

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

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.

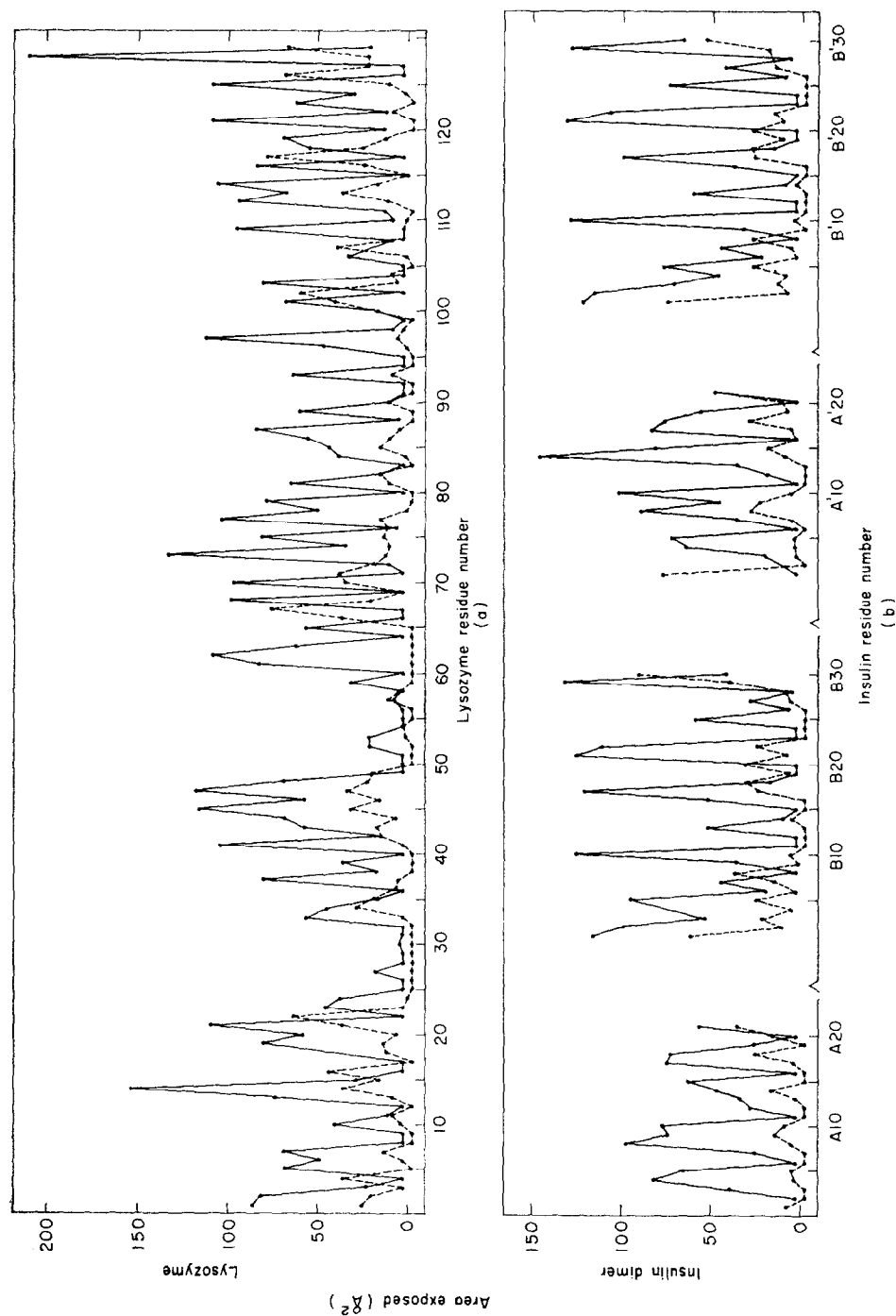


FIG. 1. Area exposed to solvent for backbone atoms (---●---) and side chain atoms (—). (a) Results for lysozyme; (b) results for insulin dimer.

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
G	0	22	23	13 LYS	67	208	237	CB	0	35	36	CZ	22	17	36	NE	2	29	10
CB	8	17	28	CA	0	33	39	CG	0	26	29	CE2	13	27	26	CZ	2	23	8
CG	11	5	27	N	0	31	15	CD1	4	22	28	CD2	6	26	26	NH2	32	25	0
CG	8	17	38	C	0	12	3	CE1	14	10	33					NH1	28	22	23
CE	35	21	19	O	9	13	16	CZ	6	10	40	35 GLU	33	151	319				
NZ	24	18	13	CB	2	22	35	CE2	6	23	43	CA	2	30	32	46 ASN	73	210	61
				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	C	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	14 ARG	189	153	58	N	1	26	16	GO	1	1	44	CG	0	22	0
N	1	7	23	CA	11	27	13	C	0	21	21	OE2	1	11	50	NOD2	26	18	7
G	0	9	22	N	0	22	16	O	0	21	29	OE1	14	0	35	NOD1	2	47	3
O	18	3	16	C	0	15	7	CB	35	32	22								
CB	9	16	14	O	25	7	0	OG	2	14	34	36 SER	7	206	195	47 THR	151	36	36
CG2	38	17	22	CB	0	30	8					CA	6	44	36	CA	8	13	6
CG1	35	28	14	CG	32	13	9	25 LEU	3	178	377	N	0	26	18	N	0	2	14
				CD	41	0	0	CA	0	32	38	C	0	22	24	C	0	13	6
3 PHE	15	161	500	NE	7	7	0	N	0	29	20	O	1	17	41	O	26	9	0
CA	0	22	39	CZ	1	10	1	C	0	16	22	CB	0	60	46	CB	16	0	6
N	0	16	22	NH2	37	20	1	O	0	16	32	OG	0	36	30	CG2	74	0	0
C	0	5	21	NH1	36	2	1	CB	2	24	41					OG1	28	0	4
O	2	4	40					CG	0	24	71	37 ASN	86	119	137				
CB	0	14	51	15 HIS	43	240	249	CD2	2	22	81	CA	0	32	28	48 ASP	81	79	114
CG	0	16	57	CA	9	27	16	CD1	0	16	73	N	0	31	10	CA	13	8	9
CD1	0	19	45	N	0	28	11					C	0	15	30	N	0	10	15
CE1	4	22	50	O	0	22	10	26 GLY	0	103	182	O	6	4	37	C	0	11	13
CZ	9	19	55	O	5	20	22	CA	0	38	73	CB	24	14	9	O	11	11	13
CE2	0	14	63	CB	5	24	38	N	0	21	33	CG	0	5	0	CB	38	9	9
CG2	0	10	58	CG	0	13	36	C	0	24	36	NOD2	49	0	0	CG	1	6	11
				ND1	0	16	26	O	0	20	41	NOD1	6	18	21	OD2	12	13	28
4 GLY	35	79	132	NE2	3	24	16	27 ASN	18	163	332	38 PHE	17	215	504	OD1	6	13	26
CA	35	27	36	CD2	14	19	39	CA	0	32	52	CA	0	27	46	49 GLY	20	116	60
N	0	20	28					N	0	22	20	N	0	24	16	CA	19	38	27
C	0	17	30					C	0	20	31	C	0	17	38	N	1	24	5
O	6	15	39					O	0	16	37	O	0	20	45	C	0	18	15
5 ARG	71	204	270					CB	0	33	60	CB	0	24	35	O	0	35	14
CA	0	2	44	17 LEU	0	120	434	CG	0	21	40	CG	0	22	38				
N	1	8	17	CA	9	27	16	NOD1	6	9	48	CD1	0	15	50	50 SER	3	232	112
C	0	10	15	N	0	16	15					CE1	0	14	68	CA	0	52	22
O	6	15	32	C	0	19	24	28 TRP	0	299	680	CZ	0	13	78	N	0	23	15
CB	16	3	39	O	0	21	30	CA	0	28	41	CE2	6	17	58	C	0	22	16
CG	0	13	51	CB	0	19	60	N	0	19	37	CD2	12	20	33	O	0	27	19
CD	16	16	43	CG	0	13	84	C	0	23	22	39 ASN	37	181	270	CB	2	66	33
NE	10	13	2	CD2	0	11	96	O	0	19	37	CA	0	36	47	OG	1	42	6
CZ	0	40	6	CD1	0	6	92	CB	0	22	51	N	0	24	26				
NH2	28	36	7					CG	0	23	48	C	0	25	23	51 THR	3	275	194
NH1	0	49	14	18 ASP	47	114	255	CD1	0	42	46	O	0	27	29	N	0	36	10
				CA	0	11	31	NE	0	33	29	CB	0	27	49	C	0	33	20
6 CYS	52	97	103	N	0	16	10	CE1	0	25	48	CG	0	11	29	O	0	34	27
CA	0	14	22	C	0	16	10	CZ1	13	69	69	CE	0	5	37	CB	0	54	32
N	1	9	23	O	0	11	21	CH	0	10	84	NOD2	30	5	37	CG2	3	36	57
C	0	15	17	C	0	11	21	CZ2	0	10	81	NOD1	2	26	39	OG2	0	33	19
O	1	16	26	O	11	14	21	CE2	0	6	69					OG1	0	33	19
CB	36	3	11	CB	11	6	62	CD2	0	12	38	40 THR	0	257	234				
SG	13	39	4	OG	0	15	24					CA	0	47	32	52 ASP	22	276	209
				OD2	19	5	36	29 VAL	0	165	307	N	0	30	13	CA	0	49	36
7 GLU	82	131	233	OD1	6	22	22	N	0	32	11	C	0	26	25	N	0	17	25
CA	11	24	32					O	0	33	36	O	0	28	30	C	0	28	37
N	1	17	22	19 ASN	94	175	99	C	0	17	23	CB	0	58	39	O	0	27	39
C	0	20	22	CA	0	39	17	O	0	12	35	CG2	0	35	74	CB	0	49	30
O	1	14	30	N	0	36	10	CB	0	21	55	OG1	0	32	20	O	0	39	15
CB	14	16	24	O	0	14	12	CG2	0	19	77	41 GLN	107	147	153	OD2	21	32	13
CG	21	9	27	CG	0	22	8	CG1	0	32	68	N	0	28	33	OD1	1	35	14
CD	2	8	25	NOD2	32	20	3					C	0	30	16	53 TYR	23	232	506
OE2	15	16	25	NOD1	24	12	21	30 CYS	3	127	205	O	0	5	28	CA	0	25	44
OE1	17	7	28					N	0	28	44	O	2	2	43	N	0	24	15
				20 TYR	63	204	412	C	0	25	21	CB	11	36	14	C	0	21	24
8 LEU	0	154	421	CA	5	38	19	O	0	14	27	CG	8	35	16	O	1	21	29
CA	0	28	43	N	0	37	8	CB	0	24	52	CD	1	1	0	CB	0	22	47
N	0	29	24	C	0	11	21	SG	3	17	45	NOD2	48	0	3	CG1	0	17	45
C	0	18	22	O	0	17	36					NOE1	36	10	0	CD1	0	9	58
O	0	17	26	CB	0	32	36	31 ALA	0	143	186					CE1	16	3	62
CB	0	21	57	CG	0	13	46	CA	0	38	44	42 ALA	30	139	126	CZ	0	12	58
CG	0	19	74	CD1	10	13	23	N	0	41	10	CA	16	32	25	CE2	0	33	46
CD2	0	14	82	CE1	23	3	32	C	0	25	24	N	0	26	22	CD2	0	30	35
CD1	0	17	93	CZ	0	14	42	O	0	16	34	O	0	20	30	OH	6	15	44
				CE2	0	14	58	CB	0	28	73	CB	14	51	28				
9 ALA	0	119	235	CD2	0	3	68	32 ALA	0	167	199					54 GLY	2	143	126
CA	0	33	58	OH	25	14	24	CA	0	44	52	43 THR	75	119	135	CA	2	57	49
N	0	28	23					N	0	24	28	N	1	16	11	N	0	28	29
C	0	16	33					O	0	33	20	C	0	24	8	C	0	32	23
O	0	11	43					CB	0	33	74	O	0	28	26	O	0	27	26
CB	0	32	77																
10 ALA	44	111	147					33 LYS	58	190	349					55 ILE	0	150	398
CA	3	22	43	21 ARG	142	124	216	CA	0	43	39	44 ASN	76	138	118	CA	0	27	27
N	0	24	26	CA	14	11	19	N	0	31	24	N	0	17	30	N	0	13	28
C	0	20	20	N	2	10	20	C	0	22	16	N	0	24	21	C	0	11	31
O	0	18	29	NE	8	8	10	O	2	25	16	C	0	13	24	O			

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	296	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	C	0	23	29	NE	6	15	23	129 LEU	91	115	252
CA	2	21	36	OD1	3	32	14	D	0	21	37	CZ	1	23	24	CA	11	13	14
N	0	19	28	120 VAL	13	206	208	CR	0	22	51	NH2	14	32	29	N	1	11	7
C	0	5	6	CA	0	24	33	CG	0	16	32	NH1	21	23	26	C	5	10	2
O	19	10	9	N	0	16	8	CD1	1	23	42	126 GLY	70	64	32	OEND	41	0	0
CB	0	22	38	C	0	11	26	NE	10	13	15	CA	47	24	6	CR	14	14	24
CG	16	8	35	O	0	17	36	CE1	4	7	30	N	6	10	10	CG	0	13	47
CD	0	19	46	CR	0	38	24	CZ1	14	7	38	C	0	16	2	CD2	8	17	76
CE	49	21	21	CG2	2	60	41	CH	16	1	43	O	17	14	13	CD1	0	19	65
NZ	20	26	9					CZ2	13	1	56					OEND	12	17	17

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.

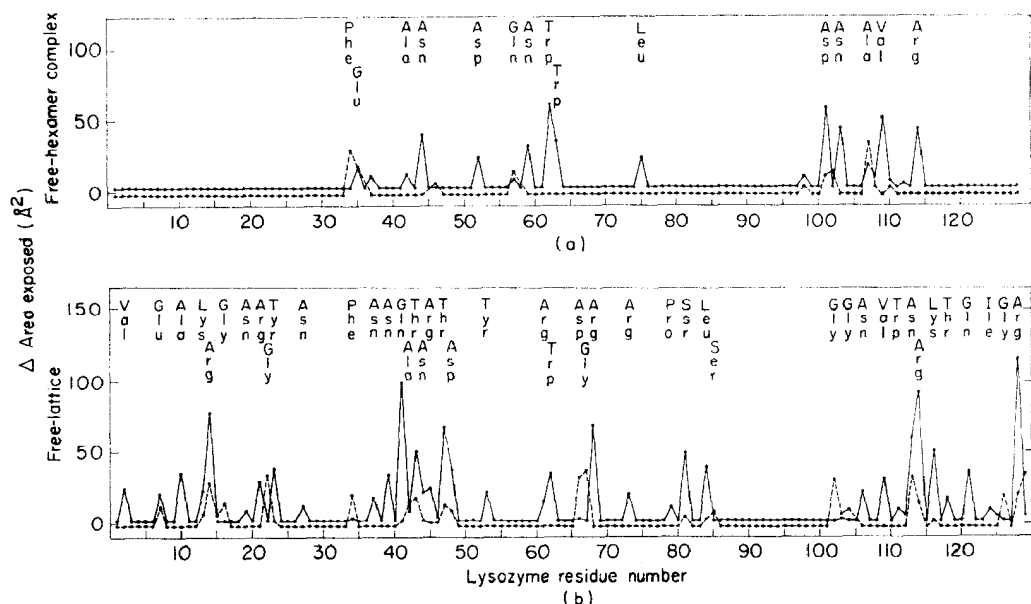


FIG. 2. Changes in area exposed to solvent for backbone atoms (---●---●---) and side chain atoms (—●—●—) for (a) binding of *N*-acetylglucosamine hexasaccharide to lysozyme; and (b) incorporation into the crystal lattice.

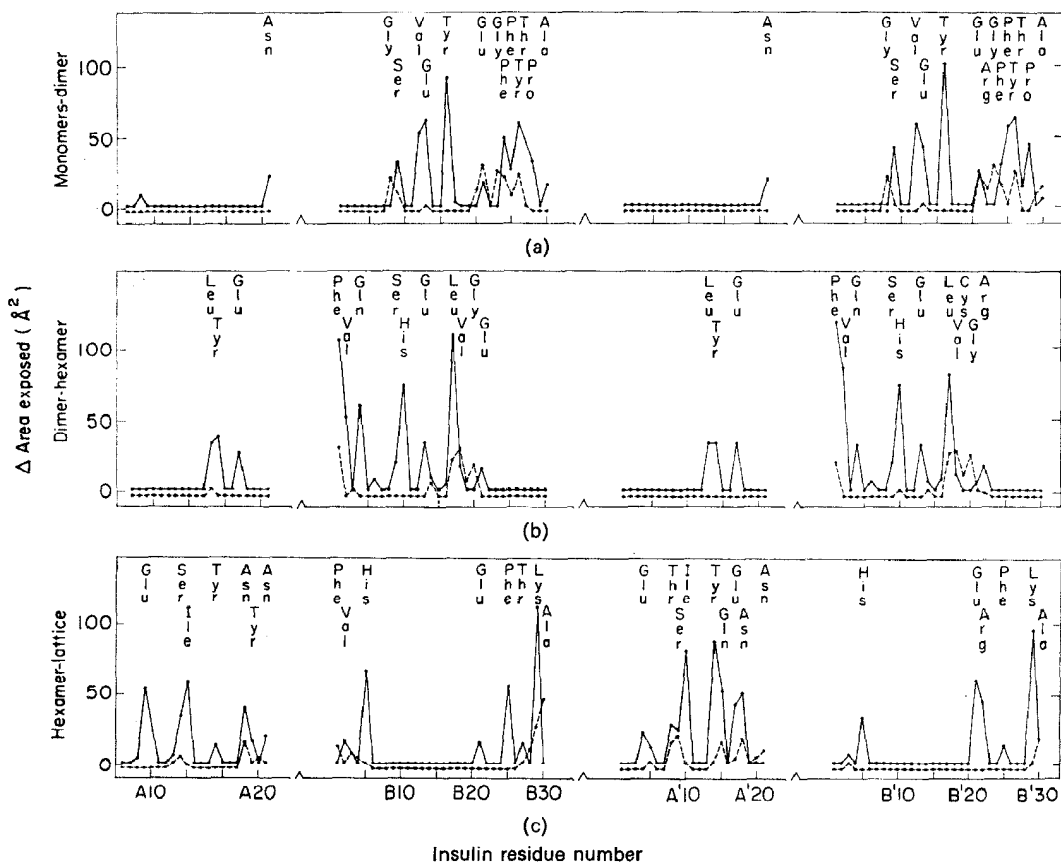


FIG. 3. Changes in area exposed to solvent for backbone atoms (---●---●---) and side chain atoms (—●—●—) for (a) association of insulin monomers to the dimer; (b) incorporation of dimer into the hexamer; and (c) incorporation of hexamer into the crystal lattice.

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		
O	3 33 24	CD1	27 3 66	O	1 25 29	OE1	14 24 15	925 PHE	58 174 395
		CD2	3 38 38	CB	6 0 0	OE2	24 14 11	CA	0 15 24
A 2 ILE	0 95 440	A14 TYR	169 108 204	CG1	39 0 14	B14 ALA	14 138 208	N	0 15 24
CA	0 21 41	C	2 22 24	CG2	52 0 0	CA	5 25 49	C	0 13 23
N	0 8 22	N	0 17 13	B 3 ASN	75 53 233	N	0 24 16	CB	0 14 57
C	0 11 25	C	0 16 16	C	0 4 43	C	0 20 36	CG	2 13 47
O	0 21 28	CB	17 12 16	N	0 2 25	O	0 18 34	CD1	0 9 42
CB	0 6 58	CB	21 21 5	C	0 3 33	CB	9 21 73	CE1	7 17 42
CG2	0 9 77	CG	3 6 10	O	0 19 0 34	B15 LEU	6 192 484	CE2	25 7 36
CG1	0 9 84	CD1	19 1 16	CB	21 11 35	CA	0 13 51	CZ	17 25 22
CD1	0 8 104	CE1	16 1 12	CG	1 8 15	N	0 25 23	CD2	7 25 29
		CZ	6 0 23	NOD1	28 2 22	C	0 24 24	B26 TYR	6 87 647
A 3 VAL	41 120 236	CE2	22 0 29	NOD2	5 18 25	O	0 19 24	CA	0 14 49
CA	0 16 43	OH	17 12 20	B 4 GLN	77 118 223	CB	0 19 24	N	0 14 30
N	0 13 28	CD2	29 0 21	N	0 19 24	CG	0 24 62	C	0 13 24
C	0 20 18	A15 GLN	64 263 122	C	0 15 24	CD1	0 35 71	O	0 18 34
O	0 21 28	CA	2 43 8	O	0 22 32	CD2	0 13 95	CB	0 9 58
CB	5 9 33	N	0 24 10	CB	0 22 24	B16 TYR	58 198 392	CG	0 3 46
CG1	19 13 57	C	0 29 17	CG	22 11 27	CA	0 28 44	CD1	0 7 59
CG2	17 28 33	O	0 29 20	CD	0 6 20	N	0 28 22	CE1	0 6 71
		CB	0 46 13	NOD1	3 6 40	C	0 16 5	CZ	0 6 72
A 4 GLU	85 183 170	CD	2 13 10	NOD2	0 13	O	0 20 15	CE2	0 0 76
CA	2 28 25	NOD1	10 22 21	B 5 HIS	118 139 192	CB	0 19 57	CD2	0 0 71
N	0 20 28	NOD2	17 28 13	CA	0 22 32	CG	0 7 42	OH	6 3 56
C	0 18 18	A16 LEU	0 165 420	N	2 1 25	CD1	0 12 59	B27 THR	33 84 259
O	1 25 20	CA	0 28 46	C	0 13 18	CE1	3 25 46	CA	0 13 54
CG	17 13 21	N	0 20 18	O	21 3 16	CZ	4 12 33	N	0 9 39
CG	41 14 6	C	0 13 26	CB	25 21 11	CE2	27 12 26	C	0 28
CD	0 23 13	O	0 14 41	CG	3 19 14	OH	24 7 21	O	0 6 25
OE1	22 11 25	CB	0 28 41	NOD1	1 29 10	B17 LEU	142 130 142	CB	0 25 47
OE2	0 36 22	CG	0 25 76	NE2	11 0 16	CA	5 24 25	OG1	13 16 16
		CD1	0 14 99	CD2	29 0 20	N	0 21 21	CG2	14 21 51
A 5 GLN	68 293 129	CD2	0 22 82	B 6 LEU	22 145 329	C	0 11 10	B28 PRO	12 107 265
CA	5 49 22	A17 GLU	79 206 153	CA	3 22 32	O	0 17 6	CA	0 17 41
N	0 20 15	CA	0 44 32	N	0 25 16	CB	0 19 32	N	0 6 11
C	0 42 11	N	0 28 23	C	0 25 16	CG	6 9 19	C	0 13 18
O	0 43 11	C	0 29 7	O	0 10 30	CD1	47 9 21	O	0 19 11
CB	2 49 14	O	0 20 15	CB	0 22 44	CD2	65 0 8	CB	5 13 60
CG	14 39 21	CB	2 33 32	CG	0 21 57	B18 VAL	44 173 171	GD	0 27 52
CD	0 20 5	CG	17 28 25	CD1	19 25 30	CA	8 36 9	CG	0 13 71
NOD1	5 24 26	CD	1 2 3	CD2	0 11 88	N	0 24 13	B29 LYS	166 99 125
NOD2	43 9 5	OE1	11 21 13	B 7 CYS	57 31 73	C	0 24 1	CA	0 11 21
		OE2	43 0 3	CA	2 0 22	O	0 20 7	N	0 13 28
A 6 CYS	0 125 242	A18 ASN	97 142 117	N	0 9 13	CB	0 22 38	C	0 25 10
CA	0 28 41	CA	8 27 16	C	0 3 10	CG1	6 38 43	CR	9 16 11
N	0 25 9	N	0 26 13	O	12 1 19	CG2	9 16 63	CG	28 9 17
C	0 22 28	C	0 16 15	CB	24 0 0	B19 CYS	8 114 183	CO	19 21 17
O	0 20 26	O	0 17 14 16	SG	25 17 9	CA	0 21 32	CE	52 14 5
CB	0 16 71	CB	17 21 21	B 8 GLY	35 66 122	N	0 20 17	NZ	21 15 6
SG	0 13 68	O	0 13 7	N	1 8 24	C	0 10 8	B30 ALA	131 58 31
		NOD1	22 13 24	C	0 18 23	O	0 12 24	CA	11 11 3
A 7 CYS	25 158 113	NOD2	32 13 7	O	3 14 34	CB	2 27 49	N	0 16 77
CA	0 39 22	A19 TYR	24 235 468	B 9 SER	36 64 196	SG	0 25 53	O	1 2 0
N	0 29 13	CA	0 19 38	CA	0 14 49	B20 GLY	36 84 96	OEND	34 5 0
C	0 28 13	N	0 15 16	C	0 17 17	CA	28 36 28	OEND144	0 0
O	0 24 19	C	0 21 25	O	0 19 26	N	0 30 11	CB	41 24 21
CB	0 30 32	O	0 18 34	CB	16 33 16	C	1 9 18	A* 1 GLY	76 60 93
SG	25 9 16	CB	0 21 84	CG	0 17 4	O	0 9 37	CA	36 27 28
		CG	0 14 45	NOD1	8 15 0	B21 GLU	131 62 86	NEND	23 13 18
A 8 THR	131 163 66	CD1	0 10 55	NE2	16 7 0	CA	5 16 21	C	0 13 22
CA	0 33 14	CE1	10 10 93	CD2	22 22 10	N	2 9 1	O	17 7 25
N	0 30 9	CZ	6 23 38	B10 HIS	129 193 121	C	0 6 10	A* 2 ILE	11 186 333
C	0 23 10	CE2	1 36 45	CA	3 35 22	O	0 13 28	CA	0 22 24
O	0 29 2	GD2	0 16 42	N	0 16 17	CB	22 8 11	N	0 9 15
CB	16 17 37	OH	6 32 19	C	0 17 25	CG	25 3 5	C	0 24 21
OE1	15 18 15	A20 CYS	16 153 163	O	0 19 26	CD	5 0 0	O	0 27 25
CG2	65 14 0	CA	0 30 35	CB	16 33 16	OE1	28 5 4	CB	0 14 35
		N	0 21 20	CG	0 17 4	OE2	44 0 0	CG2	2 13 81
A 9 SER	89 72 89	O	0 13 9 24	NOD1	8 15 0	B22 ARG	134 252 133	CG1	0 24 49
CA	5 21 21	CB	3 44 24	CE1	58 12 0	CA	11 22 14	CD1	9 33 65
N	0 20 11	SG	0 35 42	NE2	16 7 0	N	0 17 16	A* 3 VAL	21 135 264
C	0 10 21	A21 ASN	90 171 149	CD2	22 22 10	C	0 15 3	CA	2 14 36
O	10 7 24	CA	0 25 19	B11 LEU	0 178 341	O	0 13 20 9	N	0 9 26
CB	38 14 13	C	0 15 24	N	0 35 35	CB	19 32 8	C	0 11 15
OG	36 0 0	C	0 15 14	C	0 28 21	CG	0 35 24	O	0 13 26
		OEND	25 13 7	O	0 17 21	CD	19 36 24	CB	0 24 47
A10 ILE	84 232 118	OEND1	0 31 27	CB	0 15 30	NE	0 23 14	CG1	0 32 68
CA	0 44 11	CB	9 24 25	O	0 32 27	CZ	2 17 8	CG2	19 32 44
N	0 30 5	CG	0 12 3	CG	0 10 98	NH1	33 21 14	A* 4 GLU	67 126 275
C	0 25 9	NOD1	6 20 29	CD1	0 16 76	NH2	37 14 0	CA	3 14 39
O	9 24 3	NOD2	40 2 0	CD2	0 21 74	B23 GLY	0 93 165	N	0 13 18
CB	0 28 13	B 1 PHE	176 25 242	B12 VAL	0 198 388	CA	0 36 63	O	0 14 22
CG2	28 21 14	CA	13 0 17	N	0 22 49	N	0 22 31	O	0 17 27
CG1	31 33 33	NEND	23 0 24	C	0 18 21	C	0 18 31	CB	21 11 21
CD1	47 27 30	O	0 0 20	O	0 17 36	O	0 16 40	CG	35 3 33
A11 CYS	0 180 136	C	0 26 0 19	CB	0 11 73	B24 PHE	1 182 578	CD	0 8 32
CA	0 39 19	CB	9 2 33	CG1	0 0 103	CA	0 25 44	OE1	0 14 43
N	0 29 11	CG	0 6 24	CG2	0 19 74	N	0 16 28	OE2	9 12 40
C	0 24 13	CD1	0 23 0 4	B13 GLU	51 172 254	O	0 16 28	A* 5 GLN	75 206 161
O	0 32 20	CE1	0 40 0 0	CA	0 25 43	CB	0 28 46	CA	3 35 25
CB	0 28 35	CZ	32 9 14	N	0 24 23	O	0 24 49	N	0 28 14
SG	0 27 38	CE2	4 10 43	C	0 20 28	CG	0 26 49	O	0 29 20
		CE2	6 4 43			CD1	1 7 59	C	0 25 22
A12 SER	28 190 149					CE1	0 9 75	CB	24 24 14
CA	0 47 27								
N	0 34 10								
C	0 18 32								
O	0 26 29								
CB	25 49 22								
OG	3 15 29								
A13 LEU	35 108 317								
CA	0 16 36								
N	0 11 20								

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
		NOE1 17 2 24	O 0 24 27	NH2 16 37 18
		NOE2 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
	CD2 0 21 62	CD 0 5 23	CD2 0 22 79	N 0 25 17
A* 7 CYS 38 186 101	A*17 GLU 87 242 132	NOE1 25 1 14	B*15 LEU 0 182 422	C 0 29 20
CA 0 54 19	CA 0 47 21	NOE2 20 3 43	CA 0 182 422	O 0 21 27
N 0 39 14	N 0 28 15	B* 5 HIS 131 232 167	CB 0 28 44	CB 0 28 44
C 0 30 15	C 0 32 5	CA 0 17 33	CG 0 14 43	CD1 0 16 58
O 4 26 15	CB 0 32 5	N 0 1 26	CD1 0 16 58	CE1 0 13 78
CR 6 39 27	CG 5 29 27	C 0 8 20	CG 0 23 24	CZ 0 7 91
SG 27 10 12	CG 9 22 24	O 24 2 11	CB 0 27 25	CE2 0 4 79
	CD 2 0 13	CB 17 13 28	CG 0 19 76	CD2 0 7 46
	OE1 27 2 17	CG 0 26 20		
	OE2 40 1 6	ND1 15 6 6	B*16 TYR 48 158 387	B*25 PHE 74 157 402
		CE1 43 39 1	CA 0 30 35	CA 0 19 44
	A*18 ASN 104 104 100	NE2 0 43 7	N 0 12 50	N 0 16 22
	CA 0 18 14	CD2 0 48 14	C 0 15 31	C 0 15 31
	C 0 13 15	B* 6 LEU 25 148 341	CR 8 22 35	CG 5 13 35
	O 18 20 3	CA 3 17 41	CG 5 13 35	CD1 1 26 19
A* 9 SER 65 139 103	CB 24 6 27	N 0 11 21	CB 0 14 54	CE1 17 14 36
CA 14 21 13	CG 5 11 9	O 0 18 25	CG 0 1 39	CZ 0 23 40
N 0 23 9	NOE1 20 16 15	CG 0 17 41	CD1 1 4 39	CE2 17 9 48
C 1 8 29	NOE2 32 8 6	CG 0 14 52	CE1 17 12 30	CD2 1 12 42
O 5 20 29		CD1 21 21 51	CZ 3 22 21	
CB 16 25 19	A*19 TYR 62 182 355	CD2 2 27 74	OE2 0 29 38	B*26 TYR 9 139 662
OG 29 12 4	CA 6 22 25	B* 7 CYS 50 50 99	CD2 1 10 53	CA 0 19 51
	N 0 14 11	CA 2 5 19	OH 24 16 12	N 0 17 32
A*10 ILE 106 122 183	C 0 15 21	N 0 2 20	B*17 LEU 126 127 120	C 0 14 26
CA 0 24 36	O 1 19 28	C 0 6 14	CA 5 25 14	CB 0 16 58
N 0 9 17	CB 0 28 36	O 4 14 25	N 0 25 15	CG 0 3 93
C 0 17 20	CG 0 13 26	CB 32 8 8	O 1 14 8	CD1 0 1 63
CB 3 17 19	CE1 19 14 22	SG 13 23 6	CB 24 9 9	CE1 0 1 72
CD2 35 24 16	CE2 3 3 39		CG 0 21 8	CZ 0 1 75
CG1 6 13 36	CD2 3 10 56	B* 8 GLY 27 75 119	CD1 46 16 25	CE2 1 7 75
CD1 55 6 17	OH 25 9 20	CA 22 35 38	LD2 30 13 35	CD2 0 7 62
		N 5 3 18		OH 7 1 58
A*11 CYS 3 146 135	A*20 CYS 11 232 114	O 0 17 35	B*18 VAL 39 147 159	B*27 THR 56 64 224
CA 3 39 16	CA 0 51 22	B* 9 SER 32 107 189	CA 9 21 6	CA 0 11 41
N 0 18 15	N 0 25 18	CA 0 19 49	N 0 17 15	N 1 11 30
C 0 22 21	C 0 34 7	N 0 11 24	C 0 14 2	C 0 0 34
O 0 17 25	O 11 39 1	C 0 30 18	O 17 5 7	O 13 0 25
CB 0 39 16	CB 0 46 27	O 0 26 28	CB 0 32 27	CB 14 9 30
SG 0 10 43	SG 0 38 39	CB 6 17 57	CG1 3 43 35	CG1 11 3 25
		CG 26 3 13	CG2 9 16 66	CG2 17 28 39
A*12 SER 29 171 166	A*21 ASN 95 168 162	B*10 HIS 131 174 133	B*19 CYS 11 96 179	B*28 PRO 24 118 216
CA 0 39 35	CA 8 28 13	CA 2 39 28	CA 2 19 33	CA 2 9 33
N 0 29 14	C 0 13 26	N 0 32 9	N 0 15 20	N 0 6 20
C 0 19 30	O 2 16 13	C 0 16 32	C 0 2 11	C 2 7 7
CB 17 51 32	OE1 10 17 25	O 2 11 39	O 10 6 24	O 14 7 10
OG 2 20 24	OE2 126 25 2	CB 9 28 22	CB 0 22 43	CB 5 21 51
	OE2 126 25 2	CG 3 16 0	SG 0 32 49	CD 2 30 47
A*13 LEU 36 93 324	CB 32 22 8	ND1 8 7 0	B*20 GLY 27 73 102	CG 0 38 49
CA 0 13 39	CG 1 10 23	CE1 62 4 0	CA 27 27 33	B*29 LYS 147 97 101
N 0 13 22	NOE1 1 17 39	NE2 15 0 0	N 0 22 20	CA 8 19 16
C 0 11 22	NOE2 16 8 22	CD2 30 20 3	C 0 11 17	N 0 17 18
O 0 14 34	B* 1 PHE 196 30 219	B*11 LEU 3 169 401	O 0 13 32	C 0 3 16
CB 2 11 43	CA 11 0 19	CA 0 32 47	B*21 GLU 141 24 79	O 11 12 22
CG 0 5 43	OE1 31 0 6	N 0 29 22	CA 0 6 21	CB 2 28 11
CD1 16 27 51	O 1 0 11	C 0 24 25	N 3 6 3	CG 27 16 9
CD2 19 0 71	O 31 1 5	O 0 21 29	C 0 6 13	O 21 2 6
	CB 2 12 44	CB 3 28 41	O 7 3 18	CE 46 0 2
A*14 TYR 155 86 226	CG 0 5 38	CG 0 13 62	B*22 ARG 117 275 134	NZ 33 0 0
CA 2 19 22	CD1 0 13 42	CD1 0 11 87	CA 6 25 8	B*30 ALA 118 61 71
N 0 13 17	CE1 22 10 26	CD2 0 11 88	N 2 17 9	CA 2 16 19
C 0 10 11	CZ 38 0 9	B*12 VAL 0 183 397	OE1 43 0 3	N 0 6 22
O 5 17 18	CE2 39 0 6	CA 0 17 47	OE2 40 0 0	O 1 3 10
CR 24 9 13	CD2 22 0 13	N 0 20 29		OE1 36 5 0
CG 3 0 9	B* 2 VAL 122 38 73	O 0 17 28		CD 13 18
CD1 19 7 6	CA 5 3 9	CB 0 14 36		CB 66 17 2
CE1 33 4 14	N 1 0 13	CB 0 8 98		
CE2 7 0 26	O 0 6 10	CG1 0 8 98		
CE2 17 6 27	C 0 12 34	CG2 0 17 76		
CD2 14 0 33	CR 11 2 0			
OH 28 0 22	CG1 47 16 2			
	CG2 55 0 5			
A*15 GLN 85 205 176	B* 3 ASN 82 44 227			
CA 16 41 14	CA 0 8 39			
N 0 22 14	N 1 2 23			
C 0 25 13				
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.

TABLE 6
Contact information for lysozyme

1 LYS...	127 OYS 13	13 LYS 76	23 TYR 51	33 LYS 140	44 LEU 12	53 TYR...	57 GLN 17
3 PHE 110	120 ALA 8	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	68 SER...
39 ASN 32	13 LYS 5	23 TYR 12	25 LEU 20	115 CYS 12	43 THR...	57 GLN 67	69 THR 63
2 VAL 31	129 LEU 49	24 SER 9	29 VAL 11	32 ALA 9	44 ASN 57	50 ILE 66	64 CYS 63
41 GLN 27	7 GLU 40	28 TRP 5	115 CYS 11	36 SER 7	53 TYR 47	43 LEU 56	51 THR 60
38 PHE 9	14 ARG 33	12 MET 1	116 LYS 4	35 GLU 3	52 ASP 31	60 SER 23	59 ASN 42
84 LEU 3	6 CYS 21			111 TRP 2	40 LEU 30	42 ALA 17	66 ASP 31
	8 LEU 19	19 ASN...	28 TRP...	37 ASN 2	41 GLN 29	64 CYS 16	58 SER 23
2 VAL...	9 ALA 17	22 GLY 67	17 LEU 198		51 THR 22	81 SER 16	53 TYR 23
39 ASN 88	12 MET 12	18 ASP 63	105 MET 140	35 GLU...	42 ALA 22	52 ASP 10	72 SER 23
38 PHE 63	11 ALA 8	23 TYR 62	12 MET 99	57 GLN 96	57 GLN 16	68 ARG 9	62 TRP 21
1 LYS 99	128 ARG 6	24 SER 44	12 TYR 80	14 ALA 82		56 LEU 2	63 TRP 7
3 PHE 16	25 LEU 2	17 LEU 18	23 TYR 89	31 ALA 60			65 ASN 6
37 ASN 11		21 ARG 12	20 TYR 57	108 TRP 50	44 ASN...		58 ILE 5
4 GLY 4	10 ALA...	25 LEU 5	99 VAL 53	32 ALA 48	52 ASP 91	54 GLY...	73 ARG 4
	15 HIS 55	20 TYR 2	31 ALA 41	56 LEU 28	51 THR 48	57 GLN 66	
3 PHE...	8 LEU 51	28 TRP 1	27 ASN 35	109 VAL 23	57 GLN 43	84 LEU 32	
38 PHE 129	14 ARG 44		56 LEU 35	36 SER 21	46 ASN 28	42 ALA 31	
8 LEU 126	80 ILE 36	20 TYR...	8 LYS 21	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	
7 GLU 88	10 ALA 26	21 ARG 82	95 ALA 26	55 ILE 12	42 ALA 7	56 LEU 21	
88 ILE 45	3 PHE 21	17 LEU 67	32 ALA 28	37 ASN 7	35 GLU 5	36 SER 20	
40 THR 36	13 LYS 16	23 TYR 60	29 VAL 19	44 ASN 3	50 SER 1	53 TYR 14	
11 ALA 33	9 ALA 14	28 TRP 54	26 GLY 15	30 CYS 1		39 ASN 11	
95 ILE 30	12 MET 12	99 VAL 51	30 CYS 6		45 ARG...	55 ILE...	
2 VAL 28		16 GLY 43	18 ASP 4	36 SER...	58 ARG 124	41 GLN 5	
86 SER 27	12 MET...	18 ASP 30	96 LYS 3	55 ILE 75	51 THR 78	91 SER 3	
5 ARG 2	17 LEU 127	100 SER 26	21 ARG 1	39 ASN 72	49 GLY 59		
	8 LEU 92	22 GLY 24		32 ALA 52	50 SER 30	55 ILE...	
4 GLY...	7 GLU 77	19 ASN 6	29 VAL...	37 ASN 40	46 ASN 30	46 THR 82	
8 LEU 36	87 ILE 47		123 TRP 88	42 ALA 38	44 ASN 27	36 SER 84	
38 PHE 35	15 HIS 44	21 ARG...	25 LEU 50	57 GLN 31			
3 PHE 30	9 ALA 38	100 SER 120	26 GLY 49	38 PHE 31	46 ASN...	56 LEU 62	
6 CYS 19	25 LEU 32	20 TYR 73	38 PHE 42	35 GLU 24	50 SER 72	8 LEU 49	
2 VAL 12	92 VAL 26	23 TYR 63	8 LEU 38	33 LYS 19	45 ARG 44	39 ASN 44	
5 ARG 2	29 VAL 22	99 VAL 60	9 ALA 35	54 GLY 14	80 ASP 43	88 ILE 39	
	56 LEU 22	19 ASN 22	124 ILE 33	34 PHE 7	52 ASP 39	38 PHE 20	
	10 ALA 20	102 GLY 2	32 ALA 32		44 ASN 30	3 PHE 21	
5 ARG...	11 ALA 19		12 MET 31	37 ASN...	47 THR 20	91 SER 20	
123 TRP 189	32 ALA 7	22 GLY...	33 LYS 29	33 LYS 106	49 GLY 10	57 GLN 16	
38 PHE 83	14 ARG 7	49 ASN 71	5 ARG 27	34 PHE 67	51 THR 8	74 ASN 9	
29 VAL 52	16 GLY 4	23 TYR 34	28 TRP 20	39 ASN 47	59 ASN 6	35 GLU 12	
125 ARG 48	13 LYS 2	20 TYR 19	31 ALA 19	36 SER 19		92 VAL 2	
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	
8 LEU 29				35 GLU 5	48 ASP 24		
9 ALA 25	13 LYS...	23 TYR...	30 CYS...		49 GLY 24	56 LEU...	
7 GLU 17	129 LEU 101	105 MET 129	123 TRP 88	38 PHE...		108 TRP 101	
124 ILE 15	18 ASP 80	21 ARG 78	34 PHE 50	3 PHE 133	48 ASP...	58 ILE 84	
127 CYS 14	25 LEU 67	20 TYR 66	33 LYS 125	3 LYS 125	50 SER 37	58 ILE 67	
3 PHE 3	10 ALA 54	28 TRP 61	27 ASN 35	5 ARG 83	61 ARG 41	28 TRP 43	
126 GLY 2	9 ALA 31	19 ASN 57	26 GLY 31	8 LEU 70	46 ASN 31	35 GLU 37	
	16 GLY 30	27 ASN 56	29 VAL 22	2 VAL 64	47 THR 27	95 ALA 35	
6 CYS...	15 HIS 26	111 TRP 55	120 VAL 20	55 ILE 61	69 THR 4	54 GLY 26	
5 ARG 66	11 ALA 19	104 GLY 52	32 ALA 13	29 VAL 42	70 PRO 3	91 SER 25	
9 ALA 47	12 MET 18	99 VAL 41	114 ARG 10	32 ALA 32		22 ILE 17	
4 GLY 23	17 LEU 13	25 LEU 5	28 TRP 10	36 SER 32	49 GLY...	87 GLN 19	
10 ALA 22	14 ARG 4	106 ASN 4	31 ALA 5	4 GLY 31	45 GLY 47	32 VAL 13	
126 GLY 13		17 LEU 4	118 THR 1	123 TRP 16	69 THR 36	92 VAL 11	
8 LEU 10	14 ARG...	24 SER 3		40 THR 10	46 ASN 22	31 ALA 11	
127 CYS 6	15 HIS 66	18 ASP 3	31 ALA...	39 ASN 8	47 THR 16	53 TYR 9	
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	
129 LEU 2	13 LYS 24	24 SER...	28 TRP 37		68 ARG 11	105 MET 3	
123 TRP 2	16 GLY 16	27 ASN 101	27 ASN 33	39 ASN...	70 PRO 9		
	12 MET 10	19 ASN 47	115 CYS 27	2 VAL 108	50 SER 6		
7 GLU...		26 GLY 38	34 PHE 23	42 ALA 75		57 GLN...	
4 GLY 90	15 HIS...	28 TRP 30	56 LEU 23	44 THR 71		52 ASP 109	
3 PHE 87	14 ARG 85	120 VAL 26	105 MET 17	41 GLN 60	50 SER...	35 GLU 66	
1 LYS 49	92 VAL 75	18 ASP 16	33 LYS 16	55 ILE 43	46 ASN 78	53 TYR 68	
10 ALA 44	99 THR 69	22 GLY 12	29 VAL 15	37 ASN 35	69 THR 54	94 GLY 57	
11 ALA 28	11 ALA 60	121 GLN 12	30 CYS 14	1 LYS 34	59 ASN 51	42 ALA 26	
6 CYS 27	68 ILE 55	23 TYR 9	119 ALA 11	40 THR 14	60 SER 28	36 SER 25	
5 ARG 28	12 MET 53	124 ILE 6	108 TRP 8	54 GLY 5	51 THR 25	58 ILE 14	
9 ALA 14	13 LYS 31	25 LEU 6	32 ALA 6	38 PHE 5	45 ARG 20	108 TRP 12	
8 LEU 7	96 TYR 27		26 GLY 1		61 ARG 19	43 THR 12	
	17 LEU 23	25 LEU...		40 THR...	52 ASP 1	59 ASN 12	
8 LEU...	87 ASN 7	18 ASP 13	32 ALA...	55 ILE 47		55 ILE 9	
3 PHE 120	93 ASN 3	124 ILE 85	55 ILE 64	1 LYS 44	51 THR...	56 LEU 7	
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...	
55 ILE 47	18 ASP 39	29 VAL 49	36 SER 38	39 ASN 32	60 SER 97	83 LEU 87	
38 PHE 47	20 TYR 32	13 LYS 47	38 PHE 36	3 PHE 27	44 ASN 44	53 TYR 70	
28 VAL 46	13 LYS 30	24 SER 36	29 VAL 32	86 SER 26	68 ARG 43	93 ILE 68	
4 GLY 41	96 LYS 30	129 LEU 28	28 TRP 30	85 SER 26	50 SER 27	56 LEU 80	
5 ARG 57	12 MET 22	12 MET 22	31 ALA 22	8 ILE 24	49 GLY 57	63 TRP 55	
32 ALA 25	12 MET 16	27 ASN 44	8 LEU 17	83 LEU 22	59 ASN 25	94 CYS 47	
7 GLU 22	17 LEU 13	23 TYR 10	30 CYS 16	42 ALA 21	52 ASP 20	91 SER 36	
68 ILE 19	15 HIS 7	17 LEU 9	37 ASN 11	41 GLN 16	46 ASN 17	52 ASP 22	
6 CYS 16		19 ASN 4	34 PHE 7	28 PHE 13	66 ASP 16	64 CYS 18	
10 ALA 14	17 LEU...		12 MET 5		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	
28 TRP 2	12 MET 135	120 VAL 63		84 LEU 126	58 ILE 4	59 ASN 6	
	26 TYR 72	34 SER 41	33 LYS...	39 ASN 66		51 THR 7	
9 ALA...	92 VAL 75	39 VAL 38	38 PHE 115	1 LYS 35	52 ASP...	57 GLN 2	
25 LEU 55	96 LYS 38	124 ILE 34	37 ASN 109	43 THR 31	57 GLN 128	60 SER 2	
29 VAL 41	19 HIS 21	30 CYS 31	123 TRP 82	40 THR 24	44 ASN 107		
129 LEU 39	13 LYS 12	123 TRP 25	34 PHE 80	42 ALA 10	46 ASN 58	59 ASN...	
6 CYS 37	56 LEU 11	25 LEU 23	30 CYS 58	54 GLY 9	59 ASN 51	63 TRP 67	
5 ARG 31	95 ALA 11	28 TRP 17	29 VAL 32		53 TYR 46	52 ASP 56	
12 MET 30	19 ASN 9	121 GLN 6	35 GLU 20	42 ALA...	51 THR 40	68 CYS 18	
124 ILE 24	25 TYR 17	27 ASN 6	31 ALA 17	32 ALA 17	50 ARG 27	50 ARG 47	
13 LYS 21	23 TYR 8		31 ALA 15	97 GLN 30	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	
7 GLU 19		120 VAL 108		54 GLY 36	50 SER 1	51 THR 38	
11 ALA 14	18 ASP...	111 TRP 86	34 PHE...	41 GLN 22		64 CYS 26	
	25 LEU 106	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
68 ARG...	74 ASN 127	53 THR 58	89 THR 17	63 TRP 132	56 LEU 3	105 MET 9	26 GLY 8
45 ARG 120	75 ILE 35	43 THR 35	93 ASN 14	108 TRP 77	115 CYS 3	110 CYS 3	122 ALA 5
66 ASP 81	75 LEU 27	54 GLY 32	92 VAL 10	95 ALA 62	136 ASN...	103 ASN 2	122 ALA...
70 PRO 45	76 CYS 6	83 LEU 19	54 GLY 6	94 CYS 50	112 ARG 147	113 ASN...	125 ARG 174
69 THR 43	79 PRO 4	82 ALA 17	87 ASP 1	58 ILE 36	103 ASN 80	109 VAL 89	119 ASP 45
51 THR 40	63 TRP 3	80 CYS 17	92 VAL...	107 ALA 31	111 TRP 69	112 ARG 50	123 TRP 31
49 GLY 13	78 ILE...	42 ALA 11	17 LEU 79	101 ASP 29	116 LYS 57	114 ARG 44	126 VAL 19
53 THR 11	79 PRO 94	86 SER 6	88 ILE 79	100 SER 20	108 TRP 36	110 ALA 37	121 GLN 19
	76 CYS 61	85 SER 5	89 THR 62	97 LYS 19	104 GLY 19	111 TRP 12	5 ARG 18
	1 LYS 2	1 LYS 2	95 ALA 59	96 LYS 10	105 MET 6	115 CYS 5	124 ILE 14
69 THR...	74 ASN 58	85 SER...	96 LYS 57	99 VAL 6	107 ALA 6		
72 SER 72	64 CYS 52	87 ASP 84	15 HIS 55	76 CYS 5	23 TYR 5	114 ARG...	123 TRP...
66 ASP 72	82 ALA 51	82 ALA 57	91 SER 33	102 GLY 4		34 PHE 170	5 ARG 133
70 PRO 71	90 ALA 32	40 THR 24	56 LEU 25	103 ASN 3		110 ALA 82	30 CYS 128
60 SER 43	83 LEU 24	83 LEU 19	12 MET 21			113 ASN 41	33 LYS 82
50 SER 42	65 ASN 19	84 LEU 14	94 CYS 13	99 VAL...		118 THR 35	120 VAL 77
49 GLY 36	94 CYS 9	81 SER 12	90 ALA 12	21 ARG 64		115 CYS 22	29 VAL 68
68 ARG 33	63 TRP 9	88 ILE 7	93 ASN 11	104 GLY 63		111 TRP 22	34 PHE 40
61 ARG 25	83 CYS 8	86 SER 5	55 ILE 3	28 TRP 51		112 ARG 17	122 ALA 39
71 GLY 20	77 ASN 3	90 ALA 1		96 LYS 47		30 CYS 7	124 ILE 39
51 THR 7				96 ALA 44		116 LYS 4	121 GLN 25
48 ASP 5	79 PRO...			20 TYR 44			121 GLN 23
67 GLY 3	78 ILE 34	86 SER...	93 ASN...	108 TRP 37		115 CYS...	125 ARG 17
	65 ASN 62	1 LYS 38	89 THR 96	98 ILE 26		111 TRP 112	38 PHE 16
70 PRO...	81 SER 57	3 PHE 30	96 LYS 71	23 TYR 24		118 THR 82	127 CYS 5
72 SER 43	62 ALA 51	87 ASP 26	92 VAL 38	100 SER 23		120 VAL 49	119 ASP 3
68 ARG 40	84 CYS 15	85 SER 22	91 SER 19	105 MET 21		98 ILE 95	9 ALA 1
49 THR 19	74 ASN 10	40 THR 22	97 LYS 15	101 ASP 14		117 GLY 19	
69 GLY 11	83 LEU 5	88 ILE 13	95 ALA 12	97 LYS 8		106 ASN 45	124 ILE...
48 ASP 2	80 CYS 5	84 LEU 12	94 CYS 3	102 GLY 7		27 ASN 12	121 GLN 86
	77 ASN 9		15 HIS 1			111 TRP 44	129 LEU 75
71 GLY...	66 ASP 2	87 ASP...	94 CYS...	100 SER...		28 TRP 36	127 CYS 73
61 ARG 69		85 SER 97	98 ILE 49	21 ARG 118		107 ALA 34	25 LEU 51
70 PRO 36	80 CYS...	89 THR 73	91 SER 45	97 LYS 40		116 CYS 18	123 TRP 28
73 ARG 28	55 THR 16	88 ILE 24	58 ILE 39	99 VAL 27		113 ASN 2	9 ALA 27
73 ARG 27	65 ASN 52	90 ALA 10	97 LYS 42	102 GLY 26		112 ARG 19	5 ARG 26
72 SER 5	66 ASP 48	15 HIS 5	90 ALA 41	20 TYR 23		109 VAL 14	29 VAL 22
	70 SER 22	82 ALA 3	96 LYS 21	98 ILE 18		58 ILE 13	122 ALA 14
69 THR 64	78 ILE 16	83 LEU 1	83 LEU 24	101 ASP 7		104 GLY 12	125 ARG 12
61 ARG 48	82 ALA 16	91 SER 1	63 TRP 20			57 GLN 10	120 VAL 7
70 PRO 47	64 CYS 15		93 ASN 19	101 ASP...		97 LYS 37	118 THR 28
65 ASN 36	84 LEU 13	88 ILE...	92 VAL 18	97 LYS 37		113 ASN 69	27 ASN 3
60 SER 30	79 PRO 10	92 VAL 96	78 ILE 9	98 ILE 33		112 ARG 50	120 VAL 2
73 ARG 28	81 SER 5	91 SER 66	95 ALA 4	63 TRP 27		110 ALA...	118 THR...
74 ASN 4		15 HIS 40	64 CYS 1	100 SER 14		113 ASN 33	36 CYS 3
66 ASP 4	81 SER...	3 PHE 39	89 THR 1	103 ASN 10		112 ARG 16	123 TRP 1
	84 LEU 74	55 ILE 33		99 VAL 5		34 PHE 26	120 VAL...
73 ARG...	79 PRO 65	8 LEU 30	95 ALA...			114 ARG 68	119 ASP 64
62 TRP 86	80 CYS 19	11 ALA 25	98 ILE 64			109 VAL 59	114 ARG 25
75 LEU 68	83 LEU 17	12 MET 24	92 VAL 52			35 GLU 21	119 ASP 18
71 GLY 55	53 THR 16	90 ALA 18	108 TRP 46			108 TRP 43	116 LYS 17
61 ARG 43	85 SER 9	40 THR 16	56 LEU 32			113 ASN 33	36 CYS 3
72 SER 28	82 ALA 7	87 ASP 15	99 VAL 31			112 ARG 16	123 TRP 1
74 ASN 21		85 SER 13	91 SER 29			110 ALA...	120 VAL 61
60 SER 3	82 ALA...	86 SER 9	58 ILE 23			109 VAL 59	114 ARG 25
64 CYS 3	79 PRO 51	89 THR 6	28 TRP 21			97 LYS 37	113 ASN 69
	85 SER 43	83 LEU 3	94 CYS 20			112 ARG 50	120 VAL 2
74 ASN...	90 ALA 39	17 LEU 2	97 LYS 14			110 ALA...	118 THR...
77 ASN 102	81 SER 36	93 ASN 1	93 ASN 10			113 ASN 33	36 CYS 3
65 ASN 70	78 ILE 33		17 LEU 9			112 ARG 16	123 TRP 1
78 ILE 38	86 CYS 27	89 THR...	96 LYS 8			115 CYS 10	119 ASP...
64 CYS 33	83 LEU 20	87 ASP 74	90 ALA 1			31 ALA 9	125 ARG 69
75 LEU 24	84 LEU 18	93 ASN 77				111 TRP 7	121 GLN 58
76 CYS 22	87 ASP 4	15 HIS 74	96 LYS...			112 ARG 2	122 ALA 48
62 TRP 18		92 VAL 34	20 TYR 152			111 TRP...	118 THR 36
63 TRP 13	83 LEU...	91 SER 21	93 ASN 94			105 MET 217	120 VAL 14
72 SER 11	91 SER 83	88 ILE 19	92 VAL 82			116 LYS 174	123 TRP 4
73 ARG 10	56 ILE 74	90 ALA 17	99 VAL 59			115 CYS 129	124 ILE 21
79 PRO 3	80 CYS 69		17 LEU 43			27 ASN 92	6 CYS 16
	53 TYR 55	90 ALA...	100 SER 28			106 ASN 59	5 ARG 13
75 LEU...	90 ALA 42	93 ASN 53	94 CYS 28			23 TYR 44	123 TRP 82
63 TRP 85	82 ALA 34	83 LEU 49	15 HIS 25			107 ALA 21	31 ALA 54
62 TRP 81	64 CYS 33	89 THR 47	97 LYS 24			108 TRP 43	116 LYS 96
73 ARG 39	78 ILE 30	94 CYS 33	16 GLY 20			110 ALA 30	121 GLN 32
77 ASN 26	84 LEU 30	78 ILE 23	95 ALA 20			112 ARG 16	36 CYS 23
97 LYS 22	54 GLY 27	88 ILE 21	98 ILE 12			114 ARG 16	115 CYS 21
74 ASN 13	81 SER 26	82 ALA 21	28 TRP 6			109 VAL 11	24 SER 19
76 CYS 8	85 SER 24	87 ASP 13				113 ASN 11	122 ALA 14
	94 CYS 21	92 VAL 16				120 VAL 4	124 ILE 5
	40 THR 13	91 SER 8				34 PHE 2	119 ASP 3
63 TRP 96	79 PRO 7	85 SER 2				116 LYS 2	127 CYS 92
78 ILE 67	55 ILE 4					111 TRP 2	124 ILE 47
74 ASN 42	87 ASP 3	91 SER...				117 GLY 2	25 LEU 43
77 ASN 35	63 TRP 2	83 LEU 86				106 ASN 164	9 ALA 35
97 LYS 32	88 ILE 1	88 ILE 71				116 LYS 96	10 ALA 27
94 CYS 12		91 ILE 54				109 VAL 70	124 ILE 89
75 LEU 11	84 LEU...	94 CYS 42				113 ASN 36	125 ARG 81
98 ILE 4	41 GLY 102	95 ALA 37				108 TRP 22	119 ASP 60
62 TRP 2	81 SER 93	56 LEU 24				110 ALA 21	120 VAL 42
		90 ALA 26				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.