Trupin, J., Dickerman, H., Nirenberg, M., & Weissbach, H. (1966) *Biochem. Biophys. Res. Commun. 24*, 50.
von der Haar, F., & Cramer, F. (1975) *FEBS Lett. 56*, 215.
von der Haar, F., & Cramer, F. (1976) *Biochemistry 16*, 4131.
Walker, J. E., Shaw, D. C., Northrop, F. D., & Horsenell, T. (1978) *Solid Phase Methods in Protein Sequencing Analysis* (Previero, E., & Coletti-Previero, M. A., Eds.) p 69, North-Holland Publishing Co., Amsterdam.

Side-Chain Torsional Potentials: Effect of Dipeptide, Protein, and Solvent Environment†

B. R. Gelin‡ and M. Karplus*

ABSTRACT: Side-chain torsional potentials in the bovine pancreatic trypsin inhibitor are calculated from empirical energy functions by use of the known X-ray structure of the protein and the rigid-geometry mapping technique. The potentials are analyzed to determine the roles and relative importance of contributions from the dipeptide backbone, the protein, and the crystalline environment of solvent and other protein molecules. The structural characteristics of the side chains determine two major patterns of energy surfaces, $E(\chi_1,\chi_2)$: a γ-branched pattern and a pattern for longer, straight side chains (Arg, Lys, Glu, and Met). Most of the dipeptide potential curves and surfaces have a local minimum corresponding to the side-chain torsional angles in the X-ray structure. Addition of the protein forces sharpens and/or selects from these minima, providing very good agreement with the experimental conformation for most side chains at the surface or in the core of the protein. Inclusion of the crystalline environment produces still better results, especially for the side chains extending away from the protein. The results are discussed in terms of the details of the interactions due to the surrounding, calculated solvent-accessibility figures and the temperature factors derived from the crystallographic refinement of the pancreatic trypsin inhibitor.

Although much progress has been made in determining the factors responsible for the equilibrium conformation of a protein (Richards, 1977; Chothia et al., 1977; Anfinsen & Scheraga, 1975), a detailed understanding has not yet been achieved. One important question concerns the interactions that hold a given part of a protein in the position observed in the native structure. Since fluctuations relative to the native structure do occur, it is also of interest to evaluate the restoring forces which arise upon small displacement from equilibrium. This paper is concerned with an empirical energy function analysis of the interactions which determine the side-chain torsional potentials, given the conformation of a protein. We examine all the side chains of a single protein, the bovine pancreatic trypsin inhibitor (PTI), and focus on the relative importance of the energy contributions due to short-range (side-chain, backbone) interactions, interactions with the rest of the protein, and interactions with solvent and neighboring protein molecules of the crystalline environment.

A complementary approach to side-chain orientations is a statistical treatment of the torsional angles observed in series of proteins. Such an analysis has been made by Chandrasekaran & Ramachandran (1970). They used contact criteria for a dipeptide to determine the allowed side-chain torsional angles. Estimates of the relative probabilities of different side-chain orientations were made and found to agree well in an overall sense with the observed distributions. No attempt was made to examine individual cases nor to evaluate interactions beyond the dipeptide (i.e., from other parts of the protein and the surroundings).

The bovine pancreatic trypsin inhibitor is used in the present study for several reasons. Its X-ray crystal structure has been determined to high resolution (Deisenhofer & Steigemann, 1975) and shows many localized solvent molecules, whose steric effects on side-chain conformation can be evaluated. The small size of PTI (58 residues, 454 nonhydrogen atoms) has made possible extensive calculations including energy minimization (Gelin & Karplus, 1975), a study of aromatic side-chain rotations ("ring flips") (Gelin & Karplus, 1975; Snyder et al., 1975; Hetzel et al., 1976), and a molecular dynamics calculation of the fluctuations about the equilibrium conformation (McCammon et al., 1977).

The method used for calculating protein conformational energy and the procedures for evaluating the various contributions to side-chain torsional potentials are described under Methods. The results of the calculations are presented under Results along with their structural interpretations and comparisons with the observed torsional angles. In the final section the results are discussed and related to measures of solvent accessibility and thermal mobility of different parts of the protein.

Methods

The energy function and protein structure definition are outlined below, followed by a description of the rigid mapping procedure used to determine the side-chain torsional potentials.

A. Energy Function and Protein Structure Definition. The conformational energy of a macromolecule in vacuo or in surroundings representing the crystal environment has been calculated using empirical energy functions. The functional forms chosen for the various types of interactions, the methods of specification of the covalent structure of the protein, and the evaluation and application of the resulting energy expression are outlined below.

The potential energy is written as a sum of terms, one set of which arises from the isolated protein and another from the interactions of the protein with its environment (solvent and other proteins) in the crystal. The isolated protein terms correspond to bonds, bond angles, torsional angles, van der Waals interactions, electrostatic interactions, and hydrogen

† From the Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138. *Received July 13, 1978.* Supported in part by the National Institutes of Health.
‡ Present address: Digital Equipment Corp., Meriden, CT 06450.

0006-2960/79/0418-1256$01.00/0

Table I: Extended Atoms and Their Nonbonded Parameters

atom symbol	α[a]	N_{eff}[b]	r_{vdW}[c,d]	groups represented
O	0.84	6	1.60	carbonyl oxygen
OH	1.20	7	1.70	alcoholic hydroxyl (Ser, Thr, Tyr)
OM	2.14	6	1.60	carboxyl oxygen (Asp, Glu, carboxy terminus)
NH	1.40	7	1.65	peptide nitrogen, whether -N- or -NH; other -NH- groups
N(2)	1.70	8	1.70	$-NH_2$ terminals (Asn, Gln, Arg)
N(3)	2.13	9	1.75	$-NH_3^+$ terminals (Lys, amino terminal)
CH	1.35	6	1.85	aliphatic CH group
C(2)	1.77	7	1.90	aliphatic $-CH_2-$ group
C(3)	2.17	8	1.95	methyl terminal $-CH_3$
C	1.65	5	1.80	aromatic or carbonyl carbon
CR	2.07	6	1.90	aromatic -CH- group
S	0.34	16	1.90	sulfur (Cys, Met)

[a] Polarizability (Å^3). [b] "Effective" outer-shell election number. [c] van der Waals radius, Å. [d] In the Lennard-Jones potential function, $Ar^{-12} - Cr^{-6}$, the parameter C is determined from the Slater-Kirkwood formula for a given pair of atoms i and j: $C = (3e\hbar/2\sqrt{m})\alpha_i\alpha_j/[(\alpha_i/N_j)^{1/2} + (\alpha_j/N_j)^{1/2}]$. The value of $3e\hbar/2\sqrt{m}$ is 361.67 in units such that C has dimensions of (kcal A^6)/mol. A is then calculated as $A = {}^1/_2 C(r_i + r_j)^6$, where r_i and r_j are van der Waals radii.

bonds. The interactions between the protein and the environment consist of van der Waals, electrostatic, and hydrogen-bond terms. The resulting expression for the energy is

$$E(\mathbf{R}) = \frac{1}{2}\sum_{\text{bonds}} K_b(b - b_0)^2 + \frac{1}{2}\sum_{\substack{\text{bond}\\ \text{angles}}} K_\theta(\theta - \theta_0)^2 + \frac{1}{2}\sum_{\substack{\text{torsion}\\ \text{angles}}} K_\phi[1 + \cos(n\phi - \delta)] + \sum_{\substack{\text{nb pairs,}\\ r<8\text{Å}}}\left[\frac{A}{r^{12}} - \frac{C}{r^6} + \frac{q_1q_2}{Dr}\right] + \sum_{\text{H bonds}}\left[\frac{A'}{r^{12}} - \frac{C'}{r^{10}}\right] + \sum_{\substack{\text{int–ext}\\ \text{pairs, } r<8\text{Å}}}\left[\frac{A}{r^{12}} - \frac{C}{r^6} + \frac{q_1q_2}{Dr}\right] + \sum_{\substack{\text{int–ext}\\ \text{H bonds}}}\left[\frac{A'}{r^{12}} - \frac{C'}{r^{10}}\right] \quad (1)$$

The first five sums correspond to the energy of the isolated protein and the last two sums represent the interactions between its atoms (the "interior") and the atoms of solvent and other protein molecules in the crystal (the "exterior" or "surroundings").

The energy as given by eq 1 is a function of the Cartesian coordinate set, {**R**}, specifying the positions of all the atoms involved, but the calculation is carried out by evaluating the internal coordinates for bonds (b), bond angles (θ), dihedral angles (ϕ), and interparticle distances (r) for any given geometry, **R**, and determining the contributions using the energy parameters K_b, K_θ, and K_ϕ, Lennard-Jones parameters A and C, atomic charges q_i, dielectric constant D, hydrogen-bond parameters A' and C', and geometrical reference values ("zero values") b_0, θ_0, n, and δ. As in previous work, the protein is represented by extended atoms; one extended atom replaces a nonhydrogen atom and any hydrogens bonded to it. Table I gives the extended-atom set used in this work, along with their van der Waals parameters; how the atom codes are combined with internal coordinate specifications to obtain parameter codes is described separately (B. R. Gelin and M. Karplus, unpublished results; Gelin, 1976). Table II gives the complete set of parameters used for the calculations, except for the charges which are identical with those listed (B. R. Gelin and M. Karplus, unpublished results; Gelin, 1976). The dielectric constant in eq 1 has been taken as $D(r) = r$; this choice makes a computationally convenient expression for nonbonded energies, as only even powers of r are required; also, it is numerically reasonable in the regions of the most important interactions from 2 to 5 Å where D varies from 2 to 5. Although no quantitative theory for the dielectric constant in the interior of a protein is available, some justification of the present choice is provided by the work of Warshel & Levitt (1976).

As is evident from the use of eq 1, the protein environment is introduced so as to account for the purely steric aspects of the interactions. Solvent molecules are represented as extended atoms which interact by van der Waals and hydrogen-bond forces with protein atoms (the solvent atoms are electrically neutral). No attempt has been made to introduce the entropy of solvation or other statistical mechanical aspects of the solvent. The approximate dielectric function used in eq 1 is appropriate for the protein interior or the contact region of two proteins. Its use for interactions involving solvent–water molecules would not be valid, but since they are treated as electrically neutral, the effective dielectric constant is infinite; only the hydrogen-bond terms are important for attractive interactions with the solvent.

Specification of the covalent structure of a single PTI molecule (the "central" molecule) was accomplished with the sequential generation procedure described (B. R. Gelin and M. Karplus, unpublished results; Gelin, 1976); from a knowledge of the structure, the first five sums of eq 1 can be enumerated. Four solvent molecules which are internally bound were included as part of the central molecule. To define the surroundings in the crystalline environment, the coordinates of the central molecule and all its other localized solvent molecules were subjected to the spatial transformations characteristic of the $P_{2_12_12_1}$ symmetry of the PTI crystal. (The required computer program was supplied by J. O. Deisenhofer.) The transformations are listed in Table III. Each unique "exterior" atom thus obtained was kept if it was within 8 Å of some atom of the central molecule and discarded otherwise. The exterior atom list obtained in this way contained 893 atoms (125 solvent atoms and 768 atoms in protein molecules other than the central one). These interact with the central protein through a total of 7516 nonbonded interactions whose distance in the X-ray conformation is less than 8 Å; the next-to-last summation in eq 1 represents these 7516 terms. The corresponding term for the internal nonbonded interactions of the central protein molecule represents just under 16000 terms. The hydrogen-bond interactions between the central molecule and surroundings (the last term in eq 1) were enumerated by searching out the donor–acceptor pairs which meet certain geometric criteria (B. R. Gelin and M. Karplus, unpublished results; Gelin, 1976); Table IV lists the 38 central molecule, exterior atom hydrogen-bond pairs included in this work.

B. Rigid-Geometry Mapping Procedures. Side-chain torsional potentials express the energy of a side chain over its range of conformations. In the simplest approximation, changes of conformation are limited to rotations about bonds, and the potential is determined by calculating the energy after successive rotations about one or more bonds of interest. This method is referred to as "rigid-geometry" mapping; in it, all bond lengths, bond angles, all but the specific torsion angles, and a very large fraction of the nonbonded interactions remain unchanged during the calculation (Gelin & Karplus, 1975). Thus, only a small subset of eq 1 has to be recalculated, and the procedure is very fast. For example, if $\chi_1 = \chi(N\text{–}C_\alpha\text{–}C_\beta\text{–}C_\gamma)$ is to be studied, the only "moving portion" of the

Table II: Parameters for Bond, Bond-Angle, Torsional, and Hydrogen-Bonded Interactions

bond energy function, $^1/_2K_b(b - b_0)^2$

bond type	code	$^1/_2K_b$ [kcal/(mol Å^2)]	b_0 (Å)
CH-OH	23	400	1.420
CH-NH	25	450	1.461
CH-N(3)	27	450	1.469
CH-CH	28	400	1.531
C(2)-OH	30	400	1.420
C(2)-NH	32	450	1.458
C(2)-N(3)	34	450	1.479
C(2)-CH	35	400	1.523
C(2)-C(2)	36	400	1.518
C(3)-CH	43	400	1.518
C(3)-C(2)	44	400	1.541
C-O	46	600	1.234
C-OH	47	450	1.378
C-OM	48	450	1.222
C-NH	49	500	1.317
C-N(2)	50	450	1.333
C-CH	52	400	1.517
C-C(2)	53	400	1.517
CR-C	65	500	1.352
CR-CR	66	500	1.388
S-C(2)	74	450	1.814
S-C(3)	75	450	1.769
S-S	78	500	2.015

torsional energy function, $^1/_2K_\phi[1 + \cos(n\phi - \delta)]$

torsion type	code	$^1/_2K_\phi$ (kcal/mol)	n (dimensionless)	δ (deg)
X-CH-NH-Y	25	0.3	3	180
X-CH-CH-Y	28	0.5	3	0
X-C(2)-NH-Y	32	0.3	3	180
X-C(2)-CH-Y	35	0.5	3	0
X-C(2)-C(2)-Y	36	0.5	3	0
X-C-NH-Y	49	7.0	2	180
X-C-CH-Y	52	0.1	3	0
X-C-C(2)-Y	53	0.1	3	0
X-CR-C-Y	65	10.0	2	180
X-CR-CR-Y	66	10.0	2	180
X-S-C(2)-X	74	0.5	2	0
X-S-S-X	78	4.0	2	0

hydrogen-bond function, $A'r^{-12} - C'r^{-10}$

bond type	code	E_{min} (kcal/mol) = $-0.067C'^6/A'^5$	R_{min} (Å) = $(1.2A'/C')^{1/2}$
OH···O	2	−3.5	2.80
OH···OH	3	−3.5	2.75
OH···OM	5	−3.5	2.85
NH···O	7	−3.0	2.95
NH···OH	8	−3.0	3.08
NH···OM	9	−3.0	3.10
N(2)···O	11	−2.5	2.87
N(2)···OH	12	−2.5	2.87
N(2)···OM	13	−2.5	2.87
N(3)···OH	17	−2.5	3.00

bond-angle energy function, $^1/_2K_\theta(\theta - \theta_0)^2$

angle type	code	sequence no.	$^1/_2K_\theta$ [kcal/(mol rad^2)]	θ_0 (deg)
OM-C-OM	610	1	50	129.0
NH-C-O	710	2	60	124.5
N(2)-C-O	1110	3	60	120.6
N(2)-C-NH	1410	4	60	117.0
N(2)-C-N(2)	1510	5	60	120.3
CH-C-O	2210	6	50	122.3
CH-CH-OH	2307	7	50	101.5
CH-C(2)-OH	2308	8	50	109.3
CH-C-OM	2410	9	40	118.0
CH-CH-NH	2507	10	50	107.1
CH-C-NH	2501	11	35	117.5
CH-C(2)-CH	2808	12	30	113.2
C(2)-C-O	2910	13	50	123.5
C(2)-C-O	3110	14	40	115.9
C(2)-CH-NH	3207	15	30	109.0
C(2)-C(2)-NH	3208	16	50	109.6
C(2)-C-NH	3210	17	35	115.0
C(2)-C-N(2)	3310	18	50	116.0
C(2)-CH-N(3)	3407	19	50	108.9
C(2)-C(2)-N(3)	3408	20	30	109.8
C(2)-NH-CH	3504	21	70	117.4
C(2)-CH-CH	3507	22	30	108.9
C(2)-C(2)-CH	3508	23	30	111.8
C(2)-C(2)-C(2)	3608	24	30	109.5
C(3)-CH-OH	3807	25	30	110.5
C(3)-CH-NH	4007	26	30	108.6
C(3)-CH-CH	4307	27	50	107.8
C(3)-C(2)-CH	4308	28	30	108.8
C(3)-CH-C(2)	4407	29	30	108.9
C(3)-S-C(2)	4412	30	50	97.2
C(3)-CH-C(3)	4507	31	30	110.7
C-CH-NH	4907	32	60	108.6
C-C(2)-NH	4908	33	90	110.5
C-CH-N(3)	5107	34	50	118.3
C-NH-CH	5204	35	60	119.1
C-CH-CH	5207	36	30	107.5
C-C(2)-CH	5208	37	50	110.4
C-NH-C(2)	5304	38	60	117.8
C-CH-C(2)	5307	39	30	107.1
C-C(2)-C(2)	5308	40	30	110.4
C-CH-C(3)	5407	41	50	104.3
CR-C-OH	5710	42	50	118.3
CR-C-C(2)	6310	43	50	122.6
CR-CR-C	6511	44	50	121.3
CR-CR-CR	6611	45	50	123.5
S-C(2)-CH	7308	46	50	117.2
S-C(2)-C(2)	7408	47	50	110.2
S-S-C(2)	7412	48	50	104.2

protein is the atom C_γ and other side-chain atoms farther from the C_α–C_β bond. For a leucine residue, this includes only the atoms C_γ, $C_{\delta 1}$, and $C_{\delta 2}$; the energy of interest comes from one torsional component and the nonbonded interactions of these three atoms with the rest of the protein (and the surroundings, if included in the calculation). Variation of a single torsion angle thus leads to an energy curve, say $E = E(\chi_1)$. Variation of two angles, say χ_1 and χ_2, produces the energy surface $E = E(\chi_1,\chi_2)$.

The full set of interactions defined by eq 1 was divided so as to be able to study the relative importance of short-range, long-range, and exterior forces in determining the side-chain torsional potentials. Three types of calculations are reported in this paper.

(*1*) *Interaction of Side Chain with Dipeptide.* The complete lists of coordinate and interaction specifications can be edited so as to discard all interactions not involving the dipeptide backbone to which the side chain is attached. The interaction energy then consists of torsional terms and a relatively small number of nonbonded interactions between the moving atoms of the side chain and the stationary atoms of the backbone. For the dipeptide fragment [$-CH_2-CONH-CHR-CONH-CH_2$], the calculated backbone, side-chain energy arises from interaction of the moving part of the side chain R with the

Table III: Symmetry Transformations[a] Applied to Central Molecule to Generate Crystal Environment

symmetry	M	V
x, y, z (identity)	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{matrix} 0 \\ 0 \\ 0 \end{matrix}$
$^1/_2 - x, -y, ^1/_2 + z$	$\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{matrix} ^1/_2 \\ 0 \\ ^1/_2 \end{matrix}$
$^1/_2 + x, ^1/_2 - y, -z$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$	$\begin{matrix} ^1/_2 \\ ^1/_2 \\ 0 \end{matrix}$
$-x, ^1/_2 - y, ^1/_2 - z$	$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$	$\begin{matrix} 0 \\ ^1/_2 \\ ^1/_2 \end{matrix}$

[a] For any atom with a coordinate vector in atom fractional coordinates $\mathbf{r} = (x/a, y/b, z/c)$ (where x, y, and z are Cartesian coordinates in angstroms and $a = 43.1$, $b = 22.9$, and $c = 48.6$ Å are the unit-cell parameters), new coordinates $\mathbf{r}' = \mathbf{M}\cdot\mathbf{r} + \mathbf{V}$. Also, the translations $x' = x \pm a$, $y = y \pm b$, and $z' = z \pm c$ can be applied after any of the above transformations.

Table IV: Hydrogen Bonds between the Central Protein Molecule and Its Crystal Environment

donor	acceptor	heavy-atom distance (Å) (X-ray coordinates)
A. Donor, Central PTI Atom		
N(1)	solvent	3.31
$N_{\eta 2}$(17)	solvent	2.94
N_{ϵ}(20)	solvent	3.07
$N_{\eta 1}$(20)	solvent	3.16
$N_{\eta 2}$(20)	solvent	2.72
$N_{\delta 2}$(24)	solvent	2.98
$O_{\gamma 1}$(32)	solvent	3.02
N_{ϵ}(39)	$O_{\epsilon 1}$(49)	3.11
$N_{\eta 1}$(39)	O_{η}(21)	2.78
$N_{\eta 2}$(39)	solvent	2.94
O_{γ}(47)	solvent	2.86
B. Donor, Solvent; Acceptor, Central PTI Atom		
	O	2.93
	$O_{\epsilon 1}$(7)	3.08
	O(16)	3.01
	O_{η}(21)	3.13
	O_{η}(23)	3.11
	O(23)	2.72
	O(28)	2.81
	O(30)	2.95
	O(30)	3.11
	$O_{\gamma 1}$(31)	3.02
	O(32)	3.18
	O(32)	2.81
	O_{η}(35)	2.83
	O(37)	2.52
	$O_{\epsilon 1}$(49)	3.02
	O(49)	2.72
	$O_{\delta 1}$(50)	2.84
	$O_{\delta 2}$(50)	2.84
	$O_{\delta 2}$(50)	2.96
	O(52)	2.85
	O(54)	3.01
	O(56)	3.00
	$O_{\text{carboxyl},1}$(58)	2.66
C. Donor, Exterior Protein Atom; Acceptor, Central PTI Atom		
O	$N_{\eta 1}$(39)	2.78
$O_{\epsilon 1}$(49)	N_{ϵ}(39)	3.11
$O_{\epsilon 2}$(49)	$N_{\eta 2}$(39)	3.06
$O_{\text{carboxyl},2}$(58)	$N_{\eta 2}$(17)	2.91

backbone in the conformation found in the protein.

(2) *Interaction of Side Chain with Protein.* In this level of calculation, the side-chain atoms under consideration interact with all protein atoms within 8 Å; the interaction list

Table V: Classification of PTI Side Chains

A. Straight-Chain: Unbranched before δ Carbon (14 Residues)	
6 arginine	(1, 17, 20, 39, 42, 53)
4 lysine	(15, 26, 41, 46)
2 glutamate	(7, 49)
1 glutamine	(31)
1 methionine	(52)
B. Branched at γ Carbon (15 Residues)	
2 aspartate	(3, 50)
3 asparagine	(24, 43, 44)
2 leucine	(6, 29)
4 phenylalanine	(4, 22, 33, 45)
4 tyrosine	(10, 21, 23, 35)
C. Short and β-Branched (7 Residues)	
3 threonine	(11, 32, 54)
2 isoleucine	(18, 19)
1 valine	(34)
1 serine	(47)
D. Others: Side Chains Not Rotatable (22 Residues)	
6 glycine	(12, 28, 36, 37, 56, 57)
6 alanine	(16, 25, 27, 40, 48, 58)[a]
6 cystine	(5-55, 14-38, 30-51)[b]
4 proline	(2, 8, 9, 13)

[a] In the extended atom model, the CH_3 group of alanine is represented as a sphere. [b] Disulfide bonds.

thus includes the dipeptide interactions as a subset. The nonbonded interaction set is chosen on the basis of the X-ray coordinates and is kept fixed in the calculation of the energy curves and surfaces. In addition to torsional and nonbonded components, the list contains hydrogen-bonded interactions in certain cases. Reorientation of each side chain in the protein is studied independently; that is, after a side chain is studied, it is restored to its original coordinates before another is examined.

(3) *Interaction of Side Chain with Protein and Surroundings.* This highest level of the hierarchy includes the two previous levels as subsets. The side chains are again considered independently, and the surroundings (other protein molecules, plus solvent) are kept stationary. Only the solvent molecules which are localized enough to be clearly observed in the X-ray structure are included in this calculation. The assumption that these contribute to stabilizing a specific conformation is probably more valid than it would be for solvent molecules in general.

The rigid-mapping procedure at all three levels of calculations expresses the interaction energy in the approximation that the surroundings represent a rigid structure. Adiabatic mapping, in which the system is energy minimized for each new side-chain conformation, would introduce flexibility of the central protein molecule and of the solvent and the other protein molecules. Such calculations are much more complicated (Gelin & Karplus, 1975) and appear not to be required as long as the primary concern is with the neighborhood of the energy minimum rather than with the heights of the barriers.

Results

Before describing the potential curves and surfaces obtained by the empirical energy function calculations, it is useful to summarize some general characteristics of the expected results. The 58 residues of PTI can be classified according to the branching properties of their side chain (see Table V). For the first two categories in Table V, the "straight-chain" (A) and "γ-branched" (B) groups, the nature of backbone–side-chain interactions as a function of χ_1 rotation can be explained with the aid of Figure 1a. The figure is a projection along the C_α–C_β bond, the axis of χ_1 rotation, viewed from

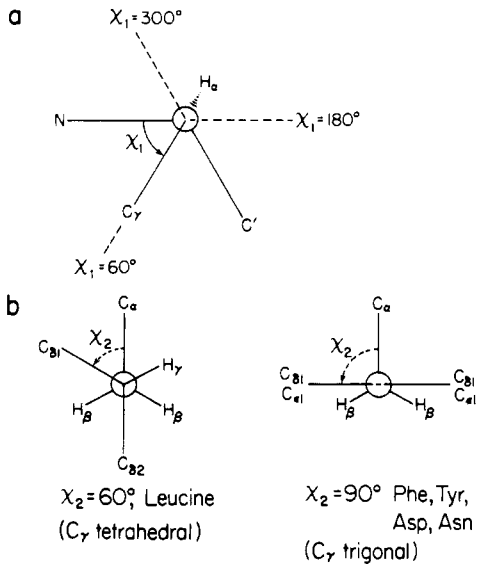

FIGURE 1: Projection diagrams for χ_1 and χ_2 rotations. (a) χ_1 axis viewed from C_β toward C_α; (b) χ_1 axis viewed from C_γ toward C_β; left, C_γ tetradedral; right, C_γ trigonal.

C_β toward C_α. The alkane-like staggered conformations occur at χ_1 = 60, 180, and 300°, where C_γ makes the smallest 1,4 interactions with the peptide nitrogen N, the carbonyl carbon C′, and the hydrogen H_α of the backbone. The curve $E(\chi_1)$ is not expected to have exact threefold periodicity because of differences among the interactions $C_\gamma \cdots N$, $C_\gamma \cdots C'$, and $C_\gamma \cdots H_\alpha$; in the extended-atom representation, the latter interaction, which is the smallest of the three, is not explicitly included because H_α and C_α are combined into a single extended atom. Similar local behavior is expected for category C of Table V, the short and β-branched side chains. The essential feature in all categories is that both C_α and C_β are tetrahedrally bonded, leading to alkane-like approximately threefold potentials. However, in all cases the minima at 60, 180, and 300° can be perturbed by interactions involving atoms not pictured in Figure 1a (i.e., the rest of the side chain, the rest of the protein, and solvent and adjacent protein units surrounding a given protein).

The χ_2 torsion gives rise to two cases, as shown in Figure 1b. When C_γ is tetrahedral, as it is for leucine and all the straight-chain residues, the torsional potentials for χ_2 are expected to have approximate threefold minima at 60, 180, and 300°, with the same limitations as stated above for χ_1; in particular, the steric interaction of C_δ with the β hydrogens (neglected in the extended-atom model) is small so that the dominant repulsion is between C_δ and C_α. A trigonal C_γ, as in the aromatic chains and Asp and Asn, should have a potential $E(\chi_2)$ with two minima at an interval of 180°; His and Trp, which do not occur in PTI, are expected to be similar. Ignoring perturbation due to interactions involving atoms not shown in Figure 1, $E(\chi_2)$ is most likely to have minima near χ_2 = 90 and 270°, due to the 1,4 interactions $C_\alpha \cdots X_\delta$ (or $C_\alpha \cdots X_{\delta 1}$ and $C_\alpha \cdots X_{\delta 2}$) which are most important in the sketched dipeptide segment.

Torsional potentials for rotation about bonds farther from C_α, such as $\chi_3(C_\gamma - C_\delta)$, $\chi_4(C_\delta - C_\epsilon)$, and $\chi_5(C_\epsilon - N_\zeta)$, can be understood by similar considerations. Chandrasekaran & Ramachandran (1970) have discussed some of these angles and the different cases that arise with tetrahedral and trigonal bonding. In the remainder of this paper, only the χ_1 and the χ_2 potentials are considered in detail, although all possible torsional angles are calculated and compared with experiment. The angles χ_1 and χ_2 account for 65 of the 95 side-chain torsional angles in PTI (categories A–C of Table V). Their potentials tend to be steeper than $E(\chi_3)$, $E(\chi_4)$, etc., since rotations about bonds farther from C_α cause smaller changes in backbone–side-chain potentials; $E(\chi_1)$ and $E(\chi_2)$ are thus best suited to studying the balance of the short-range dipeptide and the long-range protein interactions. When the crystal surroundings of the protein are taken into account, significant effects are observed in certain residues for torsion angles both near to the backbone and far from it. Since the angles χ_1 and χ_2 produce the largest reorientation of the side chain, the most striking results are obtained by examining their behavior.

A. Dipeptide Potentials: Short-Range Interactions. The energy surfaces $E(\chi_1,\chi_2)$ for the dipeptides of all the residues in categories A and B of Table V display a striking correlation between surface characteristics and category. A number of examples are shown in Figures 2–4; in each of the (χ_1,χ_2) maps, the position of the residue in the protein is indicated by a dot. The straight-chain pattern is exemplified by the surfaces for Arg-20 and Gln-31 in Figure 2. The energy is nearly independent of χ_2 over much of the map; in some cases this is true for large regions of χ_1 as well. The threefold minima described above (χ_1 and χ_2 each at 60, 180, and 300°) do not emerge clearly; in fact, other than the χ_1 minimum region near 60°, the minima are broad and not well-defined. The γ-branched pattern typical of aspartic acid, asparagine, leucine, phenylalanine, and tyrosine is exemplified by Asn-43, Asp-3, Phe-4, and Tyr-35 whose dipeptide surfaces are shown in Figure 3. For Phe-4 [Figure 3d (i)], for example, the minima appear rather clearly with those of χ_1 near 60, 180 (actually ~200°), and 300° (actually ~290°) and χ_2 near 90 and 270°; similar results are seen for Tyr-35. A detailed examination of the map for Asn-43 [Figure 3a (i)] shows that the pattern of minima is rather complicated; there are low-energy regions around χ_1,χ_2 equal to 45,90° and 45,270°, multiple minima in the regions $190° < \chi_1 < 290°$, $0 < \chi_2 < 90°$, and the corresponding regions around 270°. For Asp-3 [Figure 3b (i)] the map is also rather complicated with multiple minima in the various regions. In all cases the dipeptide maps are rather permissive and leave wide regions of low energy available. In accordance with the introductory discussion, the similar surfaces for Leu-6 and -29 (not shown) have their χ_2 minima spaced at 120° intervals since C_γ of leucine is tetrahedral. In all of these amino acids, the presence of two δ atoms doubles the number of interactions affected by χ_2 rotation, and a clearer pattern of minima is obtained.

To clarify the map results, Figure 4 shows calculated dipeptide potential energy curves (dashed lines) as a function of χ_1; Gln-31 (Figure 4a) and Glu-49 (Figure 4b) represent the straight-chain case and Phe-4 (Figure 4c) and Asp-50 (Figure 4d) the γ-branched case. In all of the figures there is a flat or nearly flat minimum between 180 and 290°; if H_α had been included, there would be a small maximum at ~240°. The curves for Gln-31, Phe-4, and Asp-50 rise to a maximum at 360°, while Glu-49 has its maximum shifted to ~40°. The behavior between 0 and 180° is generally different for the four amino acids, with Gln-31, Glu-49, and Phe-4 having a relative maximum near 60° and a maximum near 120°; Asp-50 has its minimum shifted to 40°. The important effects of the protein and the environment are also indicated in Figure 4 and will be discussed subsequently (see Results, sections B and C).

The third class of rotatable side chains (class C in Table V) includes the short serine side chain and the β-branched residues, threonine, isoleucine, and valine. Three examples of $E(\chi_1)$ potentials for this category are shown in Figure 5.

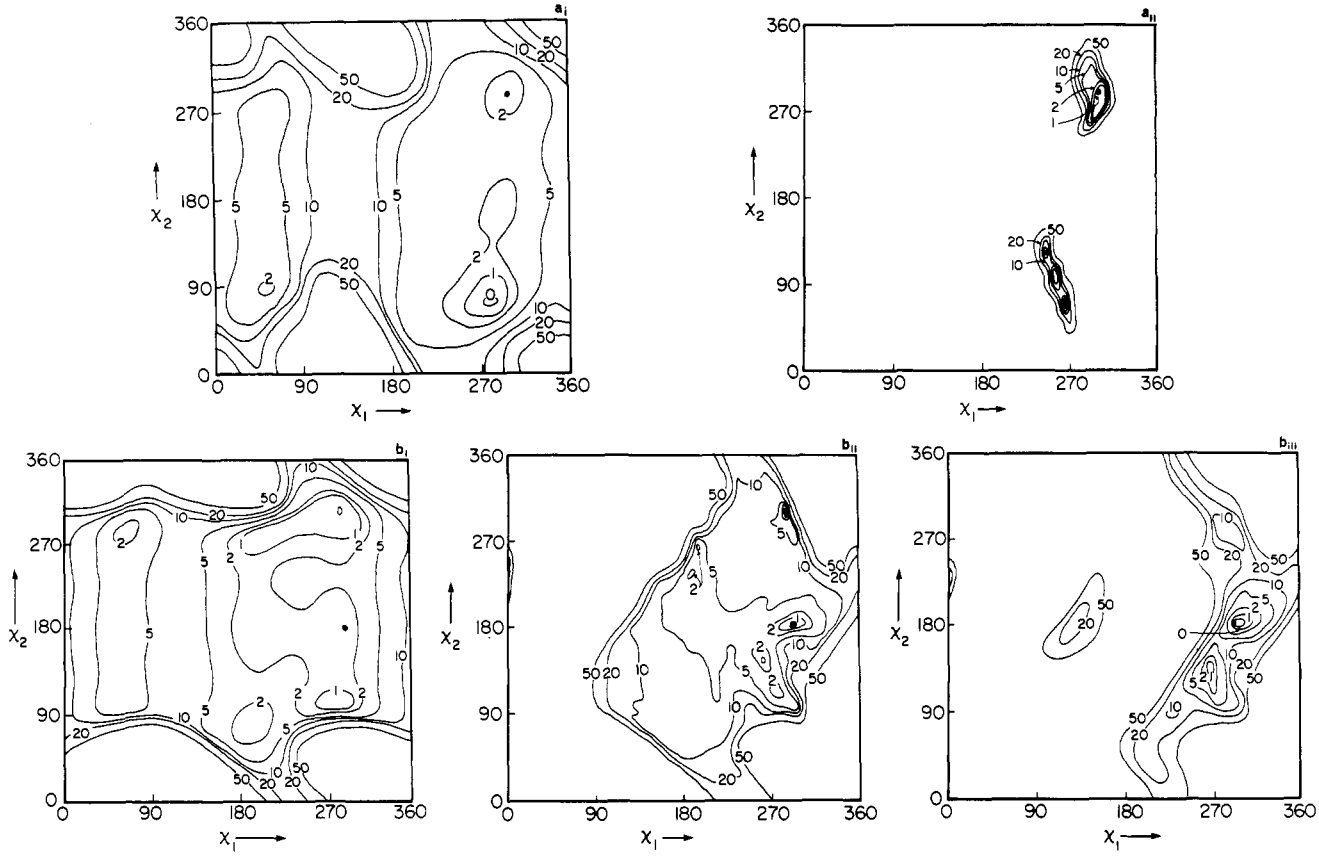

FIGURE 2: Side-chain torsional potential map (χ_1 vs. χ_2) of straight-chain group (angles in degrees and energies in kilocalories per mole; the dot indicates the X-ray value). (a) Arg-20 in the dipeptide (i) and the protein (ii). (b) Gln-31 in the dipeptide (i), the protein (ii), and the protein plus surroundings (iii).

The clearer threefold patterns for χ_1 are those of the β-branched side chains Thr-11 (Figure 5a) and Thr-54 (Figure 5c), which have barriers at $\chi_1 = 0$, 120, and 240°; the other potentials (e.g., Ser-47, Figure 5b) have a large barrier near $\chi_1 = 120°$, and a smaller one near $\chi_1 = 0°$, but lack one near 240°. As discussed at the beginning of the section, H_α is not explicitly represented in the extended-atom model and the barrier at $\chi_1 = 240°$ is missing or very small. There is an intrinsic torsional component of $E(\chi_1)$ which contributes a barrier of 1 kcal/mol at $\chi_1 = 0$, 120, and 240°. For the β-branched threonine, however, there are always two 1,4 interactions, one involving $O_{\gamma 1}$ and the other, $C_{\gamma 2}$; when $O_{\gamma 1}$ eclipses H_α, $C_{\gamma 2}$ eclipses N, producing a barrier at $\chi_1 = 240°$ somewhat more comparable to those at $\chi_1 = 0$ or 120°.

From the dipeptide maps and potential curves it is clear that the local side-chain–backbone interactions produce side-chain torsional potentials that can account for the observed angles in many cases. This is in agreement with the statistical analysis of Chandrasekaran & Ramachandran (1970). In what follows we go beyond the dipeptide model and consider the modifications of the side-chain torsional potentials due to the nonbonded interactions between a side chain and its surroundings.

B. Protein Potentials. When the energy expression for rigid mapping includes the interactions of a side chain with all other protein atoms inside an 8-Å radius, the resulting curves $E(\chi_1)$ and surfaces $E(\chi_1,\chi_2)$ have calculated minima which should in most cases agree with the experimental torsion angles obtained from the X-ray coordinates. Table VI shows the results obtained for the 36 residues in PTI with rotatable side chains (classes A–C of Table V). The numbers without parentheses in each χ_i column of the table give the torsion angles obtained from the X-ray coordinates. If there are no numbers in parentheses next to a given number, the observed torsion angle is within 0.5 kcal/mol of the primary minimum of the curve $E(\chi_1)$. For steep potential curves (narrow minima) the criterion of 0.5 kcal/mol relative energy requires agreement of the calculated and experimental angle to within a few degrees; in most cases the potentials are such that the agreement is better than 10–15°. If the calculated and experimental minima do not agree, the range of $E(\chi_i)$ over which the relative energy is 0.5 kcal/mol or less is given in parentheses next to the experimental angle. These cases usually represent broad, shallow minima (see next paragraph) not typical of most of the curves which agree well with experiment. For χ_1 of Asp-3, χ_2 of Asp-50, and χ_2 and χ_3 of Arg-53, secondary minima are used for the comparisons, and the relative energy of the second minimum is indicated along with the range in parentheses.

The 36 residues in Table VI are considered in two groups: 27 residues for which 42 of the 58 torsion angles meet the criterion described above and nine residues which give generally poor results. The latter nine are the two glutamic acids, the four lysines, and three of the arginines (17, 39, and 42); their polar or charged side chains are directed away from the protein core and are expected to be influenced more by the surroundings than by the central protein, as shown in section C below. For six of the χ angles of the group of nine residues, the table lists the calculated minima as agreeing with the experimental angle only because the potentials are flat over large regions; they are χ_2 of Glu-49 and Lys-41; χ_3 of Arg-39 and Arg-42; and χ_4 of Arg-17 and Lys-41. Other entries involving the nine residues indicate the broadness of the minima by the range of angles given in parentheses.

The potential curves for the 27 residues that give satisfactory results generally have relatively narrow minima, i.e., the 0.5

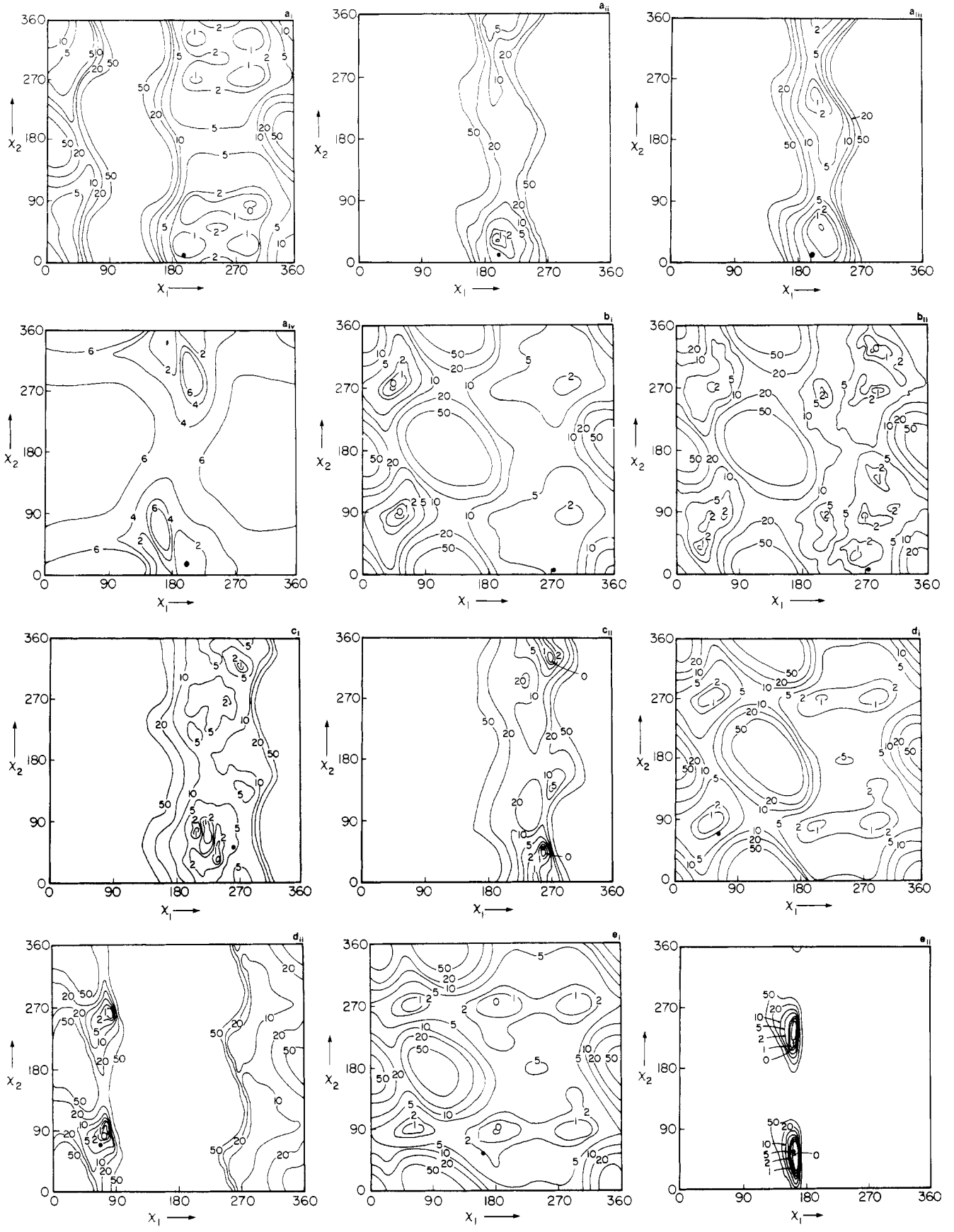

FIGURE 3: Side-chain torsional potential map (χ_1 vs. χ_2) of γ-branched group (see Figure 2 caption). (a) Asn-43 in the dipeptide (i), the protein (ii), van der Waals contribution (iii), and H-bond contribution (iv). (b) Asp-3 in the dipeptide (i) and the protein (ii). (c) Asp-50 in the protein (i) and the protein plus surroundings (ii). (d) Phe-4 in the dipeptide (i) and the protein (ii). (e) Tyr-35 in the dipeptide (i) and the protein (ii).

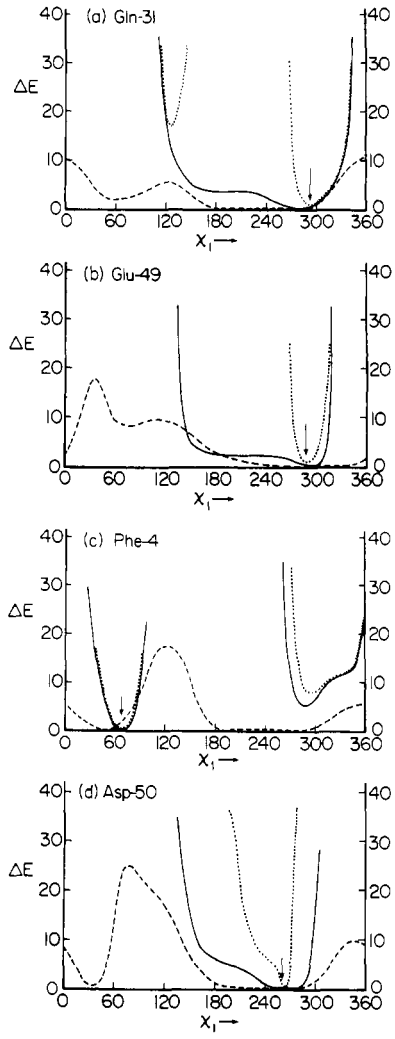

FIGURE 4: Comparison of torsional potentials, $E(\chi_1)$, based on interactions from dipeptide, protein, and protein plus surroundings for γ-branched and straight side chains [angles in degrees and energies in kilocalories per mole; the arrow indicates the X-ray value; (---) dipeptide; (—) protein; (···) protein plus surroundings]. (a) Gln-31; (b) Glu-49; (c) Phe-4; (d) Asp-50.

kcal/mol cutoff corresponds to an angular range of 20° or less. Of the 16 torsion angles not meeting the agreement criterion, eight are very minor discrepancies; i.e., χ_2 of Leu-6 (experimental angle 186°, calculated 0.5 kcal/mol range 144–180), χ_1 of Ile-18 (295 vs. 306–321), χ_1 of Ile-19 (303 vs. 312–327), χ_1 of Phe-22 (76 vs. 63–72), χ_2 of Asn-24 (349 vs. 309–336), χ_1 of Phe-33 (74 vs. 63–69), and both χ_1 and χ_2 of Asn-43 (χ_1, 199 vs. 204–213; χ_2, 7 vs. 15–39). Asn-43 is considered further near the end of this section. In two other cases a calculated secondary minimum is close to the observed torsional angles: χ_1 of Asp-3 (observed, 274°; secondary minimum at 1.5 kcal/mol, 276–294°) and χ_2 of Asp-50 (52; 1.0 kcal/mol, 51–57).

Of particular interest is the effect of the protein in restricting or excluding some of the torsion angle space which was accessible to the side chain in the presence of the dipeptide alone. Figure 5 shows examples of the reduction of χ_1 space for three of the short side chains. In Thr-11 (Figure 5a) the added interactions change the position and shape of the minimum near 300° and raise the other two minima with respect to it. Near χ_1 = 240°, $O_{\gamma 1}$ of Thr-11 is close to the carbonyl oxygen of Val-34, and near χ_1 = 0°, $C_{\gamma 2}$ of Thr-11 has the same interaction. These two interactions account for virtually all of the difference between the protein and dipeptide potentials

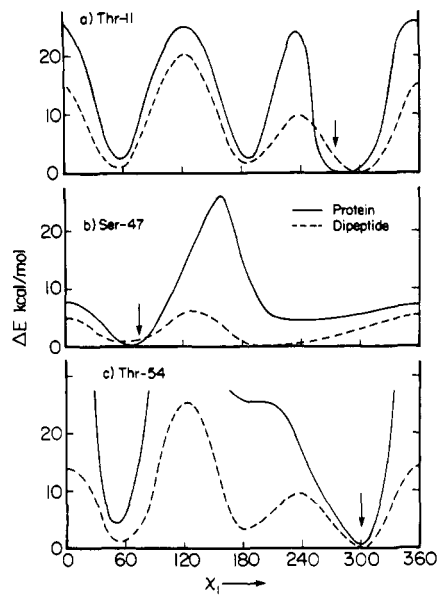

FIGURE 5: Comparison of dipeptide and protein torsional potentials, $E(\chi_1)$, for three short side chains [see Figure 4 caption; (---) dipeptide, (—) protein]. (a) Thr-11; (b) Ser-47; (c) Thr-54.

in Figure 5a. In the case of Ser-47 (Figure 5b) the protein reverses the energy ordering of the two minima, favoring by 5 kcal/mol the one corresponding to the observed torsion angle. The main difference between the curves is due to a repulsive peak near χ_1 = 150°, where O_γ of Ser-47 approaches $O_{\epsilon 2}$ of Glu-47. In Thr-54 (Figure 5c) the minimum near χ_1 = 180° is almost eliminated by a repulsive interaction between $C_{\gamma 2}$ of Thr-54 and the carbonyl oxygen of Asp-50. The slightly smaller repulsion involving $O_{\gamma 1}$ at χ_1 = 0° sharpens the remaining minima near χ_1 = 60 and 300°. In Ile-19 and Val-34 the addition of the protein potential produces only very slight interactions, so the two curves are almost identical. Thr-32 undergoes less modification than Thr-11. About half the χ_1 range of Ile-18 is eliminated by a very large repulsion between $C_{\gamma 1}$ of Ile-18 and $C_{\delta 2}$ of Tyr-35 near χ_1 = 75°.

The straight side chains, category A of Table V, show a range of effects due to the interactions with the protein. Arg-1, -20, and -53 are very severely restricted by the protein. The most extreme case is Arg-20, whose dipeptide and protein maps are compared in Figure 2a (i and ii). While the minimum within the protein corresponds to a local well on the dipeptide map, the potentials rise much more steeply on the protein map. Rotation of χ_1 moves side-chain atoms C_γ, C_δ, N_ϵ, C_ζ, $N_{\eta 1}$, and $N_{\eta 2}$, and rotation of χ_2 moves C_δ and beyond, making short side-chain–protein nonbonded interactions highly probable. Unlike some of the other long side chains, Arg-1, -20, and -53 are so situated in the protein so as to have little torsional freedom. All four lysines, Arg-17, -39, and -42, and Glu-7 have fewer restraining interactions, and addition of the protein does little to change the flat dipeptide surfaces. However, the solvent and surrounding protein units of the crystal structure exert important influences on many of these side chains, as described in the next section.

For Lys-15 (not shown) there is little change in the map due to the protein, because the residue is at the surface of the protein and its side chain is directed away from it. Gln-31 [compare Figures 3b (i and ii), as well as Figure 4a] and Met-52 [Figure 6a (i)] are significantly more restricted in the protein than the dipeptide. For Gln-31, a large repulsion between $N_{\epsilon 2}$ of Gln-31 and C_γ of Phe-22 when χ_1 is near 90° eliminates much of the left half of the potential surface. The minimum at χ_1 = ~280°, χ_2 = ~290° from the dipeptide

Table VI: Comparison of Calculated Energy Minima[a,b] with Experimental Side-Chain Torsional Angles

residue	χ_1	χ_2	χ_3	χ_4	χ_5
Arg-1	82	181	56	82 (102-120)	1
Asp-3	274 (276-294, *1.5*)	5 (30-100)			
Phe-4	66	71			
Leu-6	306	186 (144-180)			
Glu-7	17 (267-285)	177	73		
Tyr-10	178	77			
Thr-11	282				
Lys-15	249 (279-306)	92 (57-84)	141 (225-231)	52 (156-204)	
Arg-17	298 (258-297)	162 (120-126)	223 (163-210)	113	348 (354-356)
Ile-18	295 (306-321)				
Ile-19	303 (312-327)				
Arg-20	296	284	177	89	357
Tyr-21	288	77			
Phe-22	76 (63-72)	99			
Tyr-23	189	252			
Asn-24	173	349 (309-336)			
Lys-26	272 (195-228)	239 (243-297)	105 (66-102)	179 (255-288)	
Leu-29	44	72			
Gln-31	293	178	328		
Thr-32	53				
Phe-33	74 (63-69)	114			
Val-34	172				
Tyr-35	161	54			
Arg-39	314 (228-309)	308 (183-204)	182	186 (111-132)	2
Lys-41	272 (213-216)	195	185 (60-75)	180	
Arg-42	277 (261-270)	201 (60-96)	110	179 (291-294)	352
Asn-43	199 (204-213)	7 (15-39)			
Asn-44	176	330			
Phe-45	304	90			
Lys-46	280 (0-15)	160 (177-276)	218 (276-294)	91 (135-189)	
Ser-47	76				
Glu-49	286 (165-168)	87	173 (75-100)		
Asp-50	263	52 (51-57, *1.0*)			
Met-52	290	298	289 (246-282)		
Arg-53	178	156 (156, *2.5*)	37 (36-39, *7.0*)	234 (249-264)	7
Thr-54	300				

[a] Angles in degrees; energies in kilocalories per mole. [b] For each angle reported, the following format is used: A (B-C, *D*), where A is the torsional angle (in degrees) in the X-ray coordinates. If the parenthesized field is not given, the calculated curve $E(\chi_i)$ has a primary minimum such that A falls within the range of χ_i for which $E(\chi_i) \leq 0.5$ kcal/mol relative to that minimum. Otherwise, B-C is the calculated range over which $E(\chi_i) < 0.5$ kcal/mol unless *D* is given, in which case the range B-C represents a secondary minimum at *D* (kcal/mol) relative to the primary minimum.

interaction remains, but a new and deeper minimum appears which agrees well with the experimental position. Figure 4a can be interpreted as follows: the large repulsion near $\chi_1 = 90°$ eliminates some of the dipeptide curve but leaves the broad, flat region from $\chi_1 = 180$ to 300°, while an attractive interaction with the protein (the hydrogen bond between Gln-31 and Asn-24) introduces a well near $\chi_1 = 290°$. Glu-49 (Figure 4b) can be understood by similar analysis. Met-52 cannot form hydrogen bonds, so all protein influences are repulsive (the very small attractive van der Waals interactions are dominated by much larger repulsions in this and most other cases); the protein and dipeptide minima coincide and have the same energy ordering, as shown in Figure 6a (i).

The γ-branched (group B) side-chain torsional potentials exhibit a similarly wide range of effects due to addition of the protein interactions to those of the dipeptide. Two typical examples are provided by Asn-43 and Phe-4, whose protein potentials [Figures 3a (ii) and 3d (ii), respectively] resemble their dipeptide potentials [Figures 3a (i) and 3d (i)] except that large portions of χ_1 space are removed by repulsions. For Asn-43 the ranges with $\chi_1 < 90°$ and $\chi_1 > 270°$ (i.e., $-90° < \chi_1 < 90°$) represent high-energy regions because of repulsions between $O_{\delta 1}$ or $N_{\delta 2}$ of Asn-43 and C_γ or $O_{\epsilon 2}$ of Glu-7; for $\chi_1 < 90°$, an interaction of $O_{\delta 1}$ with one of the four internal water molecules is also important, while for $\chi_1 > 270°$, $N_{\delta 2}$ makes a short, repulsive contact with the carbonyl oxygen in the backbone of Phe-4. The protein excludes the region $90° \lesssim \chi_1 \lesssim 230°$ for the side chain of Phe-4 through repulsions involving the backbone peptide linkage from Arg-42 to Asn-43 and the side chain of Glu-7.

Figures 3b (i and ii) compare the dipeptide and protein surfaces for Asp-3. Of all the γ-branched residues, Asp-3 shows the smallest differences between the two types of calculation since it lies on the surface of the protein and makes few interactions with it, regardless of the values of the torsional angles. There is, however, some restriction of the conformational space. The experimental coordinates, denoted by the dot in Figure 3b, lie just outside a 2 kcal/mol contour on the protein potential surface [Figure 3b (ii)] between a shallow secondary minimum and the primary minimum; the protein potential slightly lowers the relative energy in that region as compared with the dipeptide. Even in this case, which is scored as a disagreement in Table VI, the calculated torsional angles are near the experimental values.

The rest of the γ-branched side chains give results like Asn-43 and Phe-4, with considerable narrowing of the χ_1 space; this phenomenon is more pronounced for larger side chains since the rotation of χ_1 moves more atoms. Thus, the two leucines show only slight narrowing of their dipeptide potentials upon addition of protein interactions; Asp-50, Asn-24, and Asn-44 behave much like Asn-43; and the eight aromatic side chains show the most reduction in accessible χ_1 space. Of the aromatics, Phe-4 remains the least restricted [Figures 3d (i and ii)], while Tyr-35 is the most [Figures 3c (i and ii)]. (For another comparison, see the Tyr-21 maps in Gelin & Karplus, 1975.) All eight aromatics show very sharp, well-defined

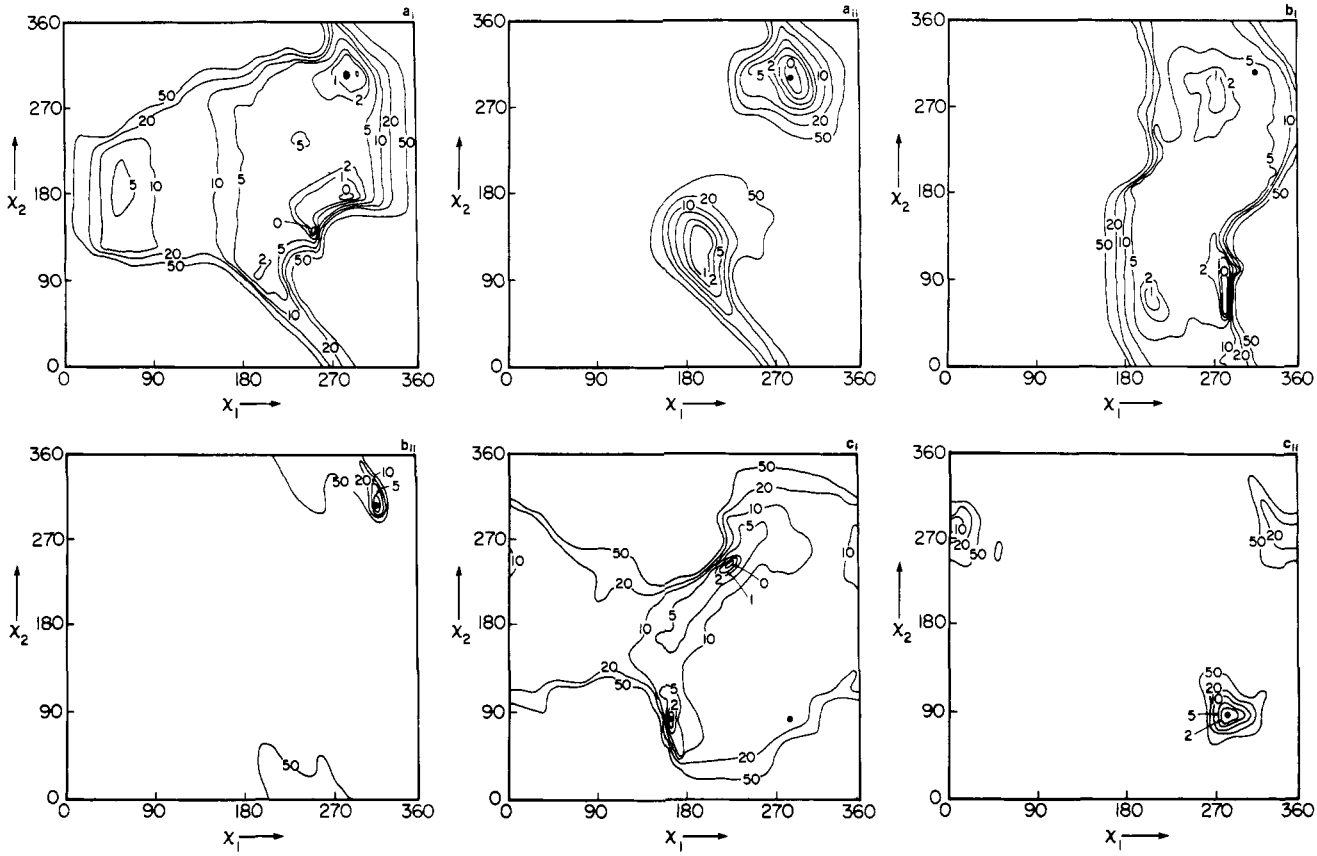

FIGURE 6: Side-chain torsional potential map (χ_1 vs. χ_2) for straight-chain groups (see Figure 2 caption). (a) Met-52 in the protein (i) and in protein plus surroundings (ii); (b) Arg-39 in the protein (i) and in protein plus surroundings (ii); (c) Glu-49 in the protein (i) and in protein plus surroundings (ii).

minima, which are in excellent agreement with experiment. The large size and rigid structure of the aromatic rings account for the importance of long-range forces in selecting a narrow range of possible χ_1 values, as compared with the flatter dipeptide maps. The same factors are also involved in placing the aromatic rings among the most precisely located features in the electron-density map; the sharpness of the calculated potentials thus correlates with the clarity of features in the map.

Asn-43 provides a particularly interesting example of the balance of forces contributing to the total-energy map, Figure 3a (ii). The van der Waals and hydrogen-bonded potentials in the protein are shown in Figures 3a (iii and iv). The range of χ_1 is selected by strong repulsive van der Waals interactions; in the remaining degree of freedom, χ_2, the primary minimum of the van der Waals potential is at $\chi_2 = 50°$. The electrostatic energy is fairly flat in the neighborhood of this region and does not affect the overall energy. The three hydrogen bonds $N_{\delta 2}(43)\cdots O(7)$, $N_{\delta 2}(43)\cdots O(23)$, and $N(23)\cdots O_{\delta 1}(43)$ produce an irregular region of low energy for χ_2 between about +15 and −15° [Figure 3a (iv)]. The total energy, which is the sum of electrostatic, van der Waals, and hydrogen-bond contributions, plus the "intrinsic" torsional potential of χ_2 (it is small enough not to matter in the region of interest), has its primary minimum at $\chi_2 = 30°$. The observed value is $\chi_2 = 7°$. It is clear from Figure 3a (ii) that the well around the calculated mininum corresponds rather closely to the experimental angular coordinates, although the discrepancy is large enough to classify both angles of Asn-43 as disagreements in Table VI. Agreement would be improved if larger hydrogen-bond strength parameters had been used. In fact, the parameters used, −2.5 and −3.0 kcal/mol [types N(2)···O and NH···O in Table II], are somewhat smaller than other estimates (McGuire et al., 1972). Clearly these internal hydrogen bonds are important in stabilizing the polar Asn side chain in the interior of the protein and determining the balance of energy contributions which leads to the observed side-chain torsional angles.

The results of this section demonstrate the possible effects of the protein matrix in modifying the dipeptide energy surfaces, $E(\chi_1,\chi_2)$. In 15 cases, the protein has essentially no effect on the dipeptide potentials; these are the nine externally directed side chains, three short chains (Ile-19, Thr-32, and Val-34), Asp-3, and two leucines, Leu-6 and -29. More commonly, the dipeptide potentials exhibit several minima, from which the protein forces "select" one or two, often narrowing them at the same time. Figures 5b and 5c show examples of the selection of a single minimum for Ser-47 and Thr-54; Ile-18 is another example. For the γ-branched chains (all three Asn, all four Phe and Tyr, compare Figures 3) and the straight chains (Arg-1 and -20, Gln-31, and Met-52), the narrowing of the available χ_1 space is an important aspect of this selection process. In some cases the added protein interactions give rise to repulsions even when the side chain is at the minimum of the dipeptide interactions and thus shift the minimum; Thr-11 (Figure 5a) shows an example in which the small shift of the minimum is easy to understand since the repulsive barrier is near one already present in the dipeptide. In other cases, more and larger repulsions lead to less obvious shifts, and in Gln-31, an unusual attractive (hydrogen-bond) interaction between the side chain and the protein creates a new minimum in a location not expected on the basis of the dipeptide surface.

C. Crystalline Environment. A given PTI molecule in the crystal is surrounded by other protein molecules, whose atomic coordinates relative to the central molecule can be calculated

from the known symmetry elements, and by solvent molecules, some of which are sufficiently localized to appear unambiguously in the electron density map. The addition of the potentials from the surrounding molecules to those produced by the protein itself should result in a more realistic description of the forces which give rise to side-chain torsional potentials; this is done by including the final two terms in eq 1.

As would be expected from the protein geometry and the calculations of the preceding section, the short and γ-branched side chains, most of which are hydrophobic and internally directed, are very little affected by the addition of surroundings. An interesting exception is Asp-50 [Figures 3c (i and ii)], whose carboxylic acid function is involved in several hydrogen bonds with solvent molecules (see Table IV). The total accessible region of χ_1,χ_2 space is restricted very little by the exterior forces, but the three hydrogen bonds (with total binding energy of 10.3 kcal/mol) create a well in the energy map corresponding perfectly with the observed values of χ_1 and χ_2.

Among the γ-branched chains, the residue whose calculated minima agree least well with experiment is Asp-3. The energy map including the surroundings is not very different from that based on the protein alone [Figure 3b (ii)]; in both cases the experimental angular coordinates lie in the flat region a short distance from a shallow minimum. The carboxyl groups $O_{\delta 1}$ and $O_{\delta 2}$ of Asp-3 have some interactions with solvent and with residues Glu-49 and Arg-53 of adjacent protein units. However, all of them are fairly weak (the shortest, $O_{\delta 1}\cdots$C 49, has a distance of 3.68 Å in the X-ray coordinates, corresponding to a van der Waals energy of only −0.22 kcal/mol), and none qualify as hydrogen bonds. While the χ_1 space is somewhat reduced by the added forces, there are no decisive interactions similar to the three hydrogen bonds which stabilized Asp-50. Thus, the χ_1,χ_2 map is very open, with the observed geometry close to one of the many shallow minima. The temperature factors of the terminal atoms C_γ, $O_{\delta 1}$ and $O_{\delta 2}$ of Asp-3 are somewhat larger than those of most backbone atoms or the atoms of well-localized side chains such as the aromatics. This suggests that the side chain of Asp-3 is undergoing substantial thermal fluctuations in the crystal or moving between two or more alternative preferred conformations. Both of these possibilities are consistent with the potential energy surface [see Figure 3b (ii)] which has a broad region of low energy (5 kcal/mol or less) surrounding the two shallow minima between which the experimental result occurs.

Of the 14 long, straight side chains (group A of Table V), five were calculated to have minima corresponding closely to the X-ray conformation when the interactions with the entire protein were included. The exterior has little effect on Arg-1 and Arg-20; in the latter case, the protein forces had already greatly restricted the accessible portion of the χ_1,χ_2 plane (Figure 2a). Arg-53 undergoes some change, but not a great deal, since again the protein potentials are quite steep and give a fair account of the experimental torsional minima, if secondary minima are considered (see Table VI). Gln-31 and Met-52 are two side chains that undergo very significant sharpening of their potentials when the exterior interactions are introduced. In the case of Gln-31 [Figures 2b (i–iii)] a single solvent molecule is situated so as to exclude much of the formerly accessible χ_1 range (see Figure 4a), while in Met-52 [Figures 6a (i and ii)], this is accomplished by atoms in the Cys-14–Cys-38 disulfide bond and adjoining residues in one of the neighboring protein molecules.

Nine of the long side chains had rather flat, featureless $E(\chi_1,\chi_2)$ maps when only the dipeptide and protein forces were included. Five of these—Arg-42, Lys-15, Lys-41, Lys-46, and Glu-7—are not significantly altered by the addition of the crystalline environment. In most cases some portions of the χ_1,χ_2 plane are excluded, but what remains has quite shallow minima surrounded by flat regions. Four other side chains show very important changes; they are Arg-17, Arg-39, Lys-26, and Glu-49. In all four cases, the most important atoms contributing to the narrowing of the potentials are atoms in adjacent protein units (not solvent atoms). Arg-39 [Figures 6b (i and ii)] interacts with Glu-49 [Figures 6c (i and ii)] of an adjacent molecule, and of course the reciprocal relation is true; one hydrogen bond is involved, and there is a large ionic contribution due to the positive charge on Arg-39 and the negative charge on Glu-49. Also, each side chain forms one good hydrogen bond with a solvent molecule. Figures 6b and 6c show the extreme narrowing of the allowed space in the χ_1,χ_2 plane for these two side chains, and Figure 4b represents the restriction of χ_1 space for Glu-49. Arg-17 and Lys-26 interact with the carboxy-terminal residue, Ala-58, of an adjacent protein molecule; Arg-17 has the most solvent interactions of these four side chains.

Table VII lists the most important interactions with the crystalline environment for the seven significantly improved side chains. The table also briefly describes the surroundings of Asp-3 and the five straight-chain cases not improved by addition of the exterior forces. Asp-3 has been discussed above; for Glu-7, Lys-41, and Arg-42, it is clear that the surroundings contribute little to the side-chain potentials simply because

Table VII: Importance of Surroundings

residue	interaction[a]	distance (Å)	vdW[b]	elec[b]	H bond[b]
A. Side Chains Significantly Improved by Inclusion of Interactions with Surroundings					
Arg-17	$N_{\eta 1}\cdots$solvent	3.61	−0.17	0	0
	$N_{\eta 2}\cdots C_\alpha 58$	2.65	3.5	0.68	0
	$N_{\eta 2}\cdots$solvent	2.94	0.17	0	−2.4
Lys-26	$C_\delta\cdots C_\beta(58)$	3.67	−0.13	0.03	0
	$C_\epsilon\cdots C_\beta(58)$	3.42	0.01	0.12	0
	$N_\zeta\cdots C_\beta(58)$	3.12	0.51	0.45	0
Gln-31	$N_{\epsilon 2}\cdots$solvent	3.25	−0.17	0	0
Arg-39	$N_\epsilon\cdots O_{\epsilon 1}(49)$	3.11	−0.28	0.57	−3.0
	$C_\zeta\cdots C_\beta(48)$	3.67	−0.15	0.23	0
	$N_{\eta 1}\cdots O_\eta(21)$	2.78	0.87	0.99	−2.3
	$N_{\eta 2}\cdots$solvent	2.94	0.19	0	−2.4
	$N_{\eta 2}\cdots O_{\epsilon 2}(49)$	3.06	0.23	−3.9	0
Glu-49	$O_{\epsilon 1}\cdots N(3)$	2.95	−0.12	2.41	0
	$O_{\epsilon 1}\cdots$solvent	3.02	−0.12	0	3.0
	$O_{\epsilon 1}\cdots N_\epsilon(39)$	3.11	−0.28	0.57	−3.0
	$O_{\epsilon 1}\cdots C_\alpha(2)$	3.34	−0.20	−1.24	0
	$O_{\epsilon 2}\cdots N_{\epsilon 2}(39)$	3.06	−0.23	−0.39	0
Asp-40	$O_{\delta 1}\cdots$solvent	2.84	0.30	0	−3.5
	$O_{\delta 1}\cdots$solvent	2.85	0.26	0	−3.5
	$O_{\delta 2}\cdots$solvent	2.96	−0.04	0	−3.3
Met-52	$C_\gamma\cdots C_\gamma(14)$	3.77	−0.04	−0.04	0
	$C_\epsilon\cdots C_\beta(16)$	3.50	−0.03	0.08	0

B. Side Chains Not Significantly Improved, in Poor Agreement with X-Ray (χ_1,χ_2)

Asp-3: many exterior interactions, all >3.7 Å; no hydrogen bonds

Glu-7: very few interactions; one hydrogen bond ($O_{\epsilon 1}\cdots$solvent, 3.08 Å, −2.76 kcal/mol)

Lys-15: shortest interaction ($N_\zeta\cdots$solvent) 4.37 Å; no hydrogen bonds

Lys-41: few interactions <4 Å; no hydrogen bonds

Arg-42: very few interactions, few shorter than 5 Å

Lys-46: many interactions, but very few <4 Å; no hydrogen bonds

[a] Residue atom···atom in adjacent protein unit or solvent atom. [b] Energies in kilocalories per mole calculated from distances in X-ray coordinates.

there are few exterior atoms near the side chain. There are more exterior atoms near Lys-15 and Lys-46, but these appear to be placed so as not to contribute much to the calculated torsional potentials. For all of these side chains there is some uncertainty concerning the X-ray structure, analogous to that pointed out for Asp-3 above.

Discussion

The rigid-geometry side-chain torsional potentials in the bovine pancreatic trypsin inhibitor have been studied by means of empirical energy functions. An essential element of the present investigation is the evaluation of the various contributions to the side-chain potentials; these arise from the local dipeptide backbone, from the rest of the protein, and from the surrounding solvent and protein molecules of the crystal. The relative importance of the various contributions is different for each of the side chains. Starting with the potentials due to the dipeptide backbone, successive addition of the protein interactions, and finally the contribution from the crystal surroundings results in improved agreement of calculated potential minima with the experimental torsional angles.

The dipeptide interactions usually play the dominant role in the energy curves and surfaces for the dihedral angles χ_1 and χ_2. This is analogous to the original work of Ramachandran and others on the allowed values of the peptide backbone dihedral angles. The possible significance of the preferred side-chain orientations in relation to protein folding has been remarked previously (Gelin & Karplus, 1975). Distinct straight side-chain and γ-branched side-chain patterns are found in the potential energy surfaces for χ_1 and χ_2. The observed pattern of minima in these dipeptide potentials is a consequence of simple structural symmetries of the residues.

In most cases the protein forces do not alter the position of the minima produced by the intradipeptide potentials; instead, they sharpen the potential curves and sometimes they select one out of several minima of similar energy. However, there are cases where the protein or external interactions create new minima.

Both the dipeptide and protein calculations show that the form of the potentials is usually dominated by the van der Waals terms, whose steeply repulsive r^{-12} behavior at short distances creates the large barriers which separate the different minima. Usually the barriers arise from a small number of surrounding atoms. This is primarily a consequence of the short-range nature of the interactions and the geometric limitations on the number of atoms which can fit in the neighborhood of any side chain. Since most of the intradipeptide atom–atom distances, such as $N\cdots C_\gamma$, $C'\cdots C_\gamma$, and $C_\alpha\cdots X_\delta$ ($C_\alpha\cdots X_{\delta 1}$ and $C_\alpha\cdots X_{\delta 2}$ in the γ-branched side chains), must be relatively small, it is not surprising that they usually contribute the dominant terms to the overall potential. The interactions with the protein generally appear to act as exclusions; in no case does it appear that the weak attractive van der Waals interactions cooperate to produce new minima, although the attractive properties of hydrogen-bond interactions can be important. The few cases where the torsional minima are different in the protein can be ascribed to repulsive interactions which upset the preferred minima within the dipeptide.

Polar side chains which extend away from the interior of the protein are naturally expected to be influenced more by the surrounding solvent and protein molecules than by the interior of the protein. For these side chains, the present method of adding the surroundings is successful in explaining the observed conformations in many cases. van der Waals repulsions are again the prime influence, although the large

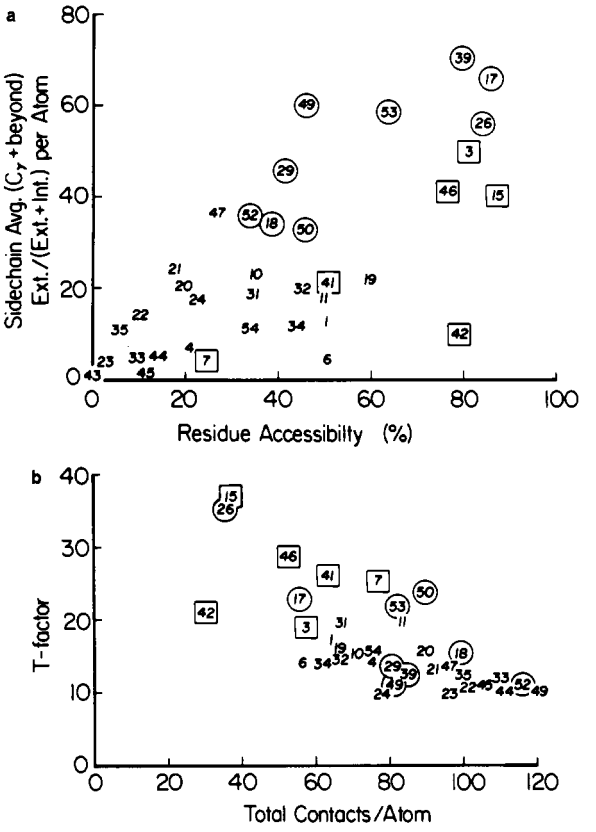

FIGURE 7: Measures of side-chain exposure, contacts, and motion. (a) Comparison of side-chain exposure to surroundings. Vertical axis: percentage of external vs. total nonbonded interactions in calculation for side-chain atoms C_γ and beyond. Horizontal axis: percentage solvent accessibility (Richards, 1977; C. Chothia, private communication). Numbers correspond to amino acid residue in sequence; for symbols, see text. (b) Comparison of X-ray temperature factor (J. O. Deisenhofer, private communication) with total number of nonbonded contacts. Vertical axis: average temperature factor for side-chain atoms C_γ and beyond. Horizontal axis: total nonbonded intactions used in the torsional potential calculations. For symbols, see text.

numbers of hydrogen bonds that are usually formed by a solvated group can be important in the detailed shape of the potential near a minimum.

The present observations concerning the importance of the environment of a side chain can be correlated with measures of the solvent exposure and the experimental uncertainty in position of the side chains. Solvent accessibility calculations (Richards, 1974, 1977; Lee & Richards, 1971) have been performed for PTI (C. Chothia, private communication). Another measure of accessibility, derived directly from the lists of nonbonded interactions appearing in the energy calculations, is the percentage of interactions for each side-chain atom due to the surroundings (as contrasted to those due to the protein itself), averaged over all of the atoms in the side chain. In Figure 7a these measures of accessibility are compared for the 36 side chains considered in this work. Although the percentage of exterior interactions on the ordinate refers to the side chains only, while the accessibility figures on the abscissa correspond to the entire residue, the correlation is clear. Low accessibility corresponds to having few exterior interactions (e.g., the asparagines and the aromatic side chains), while the polar side chains show high accessibility. In Figure 7b, the atom temperature factors derived from the X-ray analysis (Deisenhofer & Steigemann, 1975; J. O. Deisenhofer, private communication) are averaged over each residue and plotted against the average number of nonbonded interactions per

side-chain atom due to both the protein and the surroundings. There is a strong negative correlation, as expected, since a large number of contacts should reduce mobility and thereby lower the temperature factor.

To relate these results to the side-chain potentials, we indicate in Figure 7a,b by circles the nine residues whose agreement between calculated minimum and observed torsional angles is significantly improved or whose χ_1,χ_2 space is significantly restricted by the addition of the surroundings. The six residues enclosed in boxes are those which are not significantly altered by the environment. In Figure 7a it is seen that the six residues whose agreement remains poor are below the general line of correlation, while the circled ones are above. From Figure 7b it is clear that the same six residues are among those having the highest temperature factors and lowest number of nonbonded contacts. Glu-7 seems to be an exception in that it is in the interior, has an intermediate number of nonbonded contacts, but yet has a high temperature factor and a χ_1 value in disagreement with the calculations; apparently its potential surface is rather flat due to an unusual balance of forces. All of these correlations suggest that the flat potentials which persist for the six residues may not represent a fault of the calculations but, instead, suggest that it is possible for side chains to be free in the crystal and undergo large thermal motions. Fluctuations of this type could be associated with the lack of a regular solvent cage that would result in regions of disorder and low apparent electron density around the side chains. In the present model, the absence of localized solvent molecules to include in the potential functions contributes to the flat calculated potentials. It is of interest to note that the molecular dynamics studies of PTI (McCammon et al., 1977) (which are based on a potential equivalent to the protein potential used here) showed a correlation between atom positional fluctuations and temperature factors; especially large fluctuations occurred for surface residues.

Summarizing, it is clear that the rigid-geometry mapping technique based on empirical energy functions and suitably chosen energy parameters is a computationally convenient method which gives accurate and physically meaningful results when applied to the evaluation and analysis of side-chain torsional potentials. The potentials are shaped primarily by a small number of repulsive and usually local van der Waals interactions and fall into two main patterns related to the structure of the side chain, the short and γ-branched pattern and the straight side-chain pattern. Comparison of analogous calculations using a hierarchy of potentials—dipeptide, full protein, and protein plus surroundings—shows the relative importance of these contributions in introducing individual variation.

Acknowledgments

We wish to thank Dr. J. O. Deisenhofer for providing us with the necessary X-ray data for PTI and Dr. C. Chothia for the results of the accessibility calculation.

References

Anfinsen, C. B., & Scheraga, H. A. (1975) *Adv. Protein Chem. 29*, 205–300.

Chandrasekaran, R., & Ramachandran, G. N. (1970) *Int. J. Protein Res. 2*, 223–233.

Chothia, C., Levitt, M., & Richardson, D. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 4130–4134.

Deisenhofer, J. O., & Steigemann, W. R. (1975) *Acta Crystallogr., Sect. B 31*, 238–250.

Gelin, B. R. (1976) Ph.D. Thesis, Harvard University, Cambridge, MA.

Gelin, B. R., & Karplus, M. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 2002–2006.

Hetzel, R., Wüthrich, K., Deisenhofer, J., & Huber, R. (1976) *Biophys. Struct. Mech. 2*, 159–180.

Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379–400.

McCammon, J. A., Gelin, B. R., & Karplus, M. (1977) *Nature (London) 267*, 585–590.

McGuire, R. F., Momany, F. A., & Scheraga, H. A. (1972) *J. Phys. Chem. 76*, 375–393.

Richards, F. M. (1974) *J. Mol. Biol. 82*, 1–14.

Richards, F. M. (1977) *Adv. Biophys. Bioeng. 6*, 151–176.

Snyder, G., Rowan, R., III, Karplus, S., & Sykes, B. D. (1975) *Biochemistry 14*, 3765–3777.

Warshel, A., & Levitt, M. (1976) *J. Mol. Biol. 103*, 227–249.