PROTEINS: Structure, Function, and Genetics 2:330–339 (1987)

Calibration of Effective Van Der Waals Atomic Contact Radii for Proteins and Peptides

Hiroshi Iijima, James B. Dunbar Jr., and Garland R. Marshall
Department of Pharmacology, Washington University School of Medicine, St. Louis, Missouri 63110

ABSTRACT **Effective van der Waals radii were calibrated in such a way that molecular models built from standard bond lengths and bond angles reproduced the amino acid conformations observed by crystallography in proteins and peptides. The calibrations were based on the comparison of the Ramachandran plots prepared from high-resolution X-ray data of proteins and peptides with the allowed ϕ,ψ torsional angle space for the dipeptide molecular models. The calibrated radii are useful as criteria with which to filter energetically improbable conformations in molecular modeling studies of proteins and peptides.**

Key words: van der Waals radii, conformation, amino acid residue, dipeptide approximation, molecular modeling, systematic conformational search, N-alkyl amino acid

INTRODUCTION

The term "van der Waals radii" is used with several different interpretations.[1] The background common among them is that there exist attractive and repulsive forces between any pair of atoms, regardless of whether they are polar or nonpolar atoms. The attractive force arises from (induced) dipole interactions of electron clouds in which the charge distribution is not perfectly symmetric. On the other hand, close approach of the atoms results in the strong repulsive force caused by overlap of nonbonded orbitals. Thus, one can define the van der Waals radius as the distance for the minimum energy interaction, Rmin,[2] as the energetically neutral distance at which the attractive and repulsive forces are equal, Rzero, or as the closest approach distance, the so-called contact limit distance, Rcon.[3] The existence of the strong repulsive force at short distances leads us to the simplification that an atom has its own volume into which other atoms cannot intrude.[4] The third definition is related to this simplification. The purpose of this study is to determine an essential volume of atoms that sterically restricts the allowed conformations available to a molecule.

In 1963, Ramachandran and his co-workers attempted to predict the allowed conformations for amino acid residues in proteins by using the shortest atomic approach table prepared by an exhaustive examination of peptide X-ray data available at the time.[3,5,6] Although they demonstrated an elegant scientific approach, the lack of experimental data did not allow for sufficient scope. In addition, the resolution of protein X-ray analyses was not high at the time, resulting in a difficulty in comparing the theoretically predicted conformations with experimental data.

In this study, the inverse of Ramachandran's approach was utilized, assuming that the accumulated high-resolution X-ray data on proteins and peptides, which are an abundant source of amino acid conformations, could be used to calibrate the van der Waals radii of atoms, Rcon. Molecular models built with the calibrated van der Waals radii should not reject experimentally observed conformations on the basis of atomic contacts. The calibration of such van der Waals radii would be the equivalent of simplifying the van der Waals energy function,[7–8] e.g., Lennard-Jones or Buckingham potential function, into an all-or-none function.

The utility of such a set of "effective" van der Waals radii would be to quickly exclude conformations when assuming a rigid geometry molecular model. As one attempts to explore conformational space systematically, by a statistical approach such as Monte Carlo, or by molecular dynamics, an efficient procedure to eliminate conformations from further computationally expensive procedures offers considerable practical advantages.

MATERIALS AND METHODS

Software and Databases

Systematic conformational search is a computational molecular modeling module which drives the torsional angles of molecular models sequentially in a stepwise fashion and checks the validity of the generated conformations on the basis of atomic contacts.[9] This program is available in the SYBYL software package, by which the construction of molecular models, manipulation of the models, data processing, and computer graphics were also supported.[10] Some FORTRAN 77 programs were written in order to inter-

Received June 23, 1987; accepted August 20, 1987.

Address reprint requests to Dr. Garland R. Marshall, Department of Pharmacology, Washington University School of Medicine, Box 8103, 660 South Euclid Avenue, St. Louis, MO 63110.

Hiroshi Iijima is a Guest Investigator from Pharmaceutical Lab, Kirin Brewery Co., Ltd.

face the databases and their supplemental utilities to SYBYL. The databases used in this study were Brookhaven Protein Databank[11] and Cambridge Crystallographic Database. [12]

Extraction of Amino Acid Conformations From Protein X-Ray Database

The ϕ,ψ torsional angles of amino acids in proteins were extracted from the Brookhaven database by using our modification of the PHIPSI program supplied with the database. The data from the protein X-ray database were selected based on the following criteria: 1) The nominal resolution was better than 2 Å. 2) The sequence of the protein had been determined by methods other than X-ray diffraction. 3) Duplication of the data on essentially the same protein was avoided in order to prevent skewing the distribution of amino acid conformations; e.g., trypsin and the trypsin-inhibitor complex were omitted because trypsinogen and the trypsin inhibitors were already chosen. Bence-Jones protein was excluded because the immunoglobulin FAB chain was included. When the entry was a dimer, only one of the monomers was selected. 4) The parts of a protein whose coordinates were reported to be disordered were excluded. 5) The residues either preceding or following a *cis* amide bond or a *trans* amide bond more than 10° out of planarity were not used (i.e., torsional angle ω [Cα-C′-N-Cα] should be between 170 and 190°).

The filtered data were classified into two categories. The first category consisted of alanyl-type residues that preceded alanyl- or glycyl-type residues and the second category contained alanyl-type residues preceding prolyl residues.*

A total of 30 entries, about 3,500 residues, were extracted.

Extraction of Amino Acid Conformations From Peptide X-Ray Database

The conformational information about N-alkylated peptides including prolyl residues was obtained from the Cambridge X-ray database. The entries extracted are listed in Table I.

Geometries Used in Molecular Modeling Studies

An examination of the geometry, i.e., bond lengths and bond angles, of various peptides indicated that amide bonds could be categorized into the four groups summarized in Figure 1 and that the diversity of the geometry in each group was very small. The bond angles around carbon-α are taken from AMBER standard geometry (Table II).[13,14] The bond lengths for N-H, C-H, and Cα-Cβ were set to 0.960 Å, 1.020 Å, and 1.530 Å, respectively.

The molecular models were built by using a combination of these averaged geometries. For example, the molecular model for N-methyl alanyl dipeptide (C–CONCH$_3$–CHCH$_3$–CONH–C) was constructed by connecting the amide M as N-terminal unit to the amide N (Fig. 1) using the geometry of Table II for carbon-α (Fig. 2).

The geometry and ϕ,ψ torsional angles for α-amino isobutyric acid in peptides were taken from Venkataran-Prasad and Balaram.[15]

Systematic Conformational Search

The systematic conformational search was run with an angle increment of 10° for every rotatable bond. The torsional angles, ω, were rotated in the range of 170°–190° so that some distortion of planar amide bonds was incorporated into the computations. Methyl groups (χ), which have C_3 symmetry, were rotated in the range of 0°–110°. The torsional angles ϕ and ψ were rotated in the range of 0°–350° to survey all the possible conformations.

Two reduction factors were determined for special atomic contacts. (1) In order to allow intramolecular hydrogen bonding among potential hydrogen bonding acceptor and donor atoms, the reduction of the van der Waals radii by a factor of 0.75 was applied only when the atomic contacts for such pairs were considered.[16] In the case of the molecular models used in this study, this kind of hydrogen bonding corresponds to either a γ-turn or a C_5-turn.[17] (2) Anisotropic pear-shaped character of the electron clouds was taken into account by reducing the van der Waals radii by a factor of 0.87 for 1–4 van der Waals atomic contacts. Within the range of van der Waals radii considered in this study, no conformations were rejected due to 1–4 contacts when the 1–4 contact reduction was taken into account.

Calibration Procedure

The systematic conformational search of the molecular models was carried out by using a starting set of van der Waals radii approximately 80% of the van der Waals radii parameterized as the dispersion energy minimum distance in Allinger's MMII force field.[18] The allowed conformational spaces for the models were expressed as Ramachandran plots. A comparison of the calculated plots with the ones prepared from X-ray diffraction data revealed angle space for which the conformational search did not predict the existence of experimental data. The minimum atomic contact distances (D_{ij}) for an atom pair, atom type i and atom type j, were calibrated by reducing the sum of the atomic van der Waals radii, $r_i + r_j$, until the experimentally observed conformational points were

*The naturally occurring normal amino acids are classified into three categories: (1) glycyl residue [No β-carbon], (2) prolyl residue [N-alkylated], and (3) other alanyl-type residues (structurally they are alanine derivatives).

TABLE I. Peptides Which Have N-Alkylated Amino Bonds†

Peptide	R	Φ_1	Ψ_1	Φ_2	Ψ_2	Φ_3	Ψ_3	Φ_4	Ψ_4	Φ_5	Ψ_5	Code name[29]
Ac-Met-N $(Et)_2$	0.046	−97	134									ACMTDE
Ac-D-Met-N $(Me)_2$	0.084	128	−162									AMETMA
Ac-pseudo-D-Leu-N (Me_2)	0.104	120	−135									APLEUA
Ac-Val-N (Me_2)	0.059	−132	77									BEPZAX
Ac-cyclopropyl-Ala-Phe-N (Me_2)	0.077	−52	−37	−165	167							COTCIX
Ac-D,L-Val-N (Me_2)	0.069	87	−123	(D)								NACVAL
		−91	126	(L)								
Boc-Ala-Sar-NHMe	0.070	−71	153	−103	−148							CUCPAR
Z-O-Aiv-D-MeLeu-OBt	0.038	−72*	165	87	14							HIMLBU
Ala-MeTyr (OMe)^{c}Ala-OMe**	0.048		70	−117	71	−107	163*					MAMTAL
RCO-MeTyr (OMe)-D-MeVal-NH_2	0.031	−128	62	134	−99							MAJUSB
Piv-Pro^{c}MeAla-NHiPr**	0.059	−62	135	−120	60							BABYAE
Piv-Pro-MeAla-OMe	0.049	−70	153	−92	157*							BEPYUQ
Piv-Pro-D-MeAla-NHMe	0.077	−69	164	139	−35							BAFHUL
Piv-Pro-D-MeAla-NHMe	0.067	−58	136	97	−19							BAFJAT
Boc^{c}Pro^{c}Sar-OBz**	0.070	−71	170	−99	−165							BCPSBZ
Boc^{c}Pro-Sar**	0.049	−67	152	−92								BPROSA
Boc-Pro-Sar-NHiPr	0.050	−56	135	96	−17							CUCNUJ
Boc^{c}Sar-Sar**	0.054	68	−176	−107								BXSASA
iPr-Gly-Sar-NiPr	0.060	−74	163	−90	−176							CUCPEV
Z-Aib-Pro-Aib-Pro-OMe	0.051	−53	−31	−72	−4	57	47	−77	158*			BAHNON
Boc-Val-Pro-Gly-Val-Gly	0.066	−63	125	−88	145	158	−169	−127	132	134		BAKRIO
Boc-Val-Pro-Gly-Gly-OBz	0.078	−138	158	−62	136	75	3	74	177*			BATMAK
Ac-Tyr-Pro-Asp-Gly	0.078	−73	156	−87	147	−70	141	−78				BIPHIR
Boc-Val-Pro-Gly-Val	0.073	−126	157	−57	137	109	−25	−78				BUDTUP
Boc-Aib-Pro-Val-Aib-Val-OMe	0.101	−51	−46	−74	−11	−106	−52	−61	−37	−104	−57*	BIZSOS
Z-Ala-D-Phe-Pro	0.067	−84	147	133	−123	−67						BOCAPR
Boc-Val-Pro-Gly-Val	0.098	−143	165	−53	139	104	−20	−82				BUDTOJ
Boc-Leu-Aib-Pro-Val-Aib-OMe	0.069	−104	−30	−46	−41	−65	−15	−59	−38	51	43*	BUFTOL
Z-Aib-Pro-NHMe	0.054	−51	−40	−65	−26							BXABPA
Boc-Ala-Pro	0.039	−95	154	−72								BXALPR
Boc-Leu-Aib-Pro-OBz	0.046	−63	−42	53	46	−74	162*					CALFAW
Boc-Leu-Aib-Pro	0.054	−84	163	53	37	−71						CALFIE
Ac-Asn-Pro-Tyr-NHMe	0.047	−101	106	−58	−27	−80	−8					COXJII
Ac-Asn-Pro-Tyr-NHMe	0.064	−108	108	−60	−28	−90	−5					COXJOO
Ac-Tyr-Pro-NHMe	0.062	−71	149	−75	145							COXJUU
Boc-Aib-Leu-Pro-NHMe	0.060	67	23	−133	142	−69	151					CIBTOI
iBu-Ala-Pro-NHiPr	0.046	−129	76	−67	−22							IBALPR
Piv-D-Ala-Pro-NHiPr	0.041	60	−140	−89	9							PALPRI
Pyr-Glu-Ala (thienyl) -Pro-NH_2	0.155	106***	12	−97	122	−83	147					PGTAPA
		99***	18	−96	133	−86	152					
		105***	8	−92	121	−78	144					
		107***	10	−97	124	−84	142					
Pyr-Glu-His-Pro-Tert	0.048	108***	146	−70	137	−63	154*					TRHTRT

†Ac: acetyl-, Et: ethyl-, me: methyl-, Boc: butoxycarbonyl-, Piv: pivaloyl-, Z: benzyloxycarbonyl-, Bz: benzyl-, iPr: iso-propyl-, iBu: iso-butyl-, Aib: α-aminoisobutyric acid, Sar: sarcosine, Tert: tertrate.

*The torsional angle Φ is defined as C′-O-C_α-C′ for HIMLBU. The torsional angle Ψ is defined as N-C_α-C′-O for MAMTAL, BEPYUQ, BAHNON, BATMAK, BIZSOS, BUFTOL, CALFAW, and TRHTRT.

**^{c}: Cis amide bonds.

***Pyro-glutamate is involved.

recovered in the theoretically allowed conformational space. The following inequality was obtained:

$$r_i + r_j \leqslant D_{ij}$$

The minimum value for the closest contact distance, D_{ij}, was then distributed to the van der Waals radii of each atom as follows:

$$r_i = \mu D_{ij}$$
$$r_j = (1 - \mu) D_{ij}$$

where

$$\mu = \frac{R_i}{R_i + R_j}$$

and R_i is the van der Waals radius parameterized for atom i in the MMII[18] and ECEPP force fields[19] as the dispersion energy minimum distance. In other words, the relative sizes of the atoms reflected in the parameterized force fields influenced our choice of contact radii. The resolution in determining D_{ij} was 0.01 Å.

RESULTS AND DISCUSSION

The quality of this study depends mostly on two factors. The first is the atom types. As more atom types are defined, the fit of the results of the modeling studies to the experimental data improves. But it is possible that such a set of parameters is not transfer-

able to molecules which do not belong to the class of compounds on which the parameterization was undertaken. In this study, 6 atom types were defined. They were C (sp^3 carbons, Cα, Cβ, and methyl carbons), C′ (carbonyl sp^2 carbons in amide bonds), O (carbonyl sp^2 oxygen in amide bonds), N (amide nitrogen), Hc (aliphatic hydrogen), and Ham (hydrogen on amide nitrogen).

The second factor is the geometry of the molecular models. The bond length between amide nitrogen and amide hydrogen was set to 0.960 Å, and that between sp^3 carbon and aliphatic hydrogen was set to 1.020 Å. These were the averaged bond lengths obtained from X-ray analyses, which observe the electron distribution in molecules. Techniques that evaluate the position of nuclei (neutron or electron diffraction analyses) support longer bond lengths, about 1.020 Å for N-H and 1.100 Å for C-H. The reason for using the shorter bond length is that the position of atoms should be defined as the center of the electron clouds, because the van der Waals repulsive force is explained by the orbital overlap between nonbonded atoms. This is a treatment very similar to that implied in the MMII force field for aliphatic hydrogens.[18] When the van der Waals energy terms in MMII are calculated, the positions of aliphatic hydrogens are shifted toward the carbon atoms to which the hydrogen atoms are bonded by a factor of 0.915.

The geometry associated with the carbon-α is more diverse than that of the atoms composing the amide bond. It suggests that potentially unfavorable energies required for a peptide backbone to adopt a given conformation are offset by altering the geometry around carbon-α. The bond length between carbon-α and carbon-β was found to vary from 1.500 Å to 1.565 Å in the crystal structures used. The average (mean) value of 1.530 Å was used in this study. Momany, Scheraga, and co-workers had concluded in their effort to develop the ECEPP force field that the bond angle τ (<N-Cα-C′) depends on which amino acid is considered.[20,21] For example, τ for Trp was parameterized at 108.0° while τ for Gly, Met, Tyr, and Phe was parameterized at 111.0°. We did not define a different τ for each amino acid in this study because the standard deviation of bond angles observed around carbon-α was larger than 3° regardless of how the amino acids were grouped. The standard bond angles around carbon-α parameterized in the AMBER force field were used in this study.[13,14]

The theoretical computations on the conformations of amino acid residues in proteins or peptides were carried out by molecular modeling simulations based on the "dipeptide approximation,"[3,22] in which the molecular models were built with 2 peptide bonds connected at 1 carbon-α (Fig. 2). Note that the dipeptide approximation perceives all the short-range van der Waals contacts that are dependent only on ϕ, ψ, and ω.

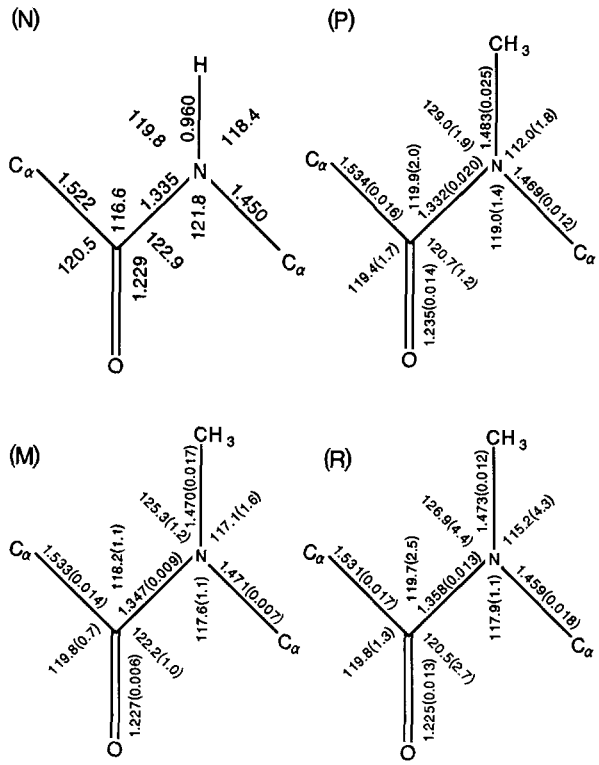

Fig. 1. Geometries of amide bonds. The geometry of the normal *trans*-amide bond (N) was taken from the work of Benedetti.[28] The geometry (P) is an averaged geometry calculated for an amide bond between an alanyl-type amino acid residue and a proline, where the N-terminal alanyl-type residue must be acylated (RCO-NH-CH (CH_3)-CO-N$\overset{(CH_2)_3}{CH}$-CO-R′). The geometry (R) was obtained for the peptides in which an alanyl-type residue or an α-hydroxy acid is preceding N-alkylated amide bonds (R-NH (or O)-$CHCH_3$-CO-NR′-$CHCH_3$-COR″). The geometry (M) was the averaged geometry of amide bonds preceding N-alkyl amino acids (alanyl- type) in which the N-terminus is acylated and the C-terminus is amidated (R-CO-NR′-$CHCH_3$-CONR″). The standard deviations are indicated in parentheses. The geometries (P), (R), and (M) were calculated by using the GEOM 78 program, which is one of the utilities provided by Cambridge. The names of the entries used to calculate the averaged geometry were as follows (see Table I and ref. 29): (P) BAHNON, BAKRIO, BATMAK, BIPHIR, BIZSOS, BOCAPR, BUDTUJ, BUDTUP, BUFTOL, BXABPA, BXALPR, CALFAW, CALFIE, COXJII, COXJOO, COXJUU, CUBTOI, IBALPR, PALPRI, PGTAPA, and TRHTRT (n = 25); (R) ACMTDE, AMETMA, APLEUA, BEPZAX, COTCIX, CUCPAR, HIMLBU, MAJUSB, MAMTAL, and NACVAL (n = 11); (M) BABYAE, BAFHUL, BAFJAT, BEPYUQ, HIMLBU, and MAJUSB (n = 8). "n" represents the number of data used to obtain the averaged geometry. As some entries have multiple molecules in a crystal cell unit, "n" is larger than the number of the entries.

The initial systematic conformational search calculations on the molecular models were carried out by using a starting set of van der Waals radii which are about 80% of the size parameterized in MMII force field as dispersion energy minimum distance, Rmin.[18] It should be mentioned that the radii in the starting set were generally large enough to prevent our model from simulating or reproducing experimentally observed conformations of molecules. These radii are

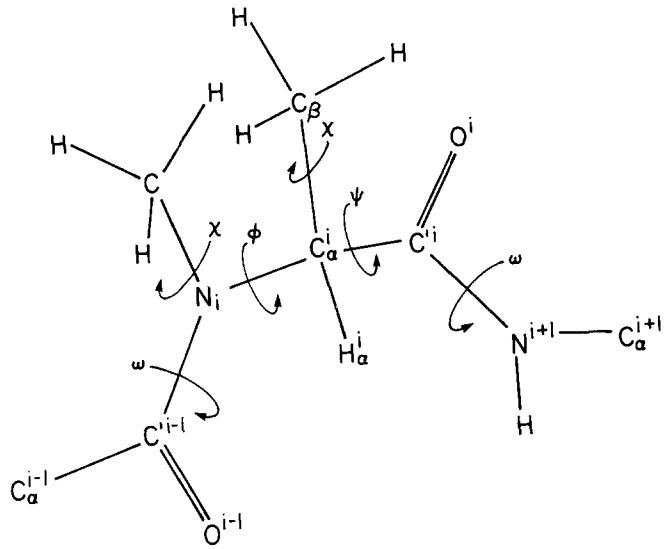

Fig. 2. Dipeptide approximation. This figure illustrates the molecular model for N-methyl alanyl-type residues in a peptide chain. The results shown in Figure 3R were obtained from this molecular model.

TABLE II. Bond Angles Around Carbon-α

<NCαC′	<NCαCβ	<CβCαC′	<NCαH	<HCαC′	<HCαCβ
110.1	109.7	111.1	109.5	109.5	109.5 (°)

TABLE III. The Calibrated van der Waals Radii (r_i)*

	C	Hc	Ham	O	N
Starting set	1.52	1.08	1.08	1.36	1.45 (Å)
ECEPP set	1.41	1.00	0.91	1.16	1.28
MMII set	1.35	1.03	0.91	1.20	1.27

*The R_i values used to calculate the calibrated radii were taken from ECEPP and MMII. ECEPP: R_c 2.06, R_{Hc} 1.46, R_{Ham} 1.34, R_O 1.56, R_N 1.89 (Å). MMII: R_c 1.90, R_{Hc} 1.50, R_{Ham} 1.33, R_O 1.74, R_N 1.82 (Å). The van der Waals radius of C′ remained at the starting value (1.53 Å), because no obvious atomic contact in which carbonyl carbons were involved was found. Thus it may be reasonable to set $r_{c'}$ to 1.42 Å, sized proportionally with r_c.

well-balanced values because, in many cases, when the van der Waals radii in the starting set are evenly reduced by a factor of 0.90–0.96, they can be satisfactorily applied in systematic conformational search analyses.[23–26] It also should be noted that the van der Waals radii in the starting set are smaller than the value corresponding to the atomic distance at which the attractive and repulsive forces are equal, Rzero. Such energetically neutral radii, Rzero, are 89% of the dispersion energy minimum radii, Rmin.

The results of systematic conformational search calculations using the starting set of radii for the molecular models, Ac-Ala-NHMe, Ac-Ala-N (Me_2) and Ac-MeAla-NHMe, are compared with the experimentally observed data extracted from the X-ray databases (upper right and upper left panels in Fig. 3N,P,R, and M). Obviously some part of the conformational angle space was not reproduced by the calculations. In the case of the molecular model Ac-MeAla-NHMe, no angle space was found to be allowed. This indicated that some of the starting van der Waals radii needed to be reduced in order to reconcile the observed conformational space available to the molecular models with that calculated.

There are many potential intramolecular atomic contacts in the molecular models employed in this study** (e.g., 21 in the model of AcNHC [CH_3] CONHMe and 28 in the 3 other models). In order to assign the pairs of atoms to the conformational spaces

**The possible intramolecular contacts must be higher than 1-4 interactions, because a reduction factor for 1-4 interactions was used as described in Materials and Methods.

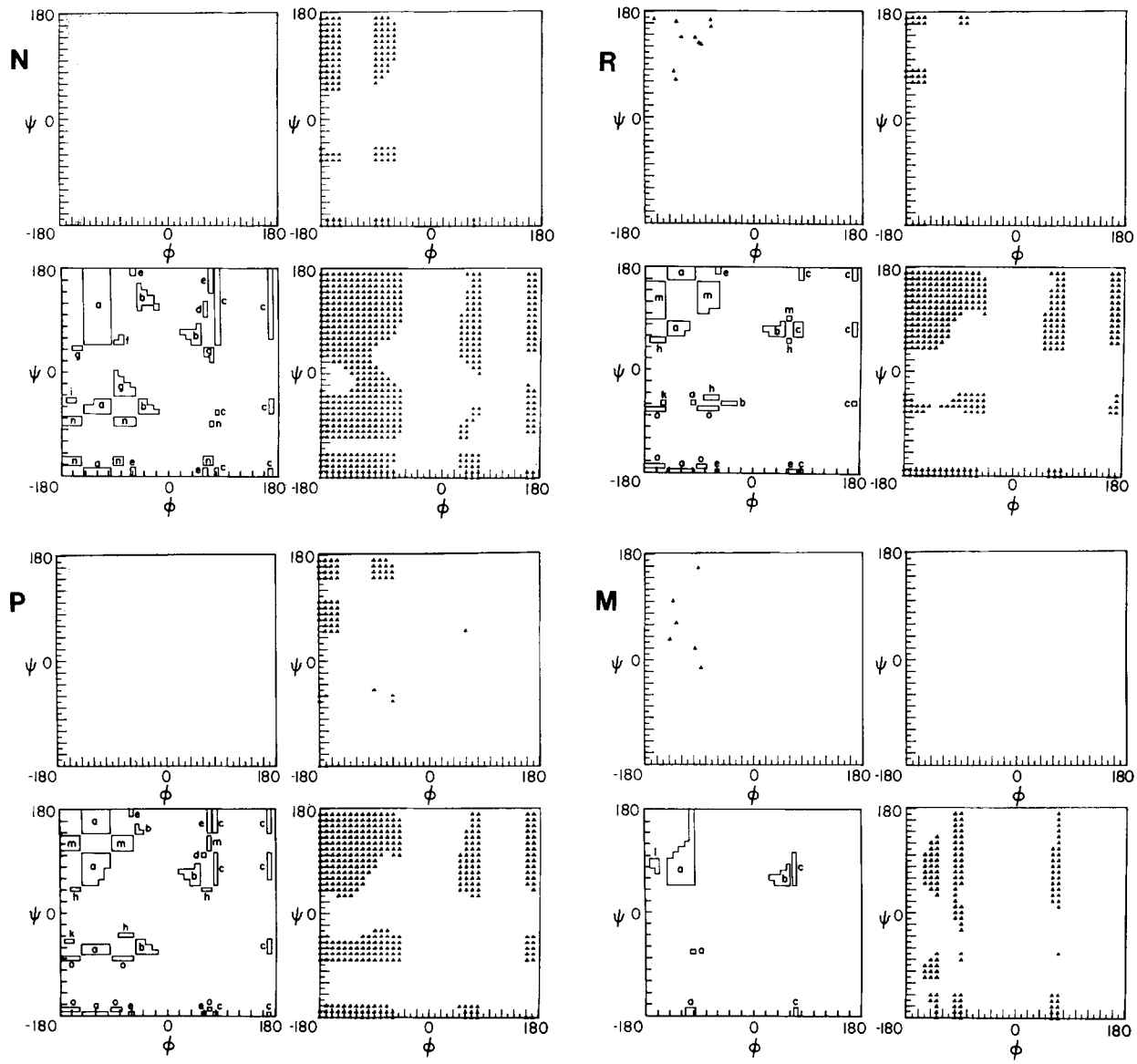

Fig. 3. **N:** Upper left: The Ramachandran plot of alanyl-type residues preceding nonprolyl residues in proteins. The entry codes of the Brookhaven data are listed in ref. 30. Upper right: The calculated angle space allowed for the molecular model, Ac-Ala-NHMe, using the starting set of van der Waals radii. Lower left: The disallowed angle space caused by single pairs of atoms. The pairs of atoms responsible for each region are a:O^{i-1}-H_α, b:O^{i-1}-C^i, c:O^{i-1}-Cβ, d:O^{i-1}-O^i, e:C'^{i-1}-O^i, f:C'^{i-1}-H(N^{i+1}), g:N^i-H(N^{i+1}), h:N^i-C(N^{i+1}Me), i:H(N^i)-N^{i+1}, j:H(N^i)-H(N^{i+1}), k:H(N^i)-C(N^{i+1}Me), l:H(N^iMe)-C'^i, m:Hα-C(N^{i+1}Me), n:Cβ-H(N^{i+1}), (o:Cβ-C(N^{i+1}Me). Lower right: The allowed angle space using the calibrated radii. The results using the MMII set are shown. The MMII set and ECEPP sets produced almost identical plots. **P:** Upper left: The Ramachandran plot of alanyl-type residues preceding prolyl residues in protein (dots) and peptides (triangles). Upper right: The calculated available angle space for the molecular model, Ac-Ala-N(Me)$_2$ where the geometry of N-methyl amide bond (P) shown in Figure 1 is employed using the starting set of radii. Lower left: The disallowed angle space caused by single pairs of atoms. (For region definitions see Fig. 3N.) Lower right: The allowed angle space after calibration. The results using the MMII set are shown. The MMII and ECEPP set produced almost identical plots. **R:** Upper left: The Ramachandran plot of alanyl-type residues preceding N-alkylated amide bonds in peptides. Upper right: The calculated allowed angle space for the molecular model, Ac-Ala-N(Me)$_2$, where the geometry of N-methyl amide bond (R) in Figure 1 was employed, using the starting set of radii. Lower left: The disallowed space caused by single pairs of atoms. (For region definitions see Fig. 3N.) Lower right: The allowed angle space after calibration. The results using the MMII set are shown. The MMII and ECEPP sets produced almost identical plots. **M:** Upper left: The Ramachandran plot of N-alkyl-alanyl-type residues preceding normal amide bonds in peptides. Upper right: The allowed angle space for the molecular model, Ac-MeAla-NH(Me), using the starting set of radii. Lower left: The disallowed space caused by single pairs of atoms. (For region definitions see Fig. 3N.) Lower right: The allowed angle space after calibration. The results using the MMII set are shown. This was the only case in which the MMII set and ECEPP set produced different plots. The allowed angle space was wider when the ECEPP set was used. The two separate regions between $\phi = -90°$ to $-150°$ merged and became a continuous region.

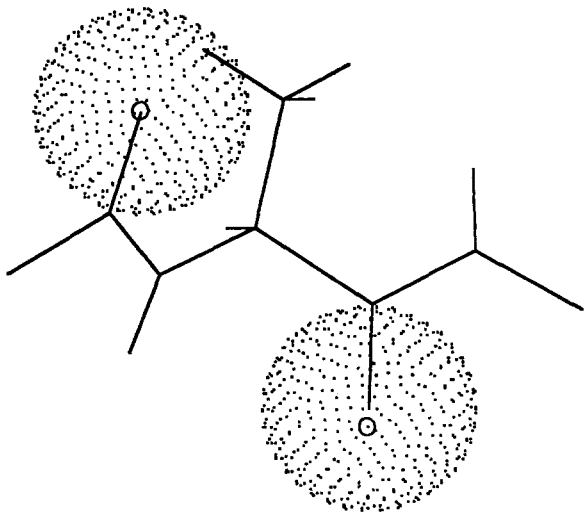

Fig. 4. An example of 2-atom molecular models. This molecular model consists of 2 oxygen atoms and dummy atoms in order to examine the effect of the O-O atomic contact. In this illustration the van der Waals radius of the oxygen atoms is 1.20 Å.

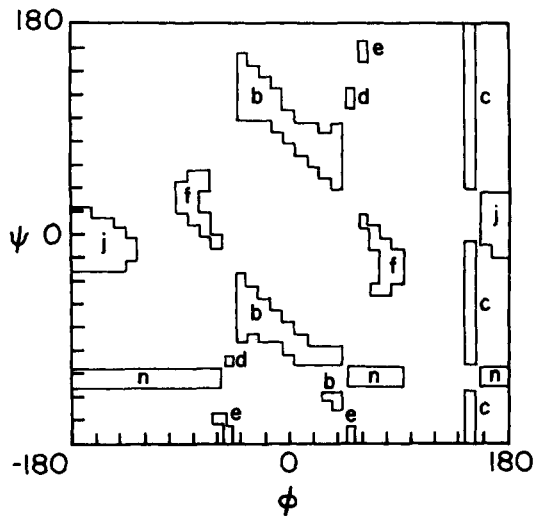

Fig. 5. Disallowed angle space for the molecular model, Ac-Ala-NHMe, because of the contacts between single pairs of atoms. (The calibrated van der Waals radii were used. For region definitions see Fig. 3N.)

where they cause unallowed atomic contacts, molecular models were built which consisted of a single pair of atoms of interest and of dummy atoms whose van der Waals radii were zero (Fig. 4). All possible "2-atom" molecular models were used in "negative" systematic conformational searches that were designed to record the disallowed conformational spaces.

Our approach to calibrating the van der Waals radii used the following two conditions: 1) It is necessary to adjust the radius of the atom type for any atomic contacts consisting of atoms of same type (homo atom contacts) to include the experimental observations. Among the 6 possible homo atom contacts (C–C, C′–C′, Hc–Hc, Ham–Ham, O–O, and N–N), the van der Waals radius of Ham had to be reduced because the negative space caused by Ham–Ham contact was found to overlap the experimental data (data not shown). This procedure gave the inequality:

$$r_{Ham} + r_{Ham} \leqslant 1.82 \text{ Å}$$

This implies that $r_{Ham} \leqslant 0.91$ Å.

2) Examinations on the conformations which were experimentally observed but were not predicted in computation led us to the fact that the majority of the unreproduced conformations were disallowed because of an atomic contact caused by a single pair of atoms (Fig. 3, lower left panels). The disallowed angle space caused by a single pair of atoms of different types (hetero atom contacts) can be recovered by adjusting the sum of their radii, i.e., either atom type. The atom pairs, Hc-O, Hc-C, and Ham-N were found to be responsible for areas where a large number of unpredicted experimental observations existed. Monitoring the maximum atomic distance needed to recover the missed angle space gave three inequalities:

$$r_{Hc} + r_C \leqslant 2.41 \text{ Å}$$

$$r_{Hc} + r_O \leqslant 2.23 \text{ Å}$$

$$r_{Ham} + r_N \leqslant 2.19 \text{ Å}$$

The four inequalities were solved by distributing the obtained distances, $D_{Hc,C}$, $D_{Hc,O}$, $D_{Ham,N}$, and $D_{Ham,Ham}$, to each atom as described in Materials and Methods. The weighting factors, μ, were calculated based on R_i values taken from MMII and ECEPP force fields, and the results are shown in Table III. Two different sets were obtained. Both sets gave practically identical Ramachandran plots that compared favorably with experimental data as shown in Figure 3 (lower right panels). All the data points obtained from peptide X-ray diffraction experiments were now included in the allowed conformational space. In contrast, a few points extracted from protein X-ray data were not included in the allowed conformational space. The negative conformational space was calculated for alanyl dipeptide by using the calibrated set of radii (Fig. 5). Most of the data points in the unpredicted conformational space are found in the negative conformational space caused by single pairs of atomic contacts. Proteins are such densely packed molecules that a few of the amino acid residues in a molecule may be forced to take energetically unfavorable conformations in order to stabilize the total structure. The correlation between the disallowed conformational space caused by single atom contacts and the unpredicted amino acid conformations supports the rationality of the calibration, because it is not reason-

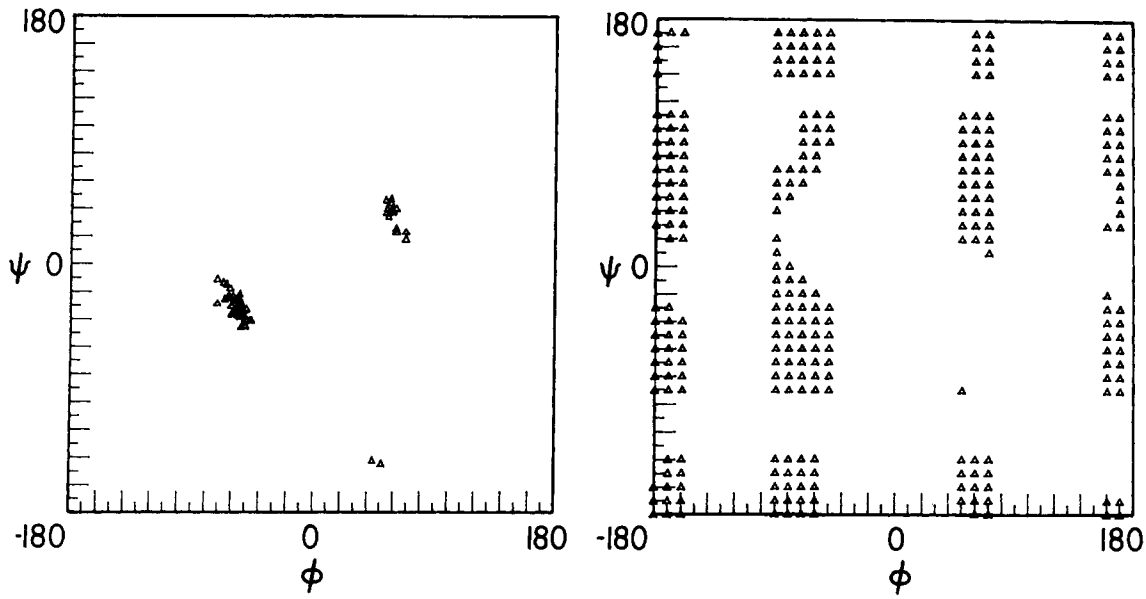

Fig. 6. The Ramachandran plot of Aib observed in peptides (left) and the angle space allowed to its molecular model using the calibrated van der Waals radii (right). The geometry of Aib dipeptide model used in the calculation was taken from the work of Venkataram Prasad and Balaram.[15]

able for an amino acid residue to take a conformation in which multiple unfavorable van der Waals contacts occur. It should be noted that rigid bond lengths and bond angles were used in this study while real molecules have some degree of flexibility to relax, thereby avoiding the bad atomic contacts.

The sets of calibrated radii were tested with α-aminoisobutyric acid (Aib), an amino acid not represented in the model peptides used to calibrate the van der Waals radii. Figure 6 shows the calculated torsional angle space for Aib with the calibrated van der Waals radii. Since Aib is a nonchiral amino acid, it was expected that its Ramachandran plot would have C_2 symmetry. However, the calculated plots are not symmetric, reflecting that the averaged geometry obtained from the X-ray data is not symmetric. The conformations of Aib determined by X-ray analyses distribute in a relatively narrow angle space, mostly in the space corresponding to right- and left-handed α-helices ($\phi = \pm 60°$, $\psi = \pm 50°$) or 3_{10} helices ($\phi = \pm 60°$, $\psi = \pm 30°$). Only 2 of 71 observed conformations were found in the space which corresponded to the extended conformation around $\phi = 50°$, $\psi = 240°$. In spite of the fact that the averaged geometry of Aib used for this calculation was weighted heavily by the contribution from helical geometries, the use of the calibrated van der Waals radii included the minor extended conformations. In contrast to the narrowly distributed experimental observations, the calculated angle space is wider. Energy calculations would be a criterion for restricting the wider conformational space.[22,27] However, we would like to emphasize that our computationally cheaper systematic search with the calibrated van der Waals radii included all the experimental observation. Long-range interactions which influence the distribution of observed conformations in the crystal data are not included in the dipeptide approximation.

In our molecular modeling study, atoms were approximated as hard sphere envelopes. The chemical interpretation of the radii calibrated in this study is that they represent the threshold distance of atomic approach; that they are the maximum size of the electron clouds that cannot be overlapped without paying considerably in energy; or that they are the values which correct for the rigidity of the geometrically fixed molecular models.

REFERENCES

1. Bondi, A.: Van der Waals volumes and radii. J. Phys. Chem. 68:441–451, 1964.
2. Burkert, U., Allinger, N.L.: "Molecular Mechanics." ACS Monograph 177. Washington, D.C.: American Chemical Society, 1982.
3. Ramakrishnan, C., Ramachandran, G.N.: Stereochemical criteria for polypeptide, and protein chain conformations. Biophys. J. 5:909–933, 1965.
4. Motoc, I., Marshall, G.R.: Van der Waals volume fragmental constants. Chem. Phys. Lett. 116:415–419, 1985.
5. Ramachandran, G.N., Ramakrishnan, C., Sasisekharan, V.: Stereochemistry of polypeptide chain configurations. J. Mol. Biol. 7:95–99, 1963.
6. Cantor, C.R., Schimmel, P.R.: Conformational analysis and forces that determine protein structure. In: "Biophysical Chemistry," Part I. San Francisco: W.H. Freeman and Co., 1980:253–309.
7. Hagler, A.T., Leiserowitz, L., Tuval, M.: Experimental and theoretical studies of the barrier to rotation about the N-C_α and C_α-C′ bonds (φ and ψ) in amides and peptides. J. Am. Chem. Soc. 98:4600–4612, 1976.

8. Kolaskar, A.S., Prashanth, D.: Empirical torsional potential functions from protein structure data. Int. J. Pept. Protein Res. 14:88–98, 1979.
9. Marshall, G.R., Barry, C.D., Bosshard, H.E., Dammkoehler, R.A., Dunn, D.A.: The conformational parameter in drug design: The active analog approach. Olson, E.C., Christoffersen, R.E., eds. In: "Computer-Assisted Drug Design." ACS Symposium series 112. Washington, D.C.: American Chemical Society, 1979:205–226.
10. Tripos Associates, Inc., 6548 Clayton Road, St. Louis, Missouri, 63117.
11. Crystallographic Data Center, University Chemical Laboratory, Lensfield Road, Cambridge, CB2 1EW, England.
12. Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, New York 11973.
13. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S., Weiner, P.: A new force field for molecular mechanical simulation of nucleic acids and proteins. J. Am. Chem. Soc. 106:765–784, 1984.
14. Weiner, S.J., Kollman, P.A., Nguyen, D.T., Case, D.A.: An all atom force field for simulations of proteins and nucleic acids. J. Computational Chem. 7:230–252, 1986.
15. Venkataram Prasad, B.V., Balaram, P.: The stereo chemistry of peptides containing α-aminoisobutyric acid. CRC Crit. Rev. Biochem. 16:307–348, 1983.
16. Schulz, G.E., Schirmer, R.H.: Noncovalent forces determining protein structure. Cantor, C.R., ed. In: "Principles of Protein Structure." New York: Springer-Verlag, 1979:33–36.
17. Rose, G.D., Gierasch, L.M., Smith, J.A.: Turns in peptides and proteins. In: "Advances in Protein Chemistry." Anfinsen, C.B., Edsall, J.T., Richards, F.M., eds. Vol. 37. Orlando: Academic Press. 1985:1–109.
18. Allinger, N.L., Yuh, Y.H.: "Quantum Chemistry Program Exchange #395," 1980.
19. Momany, F.A., Carruthers, L.M., Scheraga, H.A.: Intermolecular potentials from crystal data. III. Determination of empirical potentials and application to the packing configurations and lattice energies in crystals of hydrocarbons, carboxylic acids, amines and amides. J. Phys. Chem. 78:1595–1630, 1974.
20. Momany, F.A., McGuire, R.F., Burgess, A.W., Scheraga, H.A.: Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions and intrinsic torsional potentials for the naturally occurring amino acids. J. Phys. Chem. 79:2361–2381, 1975.
21. Némethy, G., Pottle, M.S., Scheraga, H.A.: Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions and hydrogen bond interactions for the naturally occurring amino acids. J. Phys. Chem. 87:1883–1887, 1983.
22. Pullman, B.: Proteins, nucleic acids and their constituents. In: "Quantum Mechanics of Molecular Conformations." Chichester, U.K.: John Wiley and Sons Ltd., 1976:295–383.
23. Mayer, D., Naylor, C.B., Motoc, I., Marshall, G.R.: A unique geometry of angiotensin-converting enzyme consistent with structure-activity studies. J. Comput. Aided Mol. Design. 1:3–16, 1987.
24. Naruto, S., Motoc, I., Marshall, G.R., Daniels, S.B., Sofia, M.J., Katzenellenbogen, J.: Analysis of the interaction of haloenol lactone suicide substrates with α-chymotrypsin using computer graphics and molecular mechanics. J. Am. Chem. Soc. 107:5262–5270. 1985.
25. Nelson, R., Gottlieb, D.I., Balasubramanian, T.M., Marshall, G.R.: Opioid peptides: Analysis of specificity and binding models through computer-aided drug design and structure-activity studies. In: "Opioid Peptides: Medicinal Chemistry." Rapaka, R.S., Barnet, G., Hawks, R.L., eds. NIDA Research monograph 69. Rockville: NIDA Office of Science. 1986:204–230.
26. Marshall, G.R., Humblet, C., Opdenbosch, N.V., Zabrocki, J.: Peptide bond modification and its effect on conformational mimicry. In: "Peptides." Proceedings of the Seventh American Peptide Symposium. Rich, D.H., Gross, E., eds. Rockville: Pierce Chemical Co., 1981:669–672.
27. Paterson, Y., Rumsey, S.M., Benedetti, E., Némethy, G., Scheraga, H.A.: Sensitivity of peptide conformation to geometry. Theoretical conformational analysis of oligomers of α-aminoisobutyric acid. J. Am. Chem. Soc. 103:2947–2955, 1981.
28. Benedetti, E.: Structure and conformation of peptides: A critical analysis of crystallographic data. In: "Peptides." Proceedings of the Fifth American Peptide Symposium. Goodman, M., Meienhofer, J., eds. NY: John Wiley and Sons, 1979:257–273.
29. [ACMTDE] Aubry, A., Protas, J., Cung, M.T., Marraud, M.: Acta Crystallogr. (Sect. B) 35:2634, 1979. [AMETMA] Aubry, A., Marraud, M., Protas, J., Neel, J.: C. R. Acad. Sci. (Ser. C) 273:959, 1971. [APLEUA] Aubry, A., Marraud, M., Protas, J., Neel, J.: C. R. Acad. Sci. (Ser. C) 276:579, 1973. [BABYAE] Vitoux, B., Aubry, A., Cung, M.T., Boussard, G., Marraud, M.: Int. J. Pept. Protein Res. 17:469, 1981. [BAFHUL], [BAFJAT] (1) Aubry, A., Protas, J., Boussard, G., Vitoux, B., Marraud, M.: Acta Crystallogr. (Sect. A) 37:64, 1981. (2) Aubry, A., Vitoux, B., Boussard, G., Marraud, M.: Int. J. Pept. Protein Res. 18:195, 1981. [BAHNON] Venkatachalapathi, Y.V., Nair, C.M.K., Vijayan, M., Balaram, P.: Biopolymers 20:1123, 1981. [BAKRIO] Ayato, H., Tanaka, I., Ashida, T.: J. Am. Chem. Soc. 103:5902, 1981. [BATMAK] Ayato, H., Tanaka, I., Ashida, T.: J. Am. Chem. Soc. 103:6869, 1981. [BCPSBZ] Kojima, T., Kido, T., Itoh, H., Yamane, T., Ashida, T.: Acta Crystallogr. (Sect. B) 36:326, 1980. [BEPYUQ] Aubry, A., Vitoux, B., Marraud, M.: Cryst. Struct. Commun. 11:135, 1982. [BEPZAX] Aubry, A., Cung, M.T., Marraud, M.: Cryst. Struct. Commun. 11:129, 1982. [BIPHIR] Precigoux, G., Geoffre, S., Hospital, M.: Eur. Cryst. Meet. 7:189, 1982. [BIPHIR] Cotrait, M., Geoffre, S., Hospital, M., Precigoux, G. Acta Crystallogr. (Sect. B, Str. Sci.) 39:754, 1983. [BIZSOS] Francis, A.K., Iqbal, M., Balaram, P., Vijayan, M.: J. Chem. Soc. (Perkin) 2:1235, 1982. [BOCAPR] Nair, C.M.K., Nagaraj, R., Ramaprasad, S., Balaram, P., Vijayan, M.: Acta Crystallogr. (Sect. B) 37:597, 1981. [BPROSA] Itoh, H., Yamane, T., Ashida, T.: Acta Crystallogr. (Sect. B) 34:2640, 1978. [BXSASA], [BPROSA] Benedetti, E., Ciajolo, A., Blasio, B.D.I., Pavone, V., Pedone, C., Toniolo, C., Bonora, G.M.: Macromolecules 12:438, 1979. [BUDTOJ], [BUDTUP] Yagi, Y., Tanaka, I., Yamane, T., Ashida, T.: J. Am. Chem. Soc. 105:242, 1983. [BUFTOL] Rao, C.P., Balaram, P.: Biopolymers 21:2461, 1982. [BXABPA] Prasad, B.V.V., Shamala, N., Nagaraj, R., Chandrasekaran, R., Balaram, P.: Biopolymers 18:1635, 1979. [BXALPR] Kamwaya, E., Oster, O., Bradaczek, H.: Z. Kristallogr. 149:94, 1979. [BXALPR] Kamwaya, M.E., Oster, O., Bradaczek, H., Ponnuswamy, M.N., Parthasarathy, S., Nagaraj, R., Balaram, P.: Acta Crystallogr. (Sect. B) 38:172, 1982. [CALFAW], [CALFIE] Smith, G.D., Pletnev, V.Z., Duax, W.L., Balasubramanian, T.M., Bosshard, H.E., Czerwinski, E.W., Kendrick, N.E., Mathews, F.S., Marshall, G.R.: J. Am. Chem. Soc. 103:1493, 1981. [CEWXAD] Eggleston, D.S., Jeffs, P.W., Heald, S.L., Chodosh, D.F. Am. Cryst. Assoc. (Ser. 2) 12:36, 1984. [COTCIX] Stierli, F., Prewo, V, H. Bieri, J., Heimgartner, H.: Chimia 38:435, 1984. [COXJII], [COXJOO], [COXJUU] Montelione, G.T., Arnold, E., Meinwald, Y.C., Stimson, E.R., Denton, J.B., Huang, S.-G., Clardy, J., Scheraga, H.A.: J. Am. Chem. Soc. 106:7946, 1984. [CUBTOI] Prasad, B.V.V., Balaram, H., Balaram, P.: Int. J. Pept. Protein Res. 24:135, 1984. [CUCNUJ], [CUCPAR], [CUCPEV] Aubry, A., Vitoux, B., Marraud, M.: J. Chim. Phys. Phys. Chim. Biol. 81:179, 1984. [HIMLBU] Pletnev, V.Z., Galitskii, N.M., Turkin, N.G., Tarnopolskii, B.L., Evstratov, A.V.: Bioorg. Khim. 3:869, 1977. [IBALPR] Aubry, A., Protas, J., Boussard, G., Marraud, M.: Acta Crystallogr. (Sect. B) 36:2825, 1980. [MAJUSB] Marner, F.-J., Moore, R.E., Hiortsu, K., Clardy, J.: J. Org. Chem. 42:2815, 1977. [MAMTAL] Bates, R.B., Hruby, V.J., Kriek, G.R.: Acta Crystallogr. (Sect. B) 35:188, 1979. [NACVAL] Aubry, A., Marraud, M., Protas, J., Neel, J.: C. R. Acad. Sci. (Ser. C) 278:853, 1974. [PALPRI] Aubry, A., Protas, J., Boussard, G., Marraud, M.: Acta Crystallogr. (Sect. B) 35:694, 1979. [PGTAPA] (1) Stensland, B., Castensson, S.: Eur. Cryst. Meet. 5:16, 1979. (2) Norstam, R., Stensland, B., Castensson, S.: Acta Chem. Scan. [A] 38:473, 1984. (3) Stensland, B., Castensson, S.: J. Mol.Biol. 161:257, 1982. [SARSAR] Eckle, E., Stezowski, J.J., Hughes, R.E.: Acta Crystallogr. (Sect. A) 34:S74, 1978. [TRHTRT] Kamiya, K., Takamoto, M., Wada, Y., Fugino, M., Nishikawa, M.: J. Chem. Soc. Chem. Commun. 438, 1980.
30 [1BP2] Phospholipase A_2 Dijkstra, B.W., Kalk, K.H., Hol, W.G.J., Drenth, J.: J. Mol. Biol. 147:97, 1981. [1CCR] Cytochrome C Ochi, H., Hata, Y., Tanaka, N., Kakudo, M., Sakurai, T., Aihara, S., Morita, Y.: J. Mol. Biol. 166:407, 1983.

[1CRN] Crambin Teeter, M.M. Proc. Natl. Acad. Sci. USA 81:6014, 1984. [1ECA] Hemoglobin Steigemann, W., Weber, E.: J. Mol. Biol. 127:309, 1979. [1FB4] Immunoglobulin Fab Marquart, M., Deisenhofer, J.: Immunol. Today 3:160, 1982. [1INS] Insulin Bordas, J., Dodson, G.G., Grewe, H., Koch, M.H.J., Krebs, B., Randall, J.: Proc. R. Soc. Lond. [Biol.] 219:21,1983. [1LZ1] Lysozyme Blake, C.C.F., Pulford, W.C.A., Artymiuk, P.J.: J. Mol. Biol. 167:693, 1983. [1MBD] Myoglobin Phillips, S.E.V., Schoenborn, B.P.: Nature 292:81, 1981. [1PPT] Avian Pancreatic Polypeptide Blundell, T.L., Pitts, J.E., Tickle, I.J., Wood, S.P., Wu, C.-W.: Proc. Natl. Acad. Sci. USA 78:4175, 1981. [1RN3] Ribonuclease A Borkakoti, N., Moss, D.S., Palmer, R.A. Acta Crystallogr. (Sect. B) 38:2210, 1982. [1SN3] Scorpion Neurotoxin Almassy, R.J., Fontecilla-Camps, J.C., Suddath, F.L., Bugg, C.E.: J. Mol. Biol. 170:497, 1983. [2ACT] Actinidin (sulfhydryl proteinase) Baker, E.N., Dodson, E.J. Acta Crystallogr. (Sect. A) 36:559, 1980. [2APP] Acid Proteinase (Penicillopepsin) James, M.N.G., Sielecki, A.R.: J. Mol. Biol. 163:299, 1983. [2CPV] Calcium-binding parvalbumin B Moews, P.C., Krestsinger, R.H.: J. Mol. Biol. 91:201, 1975. [2CYP] Cytochrome C peroxidase Finzel, B.C., Poulos, T.L., Kraut, J.: J. Biol. Chem. 259:13027, 1984. [2GCH] γ-Chymotrypsin A Cohen, G.H., Silverton, E.W., Davies, D.R.: J. Mol. Biol. 148:449, 1981. [2PAB] Prealbumin (human plasma) Blake, C.C.F., Geisow, M.J., Oatley, S.J., Rerat, B., Rerat, C.: J. Mol. Biol. 121:339, 1978. [2PCY] Apoplastocyanin Garrett, T.P.J., Clingeleffer, D.J., Guss, J.M., Rogers, S.J., Freeman, H.C.: J. Biol. Chem. 259:2822, 1984. [2SGA] Proteinase A (*Streptomyces griseus*) James, M.N.G., Sielecki, A.R., Brayer, G.D., Delbaere, L.T.J., Bauer, C.A.: J. Mol. Biol. 144:43, 1980. [2SNS] Staphylococcal nuclease Cotton, F.A., Hazen, E.E., Jr., Legg, M.J.: Proc. Natl. Acad. Sci. USA76:2551, 1979. [2TGT] Trypsinogen Walter, J., Steigemann, S., Singh, T.P., Bartunik, H., Bode, W., Huber, R.: Acta Crystallogr. (Sect. B) 38:1462, 1982. [3DFR] Dihydrofolate reductase Bolin, J.T., Filman, D.J., Matthews, D.A., Hamlin, R.C., Kraut, J.: J. Biol. Chem. 257:13650, 1982. [3RP2] Rat mast cell protease II Reynolds, R.A., Remington, S.J., Weaver, L.H., Fisher, R.G., Anderson, W.F., Ammon, H.L., Matthews, B.W.: Acta Crystallogr. (Sect. B) 41:139, 1985. [3RXN] Rubredoxin Adman, E.T., Jensen, L.H.: Am. Cryst. Assoc. (Abstr. Papers) 6:65, 1979. [3SGB] Proteinase B complex with third domain of the ovomucoid inhibitor Read, R.J., Fujinaga, M., Sielecki, A.R., James, M.N.G.: Biochemistry 22:4420, 1983. [3TLN] Thermolysin Holmes, M.A., Matthews, B.W.: J. Mol. Biol. 160:623, 1982. [4FXN] Flavodoxin Smith, W.W., Burnett, R.M., Darling, G.D., Ludwig, M.L.: J. Mol. Biol. 117:195, 1977. [5CPA] Carboxypeptidase A Rees, D.C., Lewis, M., Lipscomb, W.N.: J. Mol. Biol. 168:367, 1983. [5PTI] Trypsin inhibitor Wlodawer, A., Walter, J., Huber, R., Sjolin, L.: J. Mol. Biol. 180:307, 1984.