PROTEINS: Structure, Function, and Genetics 14:65–74 (1992)

Variability of Conformations at Crystal Contacts in BPTI Represent True Low-Energy Structures: Correspondence Among Lattice Packing and Molecular Dynamics Structures

Anthony A. Kossiakoff,[1,2] Michael Randal,[1] Jeanmarie Guenot,[2] and Charles Eigenbrot[1]
[1]Department of Protein Engineering, Genentech, Inc., South San Francisco, California 94080; and [2]Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, California 94143

ABSTRACT **The structures of five basic pancreatic trypsin inhibitor (BPTI) molecules are compared to establish the extent and nature of the conformational variability resulting from crystal packing effects. BPTI is an ideal system to evaluate such factors because of the availability of high resolution X-ray models of five different BPTI structures, each in a different crystal packing environment. Differences observed among the structures are found to be distributed throughout the molecule, although the regions that display most variability are associated with the loop structures (residues 14–17 and 24–29). The regions of structure that show the largest rms deviations from the mean of the five packing motifs correlate well with the presence of intermolecular contacts in the crystal lattice. For most of the molecules there is also a correspondence between a larger number of intermolecular contacts and systematically higher *B*-factors, although it is not apparent whether this is induced by the crystal contact or results from the fact that the contacts are made predominantly through surface loops. The conformational differences seen among the X-ray models constitute more than local shifts at the lattice contact surfaces, and in fact involve in some cases the making and breaking of intramolecular H-bonds. The magnitudes of the differences among packing models are significantly larger than those usually associated with changes induced by mutagenesis; for instance, the structural differences at the site of mutation observed on removing an internal disulfide from the molecule are significantly less than those associated with lattice contact effects. The crystal packing conformations are compared to representative structures of BPTI generated during a 96-psec molecular dynamics (MD) simulation. This comparison shows a high level of correspondence between the protein flexibility indicated by the X-ray and MD analyses, and specifically between those regions that are most variable. This suggests that the regions that show most variability among the crystal packing models are not artifacts of crystallization, but rather represent true low-energy conformers that have been preferentially selected by crystallization factors.**

Key words: lattice packing effects, multiple low energy structures, basic pancreatic trypsin inhibitor, comparison of X-ray structures, molecular dynamics comparisons

INTRODUCTION

Crystals of the Val/Ala 30–51 disulfide mutant of BPTI contain two molecules in the crystallographic asymmetric unit. The X-ray analysis[1] demonstrated that the two independent molecules, although chemically identical, show significant structural differences in regions adjacent to the site of mutation. In fact, there are places in their structures where they differ more between themselves than either one does from the native molecule. Different conclusions would have been drawn about the effect of the mutation if only one or the other of the molecules had been analyzed. It was further determined that the

Abbreviations: BPTI, bovine basic pancreatic trypsin inhibitor; Ala/Ala, C30A/C51A, the double mutant of BPTI with cysteines at residues 30 and 51 replaced with alanine; Val/Ala, C30V/C51A, the double mutant of BPTI with cysteines at residues 30 and 51 replaced with valine and alanine, respectively; Thr/Ala, C30T/C51A, the double mutant of BPTI with cysteines at residues 30 and 51 replaced with threonine and alanine, respectively; Val**A**, one of the two unique molecules in the Val/Ala unit cell; Val**B**, one of the two unique molecules in the Val/Ala unit cell; BPTI(I), a wild-type BPTI structure, Protein Data Bank entry 4PTI; BPTI(II), a wild-type BPTI structure; Protein Data Bank entry 5PTI; BPTI(III), a wild-type BPTI structure, Protein Data Bank entry 6PTI; Type I, the crystal lattice characterizing the BPTI(I) structure; Type II, the crystal lattice characterizing the BPTI(II) structure; Type III, the crystal lattice characterizing the BPTI(III) structure.

Received August 29, 1991; revision accepted October 21, 1991.

Address reprint requests to Dr. Anthony A. Kossiakoff, Department of Protein Engineering, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080.

conclusions would have been different again if a different native BPTI structure had been used in the comparison. These findings provide some rather sobering insight into the nature and extent of crystal packing effects, even on molecules as stable and compact as BPTI.

The observation that major alterations in the chemical structure of a protein may in some instances alter the protein's tertiary structure less than the lattice packing environment has important implications for certain approaches used to study structure–function relationships in proteins by X-ray crystallography. A widely used approach in addressing structure–function issues is to make controlled changes at specific points in the primary structure by site-directed mutagenesis, and to evaluate the resulting perturbations in the tertiary structure by X-ray crystallography. Clearly, for this approach to be useful in assessing structural changes at or near the surface, consideration of packing effects is essential. Additionally, if a structural change resulting from an "internal" mutation affects surface groups, its nature has to be examined in light of the intermolecular environment.

The above points emphasize some of the issues inherent in studying proteins in their crystalline state. On the other hand, it is equally true that crystallization by its nature presumably induces few high-energy conformations in a protein structure. There is an extensive literature that supports the overall similarities of crystal and solution structures.[2–7] Although there is a clear interdependence between the conformation of exposed segments of chain and their packing environment, it is not necessarily correct to assume that these interactions confer nonnative conformations to the polypeptide chain. This suggests a question about which of two lattice packing effects is the more important. Is the conformational perturbation at the contact point largely the result of a set of energetic compromises required to form a stable contact interaction, or does the crystal contact result mainly because the inherent flexibility of certain segments of chain allow them to conform readily to structures that are conducive to forming stable packing surfaces?

The distinction between these models has broad implications for interpreting protein crystal structures, especially since in typical protein crystals about 25–30% of the protein surface is involved in, or is in close proximity to a packing contact.[8] If the former model is dominant, it means that the protein structure around the contact points is to some degree an artifact of the packing environment, although this might not affect the overall topology of the surface. On the other hand, if the second model predominates there is real and important structural information to be gleaned from the observed contact conformations. What such a situation implies is that the conformations of lattice contact residues are close to some true low-energy conformation. Extended to the present study, it would suggest that the small structural differences seen at the contacts represent a family of low-energy conformations that constitute part of the normal "breathing" of the protein structure. While no doubt both mechanisms influence the observed structural traits at the contact sites to some degree, it is important to assess which is the dominant factor, because the regions probably provide insight into the nature of the flexibility of the molecule.

To establish the extent and nature of the conformational variability resulting from crystal packing effects, we analyzed the structural differences among several crystal forms of BPTI and compared them to representative structures of the molecule obtained from a 96-psec molecular dynamics (MD) simulation. The multiple crystal forms of BPTI are an ideal system to evaluate the principal structural factors operating at the lattice contact points because of the availability of high-resolution structures of five different BPTI structures, each in a different crystal packing environment. It was found that the various crystal packing models differ not only at the contact surfaces, but also along adjacent segments of chain. These differences are more than local perturbations, and involve in some cases the making and breaking of H-bonds, especially those involving side chain groups. Comparisons between the X-ray structures and the MD simulations show a high degree of similarity, not only in the residues involved, but also in the extent of displacement. The model that best describes the family of molecular conformations seen in the packing models is one in which each packing type represents a low-energy conformer that has been preferentially selected by crystallization factors.

METHODS

Crystal Packing Models

Packing comparisons are taken from BPTI structures obtained from five different packing environments contained in four different space groups. Structures obtained from three of the space groups have been described and their coordinates are available in the Brookhaven Protein Data Bank.[9] These structures are described as BPTI(I) (4PTI),[10] BPTI(II) (5PTI),[11] and BPTI(III) (6PTI)[12] and have been refined to 1.7, 1.0, and 1.7 Å resolution, respectively. The fourth form, Val/Ala, is a mutant where the 30–51 disulfide group is replaced by a valine at 30 and an alanine at 51. There are two molecules in the asymmetric unit giving two independent packing motifs in the crystal form. The structure was refined and analyzed at 1.6 Å ($R = 0.17$).[1] The individual molecules are referred to as Val**A** and Val**B** in the text. Although Val**A** and Val**B** are mutants, they were included in this analysis because it was established that the effects of the 30–51 mutation

are relatively small compared to the shifts induced by intermolecular contacts.[1]

Calculation of Crystal Contact Zones

Sites of crystal close contacts were established by calculating for main chain atoms the number of main chain atoms from adjacent molecules within 6 Å. The range was selected to ensure the inclusion of interactions among molecules through water molecules or extended side chains, while avoiding overweighting the contribution of one or a few large side chains on the parent molecule. This criterion does not require residues within a contact zone to make direct contact with a symmetry-related molecule; rather it assumes that if residues with main chain atoms within the 6 Å range do not make contacts directly, they will nevertheless be affected by contacts through adjacent residues.

Molecular Dynamics Simulation

The coordinates for BPTI were from the BPTI(II) structure.[11] The 63 crystallographic D_2O molecules in the coordinate file were included in the model and all deuteriums were replaced with hydrogens. The protein and crystallographic waters were then solvated with a 4 Å water shell. The all-atom force field[13] was used as implemented in the AMBER 3.0 (Revision A) molecular mechanics and dynamics software.[14] The crystallographic and AMBER-added waters were minimized to a 0.5 kcal/mol·Å rms gradient with the protein held rigid. This was followed by 12 psecs of molecular dynamics on the waters to avoid spurious protein motions at the start of the simulation due to improperly placed waters. To keep the initial protein dynamics model close to the crystal structure during minimization, the system was minimized with positional constraints on the protein $NC_\alpha CO$ backbone. A force constant of 100 kcal/mol·Å^2 was used to minimize the system to an rms gradient of 0.5 kcal/mol·Å. The constraints were gradually removed by subsequent minimizations to a 0.1 kcal/mol·Å rms gradient using force constants of 50, 15, 2, and 0 kcal/mol·Å^2.

Large changes in the energy occurred during the initial 6-psec period, and these structures were omitted from our analysis. An additional 96-psec trajectory was calculated, using a 1.5 fsec time step with only bonds to hydrogen constrained with the SHAKE algorithm.[15] In order to prevent solvent evaporation and maintain equal partitioning of kinetic energy between solvent and solute, a linear distance-dependent dielectric was used with a multiplicative constant equal to 1. Rationale for using a linear distance-dependent dielectric with a small shell of explicit waters is provided by Guenot.[16] A 8 Å residue-based cutoff was used for all nonbonded interactions. The nonbonded list was updated every 10 steps. The system was coupled to a constant temperature bath with the temperature coupling constant set to 0.2 psec^{-1}. After the initial 6 psec, there remained a slight downward trend in the average total energy and average potential energy, as seen from these quantities calculated for the 12–30 psec (−4476 ±30 and −3249 ±21 kcal/mol) and for the 60–66 psec (−4564 ±28 and −3340 ±19 kcal/mol) intervals. The coordinates were collected every 0.15 psec during the trajectory, and the coordinate average was calculated from the 12 to 66 psec interval. Average structures were calculated over 6 psec intervals and compared to the starting structure (residues 4–56, inclusive). Although the rms deviations appear to be oscillating (Fig. 1), the oscillation is small, close to that previously reported,[16] and has completed its initial cycle at 66 psec. Instantaneous structures were taken from the simulation at 6 psec intervals, beginning at 12 and ending at 66 psec. Of these 10 structures, only 6 were required to derive the basic results, and some of the following analysis addresses only them explicitly. This choice of MD structures on which detailed structural comparisons would be made was designed to provide a representative sampling of the MD structures while limiting the total number of structures under consideration. These structures were not minimized prior to the analysis.

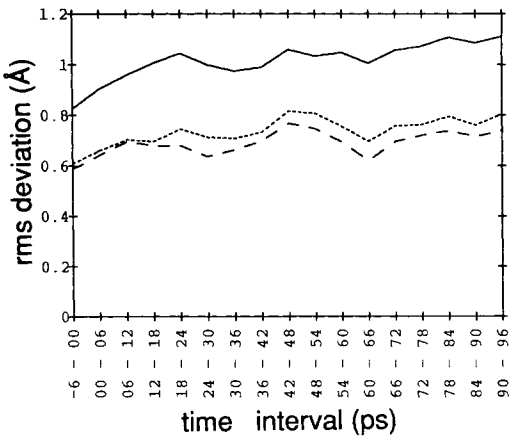

Fig. 1. The rms deviation from the starting structure as a function of time for all nonhydrogen atoms (solid line), main chain atoms (dotted line), and C_α atoms only (dashed line) for residues 4–56, inclusive, in the MD simulation. The 6 psec interval over which an average structure was calculated appears as the *x*-axis. The −6 to 0 psec interval comprises the MD equilibration.

RESULTS

Correlation of Intermolecular Contact Points With Variability in Packing Motifs

Figure 2 plots the main chain rms deviation of each residue from the mean position in the five packing motifs. A composite of the intermolecular contact zones is given in the upper curve. Discounting the very highly variable segments at the termini, the principal maxima of the rms differences

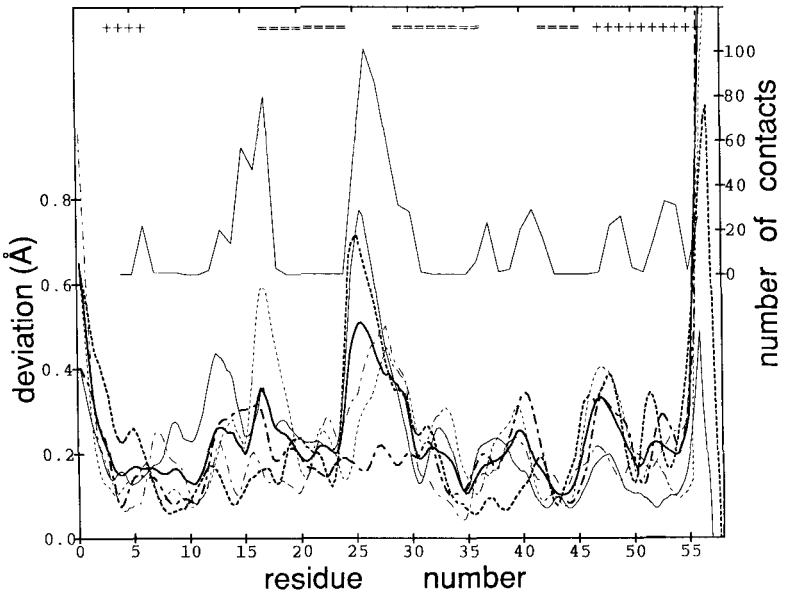

Fig. 2. Plot of main chain displacements of the five different packing types from the average structure, smoothed over a 2 residue (8 atom) window [BPTI(I) dotted line; BPTI(II) **bold** dot-dash line; BPTI(III) solid line; Val**A** dot-dash line; Val**B** **bold** dotted line]. The rms deviation of these five from the average is the solid **bold** line. The upper curve is the composite of the "crystal contacts" for all crystal packing types (solid line). The number of crystal contacts for each residue was calculated by counting the number of main chain atoms from neighboring molecules within 6 Å of each main chain atom of the reference molecule, and then summing across all the five packing types. The secondary structure is designated + α-helix, = β-sheet across the top of the figure.

are at residues 13–18, 25–30, and 46–49. The β-turn region of 23–29 displays particularly pronounced deviations with only the BPTI(II) packing having a conformation close to the average. The most prominent outliers are Val**B** and BPTI(III). When a plot of rms deviation is compared to the composite of intermolecular contacts in these cells, a striking correspondence is seen in the positions of the maxima of the two curves, indicating a correlation between the variability of the main chain and the number of packing contacts.

Variability in Temperature Factors Among the Packing Motifs

Figure 3 shows the pattern of *B*-factor variability of the main chain atoms for all the packing types; the points of intermolecular contact are identified at the top of the plots. On average, about 20% of the residues are involved to some extent in intermolecular contacts. There is clearly a preferred set of contacts, with the region 25–30 being contacted in all crystal forms. The second most highly populated contact is among residues 12–17 with Val**B** the only exception. There is a general correspondence between the higher *B*-factors and the points of contact, and although not every region of high *B*-factor is associated with a contact, almost every contact point is in a local maximum of the *B*-factor curve. This behavior is consistent with the fact that the sections of chain involved in interaction are predominantly surface loops, which are inherently more mobile.

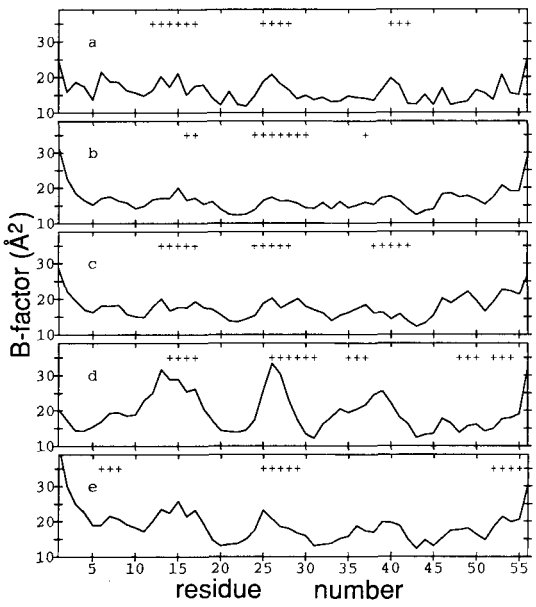

Fig. 3. Plot of main chain *B*-factors (Å^2) for all packing types: **(a)** BPTI(I), **(b)** BPTI(II), **(c)** BPTI(III), **(d)** Val**A**, and **(e)** Val**B**. The value of the *B*-factor is given by the average of the main chain atoms for each residue. The zones of intermolecular contact for each packing type are indicated by a + symbol at the top of their panel.

A distinctive feature of the plots is that among the packing motifs, Val**A** and Val**B** show the largest variations along the polypeptide chain. This may re-

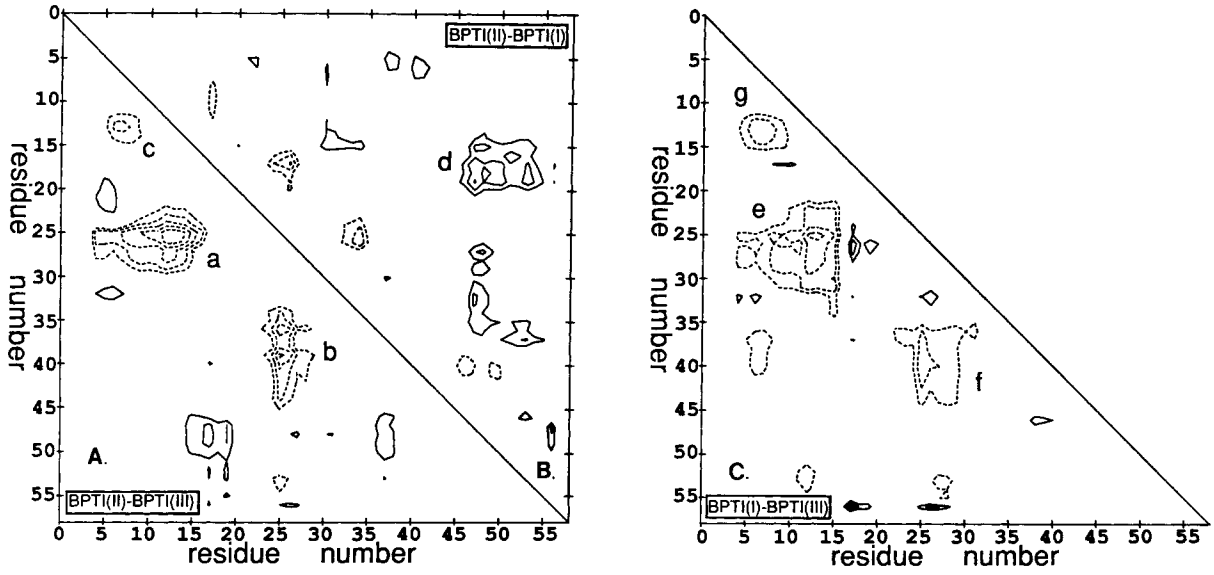

Fig. 4. Difference-distance matrix plots showing the changes in interresidue distances between crystal packings: **(A)** BPTI(II)–BPTI(III), **(B)** BPTI(II)–BPTI(I), and **(C)** BPTI(I)–BPTI(III). Solid contours indicate a decrease in an interresidue distance, dotted lines represent increased distances. Changes in distances of individual atoms in each residue were averaged to give overall residue displacements. Contour levels begin at 0.4 Å and are incremented in steps of 0.2 Å. The termini were not included in the analysis.

flect a somewhat less tightly packed environment in the Val/Ala case where the solvent content is 41% compared to the other crystal forms, which contain between 31 and 34% solvent. Because the number of contacts and their locations are similar for most of the packing motifs, the apparent dampening effect on the *B*-factors over most the polypeptide chain of the BPTI(I), (II), and (III) molecules is most likely due to the relatively tighter packing in the crystal lattice.

Figure 3 shows Val**A** to have the largest range of *B*-factors. Within the three segments of chain with the highest main chain *B*-factors (12–18, 24–29, 37–39) there are also extensive intermolecular contacts. In comparing contacts between Val**A** and Val**B**, only the 24–29 and 50–55 contacts are common; note, however, that the relative magnitudes of the *B*-factors for the three segments are essentially equivalent in these two packing types. This suggests that even though there is some difference in overall *B*-factor for the molecules [overall main chain *B*-factors: Val**A** (19 Å^2), Val**B** (15 Å^2)], the pattern of *B*-factors is not appreciably altered by the crystal contacts.

Changes in Intermolecular Packing

Difference–distance plots give a useful representation of the nature of the structural variability among the packing types, and provide information about how each region of the structure moves relative to all the others. These plots are constructed by calculating the differences in distance from each residue to all others in one packing type, compared to those of a second type. The position of a residue is taken as the average of the coordinates of its main chain atoms. A general feature of all the plots is that many of the major peaks fall proximal to lattice contacts (Fig. 2) suggesting a correlation between contact areas and regions that are most variable. Another principal feature is that structural differences extend over segments up to 10 residues in length. The extent of the effect is noteworthy since the protein contains only 58 residues.

In the difference–distance plot for BPTI(II)–(III) (Fig. 4A) the largest feature (peak a) indicates a much shorter distance between main chain residues 5–15 and 23–28 in the BPTI(II) molecule. The maximum, indicating a decrease in distance of about 1.1 Å, is centered around residues 13 and 25. Differences greater than 0.5 Å extend over 10 residues. This feature indicates a significant remodeling of the molecule. In addition, the loop 23–28 is also closer to residues 33–44 in BPTI(II) (peak b). Conversely, the sets of interatomic distances between both 15–20 and 35–38 and the C-terminal helix residues 45–52 are greater in BPTI(II) than in BPTI-(III). Smaller differences are seen between residues 4–8 and those of the three segments 11–15, 19–22, and 30–32.

The pattern of features derived in the BPTI(II) to BPTI(I) comparison (Fig. 4B) differs significantly from that described above for BPTI(II) to BPTI(III). The smaller magnitudes of the peaks also indicate that molecules BPTI(I) and BPTI(II) are more similar to each other than either is to BPTI(III). The major feature (peak d) shows that residues 12–19

TABLE I. Variable H-Bonding Interactions*

Donor		Acceptor		Average H-bond distance (Å)	Packing structure	MD occupancy
Arg-1	NH_1	Cys-55	O	3.0	II	0.02
Arg-1	NH_1	Tyr-23	OH	3.5	II, Val**A**	0.41
Arg-1	Nε	Cys-55	O	2.7	III, Val**A**, Val**B**	—
Arg-1	N	Ala-58	O(Term)	2.6	Val**A**	—
Cys-5	N	Pro-2	O	2.9	II, III, Val**B**	—
Leu-6	N	Asp-3	O	3.1	II, III, Val**A**	0.96
Thr-11	$O\gamma_1$	Val-34	O	2.7	All	0.13
Lys-15	Nζ	Pro-13	O	2.9	Val**B**	0.01
Asn-24	$N\delta_2$	Gln-31	$O\epsilon_1$	2.8	II, III	1.00
Asn-27	N	Asn-24	O	3.2	I, II, III, Val**B**	0.05
Asn-27	N	Asn-24	$O\delta_1$	3.2	All	0.78
Asn-43	$N\delta_2$	Glu-7	O	2.8	II, Val**B**	0.86
Asp-50	N	Ser-47	Oγ	2.9	I, III, Val**A**, Val**B**	—
Arg-53	NH_1	Asp-50	$O\delta_1$	2.6	I, Val**A**	0.82
Arg-53	N	Glu-49	O	2.9	I, II, III	1.00
Thr-54	$O\gamma_1$	Asp-50	O	3.0	I, III, Val**A**, Val**B**	0.94
Thr-54	$O\gamma_1$	Asp-50	$O\delta_2$	3.2	—	0.10
Gly-56	N	Met-52	O	2.8	I, II, Val**A**, Val**B**	0.04

*H-bonding interactions that are variable among the X-ray packing structures and the MD structures.—indicate no structures met the H-bonding criteria (distance <3.5 Å, >125°). "All" means all structures met the criteria. O(Term) is the C-terminal carboxyl group. The average H-bond distance is calculated from the X-ray structures. The MD occupancy is the fraction of all the MD structures generated between the 12 and 66 psec time points that satisfy the geometric criteria. Because coordinates were collected every 0.15 psec, there are 360 total MD structures considered in this analysis.

are closer to the section of C-terminal α-helix in BPTI(I), by up to 0.5 Å. On the other hand, the large differences that were seen in the BPTI(II)–BPTI(III) comparison between the 23–28 loop to both 5–15 and 33–44 (Fig. 4, peaks a and b) are not seen in the BPTI(II)–BPTI(I) comparison.

The relative degree of structural compaction between BPTI(I) and BPTI(III) can be readily inferred from Figure 4C. Essentially all the largest features are negative indicating a prevalence of shorter interatomic distances in BPTI(I). The largest of the peaks (Fig. 4C, peak e) indicates that segments 3–15 and 21–31 are significantly closer (0.4–0.7 Å) in BPTI(I) than in BPTI(III). A similar, but less extensive peak was found in the BPTI(II)–BPTI(III) comparison (Fig. 4A, peak a). The second largest feature (Fig. 4C, peak f), between 24–28 and 35–43 showing differences up to 0.6 Å, also superimposes on the BPTI(II)–BPTI(III) comparison. Overall, the number of negative contours compared to the positive ones suggests that the BPTI(I) molecule is somewhat more compact than either the BPTI(II) and BPTI(III) molecules.

Summarizing the general patterns found among the difference–distance matrices, the regions of highest variability among the packing types involve segments of the polypeptide chain that are in loops or extended conformations. In particular, residues flanking and including the β-turn region (residues 23–29) and the residues (10–18) making up the specificity loop of the inhibitor are almost always affected. It is noteworthy that the three disulfide bridges, 5–55, 14–38, and 30–51, are contained in regions that show comparatively large variability. Their role, therefore, does not appear to involve rigidifying the native form of the structure.

Hydrogen Bonding Differences

X-ray structures

The variability of the intramolecular H-bonding pattern is a good indicator of regions of local flexibility in the molecules. The making and breaking of H-bonds can give specific information about the stability of individual secondary structure units. Table I is a compilation of those main chain and side chain H-bond interactions that show most variability in either or both the MD and X-ray structures. These changes in H-bonding structure complement the difference–distance matrix map data discussed above, which identifies structural variability over longer distances but is less detailed. By and large, the interactions that deviate between packing types involve at least one side chain donor/acceptor contribution; however, there are several main chain–main chain interactions that differ, especially within the C-terminal helix. The R53 N to E49 O H-bond is formed in the BPTI(I), (II), and (III) molecules, but is absent in Val**A** and Val**B**; the G56 N to M52 O in-

teraction is found in all the molecules with the exception of BPTI(III).

Molecular dynamics structures

Molecular dynamics (MD) simulations similarly show a variable population of H-bonds, especially those that are weakest based on H-bonding energy criteria. Variability is more common when side chain atoms are involved than when only main chain atoms are involved. Because coordinates were collected every 0.15 psec, the period from 12 to 66 psec provided 360 individual structures. H-bond occupancies were calculated as the number of structures with an H-bond divided by 360. Although there was some variation in H-bonding pattern for the side chain groups within these structures, the majority of main chain H-bonds (23 out of 26) was conserved in virtually all structures (occupancy greater than 0.90). The "nonconserved" H-bonds were characteristically long (>3.5 Å) or distorted (> 60° off linear) and their absence in some of the structures results from the interaction falling just outside the accepted H-bond criterion rather than representing a different structure.

In a related MD simulation, van Gunsteren and Berendsen compared conformation differences between BPTI molecules in the crystalline (MD_{crys}) and solution (MD_{sol}) states.[17] Within the listed H-bonds there are several that differ between the two states. As in the case of the simulation described here, there is a core of conserved H-bonds in MD_{crys} and MD_{sol}. Some reasonably striking differences were found, however. The β-sheet H-bond 18 N to 35 O, which is found in MD_{sol}, is broken in the MD_{crys} model. H-bonds 21 N to 45 O and 45 N to 21 O are found in all X-ray structures, our simulation and MD_{crys}, but not MD_{sol}. The α-helix H-bond 56 N to 52 O is found in neither MD_{crys} or MD_{sol}, but is found in all X-ray structures except BPTI(III). In the simulation presented here, the 56–52 H-bond is distorted and an alternate H-bond is made from 56 N to 51 O.

MD Difference–Distance Matrices

The difference–distance matrices comparing the individual structures sampled along the trajectory to their average (MD_{ave}) display many of the same characteristics seen in the X-ray structure comparisons. One difference, however, is that the MD structures appear to be somewhat more variable, that is, if the matrix maps are contoured at the same level used for the X-ray structures (start at 0.4 Å) there are many more small features distributed throughout them. This suggests a difference between the X-ray and MD matrices of the baseline defining significance. This effect was minimized by setting the lowest contour level of the MD plots slightly higher, at 0.5 Å.

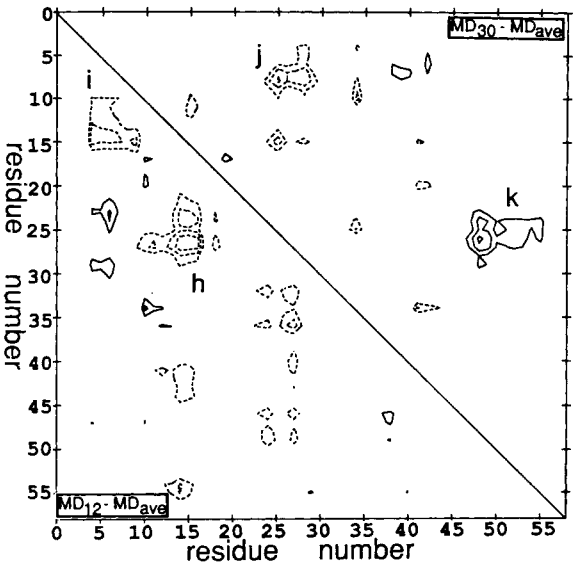

Fig. 5. Difference–distance matrix plots for two representative MD structures: **(a)** MD_{12} and **(b)** MD_{30}. The same conventions are used as in Figure 4. Contour levels begin at 0.5 Å and are incremented by 0.2 Å. MD_{12} and MD_{30} are the structures sampled after 12 and 30 psecs, respectively.

Figure 5 shows representative difference–distance matrix maps for two MD structures compared to the average of the selected MD structures (MD_{ave}). In the MD_{12} map the principal features indicate that the distances of residues 10–16 to residues 22–28 and of residues 10–16 to residues 4–10 is smaller in MD_{12} than in MD_{ave} (Fig. 5a, peaks h and i). These same two features are prominent in the BPTI(I)–BPTI(III) and BPTI(II)–BPTI(III) comparisons (Fig. 4, peaks a, c, e, and g). In MD_{30}, the segment 24–29 is closer to 5–9 than in MD_{ave} (Fig. 5b, peak j); however, 23–28 to the C-terminal helix is farther away in MD_{30} (Fig. 5b, peak k). The feature indicating different structural relationships between the β-turn residues, 24–28, and the segment containing the C-terminal helix and the loop leading to it is a common element in the comparisons of the X-ray packing structures. Another consistent feature in the X-ray packing comparisons indicates that the relationship between segments 12–20 and 45–53 is highly variable (Fig. 4). Although this is not seen in the two MD examples given above, it is a major feature of the MD_{66} matrix map (not shown). The principal conclusion drawn from these comparisons is that, taking into consideration the different experimental origins of the models, the variability in the X-ray derived packing structures and the MD simulation structures show striking similarities, not only in the residues involved, but also in the extent of displacement seen among the different models.

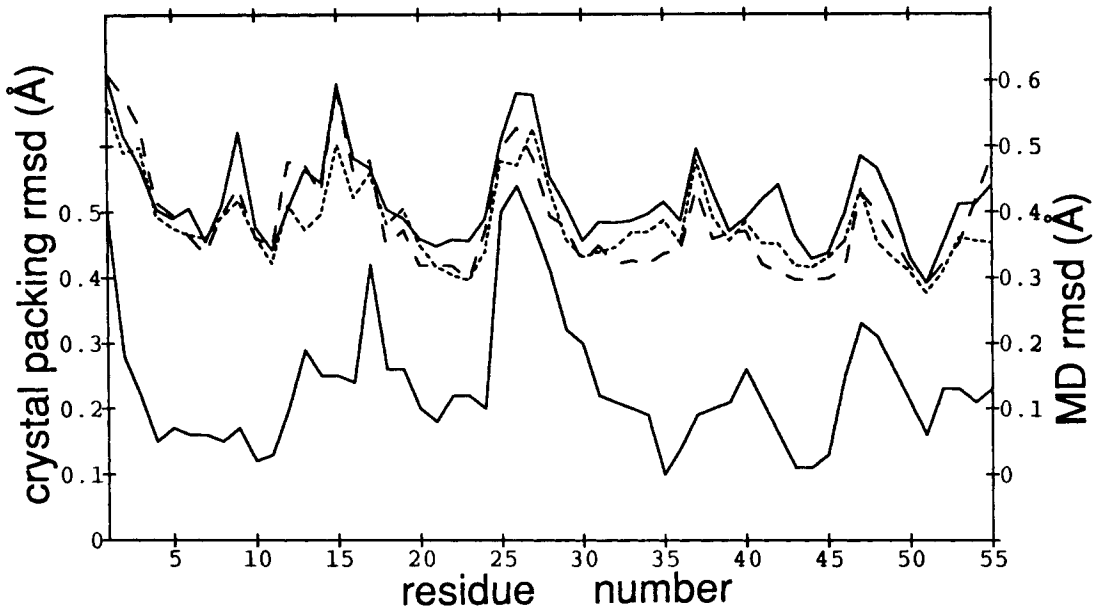

Fig. 6. Time dependence of the rms deviations of MD structures. The rms deviation from the 18 psec average structure taken over three different 18 psec time periods, 12–30 psec (solid line), 30–48 psec (dotted line), and 48–66 psec (dashed line) (all non-hydrogen atoms). Overlaid on this curve is the rms deviation of all the packing types as indicated in Figure 2 (the lower solid line). The scale for the MD rms has been shifted up by 0.1 Å for clarity.

DISCUSSION

Variability in Molecular Simulation Structures and Lattice Packing Structures Are Correlated

The H-bonding comparisons between packing types and MD structures point out that the structures being sampled, though certainly not identical, are nevertheless related (see above). A similar type of comparison is shown in Figure 6 where the variation among a set of MD structures (solid line) is plotted along with that seen in the 5 packing types. A striking characteristic of the plot is the strong correlation between the lattice packing and MD structures, not only in their general features, but also in their fine structure. In conjunction with the packing-contacts curve in Figure 2, Figure 6 again shows there is a clear correlation between variability and contact site. The major features of the curves show structural variability exists in segments 12–19, 24–28, 36–42, and 46–49. A similar MD displacement pattern has been reported by Levitt.[18] In addition, an analysis defining the anisotropic motions of BPTI by Diamond using crystallographic *B*-factors gives similar results.[19]

The MD structures determined here appear to favor the model of inherent regional flexibility. If the structures at the contact points were highly ordered, but deformed through crystallization effects, features similar to those of the packing structures should not be as strongly emphasized in the simulations as they are. Other corroborating evidence comes from changes seen in mutant structures where the affected groups were associated with a contact point. The 30–51 disulfide mutant Ala/Ala crystallizes in the BPTI(II) cell, and has contacts in the β-turn loop 24–28. It is along this loop segment in which the largest differences between the BPTI(II) native and the mutant are found. It is difficult to rationalize how this would be the case if the conformation of the loop were dictated solely by close contact energies. In fact, it is somewhat surprising that the loop can move to the extent that it does (0.3–0.5 Å) while retaining the contact interface. Although this suggests that contact points have a plasticity associated with them, there is clearly a limited number of low energy structures on which productive crystal contacts can be built. The 30–51 disulfide mutant with Thr at position 30 seemingly cannot be induced to crystallize. Thus, the progression of crystal formation goes Ala/Ala in the BPTI(II) cell, Val/Ala in a unique space group, and Thr/Ala giving no crystals at all. Because the largest changes in Ala/Ala and Val/Ala are at the 24–28 crystal contact, the apparent poor crystallizability of Thr/Ala suggests that its associated low-energy structures are not compatible with forming a productive contact surface.

Structural Differences as a Function of Distance From Intermolecular Contact Regions

Is there a radial distance beyond which a region of a protein structure is not significantly affected by its proximity to an intermolecular contact interface? If

this were the case, it would be possible to ascribe structural differences to factors other than the intermolecular environment if the regions under comparison were sufficiently remote from contacts. Clearly, this issue is extremely important for establishing criteria to interpret structural information and define the relevance of small changes from X-ray studies of site-directed mutants.

For the BPTI(II) X-ray model, the deviation from the consensus BPTI model of each main chain atom that was within 6 Å of one of the 4 major contact regions was plotted as a function of its distance to the nearest contact point. The result was a random scatter of points suggesting little or no correlation (data not shown). Although this finding indicates that within at least 6 Å of the contact surface atoms are systematically influenced by packing factors, two caveats need to be considered. First, BPTI is a small protein and there are few residues more than 8–10 Å from a contact surface. It is assumed that the larger the protein, the less packing interactions will have an effect. Second, the uniform distribution of atomic shifts may reflect differences in real low energy conformers. This being the case it would be expected to have more widely distributed changes than if the effect were due to crystal contact forces alone. This view is supported by the MD structures that had small, but measurable, differences distributed throughout the molecule.

Structural Differences at Similar Contact Surfaces

It is not unusual for protein molecules that crystallize in different space groups to share one or more identical sets of packing interactions. It might be expected that in situations where two packing motifs have a contact interface that includes equivalent sets of contact residues on the symmetry related molecules, the observed conformations of the included residues would be highly similar. In both BPTI(I) and BPTI(III) packings, the β-turn residues, 24–27, share an interface with residues 14–19. The curves shown in Figure 2 indicate that BPTI(III) has the largest displacement of all the packing types at the β-turn; the equivalent displacement for BPTI(I) is about one-half as large. Additionally, these two structures differ in the 14–19 region. The difference–distance matrix map (Fig. 4C) also shows that the largest differences between the two molecules involve the above segments of chain. These two comparisons demonstrate that having similar contact surfaces does not guarantee that the structures of the residues involved will be highly similar.

CONCLUDING REMARKS

Typically, about 25% of a protein molecule's surface in a crystal is close to or directly associated with a crystal packing contact.[8] Because many of the physiologically important functions of proteins involve surface residues, it is clearly important to understand the influence packing contacts have on the conformation of adjacent residues. It is particularly relevant to determine the extent to which packing contacts can perturb conformations. BPTI has proven to be a rather unique system with which to study these issues because structures of the molecule in 5 different packing environments are available and its small size accentuates the packing effects. Analysis of the different packing structures in combination with representative MD structures suggests the following conclusions:

1. The magnitude of the changes, the backbone conformations, and H-bonding lengths are similar in the X-ray packing and MD models. The same regions of the molecule consistently show the largest variations in both analyses. If the MD models are compared to each X-ray model independently, the differences found are greater than if the X-ray structures are combined and compared as a group. Some of the differences that have been found comparing MD and X-ray models probably arise because the X-ray structure used is a time averaged model of an envelope of structures which are a subset of a family of low-energy conformations.

2. Side chain H-bonding interactions are more variable than those of the main chain. Comparing interactions among all the structures, they sometimes involve groups having different H-bonding partners. It is apparent that if only one of the X-ray (or MD) structures had been analyzed, several H-bonding interactions would not have been identified.

3. The interplay between intermolecular contacts and their effects on the *B*-factors of groups near those contacts is unclear. Sheriff et al.,[21] in a comparison of hemerythrins, make a compelling argument for a significant reduction of atomic mobilities at contact interfaces; a conclusion also reached by Phillips[22] when comparing the pattern and magnitude of *B*-factors from two crystal forms of myoglobin. Additionally, Artymiuk et al.[23] reported smaller, but consistent perturbations with two crystal forms of lysozyme. In BPTI there are also definite effects of crystal packing in the observed *B*-factors, but it is not apparent that there exists a direct correlation between the presence of a contact and a resulting dampening of atomic mobility as was suggested in the hemerythrin and myoglobin cases.

4. The nature and extent of the changes found in mutant structures, even when compared to their native counterpart in an isomorphous cell, are affected by lattice contacts. As was seen in the Val/Ala mutant, which has two molecules in the asymmetric unit, the perturbations at the site of mutation differed depending on the packing environment.[1]

5. The structural comparisons among the X-ray packing and MD models suggest that the differences in conformations within the X-ray structures repre-

sent a family of low-energy conformations that are an inherent part of the normal "breathing" of the BPTI molecule itself. Although a majority of the differences are located at the crystal contacts, they are not found there exclusively and are distributed throughout the molecule. Consequently, the differences seen among the X-ray structures are not predominantly induced by crystal packing forces, and taken together give a considerably more accurate description of the conformation of the BPTI molecule than any one of the models alone could give.

ACKNOWLEDGMENTS

The authors are grateful for useful discussions with Peter Kollman and Steve Anderson, and support for J.G. through NIH Grant GM-29072 to Peter Kollman.

REFERENCES

1. Eigenbrot, C., Randal, M., Kossiakoff, A. A. Structural effects induced by mutagenesis affected by crystal packing factors: The structure of a 30–51 disulfide mutant of basic pancreatic trypsin inhibitor. Proteins: Struct. Funct. Genet. 14:75–87, 1992.
2. Crick, F. H. C., Kendrew, J. C. X-ray analysis and protein structure. Adv. Prot. Chem. 12:133–214, 1957.
3. Haggis, G. H. Proton-deuteron exchange in protein and nucleoprotein molecules surrounded by heavy water. Biochim. Biophys. Acta 23:494–503, 1957.
4. Richards, F. M. Structure of proteins. Annu. Rev. Biochem. 32:269–300, 1963.
5. Praissman, M., Rupley, J. A. Comparison of protein structure in the crystal and in the solution. II. Tritium-hydrogen exchange of zinc-free and zinc insulin. Biochemistry 7:2431–2445, 1968.
6. Tüchsen, E., Hvidt, A., Ottesen, M. Enzymes immobilized as crystals: Hydrogen isotope exchange of crystalline lysozyme. Biochemie 62:563–566, 1980.
7. Kossiakoff, A. A. Protein dynamics investigated by the neutron diffraction-hydrogen exchange technique. Nature (London) 296:713–721, 1982.
8. Islam, S. A., Weaver, D. L. Molecular interactions in protein crystals: Solvent accessible surface and stability. Proteins: Struct. Funct. Genet. 8:1–5, 1990.
9. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., Tasumi, M. The Protein Data Bank: A computer-based archival file for macromolecular structures. J. Mol. Biol. 112:535–542, 1977.
10. Deisenhofer, J., Steigemann, W. Crystallographic refinement of the structure of bovine pancreatic trypsin inhibitor at 1.5 Å resolution. Acta Crystallogr. Sect B 31:238–250, 1975.
11. Wlodawer, A., Walter, J., Huber, R., Sjolin, L. Structure of bovine pancreatic trypsin inhibitor. Results of joint neutron and X-ray refinement of crystal form II. J. Mol. Biol. 180:301–329, 1984.
12. Wlodawer, A., Deisenhofer, J., Huber, R. Comparison of two highly refined structures of bovine pancreatic trypsin inhibitor. J. Mol. Biol. 193:145–156, 1987.
13. Weiner, S. J., Kollman, P. A., Case, D., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. J. Am. Chem. Soc. 106:765–784, 1984.
14. Singh, U. C., Weiner, P. K., Caldwell, J., Kollman, P. A. AMBER 3.0 Revision A. Department of Pharmaceutical Chemistry, University of California–San Francisco, San Francisco, CA, 1989.
15. Van Gunsteren, W. F., Berendsen, H. J. C. Algorithms for macromolecular dynamics and constraint dynamics. Mol. Phys. 34:1311–1327, 1977.
16. Guenot, J., Kollman, P. A. Molecular dynamics studies of a DNA-binding protein: 2. An evaluation of implicit and explicit solvent models for the molecular dynamics simulation of the *Escherichia coli trp* repressor. Proteins: Struct. Funct. Genet., submitted for publication, 1991.
17. van Gunsteren, W. F., Berendsen, H. J. Computer simulation as a tool for tracing the conformational differences between proteins in solution and in the crystalline state. J. Mol. Biol. 176:559–564, 1984.
18. Levitt, M. Molecular dynamics of native protein. II. Analysis and nature of motion. J. Mol. Biol. 168:621–657, 1983.
19. Diamond, R. On the use of normal modes in thermal parameter refinement: theory and application to the bovine tpancreatic trypsin inhibitor. Acta Crystallogr. Sect. A 46: 425–435, 1990.
20. Weber, I. T. A vector-averaging method for locating small differences between nearly identical protein structures. J. Appl. Crystallogr. 20:388–393, 1987.
21. Sheriff, S., Hendrickson, W. A., Stenkamp, R. E., Sieker, L. C., Jensen, L. H. Influence of solvent accessibility and intermolecular contacts on atomic mobilities in hemerythrins. Proc. Natl. Acad. Sci. U.S.A. 82:1104–1107, 1985.
22. Phillips Jr., G. N. Comparison of the dynamics of myoglobin in different crystal forms. Biophys. J. 57:381–383, 1990.
23. Artymiuk, P. J., Blake, C. C. F., Grace, D. E. P., Oatley, S. J., Phillips, D. C., Sternberg, M. J. E. Crystallographic studies of the dynamic properties of lysozyme. Nature (London) 280:563–568, 1979.