22
62 TRP  18
63 TRP  13
72 SER  11
73 ARG  10
79 PRO   3

75 LEU...
63 TRP  85
62 TRP  81
73 ARG  39
77 ASN  26
97 LYS  22
74 ASN  13
76 CYS   8

76 CYS...
73 TRP  96
78 ILE  67
74 ASN  42
77 ASN  35
97 LYS  32
94 CYS  12
75 LEU  11
98 ILE   4
62 TRP   2

77 ASN...
74 ASN 127
78 ILE  35
75 LEU  27
76 CYS   6
79 PRO   4
63 TRP   3

78 ILE...
79 PRO  94
76 CYS  61
74 ASN  56
64 CYS  52
82 ALA  51
90 ALA  32
83 LEU  24
65 ASN  19
94 CYS   9
63 TRP   9
80 CYS   8
77 ASN   3

79 PRO...
78 ILE  89
65 ASN  62
81 SER  57
82 ALA  51
64 CYS  15
74 ASN  10
83 LEU   5
80 CYS   5
77 ASN   5
66 ASP   2

80 CYS...
53 TYR 106
83 LEU  56
65 ASN  52
66 ASP  48
60 SER  22
78 ILE  16
82 ALA  16
64 CYS  15
94 LEU  13
79 PRO  10
81 SER   5

81 SER...
84 LEU  74
79 PRO  65
80 CYS  19
83 LEU  17
53 TYR  16
85 SER   9
82 ALA   7

82 ALA...
79 PRO  51
85 SER  43
90 ALA  39
81 SER  36
78 ILE  33
80 CYS  27
83 LEU  20
84 LEU  18
87 ASP   4

83 LEU...
91 SER  83
56 ILE  74
80 CYS  69
53 TYR  55
90 ALA  42
82 ALA  34
64 CYS  33
78 ILE  30
84 LEU  30
54 GLY  27
81 SER  26
85 SER  24
94 CYS  21
40 THR  13
79 PRO   7
55 ILE   4
87 ASP   3
63 TRP   2
88 ILE   1

84 LEU...
41 GLN 102
81 SER  93

40 THR  68
53 TYR  58
43 THR  35
54 GLY  32
83 LEU  19
82 ALA  17
80 CYS  17
42 ALA  11
86 SER   6
85 SER   5
 1 LYS   2

85 SER...
87 ASP  84
82 ALA  57
40 THR  24
83 LEU  19
84 LEU  14
81 SER  12
88 ILE   7
86 SER   5
90 ALA   1

86 SER...
 1 LYS  88
 3 PHE  30
87 ASP  26
85 SER  22
40 THR  22
88 ILE  13
84 LEU  12

87 ASP...
85 SER  97
89 THR  73
88 ILE  24
86 SER  16
90 ALA  10
15 HIS   5
82 ALA   3
83 LEU   1
91 SER   1

88 ILE...
92 VAL  96
91 SER  66
15 HIS  40
 3 PHE  39
55 ILE  33
 8 LEU  30
11 ALA  25
12 MET  24
90 ALA  18
40 THR  16
87 ASP  15
85 SER  13
86 SER   9
89 THR   8
83 LEU   3
17 LEU   2
93 ASN   1

89 THR...
87 ASP  84
93 ASN  77
15 HIS  74
92 VAL  34
91 SER  21
88 ILE  19
90 ALA  17

90 ALA...
93 ASN  53
83 LEU  49
89 THR  47
94 CYS  33
78 ILE  23
88 ILE  21
82 ALA  21
87 ASP  13
92 VAL  10
91 SER   8
85 SER   2

91 SER...
83 LEU  86
88 ILE  71
58 ILE  54
94 CYS  42
95 ALA  37
56 LEU  28
90 ALA  26

55 ILE  23
89 THR  17
93 ASN  14
92 VAL  10
54 GLY   6
87 ASP   1

92 VAL...
17 LEU  79
88 ILE  79
89 THR  62
95 ALA  59
96 LYS  57
15 HIS  55
91 SER  33
56 LEU  25
12 MET  21
94 CYS  13
90 ALA  12
93 ASN  11
55 ILE   3

93 ASN...
89 THR  96
96 LYS  71
90 ALA  54
92 VAL  38
91 SER  19
97 LYS  15
95 ALA  12
94 CYS   3
15 HIS   1

94 CYS...
98 ILE  49
91 SER  45
58 ILE  43
97 LYS  42
90 ALA  41
96 LYS  24
83 LEU  21
63 TRP  20
93 ASN  19
92 VAL  18
78 ILE   9
95 ALA   4
64 CYS   1
89 THR   1

95 ALA...
98 ILE  64
92 VAL  52
108 TRP  46
56 LEU  32
99 VAL  31
91 SER  29
58 ILE  23
28 TRP  21
94 CYS  20
97 LYS  14
93 ASN  10
17 LEU   9
96 LYS   8
90 ALA   1

96 LYS...
20 TYR 152
93 ASN  94
92 VAL  82
99 VAL  59
17 LEU  43
100 SER  28
94 CYS  28
15 HIS  25
97 LYS  24
16 GLY  20
95 ALA  20
98 ILE  12
28 TRP   6

97 LYS...
94 CYS  53
100 SER  41
76 CYS  35
63 TRP  34
75 LEU  28
101 ASP  27
96 LYS  26
93 ASN  26
95 ALA  16
99 VAL  15
98 ILE  14

98 ILE...
63 TRP 132
108 TRP  77
95 ALA  62
94 CYS  50
58 ILE  36
107 ALA  31
101 ASP  29
100 SER  20
97 LYS  19
104 GLY  14
96 LYS  10
99 VAL   6
76 CYS   5
102 GLY   4
103 ASN   3

99 VAL...
21 ARG  64
104 GLY  63
28 TRP  51
96 LYS  47
95 ALA  44
20 TYR  44
108 TRP  37
98 ILE  26
23 TYR  24
100 SER  23
105 MET  21
101 ASP  14
97 LYS   8
102 GLY   7

100 SER...
21 ARG 118
97 LYS  40
96 LYS  39
99 VAL  33
102 GLY  26
20 TYR  23
98 ILE  18
101 ASP   7

101 ASP...
97 LYS  37
98 ILE  33
63 TRP  27
100 SER  14
103 ASN  10
99 VAL   5

102 GLY...
103 ASN  35
101 ASP  31
104 GLY  22
100 SER  20
98 ILE   7
99 VAL   5
21 ARG   3

103 ASN...
106 ASN  70
107 ALA  40
101 ASP   9
98 ILE   8
105 MET   7
112 ARG   2
23 TYR   1

104 GLY...
99 VAL  60
23 TYR  44
106 ASN  23
107 ALA  21
102 GLY  16
103 ASN  15
105 MET  14
98 ILE  12
108 TRP  12

105 MET...
111 TRP 177
108 TRP 131
23 TYR  99
28 TRP  78
99 VAL  33
27 ASN  29
107 ALA  22
31 ALA  17
103 ASN  14
106 ASN   8

112 ARG   4
56 LEU   3

106 ASN...
112 ARG 147
103 ASN  80
111 TRP  69
116 LYS  57
108 TRP  36
104 GLY  19
109 VAL  10
105 MET   6
107 ALA   6
23 TYR   5

107 ALA...
108 TRP  48
103 ASN  45
98 ILE  32
104 GLY  25
105 MET  25
109 VAL  17
112 ARG  14
106 ASN   7

108 TRP...
105 MET 167
56 LEU 107
98 ILE  95
35 GLU  59
106 ASN  45
95 ALA  45
111 TRP  44
28 TRP  36
107 ALA  34
110 ALA  32
99 VAL  27
31 ALA  19
112 ARG  19
109 VAL  14
58 ILE  13
104 GLY  12
57 GLN  10

109 VAL...
113 ASN  69
112 ARG  58
35 GLU  21
110 ALA  18
106 ASN  16
107 ALA  12
111 TRP   8
108 TRP   6

110 ALA...
114 ARG  68
109 VAL  59
35 GLU  47
108 TRP  43
113 ASN  33
34 PHE  21
112 ARG  16
115 CYS  10
31 ALA   9
111 TRP   7

111 TRP...
105 MET 217
116 LYS 174
115 CYS 129
27 ASN  92
106 ASN  59
23 TYR  54
31 ALA  54
108 TRP  51
110 ALA  30
112 ARG  16
114 ARG  16
109 VAL  11
113 ASN  11
120 VAL   4
34 PHE   2

112 ARG...
106 ASN 164
116 LYS  96
109 VAL  70
111 TRP  56
113 ASN  36
108 TRP  22
110 ALA  21
114 ARG  16

107 ALA   9
105 MET   9
115 CYS   3
103 ASN   2

113 ASN...
109 VAL  80
112 ARG  59
114 ARG  44
110 ALA  37
111 TRP  12
115 CYS   5

114 ARG...
 34 PHE 170
110 ALA  82
113 ASN  41
118 THR  35
115 CYS  22
111 TRP  22
112 ARG  17
 30 CYS   7
116 LYS   4

115 CYS...
111 TRP 112
118 THR  82
120 VAL  49
114 ARG  19
117 GLY  15
 31 ALA  15
 27 ASN  12
110 ALA  12
 34 PHE  10
 30 CYS  10
112 ARG   4
116 LYS   2
113 ASN   2

116 LYS...
111 TRP 145
112 ARG 117
106 ASN  61
118 THR  28
115 CYS  17
 27 ASN   3
120 VAL   2

117 GLY...
118 THR  38
116 LYS  23
115 CYS  23

118 THR...
115 CYS  72
120 VAL  61
114 ARG  25
119 ASP  18
116 LYS  17
 30 CYS   3
123 TRP   1

119 ASP...
125 ARG  69
121 GLN  58
122 ALA  48
118 THR  36
120 VAL  14
123 TRP   4

125 ARG...
121 GLN  58
122 ALA  48
124 ILE  36
128 ARG  21
 9 ALA  19
125 ARG  19
 6 CYS  16
 5 ARG  13
123 TRP   1

129 LEU...
 13 LYS 110
127 CYS  92
124 ILE  47
 25 LEU  43
  9 ALA  35
 10 ALA  27
128 ARG   7
  6 CYS   6

 24 SER  11
 26 GLY   8
122 ALA   5

122 ALA...
125 ARG 174
119 ASP  45
123 TRP  31
120 VAL  19
121 GLN  19
  5 ARG  18
124 ILE  14

123 TRP...
  5 ARG 133
 30 CYS 128
 33 LYS  82
120 VAL  77
 29 VAL  68
 34 PHE  40
122 ALA  39
124 ILE  39
 26 GLY  25
121 GLN  23
125 ARG  17
 38 PHE  16
127 CYS   5
119 ASP   3
  9 ALA   1

124 ILE...
121 GLN  86
129 LEU  75
127 CYS  73
 25 LEU  51
 26 GLY  39
123 TRP  28
  9 ALA  27
  5 ARG  26
 29 VAL  22
122 ALA  14
125 ARG  12
120 VAL   7
126 GLY   7
 24 SER   5

125 ARG...
122 ALA 190
121 GLN  86
119 ASP  64
  5 ARG  57
127 CYS  26
124 ILE  21
  6 CYS   7
123 TRP   3

126 GLY...
  6 CYS  29
128 ARG  19
124 ILE  18
125 ARG  17
127 CYS   8
  5 ARG   5

127 CYS...
129 LEU  86
124 ILE  74
128 ARG  21
  9 ALA  19
125 ARG  19
  6 CYS  16
  5 ARG  13
123 TRP   1

128 ARG...
129 LEU  45
126 GLY  31
127 CYS  24
 10 ALA   7

129 LEU...
 13 LYS 110
127 CYS  92
124 ILE  47
 25 LEU  43
  9 ALA  35
 10 ALA  27
128 ARG   7
  6 CYS   6
```

The first line of each block gives the residue number and name of the central residue. The following lines list in descending order of significance all residues containing long atoms that occlude surface of the central residue. The values are residue number, residue name, and surface area (Å^2) occluded on the central residue. Because only long atoms are considered, the area occluded by a residue adjacent to a central residue represents contacts of only side chain atoms of the adjacent residue.

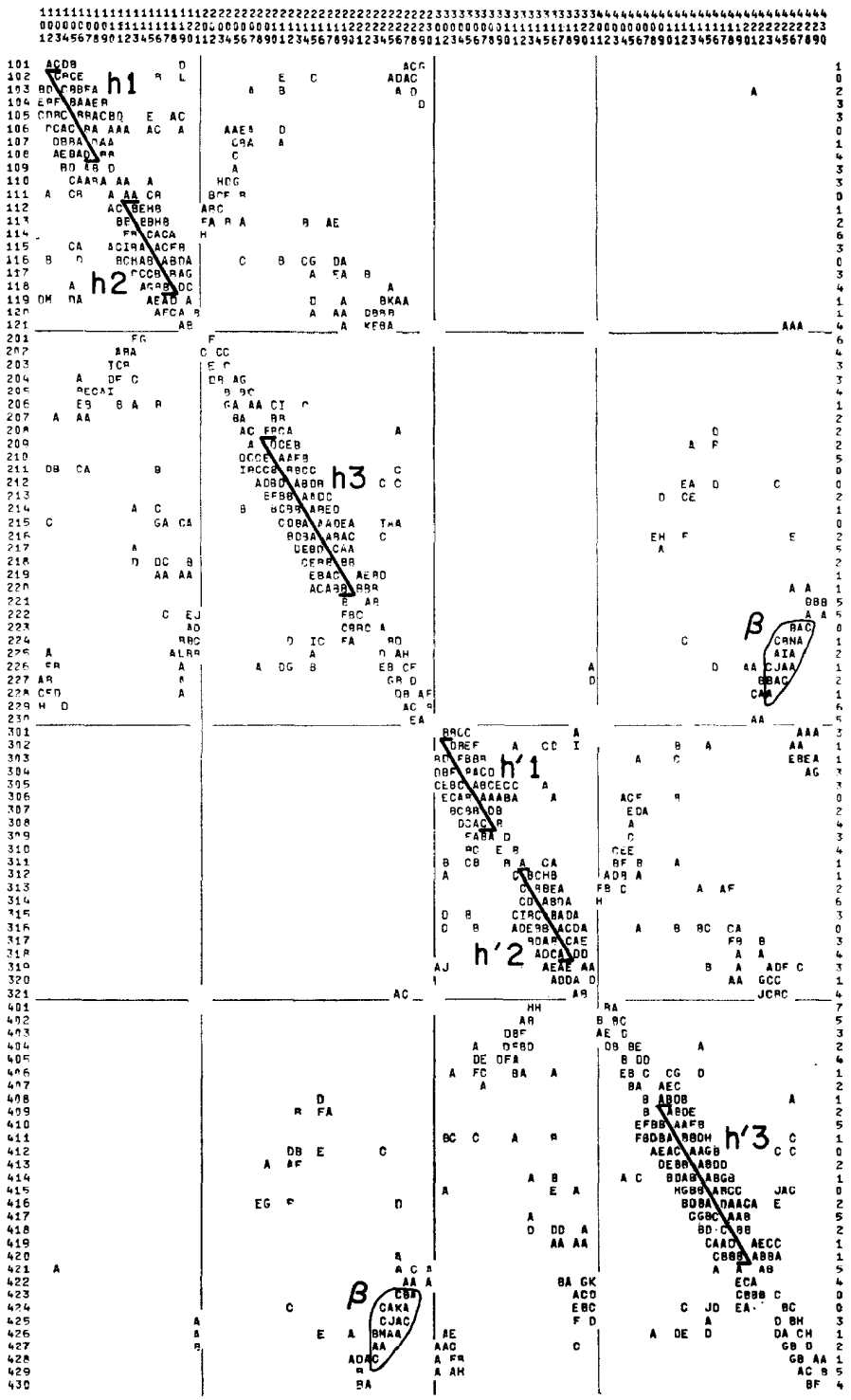

FIG. 4. *See following page for legend.*

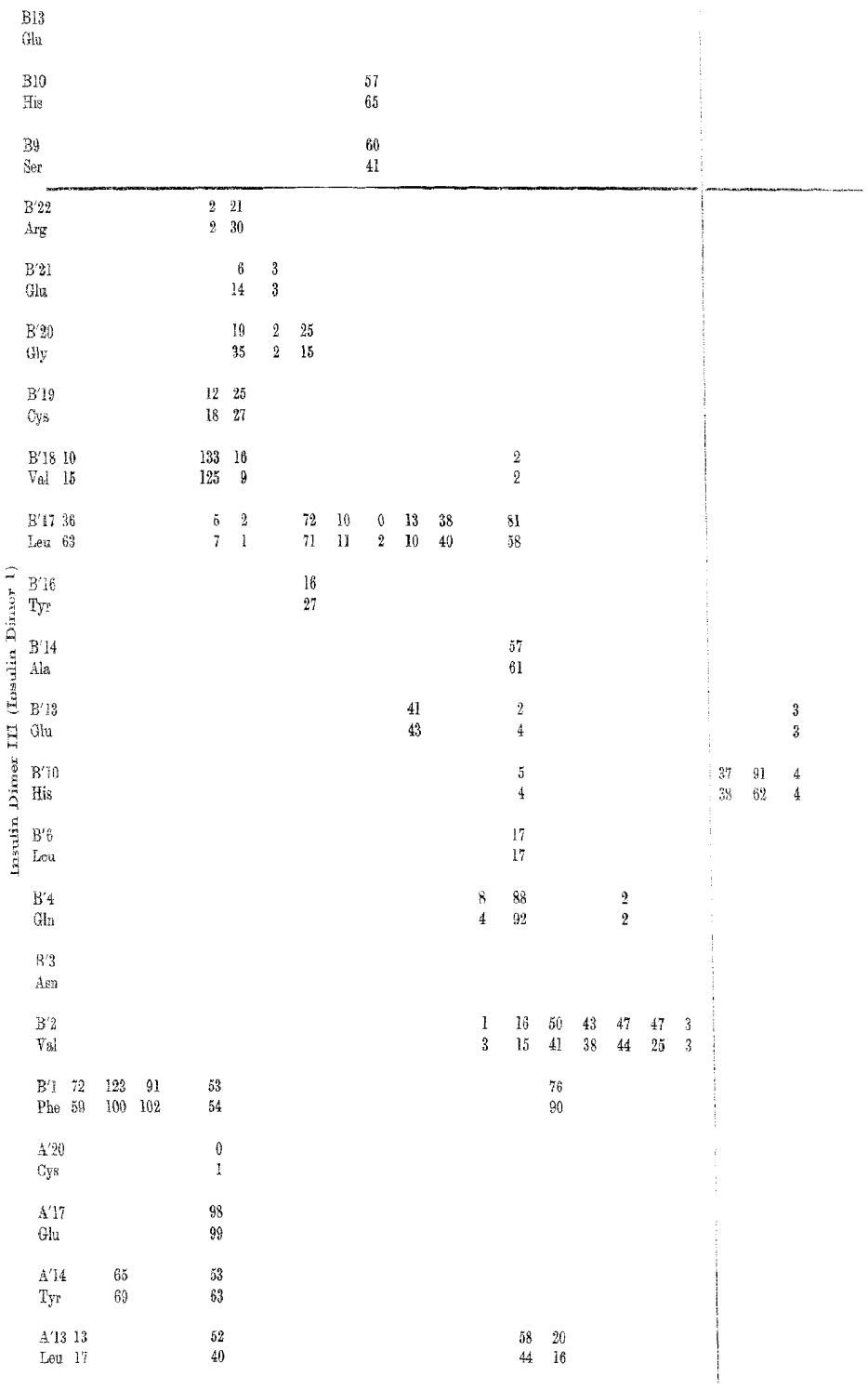

FIG. 5. Contacts at the dimer–dimer interfaces of insulin. The upper number of each pair gives the area (Å^2) of the abscissa residue occluded by the ordinate residue; the lower number gives the converse. The three dimer units of the hexamer are designated I, II and III according to right-hand rotation about the z-axis.

[facing p. 364

hexasaccharide of *N*-acetylglucosamine in the lysozyme cleft (Fig. 2(a)) and incorporation of a lysozyme molecule into the crystal lattice (Fig. 2(b)). Figure 3 shows the changes in exposed area for formation of the insulin dimer from two monomers (Fig. 3(a)), for incorporation of insulin dimer into the hexamer (Fig. 3(b)) and for incorporation of hexamer into the crystal lattice (Fig. 3(c)). The graphs of Figure 1 serve as a basis for evaluating the extent of the changes shown in Figures 2 and 3.

(c) *Contact information*

Table 6 gives the extent of contact of each residue of lysozyme with its neighbors. Tabulations of this kind give an objective description of the environment of atoms or residues within the native structure. The contact values of Table 6 are constructed from long atom information only. This seems to be the most appropriate description of the special environment of a central atom in the folded state (near atoms are present in both the folded and unfolded states).

Contact information can also be displayed graphically. Figure 4 plots data for the insulin dimer corresponding to those listed for lysozyme in Table 6. Ooi and coworkers (Nishikawa *et al.*, 1972) have used similar plots of α-carbon distances to display structures of proteins. Ooi pointed out that elements of the matrix near the diagonal reflect secondary structure and off-diagonal elements represent tertiary structure. Helical regions show as four-residue thick ribbons on either side of the diagonal. The helices of insulin are largely irregular, and this is reflected as irregularity in the patterns of Figure 4. Contacts between monomer units of the dimer are given in the upper right and lower left quadrants. The anti-parallel β-structure, involving residues B23 to B28 and B*23 to B*28, developed in the association of insulin monomers to the dimer, appears as a ribbon of unit slope orthogonal to the diagonal in both the lower left and upper right quadrants; these ribbons are encircled in Figure 4. Within each monomer the chain from residues A6 to A13 runs anti-parallel to that from B1 to B6 to give irregular β-structure, which is also seen as off-diagonal ribbons of unit slope.

Figures 5 and 6 give contacts that develop through incorporation of the insulin dimer into hexamer and through binding of hexasaccharide to lysozyme. Figure 5 describes the two different dimer–dimer interfaces of insulin. The dimer I–dimer II and dimer III–dimer I contacts are related by symmetry and therefore both may be plotted in Figure 5. The pseudo 2-fold symmetry axis of the dimer is reflected in the symmetry about the diagonal of Figure 5.

As one expects, the areas that two atoms occlude on each other are approximately the same. Thus, the pairs of numbers in Figures 5 and 6 are comparable and Figure 4 has approximate symmetry about the diagonal.

FIG. 4. Ooi plot for the insulin dimer of contacts between residues. Letter symbols give the extent that an abscissa residue occludes area of an ordinate residue. Each increase in alphabet stands for 15 Å^2 of occluded surface. The right-hand column gives the surface area (Å^2) exposed to solvent times 1/30. Only long atom contacts are included in the sums; thus, the diagonal is blank. The following gives the correspondence between the standard residue designations of Table 5 and those used in this Figure: A1 to A21 = 101 to 121; B1 to B30 = 201 to 230; A*1 to A*21 = 301 to 321; B*1 to B*30 = 401 to 430. Regions of the plot corresponding to each chain are delineated. Contacts within monomer units are given in the upper left and lower right quadrants. Contacts between the monomers are shown in the upper right and lower left quadrants. Lines along the diagonal indicate α-helical sections. The two enclosed regions off the diagonal represent the β-structure at the monomer–monomer interface.

	33 Lys	34 Phe	35 Glu	36 Ser	37 Asn	42 Ala	43 Thr	44 Asn	46 Asn	52 Asp	56 Leu	57 Gln	58 Ile	59 Asn	62 Trp	63 Trp	75 Leu	98 Ile	101 Asp	102 Gly	103 Asn	107 Ala	108 Trp	109 Val	110 Ala	112 Arg	114 Arg
NAG C′			36 43					3 2	35 35	69 98	1 2	48 74	44 37	82 97	0 3	53 51		54 34				62 88	120 85	92 122			
NAG A															20 18	10 9	49 27		143 185	38 46	50 58						
NAG B															112 93	81 70		18 19	50 59		37 57	11 20				3 3	
NAG C											5 8	21 22	68 46	56 96	68 89	68 78		52 56				90 134	95 58				
NAG D			31 32						110 118	93 120	8 7	76 85	5 3	53 63								9 17	50 39	57 63			
NAG E		3 3	100 139	24 13	2 0	28 19	12 8	139 132	5 8	33 30		83 93												75 95	29 32		
NAG F	5 4	64 94	34 26	16 7	24 14																				3 0		74 99

Lysozyme

FIG. 6. Contacts of the *N*-acetylglucosamine hexasaccharide and the α-anomer of *N*-acetylglucosamine with residues of lysozyme. The α-anomer binds "anomolously" with some contacts like those of the unit of the hexasaccharide bound at site C (Blake *et al.*, 1967). Contacts for hexasaccharide bound at sites A to F are given separately. The upper number of each pair gives the area (Å²) of the protein residue occluded by the saccharide unit; the lower number gives the converse. *NAG*, *N*-acetylglucosamine.

TABLE 7

Summations of areas exposed to solvent for lysozyme, insulin and several complexes

	Atoms included in summations								
	All	Polar	Charged	Non-polar	Backbone	Side chain	Polar side chain	Charged side chain	Non-polar side chain
A *Lysozyme*									
Unfolded	21,723	6175	2466	13,082	5840	15,884	3777	5141	6966
Native	6583	1811	1261	3511	1599	4984	1564	2302	1118
Hexasaccharide complex	5919	1586	1128	3205	1462	4457	1395	2162	900
Crystal lattice	4786	1261	944	2581	1157	3629	1064	1659	906
B *Lysozyme differences*									
Unfolded–native	15,140	4364	1205	9571	4241	10,900	2213	2839	5848
Native–hexasaccharide complex	664	225	133	306	137	527	169	140	218
Native–lattice	1797	550	317	930	442	1355	500	643	212
C *Insulin*									
Unfolded dimer	17,348	4178	1954	11,215	4507	12,841	2565	4212	6065
Monomers	7334	1642	1278	4459	1557	5777	1420	2508	1849
Dimer	6023	1345	1169	3510	1245	4778	1249	2053	1477
Dimer in hexamer	4585	1130	859	2595	1017	3568	1119	1648	801
Dimer in lattice	3057	766	506	1784	739	2317	791	978	549
D *Insulin differences*									
Unfolded dimer–native dimer	11,325	2833	785	7705	3262	8063	1316	2159	4588
Monomers–dimer	1311	297	109	949	312	999	171	455	372
Dimer–dimer in hexamer	1438	215	310	915	228	1210	130	405	676
Dimer in hexamer–dimer in lattice	1528	364	353	811	278	1251	328	670	252

The contact information focuses on the immediate protein environment of an atom or group and is complementary to the extent of exposure to solvent. Thus, Table 6 is complementary to Figure 1(a), Figure 4 to Figures 1(b) and 3(a), Figure 5 to Figure 3(b), and Figure 6 to Figure 2(a).

(d) *Summary tabulations*

Table 7 gives results for lysozyme and insulin summed over classes of atoms (all atoms; polar, charged, non-polar; backbone, side chain) and over types of side chains (polar, charged, non-polar). The side chain categories are specified as follows: charged, those containing groups that bear charge at any pH in the range 0 to 12; polar, those containing polar but no charged atoms; and non-polar, those containing only non-polar atoms, and tryptophan, methionine and cystine. The values presented in Table 7 are the areas exposed to solvent and in sections B and D changes in exposed area (areas are in $Å^2$).

4. Discussion

(a) *Comparison with results of Lee & Richards (1971)*

The van der Waals' radii used in these computations (Table 2) differ significantly from those of Lee & Richards (1971) in particular for side chain atoms for which Lee & Richards use the uniform value of 1·8 Å. The values of the static accessibility† calculated for lysozyme from column 2 of Table 4 and the surface areas of Table 2 are very close to the values for lysozyme listed by Lee & Richards. If areas rather than ratios of areas are considered, differences due to changes in radii become apparent. Nevertheless, general conclusions drawn from the computations remain essentially unaltered by the changes in radii. For example, Lee & Richards made the striking point that a large fraction of the total surface of globular proteins is comprised of non-polar atoms in the folded as well as in the unfolded state. The data of Table 7 (compare columns 1 and 4) confirm this conclusion; non-polar atoms constitute 0·53 and 0·60 of the lysozyme surface for the folded and unfolded molecules, respectively. Because in the present calculation the van der Waals' radii assigned to non-polar atoms are larger than those assigned to polar atoms, the above fractions are each about 0·1 greater than those determined by Lee & Richards. The salient point is that in spite of the crude model used in the computations, exposure values have semi-quantitative reliability and trends within self-consistent sets of results appear to be meaningful.

In explanation of the considerable non-polar surface in the folded state, examination of Table 7 (compare columns 4 and 9) shows that a relatively high proportion (approx. two-thirds) of the non-polar surface of the folded molecules is associated with non-polar atoms that are part of polar or charged side chains, e.g. the methylene carbons of lysine. The extent to which the surface exposed in the unfolded state becomes buried on folding is two to three times greater for non-polar residues than for polar residues (charged and uncharged). This observation is consistent with the "oil-drop" model of protein folding.

Cavities within the lysozyme structure located by the graphics display of Lee & Richards (1971) do not exist according to the present calculations. This reflects the

† Defined by Lee & Richards (1971) as 100 × area of solvated sphere exposed to solvent/total area of solvated sphere.

change in van der Waals' radii. Computations with the solvent radius reduced to 1·0 Å and to 0·4 Å show the cavities. These remarks do not weaken the important conclusion of Lee & Richards—the density of packing of side chains within a native protein is not uniform.

(b) *Exposure of main chain carbonyl oxygen atoms in helical regions*

The carbonyl oxygen is the major contributor to the exposed surface of the polypeptide backbone (Table 1). The computations for insulin and lysozyme show that about half of the carbonyl oxygens are exposed to solvent in both helical and non-helical regions. Thus, it is of interest that the first residue of each of the eleven helices in these two molecules has the carbonyl oxygen entirely buried and that in ten of the eleven helices the carbonyl oxygen of the last helical residue is exposed to solvent. This observation is confirmed by examination of seven other proteins for which computations have been carried out and assignments of helical regions could be made. Although this correlation cannot be used to predict secondary structure because the exposure of the carbonyl oxygen depends on long range interactions, the origin of the effect presumably is related to the stability of helical regions in macromolecules and ultimately should be explicable in terms of theories of protein folding.

(c) *Ranking of residues according to exposure of backbone or side chain atoms to solvent*

In lysozyme, glycine is among the residues having exposure least affected by folding. The relatively low effect of folding on glycine exposure was confirmed through calculations on nine other proteins (glycine, serine and glutamic acid are the residues with the highest probability for exposure of backbone atoms). The relatively high exposure of glycine residues in the folded state accords with the expectation that glycine lies at bends of the chain (Venkatachalam, 1968) and, thus, on the surface of the molecule. Lee & Richards (1971) noted that proline is more exposed than expected from the non-polar character of its side chain. The present calculations confirm this behavior, which probably also reflects the participation of proline in bends in the chain.

(d) *Comparison of folding, binding, association and crystallization reactions*

Figures 2 and 3 and Table 7 show that lattice contacts are extensive, involving about 30% of the surface of lysozyme and insulin. The types of residues participating in lattice contacts of lysozyme and insulin are not the same as those that on the average constitute the surface of the free molecule or that are involved in the contacts between saccharides and lysozyme or that constitute the monomer–monomer and dimer–dimer interfaces of the insulin polymers. A significantly smaller participation of non-polar side chains in the lattice contacts and a proportionately greater involvement of charged side chains are found (Table 7).

(e) *Protein environment of atoms and residues*

The description of the environment of an atom or residue through listing atoms or residues that occlude its surface includes all important interacting elements except for those involved in long range ionic interactions. A comparison of the tabulations for lysozyme and insulin with Kendrew-type models of these proteins shows that the listing of residues contacting other residues is substantially correct and that

hydrogen bonding or hydrophobic interactions are reflected in extensive overlap. Since solvated radii are used, the listing includes some less significant elements of the environment; thus, contacts between residues with areas less than 5 to 10 Å^2 are not important.

The list of contacting residues is altered substantially when the solvent radius is decreased, e.g. from 1·4 to 0·4 Å. However, the most important contacts remain in the list generated with the reduced solvent radius. Decreasing the solvent radius to 1·0 Å does not significantly change the nature and extent of residue–residue contacts.

The environment about each atom of native lysozyme and insulin is described in Tables 4 and 5 by giving the surface area occluded by polar and non-polar long atoms, with the intent of indicating the polarity of the environment. These numbers can be summed over groups of atoms. The carboxyl groups of Glu-35 and Asp-52 of lysozyme are both largely shielded from solvent; the ratios of exposure in the folded state to that in the unfolded state are 0·2 and 0·3, respectively. Glu-35 has a considerably less polar environment, however. The fraction of its occluded surface assigned to polar contacts is 0·1 compared with 0·7 for Asp-52. This difference in environment is reflected in the pK 6 to 6·5 found for Glu-35 and pK approx. 4 found for Asp-52 (Imoto *et al.*, 1972).

(f) *Structural assumptions*

It is assumed that computations based on the crystallographically defined co-ordinates are relevant to solution properties. Two aspects of this assumption should be discussed. First, uncertainty in the crystallographic results is generally estimated at 0·5 Å for protein molecules studied at 2 to 3 Å resolution. In order to investigate this difficulty, a random error was introduced into the Cartesian co-ordinates of each atom of lysozyme using a Gaussian probability distribution that gave an arithmetic average movement for each atom of 0·39 Å. Computations with this perturbed set of co-ordinates show no significant changes in contacts between residues, i.e. very few changes greater than 5 to 10 Å^2 in area of contact. Exposure of individual atoms to solvent is also only slightly affected by this co-ordinate perturbation; atoms that are completely buried according to the unperturbed calculations remain so and atoms exposed to solvent undergo changes in exposure of approximately 5 Å^2. The exposure summed over classes of atoms (as in Table 7) changes by less than 5%. Thus, the conclusions drawn from exposure and contact computations of the kind described here are not sensitive to considerable error in the co-ordinates.

Second, the conformation of a protein molecule in the crystal may differ from that in solution. This problem has been considered by many workers (see review by Rupley, 1969). The time-average conformation of a protein as reflected in equilibrium properties appears to be unaffected by crystallization. Surface side chains involved in lattice contacts can be expected to undergo perturbation if they are relatively unrestricted in solution.

X-ray studies of complexes of lysozyme with the $\beta(1\rightarrow4)$-linked monomer, dimer and trimer of N-acetylglucosamine have provided co-ordinate information for the saccharide moieties binding at sites A through C. The co-ordinates for the moieties binding at the remaining sites (for N-acetylglucosamine hexamer) are derived from model building. A few of the side chains of the protein residues that are involved in the binding of hexasaccharide (see Figs 2(a) and 6) also participate in lattice contacts (see Fig. 2(b)).

The co-ordinates of the free insulin monomer and dimer are assumed to be the same as those in the hexamer. In the hexamer the first few residues of the B and B′ chains are buried in the adjacent dimers. In the free dimer these residues are possibly folded back onto the surface. Thus, the actual areas exposed to solvent in the free dimer are presumably less than the computed values and the calculated changes in exposed area brought about by the association of dimers to form hexamer are only approximate.

(g) *Concluding remarks*

Exposure values for atoms can be summed in different ways, e.g. to describe exposure of chromophores or other side chain elements of a protein. Values of this kind based on the present calculations have been used (see review by Imoto *et al.*, 1972) for examining free energies of association reactions and for understanding perturbations of ionizable groups and perturbations of chromophores. Exposure and contact information define environment more precisely than terms such as "partially buried".

Lee & Richards (1971) have discussed the limitations in applying exposure computations that are based on the relatively crude model of hard-sphere atoms and on the equilibrium structure determined by X-ray diffraction. In particular, conclusions related to rate processes must be made with caution. Nevertheless, the use of exposure calculations is justified by the need to summarize structural information objectively and semi-quantitatively and by the advantages of concise tabulations and graphical display.

We are grateful to Professor D. C. Phillips and his colleagues for the hospitality they extended and the encouragement they offered in the early stages of this work. We are also indebted to Professor D. C. Hodgkin and her colleagues for giving us the co-ordinates of insulin and for discussions on this structure. This work was supported by the American Cancer Society, the National Institutes of Health and the University of Arizona Computer Center. One of us (A. S.) thanks the National Institutes of Health for support in the form of a postdoctoral fellowship from 1971 to 1973.

REFERENCES

Blake, C. C. F. (1967). *Proc. Roy. Soc., London (ser. B)*, **167**, 435–438.

Blake, C. C. F., Johnson, L. N., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1967). *Proc. Roy. Soc., London (ser. B)*, **167**, 378–388.

Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D. C., Mercola, D. A. & Vijayan, M. (1971). *Nature (London)*, **231**, 506–511.

Bondi, A. (1964). *J. Phys. Chem.* **68**, 441–451.

Diamond, R. (1966). *Acta Crystallog.* **21**, 253–266.

Dickerson, R. E. (1972). *Annu. Rev. Biochem.* **41**, 815–842.

Imoto, T., Johnson, L. N., North, A. C. T., Phillips, D. C. & Rupley, J. A. (1972). *The Enzymes*, **7**, 665–868.

Lee, B. & Richards, F. M. (1971). *J. Mol. Biol.* **55**, 379–400.

Nishikawa, K., Ooi, T., Isogai, Y. & Nobuhiko, S. (1972). *J. Phys. Soc. (Japan)*, **32**, 1331–1337.

Pauling, L. C. (1960). *The Nature of the Chemical Bond*, 3rd edn. Cornell University Press, Ithaca, New York.

Rupley, J. A. (1969). In *Structure and Stability of Biological Molecules* (Timasheff, S. N. & Fasman, G. D., eds), pp. 291–352, Marcel Dekker, New York.

Venkatachalam, C. M. (1968). *Biopolymers*, **6**, 1425–1436.