Crystal structure of an extensively simplified variant of bovine pancreatic trypsin inhibitor in which over one-third of the residues are alanines

Mohammad Monirul Islam*, Shihori Sohya*, Keiichi Noguchi†, Masafumi Yohda*, and Yutaka Kuroda*‡

*Department of Biotechnology and Life Sciences, Graduate School of Engineering, and †Instrumentation Analysis Center, Tokyo University of Agriculture and Technology, 2-24-16 Nakamachi, Koganei-shi, Tokyo 184-8588, Japan

Edited by Brian W. Matthews, University of Oregon, Eugene, OR, and approved June 3, 2008 (received for review March 18, 2008)

We report the high-resolution crystal structures of an extensively simplified variant of bovine pancreatic trypsin inhibitor containing 20 alanines (BPTI-20st) and a reference single-disulfide-bonded variant (BPTI-[5,55]st) at, respectively, 1.39 and 1.09 Å resolutions. The sequence was simplified based on the results of an alanine scanning experiment, as reported previously. The effects of the multiple alanine substitutions on the overall backbone structure were surprisingly small (C^{α} atom RMSD of 0.53 Å) being limited to small local structural perturbations. Both BPTI variants retained a wild-type level of trypsin inhibitory activity. The side-chain configurations of residues buried in the hydrophobic cores (<30% accessible surface area) were almost perfectly retained in both BPTI-20st and BPTI-[5,55]st, indicating that neither multiple alanine replacements nor the removal of the disulfide bonds affected their precise placements. However, the side chains of three partially buried residues (Q31, R20, and to some extent Y21) and several unburied residues rearranged into alternative dense-packing structures, suggesting some plasticity in their shape complementarity. These results indicate that a protein sequence simplified over its entire length can retain its densely packed, native side-chain structure, and suggest that both the design and fold recognition of natively folded proteins may be easier than previously thought.

native structure | protein design | protein folding | protein structure | sequence simplification

The protein folding problem (1), or how a protein structure is encoded in its amino acid sequence, remains fundamentally unsolved. Indeed, an astronomical number of sequences are generated by the combination of the 20 natural amino acids at each site of a sequence, but few of these can fold into native, fully functional protein structures. However, the information contained in natural amino acid sequences is redundant (2, 3); namely, a small fraction of the information can fully determine a protein structure (4, 5). Protein sequence simplification has emerged as a method of choice in deciphering the determinants of protein structures.

Information redundancy can be conceptually classified into class and site redundancy (6). Class redundancy refers to the overlapping physicochemical properties of the amino acids, and it can be removed by designing proteins with a reduced amino acid alphabet (7). Class redundancy has been reported in the Src SH3 domain, for which a functional variant with 70% of its sequence encoded by only five amino acids types was identified by phage display (8). Similarly, simplified variants of chorismate mutase (9) and orotate phosphoribosyl transferase (10), with sequences predominantly encoded by only nine amino acid types, rescued auxotrophic host cells lacking the respective enzymes.

Site redundancy refers to the occurrence of several residues in a protein's amino acid sequence that do not contribute to the formation of the structure, except that their backbone atoms act as linker residues in the amino acid sequence. Alanine is often used as a natural representative of such linker residues because its side chain is small enough to avoid steric clashes and does not introduce excessive flexibility into the backbone structure (11). Site redundancy was first demonstrated experimentally in the T4 lysozyme sequence, in which 10 consecutive residues at positions 40–49 were substituted with alanines with very little effect, if any, on its structure and function (12). Simplified variants of the Arc repressor, containing up to 11 alanine substitutions in 53 residues, were also shown by circular dichroism to be folded correctly and retained a wild-type level of DNA-binding activity (13). In line with these developments, we recently reported that bovine pancreatic trypsin inhibitor (BPTI) variants with up to 47% alanines could fold into a native BPTI-like structure (6), and that their thermodynamics were, in most aspects, very similar to those of natural globular proteins (14).

These reports, based on functional or spectroscopic assays, strongly suggest that proteins with simplified sequences can remain folded and retain some or all of their functions. However, very few if any high-resolution structures of simplified proteins are presently available, and when available, they report either modest simplifications of short sequence segments (12) or modifications to a small number of residues located in the protein core (15, 16). Especially, no high-resolution structure of a protein simplified over its entire sequence is yet available. Such high-resolution structures could truly address the question of the sequence determinants of protein structures by providing a precise picture of how side-chain packing is affected by extensive sequence simplification.

Here, we report the high-resolution crystal structure of an extensively simplified BPTI in which over one-third of the residues are alanines, spread over its entire sequence. BPTI-21 (containing 21 alanines) was first designed by simultaneously substituting the least destabilizing BPTI-[5,55] residues to alanines (6), according to an alanine scanning experiment, in which each residue was individually changed to alanine and the stability of the mutated BPTIs were experimentally measured (17). Among the substitution sites, three residues are fully buried and two are partially buried. In the present study, we stabilized BPTI-21 by introducing an A14G substitution (18) that yielded BPTI-20st. As a control for the native structure, we determined the structure of a stabilized BPTI-[5,55] variant (BPTI-[5,55]st; BPTI-[5,55] with the A14G mutation), which is a single-disulfide-bonded BPTI variant regarded as a reference native molecule in several protein folding experiments (19). The effects of multiple alanine substitutions were surprisingly few. The overall structure of BPTI-20st was very similar to those of

Author contributions: Y.K. designed research; M.M.I. and S.S. performed research; S.S. and M.Y. contributed new reagents/analytic tools; M.M.I. and K.N. analyzed data; and M.M.I. and Y.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org [PDB ID codes 3CI7 (BPTI-20st) and 2ZJX (BPTI-[5,55]st)].

‡To whom correspondence should be addressed. E-mail: ykuroda@cc.tuat.ac.jp.

This article contains supporting information online at www.pnas.org/cgi/content/full/0802699105/DCSupplemental.

Table 1. Statistics after final refinement of simplified BPTIs

	BPTI-20st	BPTI-[5,55]st
Space group	$P2_12_12_1$	$P12_11$
No. of molecules in the asymmetric unit	Four molecules	Two molecules
Matthews coefficient	2.14	2.17
Solvent content, %	42.50	43.32
Resolution in refinement, Å	44.76∼1.39	53.91∼1.09
No. of reflections used	38,836	40,917
R factor	0.1565	0.1811
R_{free}	0.2052	0.2090
B factor,* Å^2		
Protein atoms	15.16	11.84
Main chain	11.06	10.77
Side chain	18.44	12.82
Ramachandran plot statistics†		
Residues in most favored region, %	90.8 (98.2)	92.2 (95.5)
Planarity, %	100	100
RMS Deviation from ideality†		
Bond lengths/bad length, Å/%	0.008∼0.009/(0.0)	0.007∼0.015/(0.0)
Bond angles/bad angles, °/%	1.06∼2.79/(1.7‡)	0.96∼2.89/(0.0)

*Probe used in the refinement process, 5PTI, is solved at a resolution of 1.0 Å with the main-chain and side-chain crystallographic *B* factors of 12.2 and 14.5, respectively.

†The qualities of the models were assessed by the SFCHECK and PROCHECK (35, 36) modules in the CCP4 suite. Figures in parentheses refer to data obtained using MolProbity (37).

‡The bad bond angle reported by MolProbity is that of A58, the C-terminal residue of BPTI-20st.

BPTI-[5,55]st and the wild-type BPTI, and the side-chain configurations of the deeply buried residues were almost perfectly retained. Alternative or new side-chain configurations were observed for a few partially buried residues.

Results and Discussion

Structures of the Simplified BPTIs. The space group of BPTI-20st and BPTI-[5,55]st was $P2_12_12_1$ and $P12_11$ with four and two molecules per unit cell, respectively. High-resolution anisotropic models of BPTI-20st and BPTI-[5,55]st were determined at resolutions of 1.39 and 1.09 Å, respectively, and with crystallographic *R* factors of 0.1565 and 0.1811 (Table 1). All residue side chains had excellent electron density maps, in both BPTI-20st and BPTI-[5,55]st. The overall backbone fold of both BPTI-20st and BPTI-[5,55]st was very similar to that of wild-type BPTI (Fig. 1*A*), with overall backbone root mean square deviations (RMSDs) of 0.55 and 0.47 Å, respectively (when compared with the 5PTI structure), whereas the RMSDs among wild-type structures are between 0.40 and ∼0.44 Å, reflecting experimental error arising from variable crystallization conditions [see supporting information (SI) Table S1; also see Figs. S1–S5 and Tables S2 and S3]. These structural observations were corroborated by the full trypsin inhibition activity of the simplified sequence (Fig. S2).

Small, local structural perturbations in the 13–17 loop, 25–28 β-turn, and 47–54 α-helix regions were clearly demonstrated by calculating the distances between the C^α atoms of the simplified BPTI and those of the wild-type BPTIs (Fig. 1*C*). Structural perturbations in the 13–17 and 25–28 regions are also observed in several wild-type structures, suggesting that these loop regions are flexible and tend to be distorted. In contrast, no distortion is observed in the 47–54 α-helix region of the wild-type structures, except for the 4PTI wild-type structure (Fig. S3 *a* and *b*), which suggests that the distortion in the α-helix region, although it can occasionally occur in the wild-type structure, is essentially attributable to the multiple alanine substitutions. Furthermore, when separately fitted, the 47–54 α-helix structures of the simplified BPTI could be superimposed almost perfectly onto that of the wild type, indicating that the entire α-helix moved toward the 29–35 β-strand without affecting the α-helix structure, *per se* (Fig. S3*c*).

Side-Chain Packing in the Simplified BPTIs. Overall, the side-chain geometries of both BPTI-20st and BPTI-[5,55]st were similar to that of the wild-type BPTI (RMSD between heavy atoms, 0.75–1.06 Å; the RMSDs of residues substituted with alanines were calculated by using their C^β atoms). However, a classification of the side-chain configuration changes according to their side-chain accessible surface area (ASA) values yielded some interesting insights into how the protein core is packed.

First, the side-chain configurations of all of the buried residues (ASA 0–30%) were surprisingly well conserved in both BPTI-20st and BPTI-[5,55]st. The side-chain RMSDs and χ_1 deviations were 0.01–0.39 Å and 0.24–11.2°, respectively (Figs. 2 and 3). These variations are fully within the values observed among the wild-type BPTI structures and can thus be essentially attributed to differences in the crystallization conditions. This extreme conservation is all the more surprising because three buried residues (C30, C51, and T54) were substituted with alanines without apparently disturbing the side-chain packing of the other buried residues. It is also worth noting that the buried structural water molecules conserved in the wild-type BPTI structures were almost perfectly conserved in both simplified BPTIs (Fig. S5 and Table S2).

Three partially buried residues (ASA 30–50%; R20, Y21, and Q31) were observed in alternative configurations (Fig. 3) that are absent from any wild-type structures (Fig. S4) and may thus be attributed either to the removal of the disulfide bonds or to the multiple alanine substitutions, rather than to differences in the crystallization conditions. Q31 and R20 were observed in dual configurations in one of the four BPTI-20st chains (Fig. 4). Substitutions of buried and partially buried residues to alanine could have created small packing defects due to the removal of the side-chain atoms beyond C^β. However, the BPTI-20st structure indicates that all potential packing defects were filled by tiny displacement of the nearby side-chain atoms. For example, potential defects arising from the T32A and T54A substitutions were filled by small displacements of the nearby Y21 and F4 side chains, respectively (Fig. 4*D* and Table S3). Furthermore, in all four BPTI-20st chains, the Y21 side chain was displaced by 1.2–1.4 Å (Fig. 4*D*), which enabled the formation of a hydrogen bond between its OH group and a water molecule (HOH5271). This interaction does not occur in either the wild-type or the BPTI-[5,55]st struc-

BIOPHYSICS

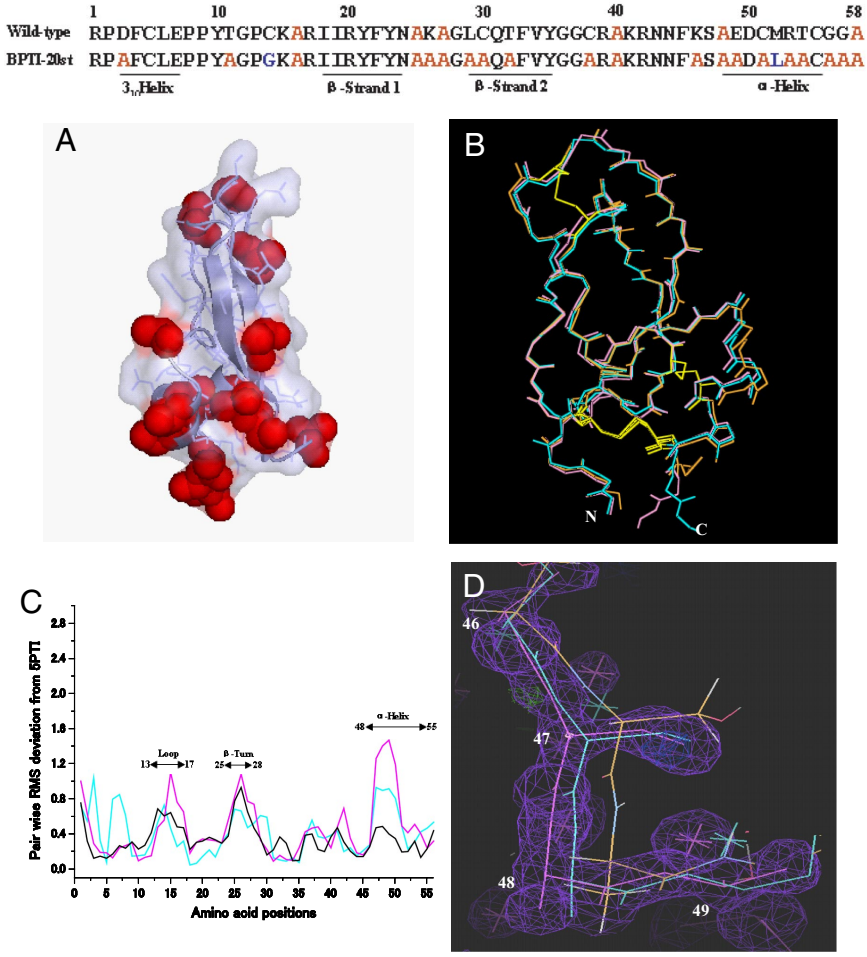

Fig. 1. Sequences and structures of the wild-type and simplified BPTIs. At the top are shown the sequences of the wild-type BPTI and the simplified BPTI-20st. Alanines in BPTI-20st are shown in red. C14G was introduced to stabilize the BPTI variants, and M52L was introduced for avoiding cleavage by cyanogen bromide during the purification protocol. BPTI-[5,55]st is a single disulfide bonded variant with the C14G, C30A, C38A, and C51A mutations. The BPTI-20st sequence contains a total of 14 alanine substitutions, of which 3 are buried and 2 are partially buried. (*A*) Surface representation of BPTI-20st. The alanine residues are shown with red spheres, and the other residues are represented with light blue transparent surfaces. The backbone structure of BPTI-20st is shown with a ribbon model. (*B*) Wire models of 5PTI (orange), BPTI-20st (pink), and BPTI-[5,55]st (cyan), superimposed using their C^{α} atoms. N and C indicate the N and C termini, respectively. (*C*) Pairwise deviations from 5PTI. The distances between the C^{α} atoms of 6PTI, BPTI-[5,55]st, and BPTI-20st are shown, respectively, in black, cyan, and violet. The three most perturbed regions are indicated by black arrowed lines with residue numbers. The average RMSD between C^{α} atoms of the wild-type structures and BPTI-[5,55]st and BPTI-20st were 0.47 and 0.55 Å, respectively (see details in Table S1). (*D*) Structural perturbation observed in the α-helix region with the electron-density map of BPTI-20st. Color codes are the same as defined in *B*.

tures. Interestingly, these changes had no effect on the melting temperatures (T_m) of the simplified BPTIs (M.M.I., *et al.*, unpublished work).

Finally, the side-chain configurations of surface residues (ASA 50–100%) were quite variable (Figs. 2 and 3), but such structural variations are also observed among the wild-type BPTI structures (Fig. S4). For example, the variations in the side-chain configurations of L6, E7, R17, R42, and D50, as assessed by RMSDs, were

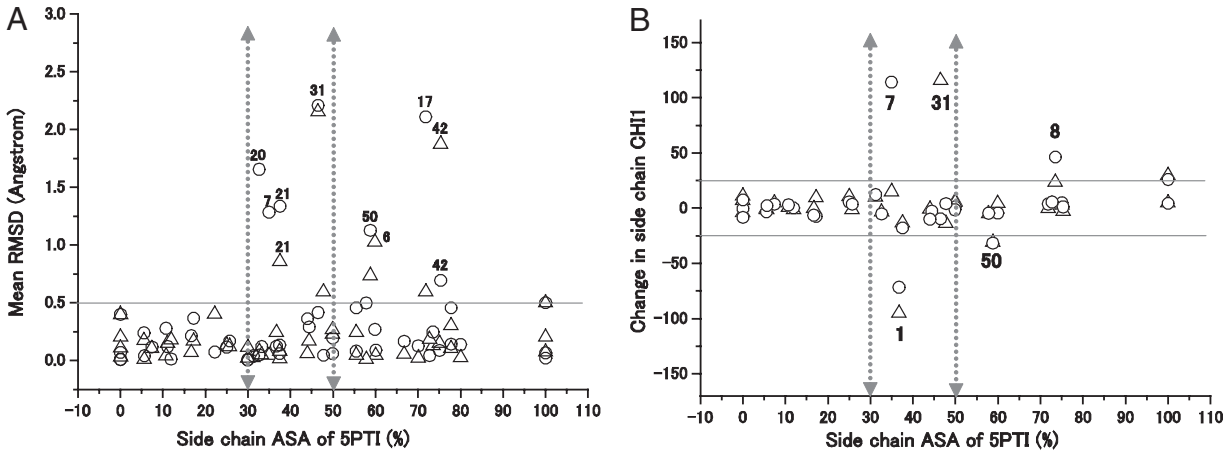

Fig. 2. Comparison of the side-chain structures. In both panels, chain 1 of BPTI-20st and chain 1 of BPTI-[5,55]st were used for the calculation, but similar results were obtained independently of the chain used. Residues showing large variations are numbered. (*A*) Side-chain displacement in RMSD versus side-chain ASA. The side-chain displacement was calculated as the heavy-atoms RMSD minus the backbone atoms RMSD between BPTI-[5,55]st and 5PTI (triangles) or BPTI-20st and 5PTI (circles). The horizontal continuous line represents an RMSD value of 0.5 Å, which can be considered as the experimental error. Arrowed dotted vertical lines are drawn between the buried (ASA < 30%) and partially buried (30% < ASA < 50%) regions, and between the partially buried and surface-exposed (50% < ASA) regions. The residues' ASAs are calculated by using 5PTI. (*B*) χ_1 Deviation versus side-chain ASA. Continuous gray lines represent a χ_1 deviation (−25° to +25°) considered as the experimental error.

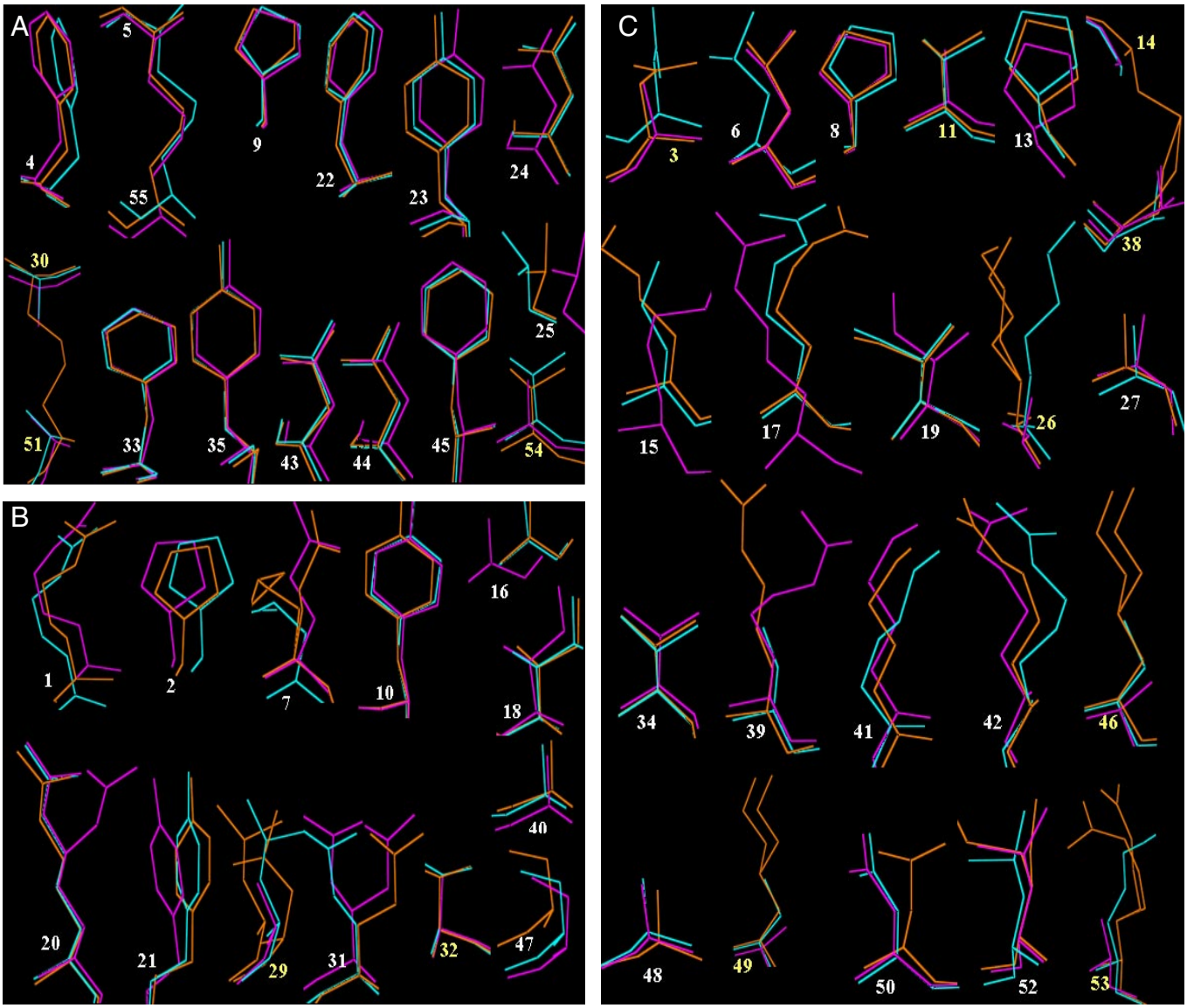

Fig. 3. Side-chain conformations. Side-chain structures of buried (ASA 0~30%) (*A*), partially buried (ASA 30~50%) (*B*), and surface-exposed (ASA 50~100%) residues (*C*). Chain 1 of BPTI-[5,55]st and chain 3 of BPTI-20st are superimposed onto that of 5PTI. Dual conformations observed in 5PTI are shown (E7, L29, C38, K26, K46, E49, and R53). Residues that are glycine in the wild sequences (G12, G28, G36, G37, G56, and G57) and the C-terminal A58 are not shown. Multiple alanine substitution sites are shown by yellow identities. Color codes are the same as in Fig. 1*B*. Significant conformational variability among the wild-type structures is observed for R1 and E7, indicating that the structural variations visible in image *C* are related to differences in crystallization conditions (see Figs. S3 and S4 for comparison among wild-type structures).

also observed in the different wild-type structures (Fig. S4). Thus, most configurational changes in the surface residues are probably unrelated to the multiple alanine substitutions.

Effect of Sequence Simplification on Core Packing. We simplified the sequences based on the individual residue's contribution to the BPTI structure stability (6), as assessed by an alanine scanning experiment (17). *A priori*, we imagined that the replacement of large side chains to the small alanine side chain's methyl group might provide spaces facilitating alternate side-chain conformations of nearby residues resulting into either an alternative densely packed protein interior or a molten globule-like protein interior (20, 21). However, our observations indicate that the replacement of three buried residues [C30, C51, and T54 (Fig. 3)] and two partially buried residues (L29 and T32) does not significantly affect the precise side-chain shape complementarity in the protein core, i.e., in a jigsaw-puzzle-like manner (22), except for the few local perturbations reported above. This observation is consistent with an alanine scanning experiment in which the substitution of any of the six core residues (F4, F22, Y23, F33, Y35, and F45) caused BPTI-[5,55] to unfold (17), probably by creating large packing defects that could not be accommodated in the finely interdigitized side-chain structure. A similar view is supported by α-helical coiled-coil structures that can switch from dimers to trimers and tetramers upon the substitution of a few core residues (23), as well as by a three-helix fold that can switch to an α/β fold by changing as few as 12% of the amino acids (24).

Previous reports suggest that the cores of some large proteins are more resilient to extensive repacking. For example, the high-resolution structure analyses of T4 lysozyme in which 10 residues in the core were simultaneously mutated to methionine residues (15) and of a staphylococcal nuclease with six core residues mutated to

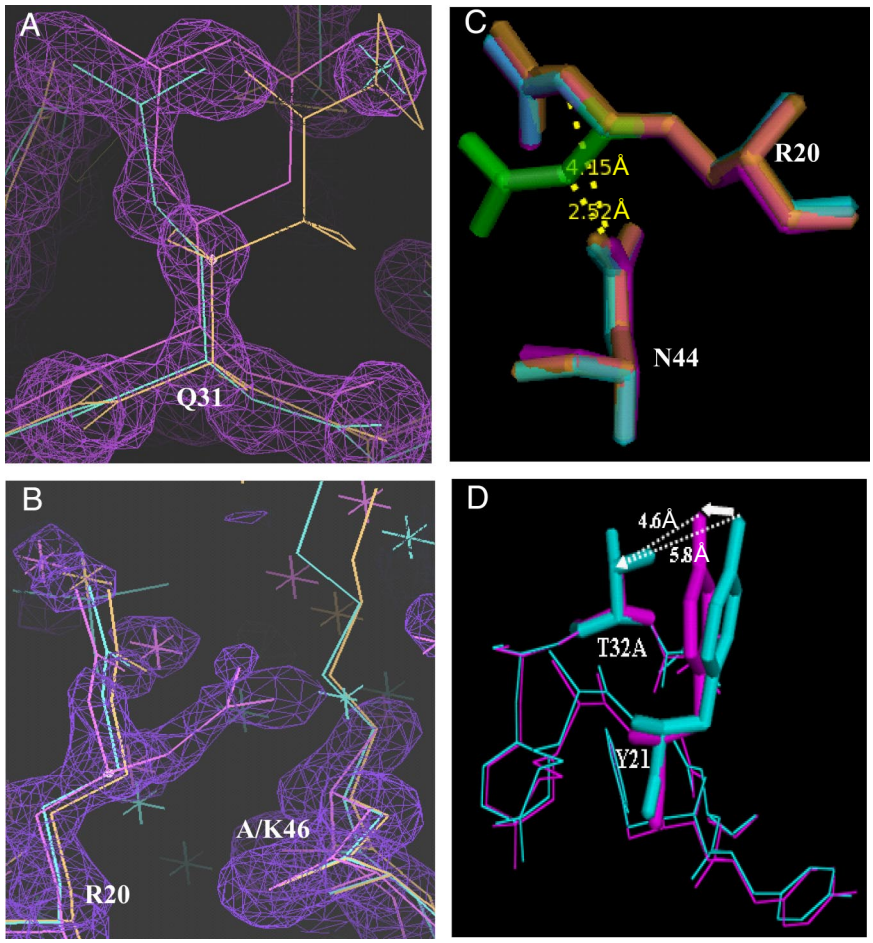

Fig. 4. Details of alternate side-chain conformations. (*A*) The side chain of Q31 beyond C^{β} was in an alternate conformation in both BPTI-20st and BPTI-[5,55]st, which is likely to originate from the neighboring C30AC51A substitutions. (*B*) An alternate conformation was also observed for the side chain of R20 beyond C^{δ} in BPTI-20st. The electron-density map of chain 3 of BPTI-20st is shown. The alternate structure was probably enabled by the K46A substitution, which provided both space for the R20 side chain and removed electrostatic repulsion between K46 and R20. (*C*) Intermolecular contact between the N^{ε} atom of R20 and O_1^{δ} of N44. In the alternate conformation (green sticks), the N^{ε} atom of R20 is hydrogen-bonded to the O_1^{δ} of N44 (2.52 Å), whereas they are far apart in the original structure (>4 Å). (*D*) Alternate conformation of Y21 side chain. A potential cavity created upon the T32A substitution is compensated for by displacement of the Y21 side chain.

leucine (16), indicating that they retain native structures, suggest some plasticity in the formation of the cores of large proteins. In BPTI, which is a small protein, we observed such flexibility, at a more modest level, for some partially buried residues of the BPTI-20st.

Possible Implications for Protein Design and Fold Recognition. The response to the packing defects introduced by extensive sequence simplification seems to involve only the local displacement of backbone atoms, accompanied by minimal side-chain configuration changes, but without any major rearrangement of the buried side-chain configurations. As an ideal model for small proteins, we may hypothesize that each backbone structure has one unique corresponding side-chain packing arrangement of the core residues, and that all sequences producing a similar backbone structure, to within a few sub-angstroms (e.g., 0.5–1.0 Å), will have either fully conserved buried residues or a small number of mutations that are easily accommodated by minimal side-chain or local backbone packing rearrangements. This hypothesis implies that fold recognition (25, 26) and *de novo* design (27–29) of native proteins can be simplified by restricting any detailed analysis to the small number of deeply buried residues of the template structure, because the physicochemical constraints at the remaining residues appears to be minor.

Conclusions

We report the high-resolution structural analysis of an extensively simplified BPTI variant, in which over one-third of the residues are alanines. The overall backbone and side-chain configurations were well retained. In particular, the side-chain configurations of deeply buried residues were fully conserved in both BPTI-20st and BPTI-[5,55]st. In contrast, there was some degree of plasticity in the conformations of a few partially buried residues, whose side chains were rearranged into novel, densely packed structures. Thus, some residues (mostly buried ones) seem to interact strongly and act like a jigsaw puzzle model, whereas some others (mostly partially buried residues) are able to rearrange into alternative side-chain conformations for accommodating structural perturbations arising from mutations of nearby residues. Together, these observations indicate that most protein structure determinants occur at a few sites deeply buried in the protein core, at least for small globular proteins.

Materials and Methods

Protein Expression and Purification. BPTI variants were expressed by using the pMMHA expression vector in the *Escherichia coli* JM109(DE3)pLysS cell line as inclusion bodies, and purified as described in ref. 14. After cell lysis by sonication, the cysteines were air-oxidized overnight in 6 M GnHCl at room temperature. The fusion partner, a His-tagged TrpΔLE leader, was cleaved by cyanogen bromide treatment of methionine and removed as precipitate upon dialysis in 20 mM

phosphate buffer (pH 6.0). Proteins were purified by reverse-phase HPLC, lyophilized, and preserved at −80°C until use. The protein identities were confirmed by electrospray ionization time-of-flight mass spectroscopy.

Crystallization and Data Collection. Initial crystallization conditions were screened with Hampton reagents by the sitting-drop technique at 20°C. Crystals of BPTI-20st and BPTI-[5,55]st were grown at 20°C in 0.1 M Tris·HCl buffer (pH 8.5) with 30% PEG-4000 and 0.2 M lithium sulfate. The x-ray diffraction data were recorded from single crystals by using a synchrotron beamline at the Photon Factory (Tsukuba, Japan). The data were processed with the HKL2000 program package, using DENZO for the integration and SCALEPACK for the merging and statistical analysis of the diffraction intensities (30).

Structure Determination. The structures of both BPTI variants were determined by molecular replacement method, using 5PTI (31) as a molecular probe and with the molrep program in the CCP4 program suite (32). The initial refinement was started with 48.2∼2.5 Å resolutions data, and the high-resolution limit was gradually extended to the maximum resolution, leading to the inclusion of all reflections. The structural refinement was performed with SHELXL-97 (33). A step-by-step improvement of the models was achieved through model optimization with the Xfit and followed by subsequent positional refinement. Solvent molecules were automatically added to the models at the lowest accepted occupancies of 0.5. The final structures were determined by including anisotropic displacement parameters in the refinement process. The structures were visualized by using COOT (34) and PyMOL (www.pymol.org).

ACKNOWLEDGMENTS. We thank Dr. Akashi Ohtaki and Mr. Kouichi Hashimoto for their help with x-ray diffraction data collection at the Photon Factory, Japan; Dr. Mark B. Swindells for discussions; Mr. Ryoh Yajima and Teppei Ebina for computational help; and Mr. Takayuki Kobayashi for installation of structure refinement software. Y.K. thanks Professor Peter S. Kim for the pMMHA expression vectors constructed while in his laboratory at the Whitehead Research Institute.

1. Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181:223–230.
2. Bowie JU, Reidhaar-Olson JF, Lim WA, Sauer RT (1990) Deciphering the message in protein sequences: Tolerance to amino acid substitutions. *Science* 247:1306–1310.
3. Rose GD (2000) Lysozyme among the Lilliputians. *Proc Natl Acad Sci USA* 97:526–528.
4. Plaxco KW, Riddle DS, Grantcharova V, Baker D (1998) Simplified proteins: Minimalist solutions to the ''protein folding problem.'' *Curr Opin Struct Biol* 8:80–85.
5. Kallenbach N (2001) Breaking open a protein barrel. *Proc Natl Acad Sci USA* 98:2958–2960.
6. Kuroda Y, Kim PS (2000) Folding of bovine pancreatic trypsin inhibitor (BPTI) variants in which almost half the residues are alanine. *J Mol Biol* 298:493–501.
7. Schafmeister CE, LaPorte SL, Miercke LJ, Stroud RM (1997) A designed four helix bundle protein with native-like structure. *Nat Struct Biol* 4:1039–1046.
8. Riddle DS, *et al.* (1997) Functional rapidly folding proteins from simplified amino acid sequences. *Nat Struct Biol* 4:805–809.
9. Walter KU, Vamvaca K, Hilvert D (2005) An active enzyme constructed from a 9-amino acid alphabet. *J Biol Chem* 280:37742–37746.
10. Akanuma S, Kigawa T, Yokoyama S (2002) Combinatorial mutagenesis to restrict amino acid usage in an enzyme to a reduced set. *Proc Natl Acad Sci USA* 99:13549–13553.
11. Cunningham BC, Wells JA (1989) High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* 244:1081–1085.
12. Heinz DW, Baase WA, Matthews BW (1992) Folding and function of a T4 lysozyme containing 10 consecutive alanines illustrate the redundancy of information in an amino acid sequence. *Proc Natl Acad Sci USA* 89:3751–3755.
13. Brown BM, Sauer RT (1999) Tolerance of Arc repressor to multiple-alanine substitutions. *Proc Natl Acad Sci USA* 96:1983–1988.
14. Kato A, Yamada M, Nakamura S, Kidokoro S, Kuroda Y (2007) Thermodynamic properties of BPTI variants with highly simplified amino acid sequences. *J Mol Biol* 372:737–746.
15. Gassner NC, Baase WA, Matthews BW (1996) A test of the ''jigsaw puzzle'' model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. *Proc Natl Acad Sci USA* 93:12155–12158.
16. Chen J, Lu Z, Sakon J, Stites WE (2004) Proteins with simplified hydrophobic cores compared to other packing mutants. *Biophys Chem* 110:239–248.
17. Yu MH, Weissman JS, Kim PS (1995) Contribution of individual side-chains to the stability of BPTI examined by alanine-scanning mutagenesis. *J Mol Biol* 249:388–397.
18. Hagihara Y, *et al.* (2002) Screening for stable mutants with amino acid pairs substituted for the disulfide bond between residues 14 and 38 of bovine pancreatic trypsin inhibitor (BPTI). *J Biol Chem* 277:51043–51048.
19. Staley JP, Kim PS (1992) Complete folding of bovine pancreatic trypsin inhibitor with only a single disulfide bond. *Proc Natl Acad Sci USA* 89:1519–1523.
20. Ohgushi M, Wada A (1983) ''Molten-globule state'': A compact form of globular proteins with mobile side-chains. *FEBS Lett* 164:21–24.
21. Kuwajima K (1989) The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins* 6:87–103.
22. Harrison SC, Durbin R (1985) Is there a single pathway for the folding of a polypeptide chain? *Proc Natl Acad Sci USA* 82:4028–4030.
23. Harbury PB, Zhang T, Kim PS, Alber T (1993) A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* 262:1401–1407.
24. Alexander PA, He Y, Chen Y, Orban J, Bryan PN (2007) The design and characterization of two proteins with 88% sequence identity but different structure and function. *Proc Natl Acad Sci USA* 104:11963–11968.
25. Fischer D, Rice D, Bowie JU, Eisenberg D (1996) Assigning amino acid sequences to 3-dimensional protein folds. *FASEB J* 10:126–136.
26. Jones DT, Taylor WR, Thornton JM (1992) A new approach to protein fold recognition. *Nature* 358:86–89.
27. Jin W, Kambara O, Sasakawa H, Tamura A, Takada S (2003) De novo design of foldable proteins with smooth folding funnel: Automated negative design and experimental verification. *Structure* 11:581–590.
28. Alvizo O, Allen BD, Mayo SL (2007) Computational protein design promises to revolutionize protein engineering. *BioTechniques* 42:31, 33, 35 passim.
29. Kuhlman B, *et al.* (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science* 302:1364–1368.
30. Otwinowski Z, Minor W (1997) Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276(Part A):307–326.
31. Wlodawer A, Walter J, Huber R, Sjolin L (1984) Structure of bovine pancreatic trypsin inhibitor. Results of joint neutron and x-ray refinement of crystal form II. *J Mol Biol* 180:301–329.
32. Collaborative Computational Project, Number 4. (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D* 50:760–763.
33. Sheldrick GM (1997) SHELXL-97: Program for the Refinement of Crystal Structures.
34. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D* 60:2126–2132.
35. Vaguine AA, Richelle J, Wodak SJ (1999) SFCHECK: A unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr D* 55:191–205.
36. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291.
37. Davis IW, *et al.* (2007) MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35:W375–W83.