

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

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 (\AA^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 (Å ²)
All nitrogen: $\begin{array}{c} \\ -N- \\ \end{array}$, $\begin{array}{c} \\ -NH- \\ \end{array}$, $-NH_2$, $-NH_3^+$	1.5	106
All oxygen: $=O$, $-O-$, $-OH$	1.4	99
All sulfur: $-S-$, $-SH$	1.85	133
Non-aromatic carbon: $\begin{array}{c} \\ -CH- \\ \end{array}$, $\begin{array}{c} \\ -CH_2- \\ \end{array}$, $-CH_3$	2.0	145
Aromatic carbon: $\begin{array}{c} \\ =CH \\ \end{array}$, $\begin{array}{c} \\ =C- \\ \end{array}$	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 Å ²			
	Extents of contact‡		
	Dimer I	Dimer II	Dimer III
	(Å ²)	(Å ²)	(Å ²)
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 Å ²			
	Extents of contact‡		
	Dimer I	Dimer II	Dimer III
	(Å ²)	(Å ²)	(Å ²)
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.

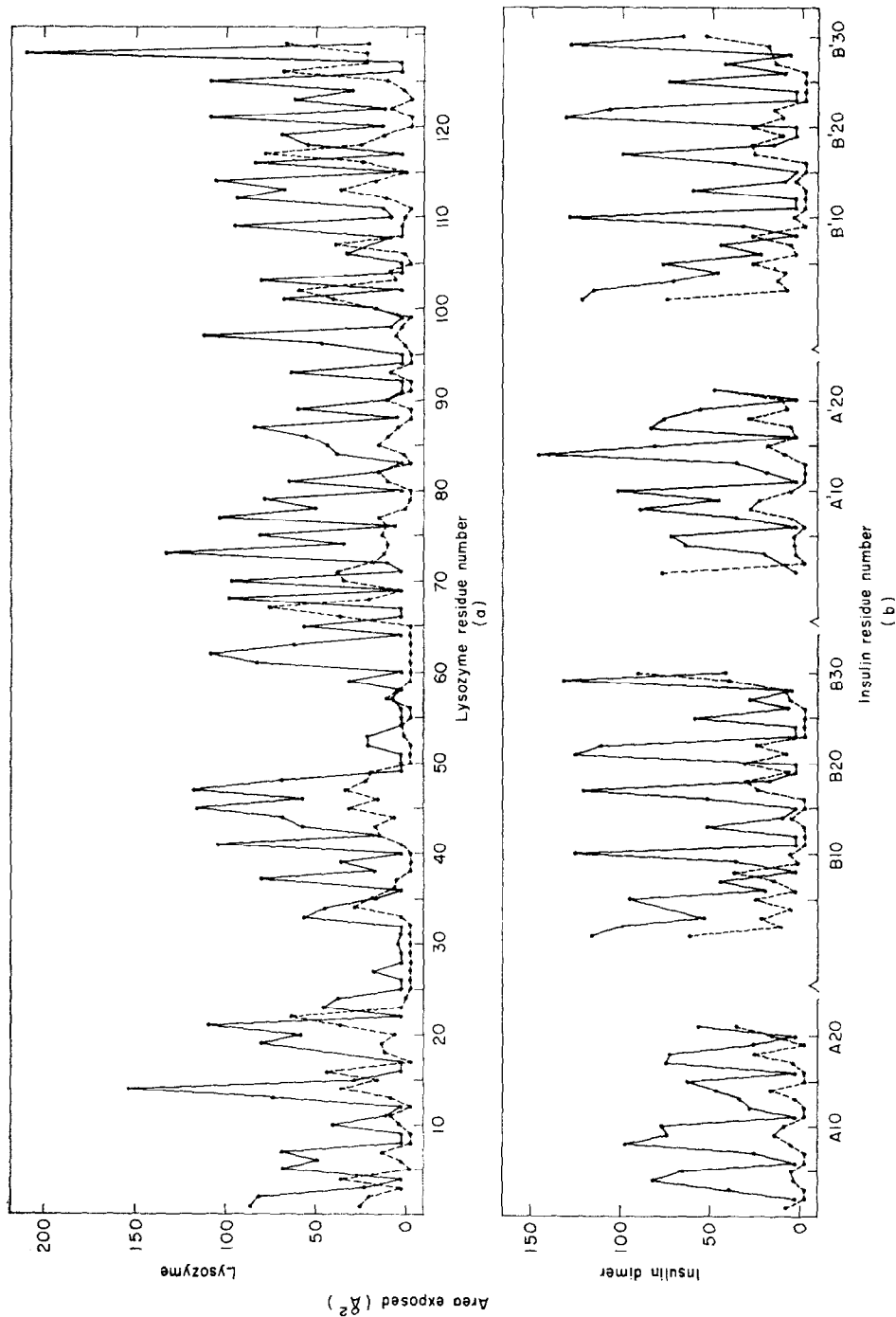


FIG. 1. Area exposed to solvent for backbone atoms (---●---) and side chain atoms (—●—). (a) Results for lysozyme; (b) results for insulin dimer.

Exposure and environment results for lysozyme

1	LYS	111	200	216		SD	0	17	76	N	0	23	7	CG	0	28	28	CR	27	22	0
	CA	21	30	25		CE	0	11	93	C	0	22	20	CG	0	36	33	CG	2	36	5
	NENQ	5	38	16						C	0	32	18	CE1	6	32	43	CD	24	38	0
	C	0	22	23		13 LYS	67	208	237	CG9	0	35	36	CZ	22	17	36	NE	2	29	10
	CB	0	32	27		CA	0	33	39	CG0	0	26	29	CE2	13	27	26	CZ	2	23	8
	CO	8	17	28		N	1	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		CR	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
						CE	35	30	21					O	15	4	20	O	1	21	9
						NZ	16	22	15	24 SER	38	166	136	CB	0	33	44	O	2	26	17
2	VAL	101	97	143						CA	0	52	14	CG	0	27	62	CB	28	32	2
	CA	0	16	32						N	1	26	16	CD	1	4	44	CG	0	10	0
	N	1	7	23		14 ARG	119	153	58	C	0	21	21	OE2	1	11	50	NOO2	26	18	7
	C	0	9	22		CA	1	27	13	O	0	21	29	OE1	14	0	35	NOD1	2	47	3
	O	10	3	16		N	0	22	16	C	0	15	7								
	CB	9	16	14		C	0	15	7	CR	35	32	22								
	CG2	38	17	22		O	25	7	0	OG	2	14	34	36 SER	7	206	195	47 THR	151	36	36
	PG1	35	28	14		CB	0	30	9					CA	6	44	36	CA	8	13	6
3	PHE	15	161	500		CG	32	13	9	25 LEU	3	178	377	N	0	26	18	N	0	2	14
	CA	0	22	39		CD	41	0	0	CA	0	32	38	C	0	22	24	C	0	13	6
	N	0	16	22		NE	7	7	0	N	0	29	20	O	1	17	41	O	26	9	0
	C	0	5	21		CZ	1	10	1	C	0	16	22	CR	0	60	46	CR	16	0	6
	O	2	4	40		NH2	37	20	1	O	0	16	32	OG	0	36	30	CG2	74	0	0
	CB	0	14	51		NH1	36	2	1	CR	2	24	41					OG1	28	0	4
	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			CA	13	30	8
	CZ	9	19	55		N	0	28	11					C	0	15	30	N	0	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	33	73	CB	24	14	9	O	11	11	13
4	GLY	35	79	132		CR	0	24	38	N	0	21	33	CG	0	5	0	CB	38	9	9
	CA	35	27	36		CG	0	13	36	C	0	24	36	NOD2	49	0	0	CG	1	6	11
	N	0	20	28		NO1	0	16	26	O	0	20	41	NOD1	6	18	21	OD2	12	13	28
	C	0	17	30		CE1	10	44	33									OD1	6	13	26
	O	0	15	39		NE2	3	24	16	27 ASN	18	163	332	CA	0	32	52	49 GLY	20	116	60
5	ARG	71	204	270		CD2	14	19	39	CA	0	22	20	N	0	24	16	CA	19	38	27
	CA	0	2	44		GLY	44	98	87	C	0	20	31	C	0	17	38	N	1	24	5
	N	1	8	17		CA	36	36	21	CB	0	16	37	O	0	20	45	C	0	18	15
	C	0	10	15		N	0	32	7	CG	0	31	60	CB	0	24	35	O	0	35	14
	O	0	15	32		C	0	16	28	CG	0	21	48	CG	0	22	38				
	CB	16	3	39		O	7	14	32	NOD2	11	15	44	CD1	0	16	50	50 SER	3	232	112
	CG	0	13	51		17 LEU	0	120	434	NOD1	6	9	48	CE1	0	14	68	CA	0	52	22
	CD	16	16	43		CA	0	19	33	28 TRP	0	299	680	CZ	0	13	78	N	0	23	15
	NE	10	13	2		N	0	16	15	CA	0	28	41	CE2	6	17	58	C	0	22	16
	CZ	0	40	6		C	0	19	24	N	0	20	18	CD2	12	20	33	O	0	27	19
	NH2	28	36	7		O	0	21	30	C	0	23	22	39 ASN	37	181	270	CR	2	66	33
	NH1	0	49	14		CB	0	19	60	C	0	19	37	CA	0	36	47	OG	1	42	6
6	CYS	52	97	103		CG	0	13	84	CB	0	22	51	N	0	24	26	51 THR	3	275	194
	CA	1	14	22		CD2	0	11	96	CG	0	23	48	C	0	25	23	CA	0	49	32
	N	1	9	23		CD1	0	6	92	CD1	0	42	46	O	0	27	29	N	0	36	10
	C	0	15	17		18 ASP	47	114	255	NE	0	33	29	CB	0	27	49	C	0	33	20
	O	0	16	26		CA	0	11	31	CG	0	25	38	CG	0	11	29	N	0	34	24
	CB	36	3	11		N	0	16	10	CE21	0	19	69	NOO2	30	5	37	CB	0	54	32
	SG	13	39	4		C	0	16	10	CH	0	10	84	NOD1	2	26	30	CG2	3	36	55
7	GLU	82	131	233		O	11	14	21	CZ2	0	10	81					OG1	0	33	19
	CA	11	24	32		CB	11	6	62	CE2	0	6	69	40 THR	0	257	234				
	N	1	17	22		CG	0	15	24	CD2	0	12	38	CA	0	47	32	52 ASP	22	276	209
	C	0	20	22		OD2	19	5	36	29 VAL	0	165	307	N	0	30	13	CA	0	49	36
	O	1	14	30		OD1	6	22	22	C	0	33	36	C	0	26	25	N	0	17	25
	CR	14	16	24		19 ASN	94	175	99	N	0	32	11	O	0	28	30	C	0	28	37
	CG	21	9	27		CA	0	39	17	C	0	17	23	CR	0	58	39	O	0	27	39
	CD	2	8	25		N	0	36	10	O	0	12	35	CG2	0	35	74	CR	0	49	30
	OE2	15	16	25		C	0	22	8	CB	0	21	55	OG1	0	32	20	CG	0	39	15
	OE1	17	7	28		O	14	16	12	CG2	0	19	77					OD2	21	32	13
8	LEU	0	164	421		CR	24	21	19	CG1	0	32	68	41 GLN	107	147	153	OD1	1	35	14
	CA	0	28	43		C	0	24	19	30 CYS	3	127	205	CA	0	28	33				
	N	0	29	24		CG	1	10	8	C	0	28	44	C	0	30	16	53 TYR	23	232	506
	C	0	18	22		NOD1	24	12	21	O	0	25	21	O	2	2	43	CA	0	25	44
	O	0	17	26						N	0	18	16	CB	11	36	14	C	0	21	24
	CB	0	21	57		23 TYR	63	204	412	C	0	14	27	CG	8	35	16	O	1	21	29
	CG	0	19	74		CA	5	38	19	O	0	24	52	CD	1	1	0	CR	0	22	47
	CD2	0	14	82		N	0	37	8	CB	0	24	52	NOE2	48	0	3	CG	0	17	45
	CD1	0	17	93		C	0	11	21	SG	3	17	45	NOE1	36	10	0	CD1	0	9	58
9	ALA	0	119	235		O	0	17	36									CE1	16	3	62
	CA	0	33	58		CR	0	32	36	31 ALA	0	143	186	42 ALA	30	139	126	CZ	1	12	58
	N	0	28	23		CG	0	7	46	CA	0	38	44	CA	16	32	25	CE2	0	33	46
	C	0	16	33		CD1	10	13	23	N	0	41	10	N	0	20	22	CD2	0	30	35
	O	0	11	43		CE1	23	3	32	C	0	20	24	O	0	17	21	OH	6	15	44
	CB	0	32	77		CZ	0	14	42	O	0	16	34	O	0	24	30				
						CE2	0	14	58	CB	0	28	73	CB	14	51	28	54 GLY	2	143	126
10	ALA	44	111	147		CD2	0	3	68	32 ALA	0	167	199	43 THR	75	119	135				
	CA	3	22	43		OH	25	14	24	CA	0	44	52	CA	0	27	32	C	0	28	23
	N	0	24	26		21 ARG	142	124	216	N	0	24	28	N	1	16	11	O	0	27	26
	C	0	20	20		CA	14	11	19	O	0	32	26	C	0	24	8				
	O	0	18	29		N	2	10	20	CB	0	33	74	O	0	16	20	55 ILE	0	150	398
	CB	41	27	30		C	0	9	8</												

357

[illegible]

114 ARG	126	131	269	117 GLY	79	23	62	121 GLN	108	110	203	CE2	1	1	50	127 CYS	26	113	137
CA	9	30	13	CA	43	8	19	CA	0	30	41	CO2	1	1	36	CA	2	17	24
N	0	28	11	N	3	11	16	N	0	14	22	124 ILE	33	164	307	N	0	20	0
C	0	20	15	C	0	3	15	C	0	21	26	CA	0	33	25	C	0	7	20
O	11	12	28	O	33	0	12	O	0	19	37	N	0	23	16	O	22	11	15
CB	6	19	28	118 THR	81	122	77	CB	0	19	37	C	0	25	14	CB	2	24	43
CG	0	9	44	CA	14	16	8	CG	0	19	33	O	3	25	18	SG	0	35	36
CD	14	9	41	N	0	13	0	CD	2	0	1	CB	4	16	46	128 ARG	233	31	77
NE	9	0	17	C	0	13	10	NOE2	41	2	6	CG2	0	14	79	CA	9	5	11
CZ	3	4	22	CB	17	21	13	NOE1	45	0	0	CG1	5	14	43	N	0	8	18
NH2	31	0	31	CG2	33	38	9	122 ALA	33	189	132	CD1	25	14	66	C	0	0	18
NH1	33	5	18	OG1	3	16	21	CA	0	49	38	125 ARG	121	236	156	O	15	0	29
115 CYS	0	140	205	119 ASP	81	120	109	N	0	33	25	CA	11	30	8	CB	32	0	0
CA	0	22	38	CA	9	8	22	C	0	21	22	N	0	26	10	CG	24	13	0
N	0	20	20	N	0	6	9	O	11	22	21	C	0	21	0	CD	35	5	0
C	0	24	23	C	0	11	17	CB	22	63	25	O	2	42	0	NE	11	1	0
O	0	26	27	O	3	15	21	123 TRP	62	185	510	CR	28	24	5	CZ	5	0	0
CB	0	25	43	CB	36	19	9	CA	0	22	16	CG	16	27	17	NH2	55	0	0
SG	0	23	55	CG	1	20	6	N	0	23	29	CD	22	33	16	NH1	47	0	0
116 LYS	105	146	226	OD2	28	10	10	CA	0	25	35	NE	6	15	23	129 LEU	91	115	252
CA	2	21	36	OD1	3	32	14	C	0	23	29	CZ	1	23	24	CA	11	13	14
N	0	19	28	120 VAL	13	206	208	D	0	21	37	NH2	14	32	29	N	1	11	7
C	0	5	6	CA	0	24	33	CR	0	22	51	NH1	21	23	26	C	5	10	2
O	19	10	9	N	0	16	8	CG	0	16	32	126 GLY	70	64	32	OEND	41	0	0
CB	0	22	38	C	0	11	26	CD1	1	23	42	CA	47	24	6	CR	14	14	24
CG	16	8	35	O	0	17	36	NE	10	13	15	N	6	10	10	CG	0	13	47
CD	0	19	46	CR	0	38	24	CE1	4	7	30	C	0	16	2	CD2	8	17	76
CE	49	21	21	CG2	2	60	41	CZ1	14	7	38	O	17	14	13	CD1	0	19	65
NZ	20	26	9					CH	16	1	43					OEND	12	17	17
								CZ2	13	1	56								

The values given for each atom are from left to right (columns 2 to 4): the area (\AA^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 \AA^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.

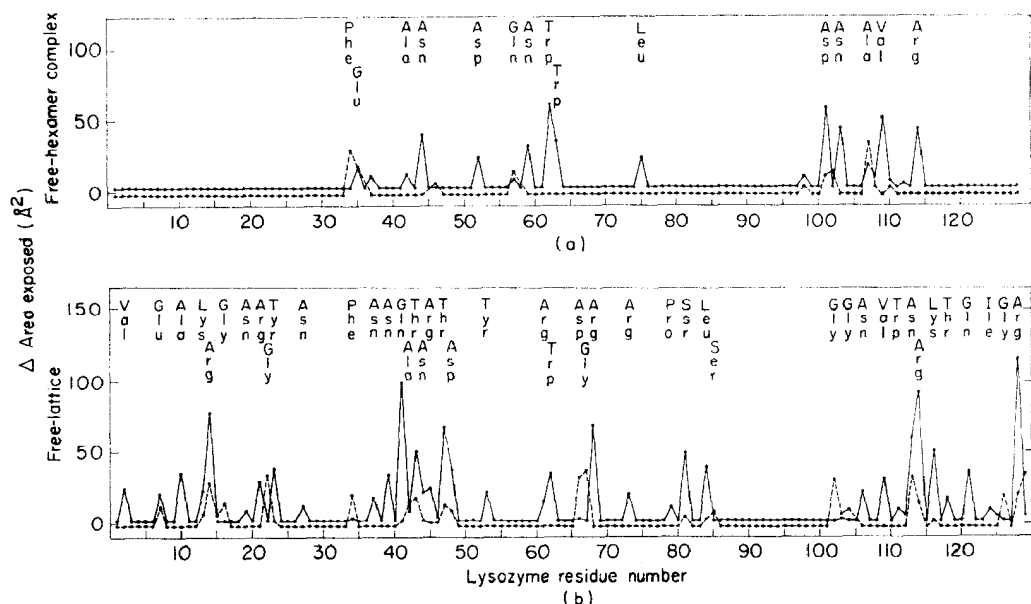


FIG. 2. Changes in area exposed to solvent for backbone atoms (---●---●---) and side chain atoms (—●—●—) for (a) binding of *N*-acetylglucosamine hexasaccharide to lysozyme; and (b) incorporation into the crystal lattice.

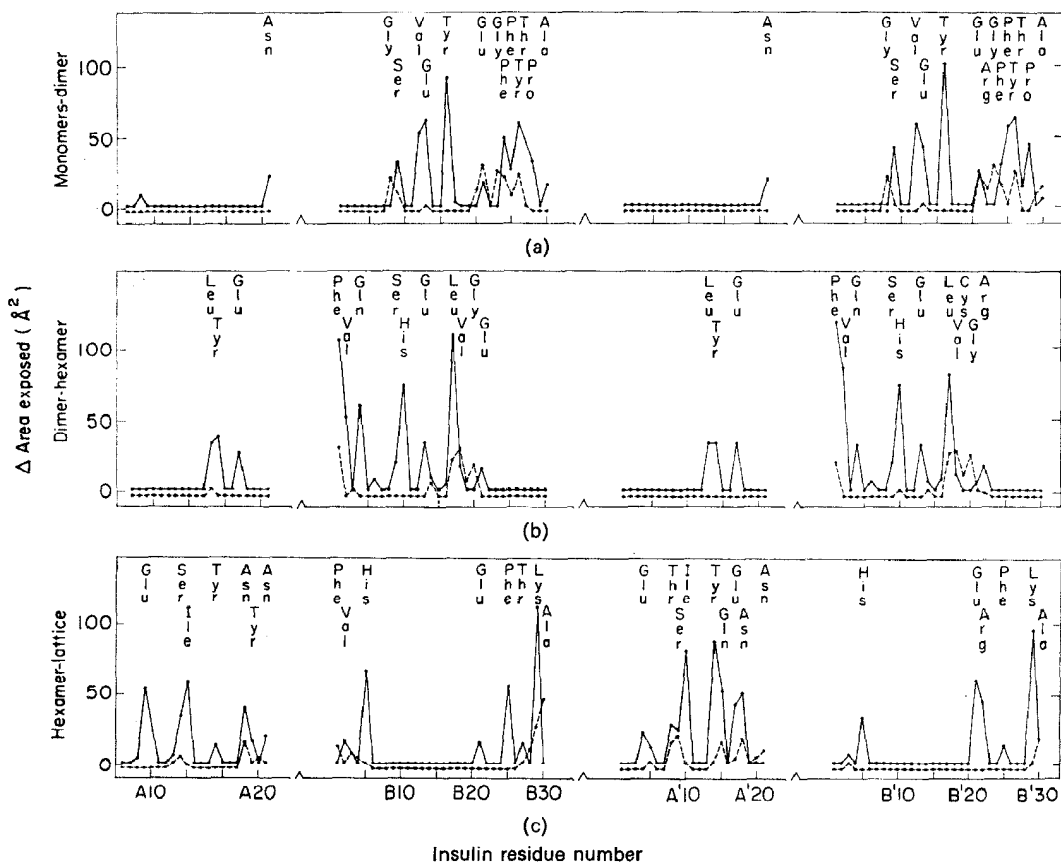


FIG. 3. Changes in area exposed to solvent for backbone atoms (---●---●---) and side chain atoms (—●—●—) for (a) association of insulin monomers to the dimer; (b) incorporation of dimer into the hexamer; and (c) incorporation of hexamer into the crystal lattice.

TABLE 5

Exposure and environment results for insulin dimer

A 1 GLY	8 129 172	C	0 16 18	B 2 VAL	168 54 90	O	0 12 36	CZ	0 12 87
	CA 5 41 73	O	2 19 30		CA 9 9 16	CB	11 19 35		CE2 0 14 74
	NEND 0 29 46	CB	3 3 58		N 0 0 21	CG	0 24 51		GD2 0 12 55
	C 0 26 30	CG	0 5 51		C 0 20 10	CD	2 11 14		
A 2 ILE	0 95 440	CD1	27 3 66	B 3 ASN	75 53 233	OE1	14 24 15	925 PHE	58 174 395
	CA 0 21 41	CD2	3 38 38		CB 6 0 0	OE2	24 14 11		CA 0 15 24
	N 0 8 22	A14 TYR	169 108 204		CG1 39 0 14	B14 ALA	14 138 208		N 0 13 23
	C 0 11 25	CA 0 22 24			CG2 52 0 0		CA 5 25 49		C 0 11 28
A 3 VAL	41 120 236	N 0 17 13		B 4 GLN	77 118 223		N 0 24 16	B26 TYR	6 87 647
	CA 0 16 43	C 0 16 16			CA 5 19 24		C 0 20 36		CA 0 14 49
	N 0 13 28	O 17 12 16			C 0 15 24		O 0 18 34		N 0 14 30
	C 0 20 18	CB 21 21 5			N 0 2 25		CB 9 21 73		C 0 13 24
A 4 GLU	85 183 170	CG 3 6 10		B 5 HIS	118 139 192	B15 LEU	6 192 484	B27 THR	33 84 259
	CA 2 28 25	CD1 19 1 16			CA 2 22 32		CA 0 133 51		CA 0 13 54
	N 0 20 20	CE1 16 1 12			N 2 1 25		N 0 25 23		N 0 9 39
	C 1 18 18	CE2 22 0 23			NOE1 28 2 22		C 0 24 24		C 0 28 28
A 5 GLN	68 293 129	OH 29 0 21		B 6 LEU	22 145 329	B16 TYR	58 198 392	B28 PRO	12 107 265
	CA 5 49 22	A15 GLN	64 263 122		CA 2 12 32		CA 0 28 44		CA 0 17 41
	N 0 20 15	CA 2 43 8			C 0 13 18		N 0 28 22		N 0 16 11
	C 0 42 11	N 0 24 10			O 21 3 16		C 0 16 5		C 0 13 18
A 6 CYS	0 125 242	C 0 29 17		B 7 CYS	57 31 73	B17 LEU	142 130 142	B29 LYS	166 99 125
	N 0 25 9	O 0 29 20			CA 2 0 22		CA 5 24 25		N 0 13 28
	C 0 22 28	CB 0 46 13			N 0 9 13		N 0 21 21		C 0 25 10
	O 0 20 26	CG 33 30 9			C 0 9 24		C 0 11 10		N 0 13 28
A 7 CYS	25 158 113	CD 2 13 10		B 8 GLY	35 66 122	B18 VAL	44 173 171	B30 ALA	131 58 31
	CA 0 39 22	NOE1 10 22 21			CA 32 25 41		CA 8 36 9		CA 0 11 11
	N 0 29 13	NOE2 17 28 13			N 1 8 24		N 0 24 13		N 0 16 7
	C 0 28 11	A16 LEU	0 165 420		C 0 18 23		C 0 24 1		C 0 12 0
A 8 THR	131 163 66	OE2 47 0 13		B 9 SER	36 64 196	B19 CYS	8 114 183	A* 1 GLY	76 60 93
	CA 0 33 14	A17 GLU	79 206 153		CA 2 0 22		CA 0 21 32		CA 36 27 28
	N 0 30 9	CA 0 44 32			N 0 9 13		N 0 20 17		NEND 23 13 18
	C 0 23 10	N 0 28 23			C 0 9 24		C 0 10 8		C 0 13 22
A 9 SER	89 72 89	C 0 29 7		B 10 HIS	129 193 121	B20 GLY	36 84 96	A* 2 ILE	11 186 333
	CA 5 21 21	O 0 20 15			CA 3 35 22		CA 28 36 28		CA 0 22 24
	N 0 20 11	CB 17 28 25			N 0 16 17		N 0 30 11		N 0 9 15
	C 0 10 7 24	CD 1 1 2 3			C 0 17 25		C 1 9 18		C 0 24 21
A10 ILE	84 232 118	OE2 43 0 3		B11 LEU	0 178 341	B21 GLU	131 62 86	A* 3 VAL	21 135 264
	CA 0 44 11	A18 ASN	97 142 117		CA 2 0 22		CA 5 16 21		CA 2 14 36
	N 0 30 5	CA 0 27 16			N 0 9 13		N 2 9 1		N 0 9 26
	C 0 25 9	N 0 26 13			C 0 16 17		C 0 0 10		C 0 11 15
A11 CYS	0 180 136	O 0 17 14 16		B12 VAL	0 198 388	B22 PHE	1 182 578	A* 4 GLU	67 126 275
	CA 0 39 19	CB 0 14 15			CA 0 18 21		CA 0 25 44		CA 3 14 39
	N 0 29 11	CG 0 11 20			N 0 16 21		N 0 22 31		N 0 13 18
	C 0 24 13	O 13 9 24			C 0 17 36		C 0 18 31		O 0 17 27
A12 SER	28 190 149	CB 9 2 33		B13 GLU	51 172 254	B23 GLY	0 93 165	A* 5 GLN	75 206 161
	CA 0 47 27	CG 0 1 24			CA 0 25 43		CA 0 36 63		CA 3 35 25
	N 0 34 10	CD1 23 0 4			C 0 24 23		N 0 22 31		N 0 28 14
	C 0 18 32	CE1 40 0 0			C 0 20 28		O 0 16 40		C 0 29 20
A13 LEU	35 108 317	CE2 32 9 14		B14 VAL	0 198 388	B24 PHE	1 182 578	A* 6 THR	131 163 66
	CA 0 16 36	CEZ 4 10 43			CA 0 18 21		CA 0 25 44		CA 0 33 14
	N 0 11 20	CE2 6 4 43			N 0 16 21		N 0 22 31		N 0 30 9
					C 0 17 36		C 0 18 31		C 0 23 10

EXPOSURE CALCULATIONS

361

CG 11 28 22	CG 13 19 38	C 0 10 23	B*13 GLU 60 216 193	CD 22 36 14
CD 0 9 14	CD 0 17 15	O 12 10 26	CA 0 35 36	NE 0 23 16
NOE1 21 12 11	NOE1 7 21 26	CB 25 5 39	N 0 29 17	CZ 2 21 17
NOE2 16 22 20	NOE2 40 13 2	CG 3 3 17	O 0 26 18	NH1 34 20 0
		MOD1 17 2 24	O 0 24 27	NH2 16 37 18
		MOD2 23 3 36	CB 9 25 28	
A* 6 CYS 0 179 175	A*16 LEU 0 193 389	R* 4 GLN 55 122 273	B*14 ALA 9 107 260	R*23 GLY 0 133 178
CA 0 49 19	CA 0 28 39	CA 9 13 30	CA 2 27 44	CA 0 46 63
N 0 28 10	N 0 26 14	N 0 1 26	N 0 26 24	N 0 24 25
C 0 29 17	C 0 22 18	C 0 16 23	C 0 17 23	C 0 18 44
O 0 26 28	O 0 25 21	O 0 14 30	O 0 16 31	O 0 15 46
CR 0 33 41	CB 0 16 49	O 0 25 32	CB 8 21 77	
SG 0 14 59	CG 0 27 36	CB 2 19 32	N 0 19 76	B*24 PHE 0 205 532
	CD1 0 28 69	CG 0 27 47	CD1 0 11 96	CA 0 39 30
A* 7 CYS 38 186 101	CD2 0 21 62	CD 0 5 23	CD2 0 22 79	N 0 25 17
CA 0 54 19	A*17 GLU 87 242 132	NOE1 25 1 14		C 0 29 20
N 0 39 14	CA 0 47 21	NOE2 20 3 43		O 0 21 27
C 0 30 15	N 0 28 15			CB 0 28 44
O 4 26 15	O 0 32 5	B* 5 HIS 131 232 167		CG 0 14 43
CR 6 39 27	CB 0 32 5	CA 0 17 33	A*15 LEU 0 182 422	CD1 0 16 58
SG 27 10 12	CG 5 39 27	N 0 1 26	CA 0 30 49	CE1 0 13 78
	CG 9 22 24	C 0 8 20	N 0 26 20	CZ 0 7 91
A* 8 THR 116 98 60	CD 2 0 13	O 24 2 11	CB 0 27 25	GE2 0 4 79
CA 2 21 14	OE1 27 2 17	CB 17 13 28	CB 0 24 54	CD2 0 7 46
N 0 26 11	OE2 40 1 6	CG 0 26 20	CG 0 19 76	
C 0 11 6		ND1 15 6 6	CD1 0 11 96	B*25 PHE 74 157 402
O 26 6 2	A*18 ASN 104 104 100	CE1 43 39 1	CD2 0 22 79	CA 0 19 44
CB 9 16 14	CA 0 11 16	NE2 0 43 7		N 0 16 22
OG1 27 0 0	N 0 18 14	CD2 0 48 14	B*16 TYR 48 158 387	C 0 15 31
CG2 52 17 13	C 0 13 15		CA 0 30 35	O 1 12 50
	O 18 20 3	B* 6 LEU 25 148 341	N 0 21 24	CR 8 22 35
A* 9 SER 65 139 103	CB 24 6 27	CA 3 17 41	C 0 15 15	CG 5 13 35
CA 14 21 13	CG 0 11 9	N 0 11 21	O 1 14 29	CD1 1 26 19
N 0 23 9	O 0 11 9	C 0 18 25	CB 0 14 54	CE1 17 14 36
C 1 8 29	MOD1 20 16 15	O 0 11 36	CG 0 1 39	CZ 23 0 40
O 5 20 29	MOD2 32 8 6	CB 0 17 41	CD1 1 4 39	GE2 17 9 48
CB 16 25 19		CG 0 14 52	CE1 17 12 30	CD2 1 12 42
OG 29 12 4	A*19 TYR 62 182 355	CD1 21 21 51	CZ 3 22 21	
	CA 6 22 25	CD2 2 27 74	OE2 0 29 38	B*26 TYR 9 139 662
A*10 ILE 106 122 183	N 0 14 11		CD2 1 10 53	CA 0 19 51
CA 0 24 36	C 0 15 21	B* 7 CYS 50 50 99	OH 24 16 12	N 0 17 32
N 0 9 17	O 1 19 28	CA 2 5 19		C 0 14 26
C 0 17 20	CB 0 28 36	N 0 2 20	B*17 LEU 126 127 120	O 0 20 35
O 5 12 28	CG 0 13 26	C 0 6 14	CA 5 25 14	CB 0 16 58
CB 3 17 19	CE1 6 23 23	O 4 14 25	N 0 25 15	CG 0 3 93
CG2 35 24 16	CE1 19 14 22	CB 32 0 8	C 0 14 8	CD1 0 1 63
CG1 6 13 36	CZ 3 3 39	SG 13 23 6	CB 24 9 9	CE1 0 1 72
CD1 55 6 17	CE2 3 10 56		CG 0 21 8	CZ 0 1 75
	CD2 0 12 48	B* 8 GLY 27 75 119	CD1 46 16 25	CE2 1 7 75
A*11 CYS 3 146 135	OH 25 9 20	CA 22 35 38	CD2 30 13 35	CD2 0 7 62
CA 3 39 16		N 5 3 18		OH 7 1 58
N 0 18 15	A*20 CYS 11 232 114	C 0 20 28		
C 0 22 21	CA 0 51 22	O 0 17 35	B*18 VAL 39 147 159	B*27 THR 56 64 224
O 0 17 25	N 0 25 18		CA 9 21 6	CA 0 11 41
CB 0 39 16	C 0 34 7	B* 9 SER 32 107 189	N 0 17 15	N 1 11 30
SG 0 10 43	O 11 39 1	CA 0 19 49	C 0 14 2	C 0 0 34
	CB 0 46 27	N 0 11 24	O 17 5 7	O 13 0 25
A*12 SER 29 171 166	SG 0 38 39	C 0 30 18	CB 0 32 27	CB 14 9 30
CA 0 39 35		O 0 26 28	CG1 3 43 35	OG1 11 3 25
N 0 29 14	A*21 ASN 95 168 162	CB 6 17 57	CG2 9 16 66	CG2 17 28 39
C 0 13 26	CA 8 28 13	CG 26 3 13		
O 0 19 30	N 0 24 18	B*10 HIS 131 174 133	B*19 CYS 11 96 179	B*28 PRO 24 118 216
CB 17 51 32	C 2 16 13	CA 2 39 28	N 0 19 33	CA 2 9 33
OG 2 20 24	OEND 10 17 25	N 0 32 9	C 0 2 11	N 0 6 20
	OEND 126 25 2	C 0 16 32	O 10 6 24	O 14 7 10
A*13 LEU 36 93 324	CB 32 22 8	O 2 11 39	CB 0 22 43	CB 5 21 51
CA 0 13 39	CG 1 10 23	CB 9 28 22	SG 0 32 49	CD 2 30 47
N 0 13 22	MOD1 1 17 39	CG 3 16 0		CG 0 38 49
C 0 11 22	MOD2 16 8 22	ND1 8 7 0	B*20 GLY 27 73 102	
O 0 14 34	B* 1 PHE 196 30 219	CE1 62 4 0	CA 27 27 33	B*29 LYS 147 97 101
CB 2 11 43	CA 11 0 19	NE2 15 0 0	N 0 22 20	CA 8 19 16
CG 0 5 43	NEND 31 0 6	CD2 30 20 3	C 0 11 17	N 0 17 18
CD1 16 27 51	O 31 1 5		O 0 13 32	C 0 3 16
CD2 19 0 71	CB 2 2 44	B*11 LEU 3 169 401		O 11 12 22
	CG 0 5 38	CA 0 32 47	B*21 GLU 141 24 79	CB 2 28 11
A*14 TYR 155 86 226	CD1 0 13 42	N 0 29 22	CA 0 6 21	CG 27 16 9
CA 2 19 22	CE1 22 10 26	C 0 24 25	N 3 6 3	CG 21 2 6
N 0 13 17	CZ 38 0 9	O 0 21 29	C 0 6 13	CE 46 0 2
C 2 10 11	CE2 39 0 6	CB 3 28 41	O 7 3 18	NZ 33 0 0
O 5 17 18	CD2 22 0 13	CG 0 13 62	CB 28 0 0	
CR 24 9 13	B* 2 VAL 122 38 73	CD1 0 11 87	CG 16 3 21	B*30 ALA 118 61 71
CG 3 0 9	CA 5 3 9	CD2 0 11 88	CD 3 0 0	CA 2 16 19
CD1 19 7 6	N 1 0 13		OE1 43 0 3	N 0 6 22
CE1 33 4 14	O 0 6 10	B*12 VAL 0 183 397	OE2 40 0 0	O 1 3 10
CE2 7 0 26	CR 11 2 0	CA 0 17 47		OEND 15 13 18
CE2 17 6 27	CG1 47 16 2	N 0 20 29	B*22 ARG 117 275 134	CB 66 17 2
CD2 14 0 33	CG2 55 0 5	C 0 17 28	CA 6 25 8	
OH 28 0 22		O 0 14 36	N 2 17 9	
A*15 GLN 85 205 176	B* 3 ASN 82 44 227	CB 0 9 84	C 0 11 6	
CA 16 41 14	CA 0 8 39	CG1 0 8 98	O 6 17 21	
N 0 22 14	N 1 2 23	CG2 0 17 76	CB 25 30 5	
C 0 25 13			CG 2 38 19	
O 2 21 25				
CB 6 25 30				

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.

Contact information for lysozyme

3 LYS...	127 CYS 13	13 LYS 76	23 TYR 91	33 LYS 140	84 LEU 12	53 TYR...	57 GLN 17
3 PHE 110	10 ALA 8	18 ASN 74	105 MET 42	30 CYS 74	40 THR 11	51 THR 138	68 SER 10
86 SER 86	123 TRP 2	20 TYR 35	31 ALA 20	123 TRP 43	53 TYR 10	40 CYS 119	46 ASN 0
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	56 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...	53 ILE 66	69 THR 63
41 GLN 27	7 GLU 40	28 TRP 5	115 CYS 11	36 SER 7	44 ASN 57	84 LEU 58	64 CYS 83
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 24			114 TRP 2	45 ASP 31	60 SER 23	58 ILE 5
	8 LEU 19	19 ASN...	20 TRP...	37 ASN 2	48 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	50 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 21
1 LYS 59	128 ARG 6	24 SER 44	25 LEU 80	110 ALA 82	57 GLN 16	68 ARG 9	80 CYS 23
3 PHE 16	25 LEU 2	17 LEU 18	23 TYR 80	31 ALA 60		56 LEU 2	62 TRP 7
37 ASN 11		21 ARG 12	20 TYR 57	108 TRP 50	44 ASN...		63 TRP 21
4 GLE 4	11 ALA...	25 LEU 5	99 VAL 35	32 ALA 48	52 ASP 91	54 GLY...	65 ASN 6
	15 HIS 55	20 TYR 2	20 TRP 2	56 LEU 41	57 GLN 66	57 GLN 66	58 ILE 5
3 PHE...	8 LEU 51	28 TRP 1	27 VAL 53	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 10	45 ARG 9	83 LEU 25	61 ARG...
7 GLU 88	10 ALA 26	21 ARG 82	55 LEU 12	55 ILE 12	42 ALA 7	56 LEU 21	62 TRP 161
80 ILE 45	3 PHE 21	17 LEU 67	32 ALA 24	37 ASN 7	35 GLU 1	36 SER 20	71 GLY 94
40 THR 36	13 LYS 16	23 TYR 60	29 VAL 19	44 ASN 3	50 SER 5	53 TYR 14	59 ASN 50
11 ALA 33	9 ALA 14	20 TRP 54	26 GLY 15	30 CYS 1		39 ASN 11	69 THR 46
55 ILE 39	12 MET 12	99 VAL 51	16 CYS 8		45 ARG...	55 ILE 10	72 SER 1
2 VAL 28		18 GLY 43	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	106 SER 26	21 ARG 1	39 ASN 72	49 GLY 59		50 SER 27
	8 LEU 92	22 GLN 6		32 ALA 52	50 SER 30	55 ILE...	60 SER 14
4 GLY...	20 TRP 70	19 ASN 24	29 VAL...	37 ASN 40	46 ASN 30	44 THR 92	63 TRP 1
7 LEU 77	88 ILE 47		123 TRP 58	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	57 GLN 49	63 TRP 105
6 CYS 19	92 VAL 26	23 TYR 63	33 ARG 38	45 ARG 44	39 ASN 44	39 ASN 44	53 TRP 103
2 VAL 12	29 VAL 22	99 VAL 00	9 ALA 35	54 GLY 14	40 ASP 43	80 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	152 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		82 VAL 2	72 SER 1
6 CYS 46	10 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...	38 CYS...	38 PHE...	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...	59 ILE 88	59 ASN 82
127 CYS 14	25 LEU 67	20 TYR 66	33 LYS 47	33 LYS 125	50 SER 87	58 ILE 87	75 LEU 73
3 PHE 3	10 ALA 54	28 TRP 61	27 ASN 35	5 ARG 83	61 ARG 41	58 ILE 67	58 ILE 67
126 GLY 2	9 ALA 31	19 ASN 57	26 GLY 31	8 LEU 70	46 ASN 31	35 GLU 37	64 CYS 36
	16 GLY 30	27 ASN 56	29 VAL 22	2 VAL 27	47 THR 27	95 ALA 35	94 CYS 36
6 CYS...	11 HIS 20	15 TRP 58	120 MET 50	55 ILE 61	69 THR 4	54 GLY 26	97 LYS 33
9 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 10	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	31 ALA...	48 THR 10	46 ASN 22	31 ALA 11	77 ASN 2
127 CYS 6	15 HIS 66	18 ASP 3	35 GLU 55	39 ASN 8	46 ASP 22	17 LEU 5	83 LEU 2
129 LEU 6	11 ALA 53	103 ASN 1	111 TRP 43	1 LYS 6	59 THR 15	51 THR 14	
7 GLU 3	13 LYS 24	24 SER...	28 TRP 37	37 ASN 6	68 ARG 11	36 SER 4	64 CYS...
125 TRP 2	16 GLY 16	19 ASN 101	27 ASN 33	39 ASN...	70 PRO 9	105 MET 3	60 SER 63
123 TRP 2	12 MET 10	27 ASN 47	115 CYS 27	2 VAL 108	50 SER 6		74 ASN 48
		26 GLY 38	34 PHE 23	42 ALA 75		52 ASP 189	83 LEU 35
7 GLU...	15 HIS...	20 TRP 30	56 LEU 23	36 SER 71	50 SER...	35 GLU 86	63 TRP 30
4 GLY 90	14 ARG 85	120 VAL 26	105 MET 17	41 GLN 60	48 ASP 78	53 TYR 67	59 ASN 23
1 PHE 87	92 VAL 75	18 ASP 16	33 LYS 16	55 ILE 43	46 ASN 68	59 THR 57	68 CYS 19
3 LYS 49	89 THR 69	22 GLY 12	29 VAL 15	37 ASN 35	59 THR 51	44 ASN 52	53 TYR 23
10 ALA 44	11 ALA 28	30 CYS 12	30 CYS 14	40 THR 14	60 SER 51	42 ALA 26	79 PRO 11
6 CYS 27	68 ILE 59	23 TYR 9	119 ALA 11	40 THR 14	60 SER 28	36 SER 25	58 ILE 9
5 ARG 20	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 120	87 ASP 7	18 ASP 113	32 ALA...	55 ILE 87		55 ILE 9	94 CYS 2
3 PHE 120	93 ASP 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 79	84 LEU 108	53 TYR 108		
11 ALA 5	16 GLY...	51 GLY 51	35 GLU 40	57 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 72	79 PRO 59
29 VAL 46	13 LYS 30	24 SER 38	29 VAL 32	86 SER 26	68 ARG 43	98 ILE 68	80 CYS 59
4 GLY 41	96 LYS 30	129 LEU 28	28 TRP 30	85 SER 26	50 SER 27	56 LEU 60	72 GLY 93
5 ARG 37	14 ARG 19	12 MET 22	31 ALA 22	88 ILE 24	49 GLY 27	63 TRP 59	72 SER 34
32 ALA 25	12 MET 16	27 ASN 14	8 LEU 17	83 LEU 22	59 ASN 25	49 CYS 47	76 ILE 26
7 GLU 22	17 LEU 13	23 TYR 10	30 CYS 16	42 ALA 21	52 ASP 20	91 SER 16	68 ASP 28
68 ILE 19	15 HIS 7	17 LEU 9	37 ASN 11	43 GLN 16	52 ASP 17	62 TRP 22	76 ILE 12
6 CYS 18		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	26 TRP 147	26 GLY...	33 LYS 3	41 GLN...	69 THR 8	108 TRP 8	66 ASP...
26 TRP 2	12 MET 135	120 VAL 63		84 LEU 126	58 ILE 4	59 ASN 9	68 ARG 80
	20 TYR 77	24 SER 41	33 LYS...	39 ASN 66		51 THR 7	83 TYR 59
9 ALA...	92 VAL 75	29 VAL 30	38 PHE 115	1 LYS 35	52 ASP...	57 GLN 2	50 CYS 54
25 LEU 55	96 LYS 30	124 ILE 34	37 ASN 109	43 THR 31	57 GLN 128	60 SER 2	60 CYS 59
2 VAL 41	24 VAL 23	30 CYS 31	123 TRP 43	40 THR 24	59 ASN 107		69 THR 58
129 LEU 39	13 LYS 12	123 TRP 25	34 PHE 80	42 ALA 10	46 ASN 58	59 ASN...	60 SER 46
6 CYS 37	56 LEU 11	25 LEU 23	30 CYS 58	54 GLY 9	59 ASN 51	63 TRP 67	61 THR 11
5 ARG 31	95 ALA 11	28 TRP 17	29 VAL 32		53 TYR 46	52 ASP 56	54 CYS 10
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 38	67 GLY...
8 LEU 20	18 ASP 2	27 ASN...	36 SER 12	54 GLY 36	42 ALA 3	51 THR 39	65 ASN 38
7 GLU 19		120 VAL 108		41 GLN 22	50 SER 1	64 CYS 26	68 ASP 28
11 ALA 14	18 ASP...	111 TRP 86	34 PHE...	41 GLN 22			68 ASP 28
	25 LEU 108	24 SER 85	114 ARG 205	36 SER 21			

EXPOSURE CALCULATIONS

363

69 THR	9	77 ASN...	40 THR 68	55 ILE 23	98 ILE...	112 ARG 4	107 ALA 9	24 SER 11
		74 ASN 127	53 TYR 58	89 THR 17	63 TRP 132	56 LEU 3	105 MET 9	26 GLY 8
68 ARG...		78 ILE 35	43 THR 35	95 ALA 14	108 TRP 77		115 CYS 3	122 ALA 5
69 ARG 120		27 LEU 27	54 GLY 32	92 VAL 10	95 ALA 62	106 ASN...	103 ASN 2	
66 ASP 81		76 CYS 6	83 LEU 19	54 GLY 6	94 CYS 50	112 ARG 147		122 ALA...
70 PRO 45		79 PRO 4	82 ALA 17	87 ASP 1	58 ILE 36	103 ASN 80	113 ASN...	125 ARG 174
69 THR 43		63 TRP 3	80 CYS 17		107 ALA 31	111 TRP 69	109 VAL 80	119 ASP 45
51 THR 40			42 ALA 11	92 VAL...	101 ASP 29	116 LYS 57	112 ARG 59	123 TRP 31
49 GLY 13		78 ILE...	86 SER 6	17 LEU 79	100 SER 20	108 TRP 36	114 ARG 44	126 VAL 19
53 TYR 11		79 PRO 94	85 SER 5	88 ILE 79	97 LYS 19	104 GLY 19	110 ALA 37	121 GLN 19
		76 CYS 61	1 LYS 2	89 THR 62	104 GLY 14	109 VAL 10	111 TRP 12	5 ARG 18
69 THR...		74 ASN 58		84 ALA 59	96 LYS 10	105 MET 6	215 CYS 5	124 ILE 14
72 SER 72		64 CYS 52	85 SER...	96 LYS 57	76 CYS 6	107 ALA 6		
66 ASP 72		82 ALA 51	87 ASP 84	15 HIS 55	79 VAL 5	23 TYR 5	114 ARG...	123 TRP...
70 PRO 71		90 ALA 32	82 ALA 57	91 SER 38	102 GLY 4		34 PHE 170	5 ARG 133
66 ARG 43		85 LEU 24	40 THR 25	56 LEU 25	103 ASN 3	107 ALA...	110 ALA 82	30 CYS 128
60 SER 42		65 ASN 66	83 LEU 19	12 MET 11		108 TRP 48	113 ASN 41	33 LYS 82
49 GLY 36		94 CYS 9	84 LEU 14	94 CYS 13	29 VAL...	103 ASN 45	118 THR 35	120 VAL 77
68 ARG 33		63 TRP 9	81 SER 12	90 ALA 12	91 ARG 64	98 ILE 32	115 CYS 22	29 VAL 68
61 ARG 25		83 CYS 8	88 ILE 7	93 ASN 11	104 GLY 63	104 GLY 25	111 TRP 22	34 PHE 40
71 ALY 20		77 ASN 3	86 SER 5	55 ILE 3	28 TRP 51	105 MET 25	112 ARG 17	122 ALA 39
51 THR 7			90 ALA 1		96 LYS 47	109 VAL 17	30 CYS 7	124 ILE 39
48 ASP 5		79 PRO...		93 ASN...	95 ALA 44	112 ARG 14	116 LYS 4	26 GLY 25
67 GLY 3		78 ILE 39	86 SER...	89 THR 96	20 TYR 44	106 ASN 7		121 GLN 23
		65 ASN 62	1 LYS 88	96 LYS 71	108 TRP 37		115 CYS...	125 ARG 17
70 PRO...		81 SER 57	3 PHE 30	90 ALA 54	98 ILE 26	108 TRP...	111 TRP 112	38 PHE 16
72 SER 43		82 ALA 81	87 ASP 26	92 VAL 38	23 TYR 24	105 MET 167	118 THR 82	127 CYS 5
68 ARG 40		64 CYS 15	85 SER 22	91 SER 19	100 SER 23	56 LEU 107	120 VAL 49	119 ASP 3
69 THR 19		74 ASN 10	40 THR 22	97 LYS 15	105 MET 21	98 ILE 95	114 ARG 19	9 ALA 1
49 GLY 11		83 LEU 5	88 ILE 13	95 ALA 12	101 ASP 14	35 GLY 59	114 GLY 15	
48 ASP 2		80 CYS 5	84 ILE 12	94 CYS 3	97 LYS 8	110 ASN 45	115 ALA 5	
		77 ASN 5	15 HIS 1	15 HIS 1	102 GLY 7	95 ALA 45	27 ASN 12	124 ILE...
71 GLY...		66 ASP 2	87 ASP...			111 TRP 44	110 ALA 12	129 ILE 75
61 ARG 69			85 SER 97	94 CYS...	100 SER...	28 TRP 36	34 PHE 10	127 CYS 73
70 PRO 36		80 CYS...	89 THR 73	98 ILE 49	21 ARG 118	107 ALA 34	30 CYS 10	25 LEU 51
69 THR 28		53 TYR 106	88 ILE 24	91 SER 45	97 LYS 40	110 ALA 32	112 ARG 4	26 GLY 39
73 ARG 27		83 LEU 56	86 SER 16	58 ILE 43	96 LYS 39	99 VAL 27	116 LYS 2	123 TRP 28
72 SER 5		65 ASN 52	90 ALA 10	97 LYS 42	99 VAL 33	31 ALA 19	113 ASN 2	9 ALA 27
		66 ASP 48	15 HIS 5	90 ALA 41	102 GLY 26	112 ARG 19		5 ARG 26
72 SER...		60 SER 22	82 ALA 3	96 LYS 24	20 TYR 23	109 VAL 14	116 LYS...	29 VAL 22
69 THR 64		78 ILE 16	83 LEU 1	83 ILE 21	98 ILE 18	58 ILE 13	111 TRP 145	122 ALA 14
61 ARG 48		82 ALA 16	91 SER 1	83 TRP 20	101 ASP 7	104 GLY 12	112 ARG 117	125 ARG 12
70 PRO 47		64 CYS 15		93 ASN 19		57 GLN 10	106 SN 67	126 VAL 7
65 ASN 36		74 LEU 13	88 ILE...	92 VAL 18	101 ASP...	109 VAL...	115 CYS 17	126 GLY 5
60 SER 30		79 PRO 10	92 VAL 96	78 ILE 9	97 LYS 37	107 VAL...	115 CYS 61	24 SER 5
73 ARG 28		81 SER 5	91 SER 66	95 ALA 4	98 ILE 33	113 ASN 69	27 ASN 3	
74 ASN 4			15 HIS 40	64 CYS 1	63 TRP 27	112 ARG 50	120 VAL 2	125 ARG...
66 ASP 4		81 SER...	3 PHE 39	89 THR 1	100 SER 14	35 GLU 21		122 ALA 190
		84 LEU 74	55 ILE 33		103 ASN 10	110 ALA 18	117 GLY...	121 GLN 86
73 ARG...		79 PRO 65	8 LEU 30	95 ALA...	99 VAL 5	106 ASN 16	118 THR 38	119 ASP 64
62 TRP 86		40 CYS 19	11 ALA 25	98 ILE 64		107 ALA 12	116 LYS 23	5 ARG 57
75 LEU 68		83 LEU 17	12 MET 24	92 VAL 52	102 GLY...	111 TRP 8	115 CYS 23	127 CYS 26
71 GLY 55		53 TYR 16	90 ALA 18	108 TRP 46	103 ASN 35	108 TRP 6		124 ILE 21
61 ARG 43		85 SER 9	40 THR 16	56 LFL 32	104 ASP 31		118 THR...	6 CYS 7
72 SER 28		82 ALA 7	87 ASP 15	99 VAL 31	101 GLY 22	110 ALA...	115 CYS 72	123 TRP 3
74 ASN 21			85 SER 13	91 SER 29	150 SER 20	114 ARG 68	120 VAL 61	
60 SER 3		82 ALA...	86 SER 9	98 ILE 23	98 ILE 7	109 VAL 59	114 ARG 25	126 GLY...
64 CYS 3		79 PRO 51	89 THR 4	28 TRP 21	99 VAL 5	35 GLU 47	119 ASP 18	6 CYS 29
		83 SER 39	83 LEU 3	97 LYS 20	21 ARG 3	108 TRP 43	116 LYS 7	128 ARG 19
74 ASN...		90 ALA 39	17 LEU 2	97 LYS 14		113 ASN 33	36 CYS 3	124 ILE 18
77 ASN 102		81 SER 36	93 ASN 1	93 ASN 10	103 ASN...	34 PHE 21	123 TRP 1	125 ARG 17
65 ASN 70		78 ILE 33		17 LEU 9	106 ASN 70	112 ARG 16		127 CYS 8
78 ILE 38		86 CYS 27	89 THR...	96 LYS 8	107 ALA 40	115 CYS 10	119 ASP...	5 ARG 5
64 CYS 33		83 LEU 20	87 ASP 44	90 ALA 1	101 ASP 9	31 ALA 9	125 ARG 69	
75 LEU 24		84 LEU 18	93 ASN 77		98 ILE 7	111 TRP 7	121 GLN 58	127 CYS...
76 CYS 22		87 ASP 4	15 HIS 74	96 LYS...	105 MET 8		122 ALA 48	129 LEU 86
62 TRP 18			92 VAL 34	20 TYR 152	112 ARG 2	111 TRP...	118 THR 36	124 ILE 74
63 TRP 13		83 LEU...	91 SER 21	93 ASN 94	23 TYR 1	105 MET 217	120 VAL 14	128 ARG 1
72 SER 11		91 SER 83	88 ILE 19	92 VAL 82		116 LYS 174	123 TRP 4	9 ALA 19
73 ARG 10		56 ILE 74	90 ALA 17	99 VAL 59		115 CYS 129		125 ARG 19
79 PRO 3		80 CYS 69		17 LEU 28	104 GLY...	27 ASN 92	120 VAL...	6 CYS 13
		53 TYR 55	90 ALA...	100 SER 28	93 VAL 60	27 ASN 92	27 ASN 82	5 ARG 13
63 TRP 85		82 ALA 42	93 ASN 53	94 CYS 28	23 TRP 56	106 ASN 59	123 TRP 73	123 TRP 1
62 TRP 81		64 CYS 33	83 LEU 49	15 HIS 25	107 ALA 21	31 ALA 54	118 THR 68	
73 ARG 39		78 ILE 30	89 THR 47	97 LYS 24	102 GLY 16	108 TRP 51	120 GLY 36	128 ARG...
77 ASN 26		84 LEU 30	94 CYS 33	16 GLY 20	103 ASN 15	110 ALA 30	121 GLN 32	129 LEU 45
97 LYS 22		54 GLY 27	78 ILE 23	95 ALA 20	105 MET 14	112 ARG 16	30 CYS 23	126 GLY 31
74 ASN 13		81 SER 26	88 ILE 21	98 ILE 12	98 ILE 12	114 ARG 16	115 CYS 21	127 CYS 24
76 CYS 8		85 SER 24	82 ALA 21	28 TRP 6	108 TRP 12	109 VAL 11	24 SER 19	10 ALA 7
		94 CYS 21	87 ASP 13			113 ASN 11	122 ALA 14	
76 CYS...		40 TYR 13	92 VAL 10	97 LYS...	105 MET...	120 VAL 4	124 ILE 5	129 LEU...
63 TRP 96		79 PRO 7	91 SER 8	94 CYS 53	111 TRP 177	34 PHE 2	119 ASP 3	13 LYS 110
78 ILE 67		55 ILE 4	85 SER 2	100 SER 41	108 TRP 131		116 LYS 2	127 CYS 92
77 ASN 35		87 ASP 3		76 CYS 35	23 TYR 99	112 ARG...	111 TRP 2	124 ILE 47
97 LYS 32		63 TRP 2	91 SER...	63 TRP 34	28 TRP 78	116 ASN 164	117 GLY 2	25 LEU 43
94 CYS 12		88 ILE 1	83 LEU 86	75 LEU 28	99 VAL 33	116 LYS 96		9 ALA 35
75 LEU 11		84 LEU...	88 ILE 71	101 ASP 27	27 ASN 29	109 VAL 70	121 GLN...	10 ALA 27
98 ILE 4		41 GLY 102	95 ILE 54	96 LYS 28	125 ARG 12	111 TRP 56	124 ILE 49	128 ARG 7
62 TRP 2		*1 SER 93	94 CYS 42	93 ASN 26	31 ALA 17	113 ASN 36	125 ARG 81	6 CYS 6
			95 ALA 37	95 ALA 16	103 ALA 14	118 TRP 22	119 ASP 60	
			56 LEU 28	99 VAL 15	106 ASN 8	114 ALA 21	120 VAL 42	
			90 ALA 26	98 ILE 14		114 ARG 16	123 TRP 17	

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 (\AA^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.

[illegible]

FIG. 4. See following page for legend.

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 co-workers (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 \AA^2 of occluded surface. The right-hand column gives the surface area (\AA^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.

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

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 Å²).

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 $100 \times \text{area of solvated sphere exposed to solvent} / \text{total area of solvated sphere}$.

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 Å² 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 p*K* 6 to 6.5 found for Glu-35 and p*K* 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 Å² 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 Å². 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 β(1→4)-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.