

MOLECULAR FACTORS STABILIZING PROTEIN CRYSTALS

F.R. SALEMME, Lars GENIESER, B.C. FINZEL, R.M. HILMER and J.J. WENDOLOSKI

Central Research and Development Department, E.I. du Pont de Nemours & Company, Inc., Experimental Station, 228/320, Wilmington, Delaware 19898, USA

Received 2 November 1987; manuscript received in final form 29 February 1988

Crystal properties of two proteins, cytochrome *c'* and hen egg-white lysozyme, are described in a context emphasizing characteristics of molecular structure important in lattice formation and stability. Major factors involved in lattice formation include protein conformational flexibility and incorporation of structured solvent at crystal contacts. Although protein surfaces provide multiple bonding sites, different crystal forms of lysozyme incorporate similar molecular chains. Molecular chain formation in proteins may be related both to crystal nucleation and development of macroscopic crystal habit.

1. Introduction

X-ray crystallographic studies provide the structural basis for understanding biological processes at the molecular level. Structural studies of biological macromolecules have revealed the atomic details of interactions that stabilize the three-dimensional architecture of proteins and nucleic acids, and features responsible for binding specificity and catalysis. Although the success of crystallography requires single crystals with a high degree of internal order, surprisingly little is known about the interactions that produce well-ordered crystals of biological macromolecules. The present work describes aspects of crystal properties of two proteins, cytochrome *c'* and hen egg-white lysozyme, that emphasize some characteristics of molecular structure important in lattice formation and stability.

2. Structured solvent and lattice contacts in cytochrome *c'* crystals

Cytochrome *c'* from *R. molischianum* is a dimeric heme protein composed of identical 128 residue polypeptide chains. The crystal structure of cytochrome *c'* has been determined by X-ray

methods and crystallographically refined to an *R*-factor of 0.19 at 1.67 Å resolution [1]. Although the refined structures of the dimer subunits are nearly identical, the monomers are related by a non-crystallographic two-fold symmetry axis (fig. 1) and the crystal asymmetric unit contains the molecular dimer. Each monomer consequently has a unique lattice environment and makes different packing interactions with adjacent protein molecules, despite the overall "symmetry" of the dimeric protein. The presence of the dimer in the crystal asymmetric unit allows the interactions of the crystallographically unique monomers to be compared, and provides a useful way for detecting subtle alterations in protein structure that accompany crystallization.

Cytochrome *c'* forms fairly typical protein crystals with about 45% of the unit cell volume occupied by protein and the remainder by solvent water or salt solution. X-ray refinement of the protein structure at high resolution [1] reveals that some 20% of the solvent volume in the cytochrome *c'* crystal is occupied by water with well-defined structure. Much of the ordered water is localized in areas between adjacent protein molecules in the crystal lattice, where extended patterns of hydrogen bonds can be formed to supplement and strengthen the few direct interactions which occur

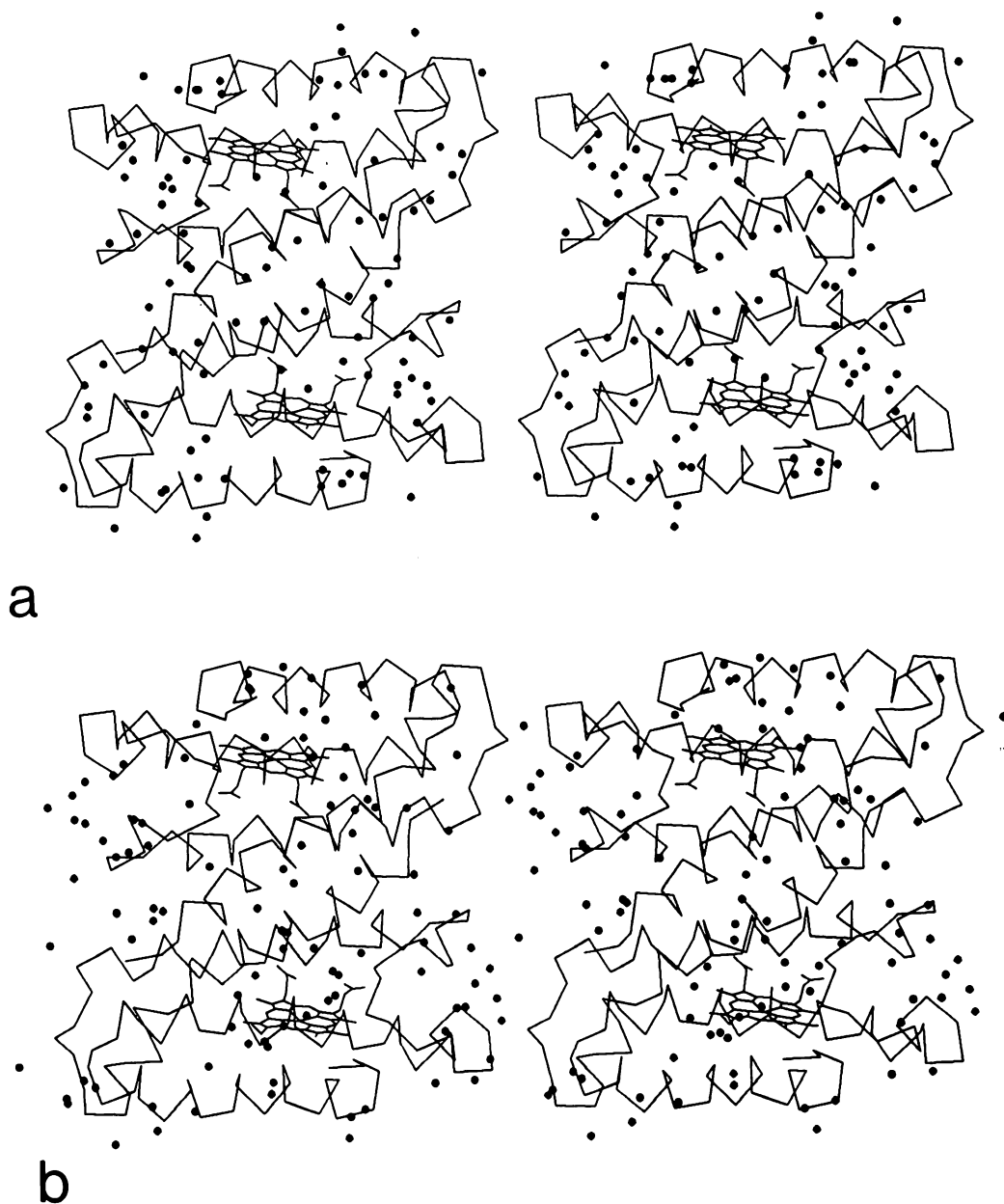


Fig. 1. Stereoview of the alpha carbon backbone of the cytochrome *c'* dimer showing positions of bound water molecules: (a) the subset of molecules which are found on both sides of the non-crystallographic molecular dyad symmetry axis; (b) asymmetrically situated "lattice" waters.

between the protein molecules. In fact, hydrogen bonding interactions mediated by intervening water molecules typically outnumber "direct" pro-

tein side chain interactions at lattice contacts by a factor of 3 or greater [2].

This pattern suggests that ordered solvent found

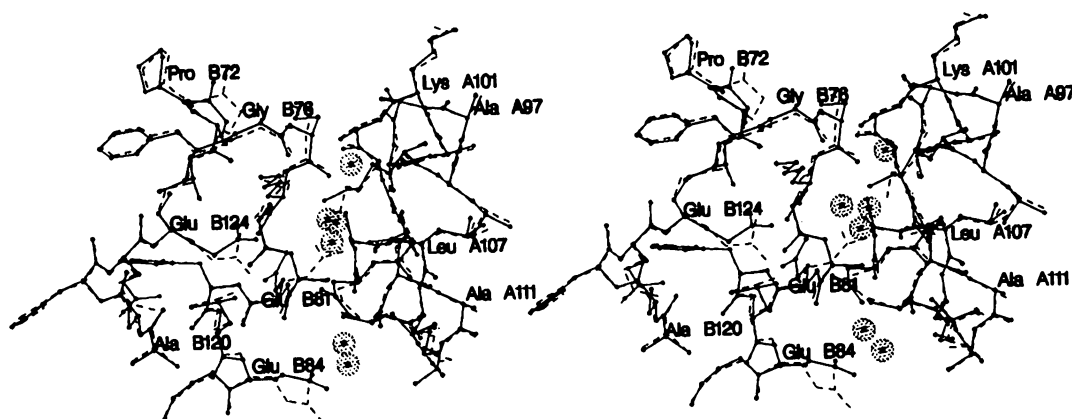


Fig. 2. Stereoview showing solvent bound at a cytochrome c' lattice contact (solid bonds and shaded spheres). The magnitude of structural changes due to lattice contacts is illustrated by the superimposed structures (dashed) of corresponding regions in the dyad-related dimer subunits.

in X-ray crystal structure serves as a molecular adhesive that fills in surface irregularities between proteins at the lattice contacts. However, physical studies have shown that proteins tightly bind substantial amounts of water that is important in maintaining structural and functional integrity [3]. It seems reasonable that a least some of this "structural" water would also manifest itself in the X-ray crystal structure. Indeed, of 194 solvent molecules localized in the refined cytochrome c' X-ray structure, 92 (46 from each monomer) occupy corresponding positions in crystallographically unique monomers (fig. 1a). These symmetry-related solvent molecules are probably representative of waters that bind tightly in solution and play a role in stabilizing the protein structure. As described elsewhere in greater detail [1], most of these structural solvent molecules make multiple hydrogen bonds with protein atoms, particularly polypeptide backbone atoms of α -helical regions of the structure. However, the remaining 104 solvent molecules are unique (fig. 1b) and clearly reflect the participation of solvent in extending networks of hydrogen bonding interactions at lattice contacts. This ordered "lattice" solvent thus does appear to supply a patchwork of interactions that structurally stabilize contacts between irregular regions of the molecular surfaces (fig. 2). An aspect of protein crystallization that differentiates it from the solution crystallization of many other compounds, is that the former structurally incor-

porate relatively large amounts of solvent at the lattice contacts during crystal growth.

3. Structural flexibility at lattice contacts

The crystallographically unique monomer subunits of cytochrome c' are related by a molecular dyad symmetry axis (fig. 1). As summarized in figs. 2 and 3, there are slight structural variations between monomers that are typical of differences observed between alternative crystal forms of the same protein (e.g. ref. [4]). These involve both small systematic shifts in polypeptide backbone positions, and some pronounced conformational shifts of amino acid side chains that participate differently in lattice contacts made by each monomer. For example (fig. 2), one of the most distinctively different side-chain conformers in cytochrome c' involves glutamic acid 84. In one monomer (B) this residue forms part of an apparent ion binding site at a lattice contact. This is inferred by participation of Glu-B84 in an extensive solvent network involving an aspartic acid side-chain (Asp-A105) from an adjacent molecule in the crystal lattice, suggesting that one of the bound solvent molecules is in reality a cation (e.g. ammonium ion) [5]. The conformation of Asp-124 is also sensitive to this difference in lattice environment. This, together with observations in other protein crystals where evidence exists for the

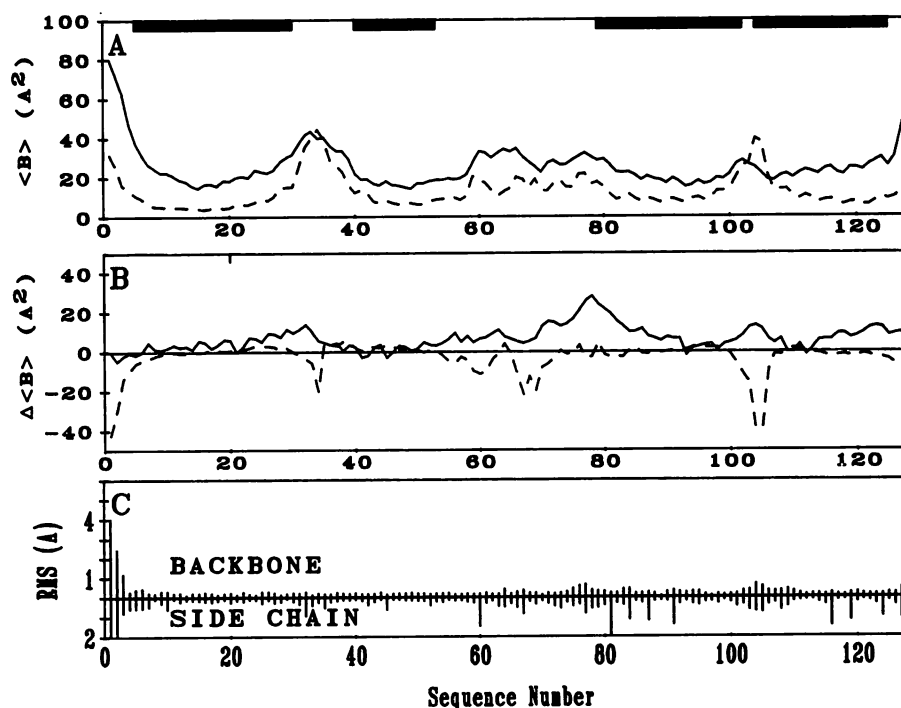


Fig. 3. Observed and simulated B -values for cytochrome c' subunits. (a) Backbone B -values as a function of sequence, taken as an average over dimer subunits. $B_n = \frac{8}{3}\pi^2\mu^2$, where μ is the RMS isotropic displacement in the atomic position of the n th atom. Solid curve shows the X-ray values [1] and dashed curve values computed from motions during 100 ps nonsolvated molecular dynamics simulation, sampled at 0.05 ps intervals (10). (b) Difference in subunit B -values from the X-ray refinement (solid) and computed from the simulation (dashed). (c) subunit structural differences in sidechain and backbone atoms from the X-ray structure (RMS on C- α is 0.4 \AA). Bars at top of figure indicate α -helical regions.

participation of alternative side chain conformations at lattice contacts [4], illustrates the role of side chain flexibility in allowing formation of specific interactions between protein side chains and/or bound ions.

More quantitative estimates of protein flexibility are experimentally determined during the X-ray refinement process. Refinement involves the variation of model atomic coordinates and mean amplitudes of atomic motion to minimize the difference between the experimental diffraction data and diffraction data computed from the current model structure. The refined atomic thermal parameters, or B -values, have magnitudes proportional to the atomic amplitudes of atomic motion in the crystal. Although many effects potentially contribute to X-ray B -value behavior, they nevertheless provide one of the few experimental measures of intrinsic protein flexibility [6–9].

Protein backbone B -values vary substantially within a given molecule, depending upon the degree of static or dynamic disordering within local regions of the structure. Typically, B -values are lowest in the protein interior or in sections with extended secondary structure stabilized by hydrogen bonds. For example, in cytochrome c' , sequence sections with low backbone B -values correspond to the α -helical regions, while chain termini or loops connecting helices are more mobile (fig. 3a). Nevertheless, crystal lattice interactions can modify solution protein mobility by trapping flexible surface loops at lattice contacts [6–8]. The magnitude of these effects can be estimated by comparing how the B -values differ for cytochrome c' subunits in their different lattice environments.

Fig. 3b shows the differences in cytochrome c' subunit B -values obtained from the crystal refine-

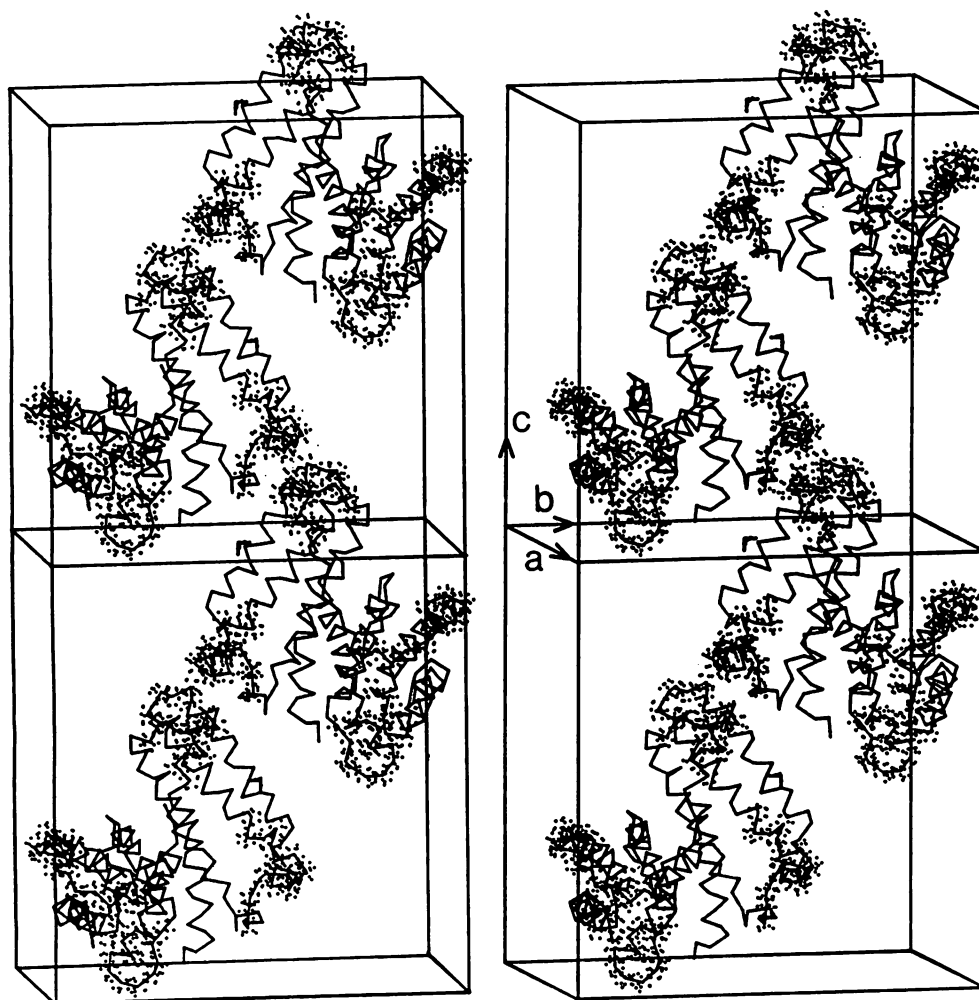


Fig. 4. Cytochrome c' lattice packing diagram. The orthorhombic $P2_12_12_1$ cell contains a dimer in the asymmetric unit. One dimer subunit forms continuous chains of molecules aligned along two-fold screw axes parallel to the c cell axis. Shaded regions show molecular lattice contacts.

ment. There is a net difference component between monomers that has been shown previously to vary systematically with distance from the crystallographic c -axis. This overall component appears to reflect some form of concerted or wagging motions of the dimer about a crystallographic screw axis along which one of the subunits forms a connected molecular chain (fig. 4) [6,7]. However, major differences in subunit B -values are localized in the loop regions that interconnect the helices in the molecule. Fig. 4 illustrates that the

major lattice contacts between cytochrome c' dimers in the crystals involve loop residues of one or the other monomer subunits. Since interactions at crystal contacts tend to immobilize otherwise flexible regions of proteins [8], the difference in observed loop mobilities can be attributed to the different lattice interactions formed at the flexible loops.

Also shown in fig. 3 are computed B -values obtained by averaging backbone motions from a 100 picosecond molecular dynamics simulation of

Table 1
Unit cell parameters for lysozyme (hen egg-white)

Form	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	α (deg)	β (deg)	γ (deg)	Space group
Triclinic	27.28	31.98	34.29	88.5	108.6	111.8	P1
Monoclinic ^{a)}	28.00	62.29	60.50	90.0	90.8	90.0	P2 ₁
Orthorhombic	59.40	68.70	30.80	90.0	90.0	90.0	P2 ₁ 2 ₁ 2 ₁
Tetragonal	79.10	79.10	37.90	90.0	90.0	90.0	P4 ₃ 2 ₁ 2

^{a)} Monoclinic cell contains a dimer in crystal asymmetric unit.

an isolated cytochrome *c'* dimer [9,10]. The average simulated behavior gives a good approximation of the experimental values (fig. 2a), a result illustrating that major features of the protein's dynamic behavior are similar both in solution and the crystal. Fig. 3b shows differences in monomer *B*-values computed from the molecular dynamics simulation. Somewhat unexpectedly, substantial differences in loop *B*-values occur in the simulation of the isolated molecule, in the absence of crystal lattice interactions. Results of numerous control simulations of protein dynamics suggest that this reflects the underlying flexibility of the loops which undergo relatively slow transitions among a large number of alternative conformational minima [9,11]. Apparently these alternative states are not uniformly sampled between monomers on the simulation timescale, so that statistical differences persist when subunit loop dynamics are compared.

The dynamics simulation of the isolated protein illustrates that many conformational substates are sampled by the loops, and that different ones are immobilized at the unique monomer crystal contacts during crystal growth. Similar phenomena are observed in the crystal structures of many other proteins determined at high resolution. In some cases alternative conformational substates are observed for side chains of the same protein in different crystal lattices, while in others, side chains of different molecules in equivalent environments of the same crystal have alternate conformations [1,12,13]. In crystals of cytochrome *c'*, the flexible loops make the dominant lattice interactions and immobilize the majority of the water molecules localized in the refined crystal structure [1].

4. Molecular chains in crystal formation: lysozyme

The pattern of molecular interactions in the cytochrome *c'* crystals shows that they are organized as an array of molecular chains. As described above, one of the dimer subunits forms head-to-tail interactions that produce continuous molecular chains. The subunits that form these chain interactions are aligned along the two-fold crystallographic screw axes oriented parallel to the *c*-axis of the P2₁2₁2₁ orthorhombic cell (fig. 4). The pendant dimer subunits, that are related to the chain-axis subunits by a noncrystallographic, molecular dyad, interconnect the *c*-axis chains by forming interactions in the *ab* crystal plane. The lattice as a whole is consequently organized as a three-connected net with one dimer subunit forming chains parallel to the cell *c*-axis, while the other subunits form links in the *ab* crystal plane between the chains. The more extensive and apparently stiffer interactions made along the *c*-axis direction suggest that aggregated chains might form as precursors to crystals. This suggests a model for crystal nucleation that involves formation of *c*-axis molecular chains, followed by their

Table 2
Recurring lysozyme molecular chains

Cell	Triclinic cell		
	<i>a</i> axis	<i>b</i> -axis	<i>c</i> -axis
Monoclinic (1) ^{a)}	<i>a</i> -axis	×	×
Monoclinic (2) ^{a)}	×	×	<i>a</i> -axis
Orthorhombic	×	<i>c</i> -axis	×
Tetragonal	×	×	<i>c</i> -axis

^{a)} Monoclinic asymmetric unit is a dimer.

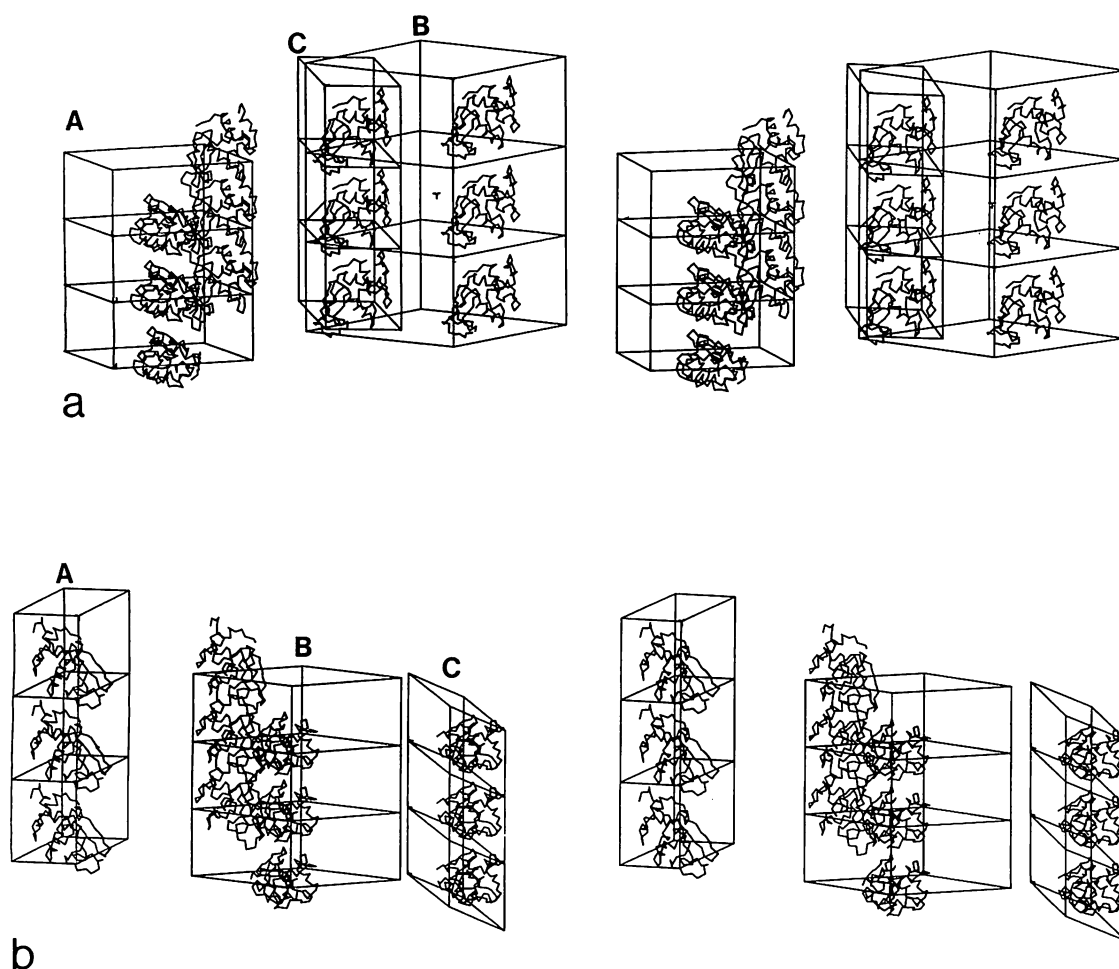


Fig. 5. Stereo views of common molecular bond chains in alternative crystal forms of hen lysozyme. (a) Views of monoclinic (A), tetragonal (B) and triclinic (C) cells illustrating the recurrence of the bond chain corresponding to the triclinic *c*-axis. (b) A shows the triclinic *c*-axis chain aligned with one subunit of the dimer asymmetric unit of the monoclinic cell (B), and C shows the triclinic *a*-axis array oriented with the other subunit of the monoclinic asymmetric unit (see also tables 1 and 2).

antiparallel association in a three-dimensional crystal nucleus.

In order to test the generality of chain organization in protein crystals, we examined several alternative crystal forms of hen egg-white lysozyme where detailed structures are known from X-ray crystallography [14]. As shown in table 1, lysozyme crystallizes in a variety of unrelated space groups. Nevertheless, careful examination revealed that several molecular chains were at least approximately shared among different crystal forms (table 2). In particular, chains corresponding to

the principal axes of the triclinic cell frequently occurred in other space groups. For example, fig. 5 shows molecular chains respectively derived from the triclinic cell *c*-axis, the tetragonal cell *c*-axis, and the *c*-axis array of one subunit of the monoclinic cell (which contains a dimer in the crystal asymmetric unit). The relative orientations of the molecules in the chains is remarkably similar despite differences in cell parameters and detailed intermolecular interactions. Interestingly, the second subunit of the monoclinic cell dimer forms *c*-axis chains that resemble the molecular chain

orientated along the triclinic cell *a*-axis (table 2). Similarly, the orthorhombic form appears to be assembled from *b*-axis chains that resemble the triclinic *c*-axis chains.

Thus far it has not been possible to trace the recurrence of molecular chains to specific molecular interactions that are conserved in different lysozyme crystal forms. Detailed patterns of intermolecular interactions involving the participation of bound solvent, that would be disrupted by only slight changes in unit cell parameters along the chain directions, do not appear to be preserved among the similar molecular chains in alternative lysozyme crystal cells. However, the variability in both the molecular orientations and axial repeats suggest that molecular attributes, such as surface loop flexibility, long range electrostatic field effects, or macrodipole interactions could be important factors in chain organization. The recurrence of the triclinic *c*-axis chain in several crystal forms may suggest that linear aggregates are im-

portant precursors of three-dimensional protein crystals [15], as does the observation that the several crystal forms can be viewed as the result of alternative ways of assembling the same primitive linear chains.

5. Molecular chains and crystal habit

Studies relating molecular crystal structure and macroscopic habit provide an alternative way of defining interactions that control crystal formation. Relationships between crystal habit and molecular structure depend upon chains of bonded interactions along principal directions in the crystal lattice [16,17]. Protein crystals are held together by a multiplicity of bonding interactions that are energetically similar and intrinsically flexible. As noted above, these typically include hydrogen bonding or ionic interactions that involve the participation of flexible amino acid

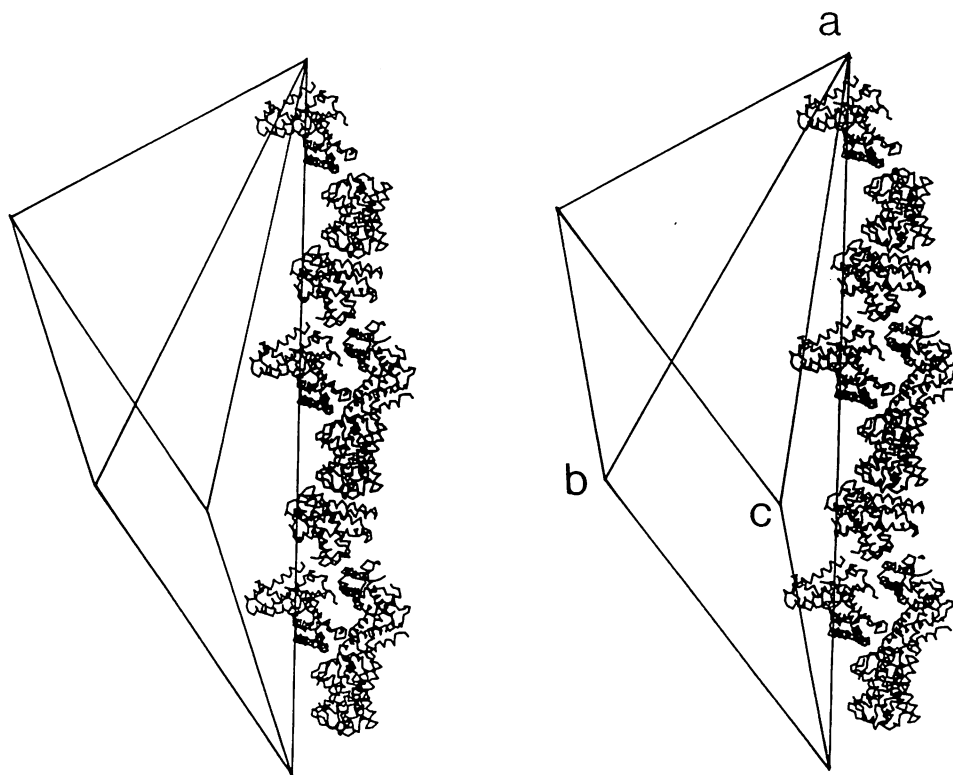


Fig. 6. A stereoscopic view of the molecular bonding chain lying along the [111] edge of the cytochrome *c'* crystal cell.

side-chains on the proteins surface, together with numerous solvent molecules or ions that are immobilized between molecules during crystal lattice formation. Despite the apparent flexibility and irregularity of the interactions that may potentially form at the surfaces of protein molecules, they commonly form crystals with well-defined and characteristic habits. Crystal geometries frequently resemble the unit cell shape, but other habits also commonly occur. For example, the orthorhombic cytochrome *c'* crystals with $a = 57$ Å, $b = 72$ Å, and $c = 75$ Å, form rectangular bipyramids with an apex formed by the a -axis, and (101), (110) developed faces. Clearly, the c -axis bond chain described above (fig. 5) appears in the [010] edge, and throughout the (110) crystal face. Other strong bond chains in the crystal appear to involve more complex interactions between molecules. For example, fig. 6 illustrates the molecular bonding chain along the [111] crystal edge. The chain incorporates a stepwise progression of links (composed of the "pendent" cytochrome *c'* subunits described above) between the c -axis bond chains, which also define the (110) and (101) crystal faces. Further studies of the bonded molecular networks that define macroscopic habits of protein crystals will undoubtedly give additional examples where both the developed crystal habit and pattern of molecular interactions appear consistent with the periodic bond chain model of Hartman and Perdok [16]. Extension of the molecular bond chain model to investigations of crystal habit variations with solution pH and ionic strength should provide additional insight into the forces stabilizing protein crystals.

6. Conclusions

Analysis of the high resolution crystal structure of cytochrome *c'* shows that several factors are involved in the formation of stable lattice contacts. These include a small number of direct, protein-protein, Van der Waals, hydrogen bonding, or ionic bonding interactions, together with the infrequent immobilization of a few solution ions (despite the frequent use of very high ionic strength conditions to grow protein crystals) and a

relatively large number of solvent water molecules. Lattice contacts are frequently formed of intrinsically flexible side chains or surface loops of the protein molecules. The surface flexibility of the protein surface groups, together with the incorporation of ordered solvent, allows the formation of extended hydrogen-bond structure networks that stabilize lattice contacts.

The surfaces of proteins are complex, and initially appear to provide a multiplicity of alternative bonding sites that could participate in lattice formation. Nevertheless, both the recurrence of similar bond chains in alternative lysozyme crystal forms, and the characteristic habits developed in many protein crystals, suggest that the effective lattice interactions between protein molecules are quite directional. This may suggest that at least one of the molecular chains manifest in the three-dimensional lattice exists as a linear precursor, whose association initiates formation of the crystal nucleus.

References

- [1] B.C. Finzel, P.C. Weber, K.D. Hardman and F.R. Salemme, *J. Mol. Biol.* 186 (1985) 627.
- [2] S. Sheriff, W.A. Hendrickson and J.L. Smith, *J. Mol. Biol.* 197 (1987) 273.
- [3] G. Careri, E. Gratton, P.-H. Yang and J.A. Rupley, *Nature* 284 (1980) 572.
- [4] P.C. Weber, S. Sheriff, D.H. Ohlendorf, B.C. Finzel and F.R. Salemme, *Proc. Natl. Acad. Sci. USA* 82 (1985) 8473.
- [5] This example points up a major difficulty in the interpretation of crystal lattice interactions: because occupancies and thermal factors of bound solvent molecules are correlated variables in X-ray refinement, it is difficult to distinguish a bound water from different bound ions that may scatter similarly to water owing to their binding with fractional occupancy or a high thermal B -value ($B = \frac{8}{3}\pi^2\mu^2$, where μ is the RMS amplitude of atomic motion in the crystal).
- [6] B.C. Finzel and F.R. Salemme, *Nature* 315 (1985) 686.
- [7] B.C. Finzel and F.R. Salemme, *Biophys. J.* 49 (1986) 73.
- [8] S. Sheriff, W.A. Hendrickson, R.E. Stenkamp, L.C. Sieker and L.H. Jensen, *Proc. Natl. Acad. Sci. USA* 82 (1985) 1104.
- [9] M. Karplus and J.A. McCammon, *Ann. Rev. Biochem.* 53 (1983) 263.
- [10] Molecular dynamics simulations of the cytochrome *c'* dimer were performed using the AMBER force field [P.K. Weiner and P.A. Kollman, *J. Comput. Chem.* 2 (1981)

- 287; U.C. Singh and P.A. Kollman, *J. Comput. Chem.* 5 (1984) 129] and a united atom model for the protein [S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta and P. Weiner, *J. Am. Chem. Soc.* 106 (1984) 765]. An initial 135 ps simulation incorporated a total of 2416 atoms, including hydrogen atoms bound to polar groups. The simulation used a distance-dependent approximation of the dielectric constant ($\epsilon_{ij} = r_{ij}$) to simulate water charge-screening effects, and included interactions between all atoms within a 9.5 Å interaction radius. Crystal coordinates of the protein dimer were energy minimized until the RMS change in coordinates was less than 0.01 Å, and the total RMS difference in C- α coordinates between the crystal and minimized structures was 0.65 Å. The trajectory was computed in time increments of 0.001 ps, with coupling to a 300 K thermal bath [H.J.C. Berendsen, J.P.M. Postman, A. di Nola, W.F. van Gunsteren and J.R. Haak, *J. Chem. Phys.* 81 (1984) 3684] and a temperature relaxation constant of 0.1 ps. The lengths of covalent bonds to hydrogen were fixed during the simulation using the SHAKE algorithm [J.P. Ryckaert, G. Cicotti and H.J.C. Berendsen, *J. Comput. Phys.* 23 (1977) 327]. *B*-values were computed from the dynamics trajectory by averaging coordinates sampled at 0.05 ps intervals.
- [11] R. Elber and M. Karplus, *Science* 235 (1987) 318.
- [12] J.L. Smith, W.A. Hendrickson, R.B. Hontzako and S. Sheriff, *Biochemistry* 25 (1986) 5018.
- [13] L.A. Svensson, L. Sjolin, G.L. Gilliland, B.C. Finzel and A. Wlodawer, *Proteins: Structure, Function, and Genetics* 1 (1986) 370.
- [14] F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer, Jr., M.D. Brice, J.R. Rodgers, D. Kennard, T. Schimanouchi and M. Tasume, *J. Mol. Biol.* 112 (1977) 535.
- [15] Z. Kam, H.B. Shore and G. Feher, *J. Mol. Biol.* 123 (1978) 539.
- [16] P. Hartman and W.G. Perdok, *Acta Cryst.* 8 (1955) 49.
- [17] G.A. Wolff and J.G. Gualtieri, *Am. Mineralogist* 47 (1962) 562.