PROTEINS: Structure, Function, and Genetics 11:223–229 (1991)

Analysis of the Steric Strain in the Polypeptide Backbone of Protein Molecules

Osnat Herzberg and John Moult
Center for Advanced Research in Biotechnology, The Maryland Biotechnology Institute, University of Maryland, Rockville, Maryland 20850

ABSTRACT **The extent to which local strain is present in the polypeptide backbone of folded protein molecules has been examined. The occurrence of steric strain associated with nonproline *cis* peptide bonds and energetically unfavorable main chain dihedral angles can be identified reliably from the well ordered parts of high resolution, refined crystal structures. The analysis reveals that there are relatively few sterically strained features. Those that do occur are located overwhelmingly in regions concerned with function. We attribute this to the greater precision necessary for ligand binding and catalysis, compared with the requirements of satisfactory folding.**

Key words: protein structure, crystal structure, dihedral angles, *cis* peptides, protein active sites

INTRODUCTION

Functional globular protein molecules are at a minimum of free energy with respect to their conformational degrees of freedom. Although the free energy difference between the folded and unfolded state of these molecules is small (typically in the range of -5 to -15 kcal/mol[1]), large favorable contributions are required to offset the configurational entropy cost of ordering the polypeptide chain, of approximately 1.3 kcal/mol per residue.[2] If protein molecules are sufficiently malleable, the required total could be achieved simply by accumulating favorable free energy contributions within the protein structure. However, the constraints of packing, and the limited flexibility of the main chain suggest that some off setting strain might be necessary to maximize favorable interactions. Functional properties, for example, binding to ligands and catalysis, require precise distribution of appropriate groups, and may result in local strain. Thus, assessing the extent of such strain should provide insight into the nature of protein architecture. A full analysis of the energetics of protein molecules is not yet possible, because of the crudeness of present models of the strength of interactions. Nevertheless, some types of strain are sufficiently straightforward that they can be assessed, given the availability of refined high resolution crystallographic structures.

An example of such strain is that present in the polypeptide backbone. Specifically, combinations of the three dihedral angles (ϕ, ψ, and ω) which lead to local close contacts between backbone atoms. There are two categories of these. When the peptide bond is in the *cis* conformation ($\omega = 0°$) there is a close contact between the α-carbon atoms of the adjacent amino acid residues, whereas in the *trans* conformation the corresponding distance between the first α-carbon atom and the following amide hydrogen is acceptable. A link containing proline in the second position is an exception: There the *cis* and *trans* conformations are more nearly equienergetic, since both involve 1–4 contacts of aliphatic carbon atoms. Nuclear magnetic resonance experiments give a value of about 2.8 kcal/mol for the energy difference between the two states for the nonproline containing link.[3] The *cis* conformation is believed to be further destabilized by longer range interactions, resulting in an expected frequency of occurrence of about 0.1% in the denatured state.[4]

Unfavorable combinations of ϕ and ψ result in a number of possible bad contacts, first enumerated by Ramakrishnan and Ramachandran.[5] With the availability of refined protein structures, it has become clear that the Ramachandran rules are generally obeyed by proteins. Accurate relative energies for different ϕ, ψ values are not available experimentally. Precise calculation of the energy as a function of the three dihedral angles is not possible, but the energy surface is relatively flat in the Ramachandran permitted regions. In contrast, excursions into the Ramachandran prohibited regions probably induce a strain of up to at least 5 kcal/mol (see later). Thus such conformations can be associated with significant strain.

It should be emphasized that errors in even the most accurate crystal structures may lead to missed occurrences of these outliers or their wrongful inclusion. The missed occurrences can be determined only

Received October 1, 1990; accepted February 22, 1991.
Address reprint requests to Dr. Osnat Herzberg, Center for Advanced Research in Biotechnology, The Maryland Biotechnology Institute, University of Maryland, 9600 Gudelsky Drive, Rockville, MD 20850.

by further crystallographic work. However, restricting the survey to the well-ordered parts of high-resolution well-refined structures minimizes the inclusion of wrongfully assigned strained conformations in the analysis.

METHODS

A set of structures was analyzed for two types of steric strain: *cis* peptide bonds, and energetically unfavorable main chain dihedral angles. The survey was restricted to a subset of structures known at 2 Å resolution or better, and refined to an R value of 0.20 or below ($R = \Sigma||F_o| - |F_c||/\Sigma|F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes, respectively). Any amino acid residue with one or more backbone atoms that had been assigned a crystallographic B factor of 20 Å^2 or higher was also excluded from consideration. The latter restriction biases the sample toward exclusion of flexible surface loops, but the data used do fairly represent all types of structural motif.

The Birkbeck Protein Structure Database (BIPED) was searched for the occurrence of nonproline *cis* peptides in structures fitting the criteria given above. Two proteins containing a total of four nonproline *cis* peptides were identified: dihydrofolate reductase (3DFR and 4DFR[6,7]) and carboxypeptidase A (5CPA[8,9]). The literature contains two more examples we are aware of: concanavalin A[10] and the subtilisin Carlsberg/eglin C complex.[11,12] A seventh *cis* peptide found during the refinement of the crystal structure of β-lactamase from *Staphylococcus aureus* PC1 (3BLM) at 2 Å resolution was also included.[13,14] That structure has been refined to an R value of 0.164.

The main chain φ and ψ dihedral angles in 24 high-resolution refined protein structures were examined (see Fig. 1 legend). All structural classes are represented in the set, with an overabundance of chymotrypsin class serine proteases, so that the extent to which strain is a common feature in evolutionarily related proteins could be examined. Figure 1 shows a plot of the observed φ, ψ angles from the low temperature factor regions of the structures. 2,930 β-carbon-containing residues other than prolines are included, and of these, 14 points lie in high energy regions of the φ,ψ surface. The 24 structures have each been refined by one of four different refinement programs.[15–18] All of these methods contain terms in the function minimized which reduce close atomic contacts. It is therefore possible that the degree of strain observed is underestimated. Nevertheless, examples of strain are found for structures refined by each of the four methods, so the identification of strain is not confined to a particular refinement method. In one case (subtilisin Carlsberg/eglin C complex[11,12]) the structure has been determined by two different groups, and refined by two different methods (restrained-parameter reciprocal-space least-squares refinement,[15] and energy minimization refinement[16]). The same steric strain features were found in each case.

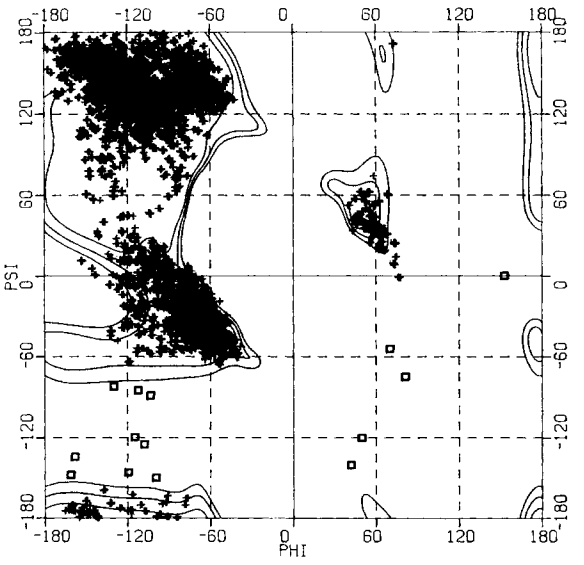

Fig. 1. Distribution of observed φ,ψ dihedral angles for 24 high-resolution, refined protein structures. Glycine and proline residues are omitted. Residues are also omitted if one or more of the four atoms defining each torsion angle has a crystallographic temperature factor greater than 20 Å^2. (+) amino acid residues that lie within the normal distribution of φ,ψ values. (□) residues that fall outside the normal clusters. The continuous lines enclose regions of equienergy according to Peters and Peters.[19] The contouring is at 4.0, 6.0, and 8.0 kcal/mol. The energy in the unfavorable regions is overestimated (see text). The left-handed α region appears to be slightly misplaced. The proteins included are β-lactamase (3BLM), actinidin (2ACT), crambin (1CRN), insulin (1INS), lysozyme (1LZ1), myoglobin (1MBO), phospholipase (1BP2), hemerythrin (1HMQ), penicillopepsin (2APP), troponin C (5TNC), α-lytic protease (2ALP), protease A (2SGA), ovomucoid inhibitor (part 3SGB), *Streptomyces griseus* trypsin (1SGT), trp repressor (2WRP), azurin (2AZA), Fab KOL (1FB4), L7/L12 50 S ribosomal protein (1CTF), subtilisin Carlsberg/eglin C complex (1CSE and 2SEC), papain (9PAP), carboxypeptidase A (5CPA), glutathione reductase (3GRS), ubiquitin (1UBQ), tonin (1TON). Codes refer to entries in the Brookhaven data bank.[38] References to the structures may be found there.

Relative energies for some dihedral φ,ψ values were obtained from ab initio quantum mechanical calculations. These rest on a better theoretical foundation than those involving empirical potentials and should represent the general shape of the energy surface reliably. They also agree qualitatively with the hard sphere calculations.[5] The energy contours in Figure 1 are taken from Peters and Peters calculations on an alanine dipeptide (*N*-formylalanyl amide).[19] Larger basis sets are required to obtain usefully quantitative energies. Relaxation of the molecular geometry for different conformations is also needed, particularly in the high energy regions.[20] Calculations were performed on *N*-acetyl-*N'*-methylalanyl amide with geometry optimization of all bonds, bond angles, and torsions, except for the central φ and ψ. The Gaussian 88[21] program and the

TABLE I. Nonproline *cis* Peptides

Protein	*cis* peptide	Remarks
β-Lactamase	Glu-166–Ile-167	On the omega loop of the active site.[13,14] Glu-166 is essential for catalysis. Ile-167 is believed to interact with bound substrate.
Concanavalin A	Ala-207–Asp-208	The carboxylate group of Asp-208 interacts with a water molecule liganded to the calcium ion.[10] In a saccharide-protein 2.9 Å resolution structure, the Asp-208 carboxylate is hydrogen bonded to the sugar.[30]
Subtilisin Carlsberg/eglin (1CSE, 2SEC)	Pro-210–Thr-211	The side chain of Pro-210 contacts Gly-63, adjacent to the catalytic His-64.[11,12] The main chain nitrogen of Asn-212 is hydrogen bonded to the side chain of Gln-36, four residues away from the catalytic Asp-32.
DHFR (3DFR, 4DFR)	Gly-98–Gly-99	The nitrogen atom of Gly-99 interacts with the 5′ phosphate group of the bound NADPH. The adjacent residues to the cis peptide are involved in the binding to both NADPH and methotrexate.[6,7]
Carboxypeptidase A (5CPA)	Ser-197–Tyr-198	Adjacent to the active site His-196.[8,9]
	Arg-272–Asp-273	Two residues after Glu-270, which has been implicated in catalysis. Also enables the simultaneous hydrogen bonding of the main chain carbonyl oxygens of Leu-271 and Asp-273 to the peptide nitrogen of Ser-199. Ser-199 has strained dihedral angles, and is located on a key loop that contacts the bound inhibitor, as seen in a 2.5 Å resolution crystal structure (4CPA[31]).
	Pro-205–Tyr-206	Not associated with the active site directly. Its functional significance, if any, is not clear.

4-21G basis set[22] were used. Geometries were obtained for three representative strained ϕ,ψ points and for a reference low energy conformation with $\phi = -92°, \psi = -5°$.[20] The energies of these geometries were then calculated with the more accurate CEP-31G* basis set,[23] which includes polarization basis functions on the nonhydrogen atoms.

RESULTS AND DISCUSSION

cis Peptides

Table I summarizes the location and functional significance of the seven nonproline *cis* peptides examined. All but one (5CPA Pro-205–Thr-206) are intimately involved in ligand binding and/or the positioning of catalytic residues. Figure 2 shows the arrangement in the case of β-lactamase, highlighting the contacts that appear to be contributing most to the stabilization of this energetically unfavorable *cis* peptide. In related β-lactamases, the second residue of the *cis* link is often a proline, and sometimes a threonine (Coulson, private communication). One could speculate that in those cases where a proline residue is found the protein stability would be greater. Denaturation studies indicate that the *Staphlyococcus aureus* enzyme is marginally stable, with a low energy of denaturation of about 6 kcal/mol.[24] A naturally occurring mutant (P54) that removes a salt bridge in the loop containing the *cis* peptide causes disorder in these residues (Kapadia, Coulson, and Herzberg, unpublished). Discussion of the role of some of the other cases can be found in the references given in Table I.

Strained ϕ,ψ Pairs

Table II lists the 14 occurrences of strained ϕ,ψ values in the selected protein set. Some of them are discussed in the references given. Nearly all of these outliers are closely associated with the function of the protein concerned, and can be rationalized in terms of constraints on the structure imposed by the function. The exceptions are in actinidin, where a surface loop remote from the substrate binding site contains the strained peptide; subtilisin Carlsberg, where the strain is only indirectly linked to the substrate binding site; and perhaps insulin, but in this case the functional role is not well understood. Figure 3 shows the role of the two strained ϕ,ψ pairs in the function of β-lactamase. Both these residues are involved in positioning catalytic components.

The ab initio quantum mechanical calculations were carried out for 3 of the observed strained ϕ,ψ points and a point in the Ramachandran permitted region. The 2 points from β-lactamase (42,−140) and (−108,−125) are both approximately 4 kcal/mol higher in energy than the reference conformation (−92,−5). The point (150,0) from carboxypeptidase is approximately 7 kcal/mol higher. These strains must be maintained by stresses exerted on the residues by the rest of the structure. In the β-lactamase example the two residues with strained ϕ,ψ angles are in interdomain regions of chain. The whole interface between the domains can be formed only when the crossing strands are folded in this manner. The energy of the interdomain interactions is thus presumably large enough to overcome the strain re-

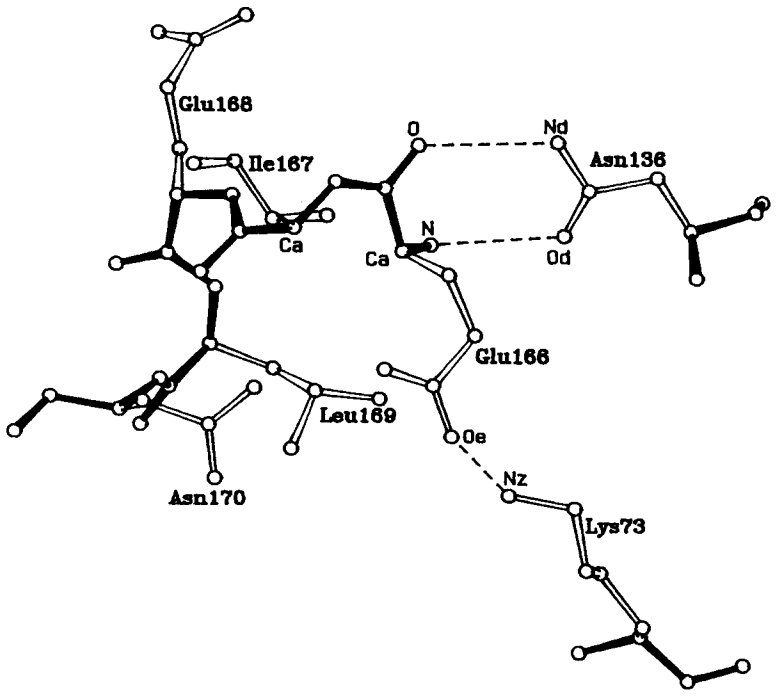

Fig. 2. The environment of the *cis* peptide bond between Glu-166 and Ile-167 in the crystal structure of β-lactamase from *Staphylococcus aureus* PC1.[13,14] Main chain bonds are filled and side chain bonds are open. Important stabilizing interactions are shown in dashed lines. The side chain of the residue before the *cis* link, Glu-166, makes a salt bridge to Lys-73. Both residues are involved in catalysis. The two charges are partially buried, so that this may be an usually strong interaction. Additionally, the main chain nitrogen and carbonyl oxygen atoms of Glu-166 are both hydrogen bonded to a conserved asparagine side chain, Asn-136, fixing the glutamate conformation further. On the other side of the *cis* link, a conserved leucine, Leu-169, packs into the interior of the protein.

quired. In other cases, possible causes of strain can also be identified. For instance, comparison of two subtilisin structures, one with a *cis* peptide between residues 210 and 211 (subtilisin Carlsberg, 1CSE, and 2SEC[11,12]) and one with *trans* there (subtilisin Novo, 2SNI[12]) suggests that different side chains at position 36 are responsible for the change. A glutamine residue in Carlsberg is replaced by an aspartate in Novo, providing space to accommodate a *trans* peptide, and forming a hydrogen bond to the main chain nitrogen of the peptide. Thus, one would expect that replacing Gln-36 by an aspartate in subtilisin Carlsberg should result in Pro-210–Thr-211 adopting the *trans* peptide conformation.

In three cases, the positions of nonproline *cis* peptides are close in sequence to those of a strained ϕ,ψ pair (two in carboxypeptidase: Ser-197–Tyr-198 with Ser-199 and Arg-272–Asp-273 with Asp-273, and in subtilisin Carlsberg, Pro-210–Thr-211 with Thr-213). The reason for this coupling is unclear. More detailed study of the energetics of all these examples would be informative.

Occasionally, proteins related to the examples given are seen to avoid strain. For example, the second residue of the *cis* peptide of the *S. aureus* β-lactamase (Glu-166–Ile-167) is sometimes replaced by a proline residue in other β-lactamases, thus reducing the strain. In protease A from *Streptomyces griseus*,[25] Asn-100, with strained ϕ,ψ angles, lies in a tight turn between two β-strands. In the equivalent mammalian serine proteases, this loop is enlarged, and the stress associated with the sharp change of direction thus avoided. Conversely, Gln-60 in actinidin is one of two residues inserted relative to papain, and its strained conformation may be needed in order not to disturb the rest of the structure.[26]

CONCLUSION

Four main points emerge from the analysis: (1) Significant steric strain of these types is rare in proteins. (2) Where it does occur, it is almost always in regions of the structures intimately involved in function. (3) Observation of strain is not the same as understanding the stresses inducing it, and further development of analytical tools is needed for that. (4) These findings indicate that the concept that proteins can always relieve steric strain by small adjustments of their atomic positions[27] requires revision.

The extent and localization of strain are relevant to the development of algorithms that model pro-

TABLE II. Strained Dihedral Angles

Protein	Residue	ϕ,ψ	Remarks
β-Lactamase	Ala-69	42,−140	Adjacent to the active site Ser-70, on a crossover between two domains. The β-carbon of Ala-69 contacts the main chain of Gln-237, maintaining the appropriate relationship between the oxyanion hole nitrogen atoms.[13,14]
	Leu-220	−108,−125	On the second crossover between domains. Makes a close contact with the β-carbon of Ser-235 on the β-strand forming part of the active site.
SGPA (2SGA)	Asn-100	81,−75	Two residues away from the active site Asp-102, on a turn assuring the proper positioning of this key residue.[25]
α-Lytic protease (2ALP)	Ala-39	−130,−82	On a loop flanking the active site.[32] The loop is linked to the catalytic His-57 by a disulfide bond between Cys-42 and Cys-58. Based on the crystal structure of a complex between the related SGPB and an ovomucoid inhibitor (3SGB[33]) Ala-39 would interact with bound inhibitors.
Tonin (1TON)	Asn-35	−158,−134	Adjacent to residue 39 (previous entry), with the same functional significance (there is a deletion of three residues relative to α-chymotrypsin, the enzyme from which the numbering scheme is derived[34]).
Subtilisin Carlsberg/eglin (1CSE, 2SEC)	Asp-32	−161,−147	The catalytic triad aspartate.[11,12]
	Thr-213	−137,−159	On a turn that includes the nonproline *cis* bond (see Table I). The *cis* bond is not observed in the subtilisin novo/CI-2 complex (2SNI[12]), and residue 213 has unstrained dihedral angles there as well.
Glutathione reductase (3GRS)	His-219	−119,−146	Interacts with a phosphate ion located in the binding site of the 2′-phosphate of NADP.[35]
Fab KOL (1FB4)	Asp-50	70,−54	A hypervariable loop residue.[36]
Insulin (1INS)	Ser-9	−115,−119	Between the intrachain disulfide bond Cys-6–Cys-11 and the intermonomer disulfide bond Cys-7–Cys-7. Involved in the hexamer contacts.[37] Aspects of the function are not well understood.
Actinidin (2ACT)	Gln-60	50,−120	On a reverse turn. No apparent relationship to the active site.[26]
Carboxypeptidase A (5CPA)	Ser-199	150,0	Adjacent to the cis peptide Ser-197–Tyr-198 associated with the active site (see Table I). Ser-199 contacts the bound inhibitor in a 2.5 Å resolution crystal structure of the enzyme–inhibitor complex (4CPA[31]).
	Ile-247	−103,−89	Interacts with the bound inhibitor.
	Asp-273	−99,−150	A nonproline *cis* peptide close to the active site (see Table I).

teins chains using restricted scanning of ϕ,ψ space.[28] Its association with function may be understood in terms of the more stringent structural requirements of function compared with folding and stability. Thus, efficient catalysis and substrate recognition and alignment may be expected to require the tuning of atomic positions to an accuracy of the order of 0.1 Å. Such precision is necessary not only for the key catalytic residues themselves, but also for residues involved in substrate binding. Apparently, peptide chains do not always possess sufficient flexibility to satisfy these requirements without introducing strain.

It has been suggested that steric strain in substrates may be induced upon binding, to increase the accessibility of transition states.[29] If we assume that the degree of strain found within the proteins themselves (up to 5 kcal/mol) could be applied to that purpose as well, then catalytic rate enhancements of approximately two orders of magnitude would be obtained.

ACKNOWLEDGMENTS

We thank Andrew Coulson for helpful discussions and Walt Stevens for advice and assistance with the calculations of peptide energies. This work was supported by NIH Grants AI27175, GM41034, and LM05102.

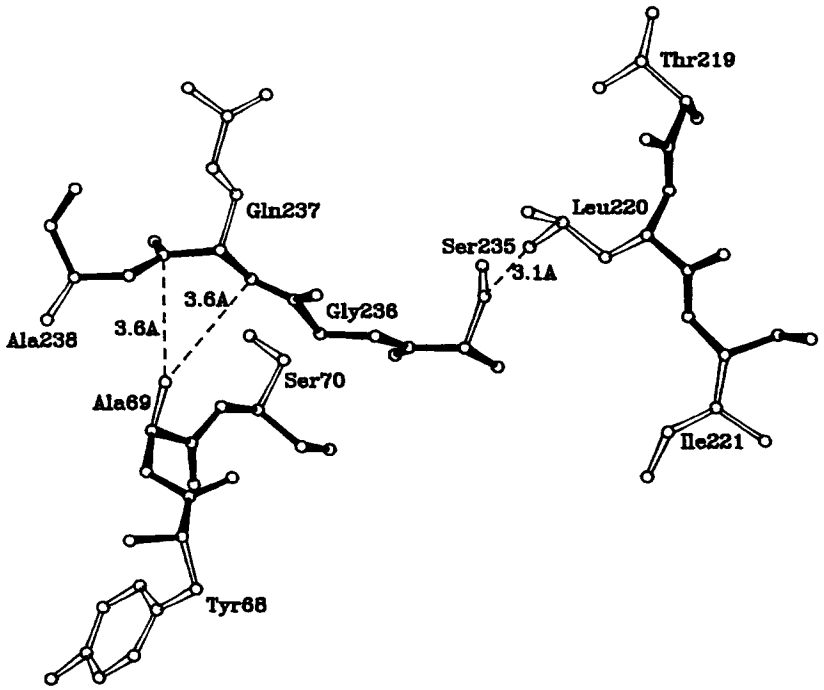

Fig. 3. The environment of the strained dihedral angles of Ala-69 and Ile-220 in the crystal structure of β-lactamase from *Staphylococcus aureus* PC1. Main chain bonds are filled and side chain bonds are open. The important van der Waals contacts to the active site β-strand Ser-235–Ala-238 are shown in dashed lines. Ala-69 is adjacent to the catalytic serine residue. Its ϕ,ψ values are such that a glycine would be acceptable here, but one is never observed in the known β-lactamase sequences (Coulson, private communication). The alanine β-carbon forms a knob which locks into a depression on the neighboring β-strand. A precise relationship between these two pieces of chain is essential for both binding and catalysis. Leu-220 forms a close van der Waals contact with the other side of this β-strand, further restricting its position.

REFERENCES

1. Privalov, P.L., Gill, S.J. Stability of protein structure and hydrophobic interaction. Adv. Protein Chem. 39:191–233, 1988.
2. Murphy, K.P., Privalov, P.L. Gill, S.J. Common features of protein unfolding and dissolvation of hydrophobic compounds. *Science* 247, 559–561 (1990).
3. Drakenberg, T., Dahlqvist, K.-I., Forsén, S. The barrier to internal rotation in amides. IV. N,N-Dimethylamides; substituents and solvent effects. J. Phys. Chem. 76:2178–2183, 1972.
4. Ramachandran, G.N., Mitra, A.K. An explanation for the rare occurrence of cis peptide units in proteins and polypeptides. J. Mol. Biol. 107:85–92, 1976.
5. Ramakrishnan, C., Ramachandran, G.N. Stereochemical criteria for polypeptide and protein chain conformation. Biophys. J. 5:909–933, 1965.
6. Bolin, J.T., Filman, D.J., Matthews, D.A., Hamlin, R.C., Kraut, J. Crystal structure of *Escherichia coli* and *Lactobacillus casei* dihydrofolate reductase refined at 1.7 Å resolution. I. General features and binding of methotrexate. J. Biol. Chem. 257:13650–13662, 1982.
7. Filman, D.J., Bolin, J.T., Matthews, D.A., Kraut, J. Crystal structure of *Escherichia coli* and *Lactobacillus casei* dihydrofolate reductase refined at 1.7 Å resolution. II. Environment of bound NADPH and implications for catalysis. J. Biol. Chem. 257:13663–13672, 1982.
8. Rees, D.C., Lewis, M., Honzatko, R.B., Lipscomb, W.N., Hardman, K.D. Zinc environment and cis peptide bonds in carboxypeptidase A at 1.75 Å resolution. Proc. Natl. Acad. Sci. U.S.A. 78:3408–3412, 1981.
9. Rees, D.C., Lewis, M., Lipscomb, W.N. Refined crystal structure of carboxypeptidase A at 1.54 Å resolution. J. Mol. Biol. 168:367–387, 1983.
10. Hardman, K.D., Agarwal, R.C., Freiser, M.J. Manganese and calcium binding sites of concanavalin A. J. Mol. Biol. 157:69–86, 1982.
11. Bode, W., Papamokos, E., Musil, D. The high resolution x-ray crystal structure of the complex formed between subtilisin Carlsberg and eglin C, an elastase inhibitor from the leech *Hirudo medicinalis*. Structural analysis, subtilisin structure and interface geometry. Eur. J. Biochem. 166:673–692, 1987.
12. McPhalen, C.A., James, M.N.G. Structural comparison of two serine proteinase—protein inhibitor complexes: Eglin-C–subtilisin Carlsberg and CI-2–subtilisin Novo. Biochemistry 27:6582–6598, 1988.
13. Herzberg, O., Moult, J. Bacterial resistance to β-lactam antibiotics: Crystal structure of β-lactamase from *Staphylococcus aureus* PC1 at 2.5 Å resolution. Science 236:694–701, 1987.
14. Herzberg, O. Refined crystal structure of β-lactamase from *staphylococcus aureus* PC1 at 2.0 Å resolution. J. Mol. Biol., 217:701–719, 1991.
15. Hendrickson, W.A., Konnert, J.H. Stereochemically restrained crystallographic least squares refinement of macromolecular structures. In: "Biomolecular Structure, Function, Conformation and Evolution" (Srinivasan, R., ed.) Vol I. Oxford: Pergamon Press, 1980, 43–75.
16. Jack, A., Levitt, M. Refinement of large structures by simultaneous minimization of energy and R factor. Acta Crystallogr. A 34:931–935, 1978.
17. Tronrud, D.E., TenEyck, L.F., Matthews, B.W. An efficient general purpose least-squares refinement program for macromolecular structures. Acta Crystallogr. A 43:489–501, 1987.
18. Agarwal, R.C. New least squares refinement technique based on a fast Fourier transform algorithm. Acta Crystallogr. A 34:791–809, 1978.
19. Peters, D., Peters, J. Quantum theory of the structure and bonding in proteins. Part 8: the alanine dipeptide. J. Mol. Struct. 85:107–123 (1981).
20. Schäfer, L., Klimkowski, V.J., Momany, F.A., Chuman, H. Conformational transition and geometry differences between low-energy conformers of N-acetyl-N-methyl alanine amide: An ab initio study at the 4-21G level with

gradient relaxed geometries. Biopolymers 23:2335–2347, 1984.
21. Frisch, M.J., Head-Gordon, M, Schlegel, H.B., Raghavachari, K., Binkley, J.S., Gonzalez, C., Defrees, D.J., Fox, D.J., Whiteside, R.A., Seeger, R., Melius, C.F., Baker, J., Martin, R., Kahn, L.R., Stewart, J.J.P., Fluder, E.M., Topiol, S., Pople, J.A. *Gaussian 88,* Gaussian Inc., Pittsburgh, PA, 1988.
22. Pulay, P., Fogarassi, G., Pang, F., Boggs, J.E. Systematic ab initio gradient calculation of molecular geometries, force constants, and dipole moment derivatives. J. Am. Chem. Soc. 101:2550–2560, 1979.
23. Stevens, W., Basch, H., Krauss, M. Compact effective potentials and efficient shared-exponent basis sets for the first- and second-row atoms. J. Chem. Phys. 81:6026–6033, 1984.
24. Craig, S., Hollecker, M., Creighton, T.E., Pain, R.H. Single amino-acid substitution blocks a late step in the folding of β-lactamase from *Staphylococcus aureus.* J. Mol. Biol. 185: 681–687, 1985.
25. Sielecki, A.R., Hendrickson, W.A., Broughton, C.G., Delbaere, L.T.J., Brayer, G.D., James, M.N.G. Protein structure refinement. *Streptomyces griseus* serine protease A at 1.8 Å resolution. J. Mol. Biol. 134:781–804, 1979.
26. Baker, E.N. Structure of actinidin after refinement at 1.7 Å resolution. J. Mol. Biol. 141:441–484, 1980.
27. Warshel, A., Levitt, M. Theoretical studies of enzymatic reactions: Dialectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. J. Mol. Biol. 103:227–249, 1976.
28. Moult, J., James, M.N.G. An algorithm for determining the conformation of polypeptide segments in proteins by systematic search. Proteins: Struct. Funct. Genet. 1:146–163, 1986.
29. Ford, L.O., Johnson, L.N., Machin, P.A., Phillips, D.C., Tjian, R. Crystal structure of a lysozyme-tetrasaccharide lactone complex. J. Mol. Biol. 88:349–371, 1974.
30. Derewenda, Z., Yariv, J., Helliwell, J.R., Kalb (Gilboa), A.J., Dodson, E.J., Papiz, M.Z., Wan, T., Campbell, J. The structure of the saccharide-binding site of concanavalin A. EMBO J. 8:2189–2193, 1989.
31. Rees, D.C., Lipscomb, W.N. Refined crystal structure of the potato inhibitor complex of carboxypeptidase A at 2.5 Å resolution. J. Mol. Biol. 160:475–498, 1982.
32. Fujinaga, M., Delbaere, L.T.J., Brayer, G.D., James, M.N.G. Refined structure of α-lytic protease at 1.7 Å resolution. Analysis of hydrogen bonding and solvent structure. J. Mol. Biol. 184:479–502, 1985.
33. Read, R.J., Fujinaga, M. Sielecki, A.R., James, M.N.G. Structure of the complex of *Streptomyces griseus* protease B and the third domain of turkey ovomucoid inhibitor at 1.8 Å resolution. Biochemistry 22:4420–4433, 1983.
34. Fujinaga, M., James, M.N.G. Rat submaxillary gland serine protease, tonin: Structure solution and refinement at 1.8 Å resolution. J. Mol. Biol. 195:373–396, 1987.
35. Karplus, P.A., Schulz, G.E. Refined structure of glutathione reductase at 1.54 Å resolution. J. Mol. Biol. 195: 701–729, 1987.
36. Marquart, M. Deisenhofer, J., Huber, R. Crystallographic refinement and atomic models of the intact immunoglobulin molecule Kol and its antigen-binding fragment at 3.0 Å and 1.9 Å resolution. J. Mol. Biol. 141:369–391, 1980.
37. Baker, E.N., Blundell, T.L., Cutfield, J.F., Cutfield, S.M., Dodson, E.J., Dodson, G.G., Crowfoot Hodgkin, D.M., Hubbard, R.E., Isaacs, N.W., Reynolds, C.D., Sakabe, K., Sakabe, N., Vijayan, N.M. The structure of 2Zn pig insulin crystals at 1.5 Å resolution. Phil. Trans. R. Soc. London Ser. B 319:369–456, 1988.
38. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Jr., 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.