(16) Appears in the microfilm edition of this volume of the journal.
(17) P. K. Ponnuswamy and V. Sasiskheran, *Int. J. Protein Res.*, **2**, 37 (1970).
(18) A. W. Burgess and S. J. Leach, *Biopolymers*, **12**, 2599 (1973).
(19) S. S. Zimmerman, L. L. Shipman, and H. A. Scheraga, *J. Phys. Chem.*, **81**, 614 (1977).
(20) G. D. Smith, W. L. Duax, D. A. Langs, G. T. Detitta, J. W. Edmonds, D. C. Rohrer, and C. M. Weeks, *J. Am. Chem. Soc.*, **97**, 7242 (1975).
(21) P. E. Hansen, J. Feeney, and G. C. K. Roberts, *J. Magn. Reson.*, **17**, 249 (1975).
(22) J. Pospisek and K. Blaha, *Collect. Czech. Chem. Commun.*, **42**, 1069 (1977).

The Effect of Refinements in Energetic Statistical Weighting on the Computed Chain Dimensions of Homopolypeptides

Yvonne Paterson and Sydney J. Leach*

Russell Grimwade School of Biochemistry, University of Melbourne, Parkville, Victoria, 3052, Australia. *Received July 11, 1977*

ABSTRACT: In order to establish more rigorously the effect of side-chain length and branching on the theoretically derived unperturbed dimensions of homopolypeptides, the most recently revised covalent geometry has been used to calculate the characteristic ratios of polymers composed solely of glycine, alanine, β-methylalanine, valine, or β-methylvaline. All atoms in the side chains were included in the calculations and different methods of weighting for multiple side-chain conformations and of incorporating side-chain flexibility were compared. The effects of weighting with the conformational space of the backbone at 30 and 10° intervals of ϕ and ψ were compared. The characteristic ratios of alanine, β-methylalanine, and valine using the most rigorous of the methods explored were found to be very similar (8.06, 7.02, and 9.21). The side chain of β-methylvaline imposed profound steric restrictions on the backbone which resulted in a slowly converging characteristic ratio (94.8 at a chain length of 500 residues). Other than those on residues with sterically restricting side chains, these results are consistent with Flory's hypothesis that the side-chain atoms further removed from the backbone than the C^{β} atom need not be considered when calculating the unperturbed dimensions of most homopolypeptides.

Since Flory established the theoretical methods for calculating the dimensions of statistically coiling noncooperative polypeptide chains[1] most workers have accepted the approximations inherent in the methods. This study is intended to establish the conditions under which these approximations are valid and when more refined techniques are justified. To this end we have used the recently revised peptide geometry of Momany et al.[2] to compare the effect of side-chain length and branching on the unperturbed dimensions of homopolypeptides using the residues of glycine, alanine, β-methylalanine,[3] valine, and β-methylvaline[3] as examples. Side-chain atoms including hydrogen atoms were individually considered when constructing the backbone conformational energy surfaces used to weight the unperturbed dimensions (the conformational space available to the residues is reported elsewhere[4]). The effect of weighting with these surfaces at 10 and 30° intervals in ϕ and ψ was also compared. Methods of weighting with multiple side-chain conformations were explored and the effect of improving the treatment of flexibility in side-chain torsion angles by including minimized energy values in the weighting procedure is examined. The results obtained in this study are compared with those obtained in earlier work using less refined techniques.

Theoretical Methods

Generation of Random Polypeptide Chains. The unperturbed dimensions of homopolypeptides were each calculated as the mean-square end-to-end distance $\langle r^2\rangle_0$ and expressed as the characteristic ratio (C_∞) as the chain length (n) tends to infinity, by the method of Brant and Flory:[5]

$$C_\infty = (\langle r^2\rangle_0/nl^2)_\infty = [(\mathbf{E} + \langle\mathbf{T}\rangle)\,(\mathbf{E} - \langle\mathbf{T}\rangle)^{-1} - (2/n)(\langle\mathbf{T}\rangle)\,(\mathbf{E} - \langle\mathbf{T}\rangle^{n})(\mathbf{E} - \langle\mathbf{T}\rangle)^{-2}]_{1,1} \quad (1)$$

where **E** is the identity matrix of order 3 and $\langle\mathbf{T}\rangle$ is given by:

$$\langle\mathbf{T}\rangle = Z^{-1} \sum_{\phi} \sum_{\psi} \mathbf{T}(\phi,\psi) \exp(-E(\phi,\psi)/RT) \quad (2)$$

where Z is the partition function for the conformational space available to the peptide residue:

$$Z = \sum_{\phi} \sum_{\psi} \exp(-E(\phi,\psi)/RT) \quad (3)$$

This expression for Z does not take into account side-chain rotational isomeric states.

$\mathbf{T}(\phi,\psi)$ is the matrix that transforms the coordinate system of the $(i + 1)$th residue to that of the ith residue.

The angles that define the peptide geometry of these transformations[5] were calculated from the geometric parameters of Momany et al.[2] and have the values $\xi = 14.9°$, $\theta = 70.7°$, and $\eta = 20.9°$. $E(\phi,\psi)$ is the conformational energy of the residue for each backbone conformation defined by the (ϕ,ψ) pair, and R is the gas constant. T, the temperature, was taken to be 293 K in this study.

Conformational Energy Calculations. The conformational energy space was calculated for the N-acetyl-N'-methylamides for each residue in this study using ECEPP[6] (Empirical Conformational Energy Program for Peptides) with a Cyber 73 computer. This program uses the empirical potential energy functions and energy parameters of Momany et al.[2] and calculates the electrostatic, nonbonded, and torsional energy components of the total conformational energy from interatomic interactions. Interactions involving side-chain atoms, including hydrogen atoms, are calculated explicitly.

The intraresidue hydrogen-bonded energy is included in these calculations (as part of the nonbonded energy compo-

0024-9297/78/2211-0415$01.00/0

nent) although the results will be compared with experimental measurements performed in randomizing solvents which disrupt hydrogen bonds. At the theta point the volume excluded by the molecule vanishes because long-range interactions between remote segments of the molecule are compensated by the mutual attractions between neighboring chain units.[1] This occurs in poor solvents; in good solvents polymer–solvent interactions are preferred over local polymer–polymer contacts. "Randomizing" solvents for poly(amino acids) break up long-range and medium-range interactions including intramolecular hydrogen bonds but at the theta point polymer–solvent interactions should be negligible and intraresidue hydrogen bonds allowed.

The conformational space of the backbone of the residues was represented by 10° increments in ϕ and ψ over their entire range. However, for side-chain bond rotations a three-state rotational isomeric model was used and only those values of the side-chain torsion angles that "stagger" the positions of substituents were considered. These are 180 ± 30°, 60 ± 30°, and $-60 \pm 30°$ and the side-chain conformations they define are designated trans, gauche$^+$, and gauche$^-$, respectively. In the case of bonds involving symmetric substituents, e.g., terminal methyl groups, only one isomeric state exists and these values of χ are equivalent; thus $\chi = 60°$ was used. The energy of the molecule for each backbone conformation was first calculated with the side-chain angles fixed at ±60 or 180°. From the distribution of conformations this produced, backbone conformations with energies within 6 kcal/mol of the global minimum were recalculated, allowing the side-chain angles torsional freedom around these positions in order to minimize the overall conformational energy. These energy minimizations were performed using the ECEPP program in conjunction with a function minimizing subroutine written by Powell.[7]

The description of the conformational energy surface and the geometry of these molecules are reported in detail elsewhere.[4]

Energetic Statistical Weighting of C_∞. Evaluations of the partition function Z and the average transformation matrix $\langle \mathbf{T} \rangle$ were carried out at equal intervals of ϕ and ψ throughout their ranges. Two intervals were tested, namely 30 and 10° for each residue and two sets of (ϕ,ψ) conformational energies, one in which χ^1 (the first side-chain torsion angle) was held fixed at the torsional minima and one in which it was allowed to vary to adopt the value that minimized the overall conformational energy.

For those molecules with three side-chain conformations determined by χ^1, $\langle \mathbf{T} \rangle$ and C_∞ were calculated for each rotamer and for the overall molecule. Two methods were used to weight $\langle \mathbf{T} \rangle$ to take into account the contributions from all three side-chain conformations:

(a) The most probable side-chain conformation for the residue at each point on the (ϕ,ψ) plane is that which gives the molecule the lowest total conformational energy. Thus, at each point on the (ϕ,ψ) surface $\langle \mathbf{T} \rangle$ was weighted using the Boltzmann factor for the lowest of the three energy values of the three rotameric states of the side chain. This will be designated as composite weighting, since conformational energy maps constructed by this procedure have been referred to as composite maps.[4,8] This method of weighting neglects entropic considerations of the side chains, since each (ϕ,ψ) pair is counted only once whereas there are values of ϕ and ψ for which more than one side-chain rotamer gives energetically allowed backbone conformations. These backbone conformations are statistically more probable than those (ϕ,ψ) pairs where only one rotamer is allowed. To overcome this problem we also used a method of statistically weighting $\mathbf{T}(\phi,\psi)$ as follows:

(b) This approach considers that the poly(amino acid) is a random copolymer of the residue with the side chain in three rotational isomeric states, defined by the first side-chain torsion angle χ^1. The average transformation matrix for a copolymer is given by[1]

$$\langle \mathbf{T} \rangle = \sum_{k=1}^{n} P(\chi_k{}^1) \langle \mathbf{T} \rangle_k \tag{4}$$

where $P(\chi_k{}^1)$ is the probability of a residue in the kth rotameric state occurring in the polymer and $\langle \mathbf{T} \rangle_k$ is the transformation matrix of the residue in that state.

The probability of the occurrence of each rotameric state, for the first side-chain torsion angle χ^1, is dependent on the statistical weight $W(\chi_k{}^1)$ of each rotamer, k. Thus:

$$P(\chi_k{}^1) = W(\chi_k{}^1)/Z \tag{5}$$

where

$$W(\chi_k{}^1) = \sum_{\phi} \sum_{\psi} \exp[-E(\phi,\psi,\chi_k{}^1)/RT] \tag{6}$$

Equation 4 is equivalent to statistically weighting the average transformation matrix over all side-chain rotational isomeric states for each (ϕ,ψ) pair using the Boltzmann factors for all three rotamers:

$$\langle \mathbf{T} \rangle = Z^{-1} \sum_{\phi} \sum_{\psi} \mathbf{T}(\phi,\psi) \sum_{k=1}^{3} \exp[-E(\phi,\psi,\chi_k{}^1)/RT] \tag{7}$$

where $\chi_k{}^1$ refers to the kth rotameric state (i.e., trans, gauche$^+$, or gauche$^-$) of the first side-chain torsion angle, χ^1, and Z in this case included the Boltzmann factors for all rotational isomeric states

$$Z = \sum_{i=1}^{36} \sum_{j=1}^{36} \sum_{k=1}^{3} \exp[-E(\phi_i,\psi_j,\chi_k{}^1)/RT] \tag{8}$$

Results

The characteristic ratio obtained in this study for a freely rotating peptide chain with no energetic weighting is 1.89. The characteristic ratios, C_∞, for the homopolymers composed of glycine, alanine, β-methylalanine, and valine are listed in Table I. The characteristic ratios of β-methylvaline for various values of n are given in Table II. In both tables values derived from minimized and nonminimized energy surfaces at 10 and 30° intervals are shown. For those residues with nonsymmetric side chains either the type of weighting used where all three rotamers are being considered or the individual rotameric state of the side-chain is indicated under the subheading "side chain".

Discussion

Energy Parameters and Covalent Geometry. The effect of using more "refined" energy parameters and new covalent geometry on the unperturbed dimensions of polypeptide chains is best demonstrated by glycine where side-chain considerations do not apply. The values for C_∞ obtained in this study (2.12 and 2.13) are slightly lower than those obtained by earlier workers (2.16[9] and 2.17[10]). Since this decrease is also reflected in the characteristic ratio of the freely rotating chain (we obtained a value of 1.89 compared with the earlier value of 1.93[1]) the new geometry, rather than the refined energy parameters, has the greater effect on the flexibility of the polymer chain.

Choice of Integration Interval. The earliest calculations on homopolypeptides by Flory's group[1,5] used discrete values of ϕ and ψ at 30° intervals to evaluate the characteristic ratios of the polymers, since a larger increment of 60° yielded comparable results.[5] Since then most workers have chosen this interval in their calculations and where smaller increments of say 20°[9] or 10°[8,11] have been used, no justification for this refinement has been given. Our calculations show that for the

Table I
C_∞ for Polyaminoacids with Different Energetic Weightings

Residue	Side chain	With energy minimization 10° grid	30° grid	Without energy minimization 10° grid	30° grid
Glycine				2.12	2.13
Alanine		8.06	8.44	8.30	8.77
β-Methylalanine	Gauche$^+$	31.29	34.64	55.99	74.39
	Gauche$^-$	8.31	8.58	9.51	9.68
	Trans	7.15	6.63	6.15	5.60
	Composite	6.80	6.68	6.12	5.82
	Statistical	7.02	6.82	6.48	6.25
Valine	Gauche$^+$	90.78	313.52	228.41	311.04
	Gauche$^-$	27.24	14.79	33.24	24.14
	Trans	10.51	11.61	12.66	13.38
	Composite	9.27	9.65	11.25	12.23
	Statistical	9.21	9.58	11.05	12.01

Table II
Characteristic Ratios for β-Methylvaline with Increasing Chain Lengths

No. of amino acid residues in the chain	With energy minimization 10° grid	30° grid	Without energy minimization 10° grid	30° grid
100	52.41	36.01	72.98	64.99
500	94.82	49.04	249.70	170.09
1000	102.22	50.76	344.99	203.03
1500	104.69	51.33	388.57	214.77
∞	109.63	52.48	483.83	238.32

majority of amino acid residues with characteristic ratios of 8 ± 2, the 30 and 10° grids of the (ϕ,ψ) surface give results that fall within 10% of each other. The exceptions to this result appear to be for residues with very restricted (ϕ,ψ) surfaces where the regions that most heavily weight the characteristic ratio are in one small area of the map. Thus the C_∞ of the gauche$^+$ rotamer for valine is greatly reduced by the choice of a 10° over a 30° grid whereas both the gauche$^-$ rotamer for this residue (see Table I) and the β-methylvaline residue (see Table II) show very significant increases in their C_∞ values when a 10° grid is used. It seems that in the majority of cases use of a 30° interval is adequate except for residues with characteristic ratios that depart from the "alanine-like" value of 8–9.

Energy Minimization. Energy minimization of the side-chain conformations of the homopolymers influence the mean-square end-to-end distance in a similar way to the choice of integration interval. Thus, for those residues with "alanine-like" chain dimensions the extra side-chain freedom affects the C_∞ values by less than 10% but where extended chain dimensions are involved, e.g., for β-methylvaline and for the gauche rotamers of valine, the effect is to dramatically decrease the characteristic ratio. Other factors involved in the calculations can compound the error arising from the use of nonminimized energy values. For example, the C_∞ for the gauche$^+$ rotamer of β-methylvaline weighted with different choices of integration interval show only a 10% difference when minimized energy surfaces are used, whereas the difference is nearly 30% where minimized energy values are not used.

The effect of using minimized energy values on all residues except β-methylalanine is to decrease the dimensions to some extent. For β-methylalanine, however, there is a slight increase in the C_∞ of the trans rotamer; since this rotamer is strongly preferred,[4] there is a net increase in the chain dimensions of the molecule weighted for all three rotamers.

Treatment of Side Chains. Experimental measurements on homopoly(amino acids) have shown that polymers composed of different residues have similar chain dimensions in various solvents[12–14] which, within the greatest quoted[13] experimental error of 16%, are remarkably similar to the theoretically predicted dimensions for alanine.[5] Thus, it was postulated[5] that substituents along the side chain beyond the C^β atom exert little influence on the conformational space available to the backbone and bond rotations beyond the C^α–C^β bond need not be taken into account in these calculations.

A later study[8] compared a "structureless" side-chain approach where the side-chain atoms were represented by enlarged C^β atoms, with an approach where the C^γ atoms, but not their adjoined atoms, were included in the calculations. This study demonstrated a significant reduction (nearly 20%) in the unperturbed dimensions of chain molecules with branched side chains when the C^γ and its hydrogen atoms were explicitly taken into account. When weighting with the energy maps of the rotamers, however, these workers[8] used a composite rather than a statistical method in which they neglected the fact that there are (ϕ,ψ) values for which more than one side-chain rotamer is equally probable. Nor did they allow flexibility about the "staggered" χ^1 values by allowing these values to vary so as to minimize the conformational energy. Thus, even using a semistructured representation, the characteristic ratios of residues with branched side chains are significantly different (27%) from those of alanine.[8]

Our results show that a composite weighting with only one energy value for each (ϕ,ψ) pair is not as inaccurate as one might expect (compare the C_∞ obtained with composite vs. statistical weightings for β-methylalanine and valine in Table I). The results obtained using either method are very close and fall well within errors involved in estimating the unperturbed dimensions experimentally. In our calculations all of the atoms in the residue were explicitly defined, and the effect of this, plus the use of the most recent[2] covalent peptide geometry and energy parameters, is to increase slightly the characteristic ratios of alanine and valine in this study (8.3 and 11.25, respectively) compared to those of Miller and Goebel[8] (8.0 and 10.7, respectively). However, by allowing the side chains more freedom by including minimization in our calculations, the chain dimensions for both β-methylalanine (7.02) and valine (9.21) fall within 14% of those for alanine, in accord with Brant and Flory's hypothesis.[12] The fact that they are not identical with that of alanine may imply that a three-state approximation does not adequately describe the rotational degrees of freedom of the side-chain bonds.

Of the residues with side chains that were chosen for this study, the only exceptional characteristic ratio was that of β-methylvaline. The presence of a *tert*-butyl substituent on the C^β atom has a profound steric effect on the backbone. Confining monomeric units in a polymer chain to any single narrowly defined region of its conformational space will inevitably lead to a large characteristic ratio.[15] In the present instance the effect is accentuated because only very extended conformations are available to this molecule on the (ϕ,ψ) energy surface.[4] The only other homopolypeptide that has been reported to have such extended chain dimensions is that of proline.[15,16] Since β-methylvaline would not be expected to undergo cis–trans isomerism across the peptide bond, a polymer composed of this residue would provide a simpler model for the hydrodynamic behavior of a randomly coiled rigid chain than does polyproline.

The severe restrictions on the number of backbone conformations available to β-methylvaline may render the random coil model inappropriate to a polymer of these residues. It is possible that poly-β-methylvaline is unable to adopt a random coil in any solvent and may exist only as a broken rod, an extended helix, or some other extended form which is not determined by the energy maps for a single residue. It should be possible to establish this by deriving experimentally the dimensions of this polymer based on a random coil model and comparing these values with the theoretical results obtained in this study.

Acknowledgments. We are indebted to Dr. A. W. Burgess for many helpful criticisms of this manuscript. We would also like to thank Professor H. A. Scheraga for making the program ECEPP available to us and to Dr. G. Némethy for many useful suggestions. This work was supported by a grant from the Australian Research Grants Committee to S.J.L. Y.P. is a recipient of a Melbourne University Postgraduate Research award.

Supplementary Material Available: Tables of the elements of the average transformation matrices $\langle \mathbf{T} \rangle$ from which the characteristic ratios shown in Tables I and II are derived (Tables III–VI) (4 pages). Ordering information can be found on any current masthead page.

References and Notes

(1) P. J. Flory, "Statistical Mechanics of Chain Molecules", Interscience, New York, N.Y., 1969.

(2) F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.*, **79**, 2361 (1975).

(3) β-Methylalanine is the trivial name of 2-aminobutanoic acid and β-methylvaline is the trivial name of 2-amino-3-dimethylbutanoic acid. All amino acids and their residues referred to in this paper are in the L configuration.

(4) Y. Paterson and S. J. Leach, *Macromolecules*, preceding paper in this issue.

(5) D. A. Brant and P. J. Flory, *J. Am. Chem. Soc.*, **87**, 2791 (1965).

(6) ECEPP was developed in the Department of Chemistry, Cornell University, and is available from the Quantum Chemistry Program Exchange, Chemistry Department, Room 204, Indiana University, Bloomington, Ind. 47401, as program number QCPE 286.

(7) M. J. D. Powell, *Comput. J.*, **7**, 155 (1964).

(8) W. G. Miller and C. V. Goebel, *Biochemistry*, **7**, 3925 (1968).

(9) D. A. Brant, W. G. Miller, and P. J. Flory, *J. Mol. Biol.*, **23**, 47 (1967).

(10) S. Tanaka and A. Nakajima, *Polym. J.* **2**, 717 (1971).

(11) A. W. Burgess, Y. Paterson, and S. J. Leach, *J. Polym. Sci., Polym. Symp.*, **No. 49**, 75 (1975).

(12) D. A. Brant and P. J. Flory, *J. Am. Chem. Soc.*, **87**, 2788 (1965).

(13) H. Fujita, A. Teramato, T. Yamashita, K. Okita, and S. Ikada, *Biopolymers*, **4**, 781 (1966).

(14) M. Terbojevitch, E. Peggion, A. Cosani, G. D'Este, and E. Scoffone, *Eur. Polym. J.*, **3**, 681 (1967).

(15) P. R. Schimmel and P. J. Flory, *Proc. Natl. Acad. Sci. U.S.A.*, **58**, 52 (1967).

(16) S. Tanaka and H. A. Scheraga, *Macromolecules*, **8**, 623 (1975).

The Random Coil Dimensions of Sequential Copolypeptides Containing N^5-(2-Hydroxyethyl)-L-glutamine. 1. Theoretical Studies

Yvonne Paterson and Sydney J. Leach*

Russell Grimwade School of Biochemistry, University of Melbourne,
Parkville, Victoria, 3052, Australia.

Received August 22, 1977

ABSTRACT: The statistically most probable backbone and side-chain conformations for the N-acetyl-N'-methylamide derivative of N^5-(2-hydroxyethyl)-glutamine (HEG) have been determined using a 12-state model for each of the backbone dihedral angles ϕ and ψ and a 3- or 6-state model for each of the side-chain torsion angles. The most probable backbone conformation is an extended one and the most probable side-chain conformation is one which facilitates the formation of a hydrogen bond between the side-chain hydroxyl group and the backbone carbonyl oxygen atom of the N-acetyl group. The characteristic ratios for the random coil forms of the homopolypeptide, $(\mathrm{HEG})_n$, and its copolypeptides $(\mathrm{HEG\text{-}X\text{-}X})_n$, where X = gly, ala, or β-methylala, have been derived using this conformational energy surface and are respectively 9.75, 2.27, 8.56, and 7.73.

The identification of conformations of minimum energy of single amino acid residues is finding increasing use in studies on the tertiary structure of peptides.[1-3] The validity of this type of calculation may be assessed by comparing the predicted structure with that derived experimentally, either by x-ray crystallographic[2] or solution[4,5] studies. For polypeptides, interactions between residues which are nearest neighbors are dominant in the computations. This simplification may invalidate the comparisons so that experimental observations fail to agree with the theoretically predicted structure of minimum energy. The theoretically predicted total conformational space available to a residue can also be tested for validity by comparing the characteristic ratio of a randomly coiled homopolypeptide (derived theoretically from the conformational energy surface of the monomer unit over the (ϕ,ψ) plane) with that deduced experimentally from the hydrodynamic behavior of the polymer. Excluding the case of amino acid residues containing a pyrollidine ring, there are few exceptions[6,7] to the close agreement between the theoretically predicted[8-10] and experimentally observed[11-13] unperturbed dimensions of polyamino acids. In such cases intraresidue interactions alone are sufficient to account for the behavior of randomly coiled unperturbed homopolypeptides. This is not the case with polymers containing the proline and hydroxyproline residues where some experimental estimates of the characteristic ratio[14,15] have been markedly lower than

0024-9297/78/2211-0418$01.00/0