TORSIONAL POTENTIALS FROM OBSERVED DIPEPTIDE CONFORMATIONS IN PROTEIN STRUCTURE DATA

A. S. KOLASKAR AND D. PRASHANTH*
Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, *India*

ABSTRACT

The (ϕ, ψ)-values obtained from the crystal structure data of twentytwo globular proteins have been used to obtain the less understood torsional potential functions $V_t(\phi)$ and $V_t(\psi)$, for rotations around the single bonds N—C^α and C^α—C respectively, in polypeptides. The method developed to obtain these functions, from an analysis of the crystal structure data and energy calculations on dipeptide units having different side chains, is discussed very briefly. The newly obtained torsional potential functions are: $V_t(\phi) = -1{\cdot}0 \cos(\phi + 60°)$ and $V_t(\psi) = -0{\cdot}5 \cos(\psi + 60°) - 1{\cdot}0 \cos(2\psi + 30°) - 0{\cdot}5 \cos(3\psi + 30°)$. The ($\phi$, ψ)-Ramachandran energy map obtained for an 'averaged' dipeptide unit, using these functions, is found to be in very good agreement with the energy map obtained from the experimental protein data.

INTRODUCTION

THE prediction of preferred conformations of polypeptide chains, obtained by using semiempirical potential energy functions, depend on the accuracy of these functions. Ramachandran[1] has indicated that the potential functions, which are commonly used to predict the conformation of biopolymers, must be tested using data obtained from various physicochemical methods on related small molecules and macromolecules. In this connection, from the studies on the packing of molecules in the crystalline state, it was shown that the nonbonded potential energy functions used in this laboratory, as well as the functions used by various other groups for the study of biopolymer conformation, give data which are in good agreement with experimental results[2, 3]. However, studies carried out in our laboratory on model compounds using semi-empirical quantum chemical methods have indicated that the torsional potential function $V_t(\phi)$, for rotations ϕ around the single bond N—C^α, is small[4], while the torsional potential function $V_t(\psi)$, for rotation around the C^α—C single bond is mainly two-fold in nature, with a large barrier of the order of 4·0 kcal/mole[5]. From these studies the function $V_t(\psi)$ was suggested to be of the form

$$V_t(\psi) = \tfrac{1}{2} V_{\psi o} (1 - \cos 2\psi)$$

with

$$V_{\psi o} = 4{\cdot}0 \text{ kcal/mole.} \quad (1)$$

This is different from the previously accepted functions both in the form of $V_t(\psi)$ and the value of its barrier. Thus there is a need for a further investigation in order to obtain the correct nature of these less understood potential energy functions, $V_t(\phi)$, and $V_t(\psi)$, for peptide units. Therefore, we have analysed the observed distribution of the quantities (ϕ, ψ) in the crystal structure data of twentytwo globular proteins and derived from them the function $V_t(\phi)$ and $V_t(\psi)$. The method developed for this purpose is discussed briefly in the next section, and the results are outlined in the succeeding section.

METHOD

In the semi-empirical conformational energy calculations on polypeptides and proteins, the total potential energy is usually partitioned as

$$V = V_{tot} = V_{nb} + V_{es} + V_t(\phi) + V_t(\psi) + V_{hb} + V_{ss} \quad (2)$$

where, V_{nb} is the nonbonded interaction energy, V_{es} the electrostatic interaction energy, $V_t(\phi)$ and $V_t(\psi)$ are torsional potential energy functions, V_{hb} is the hydrogen bond energy and V_{ss} is the solute-solvent interaction energy. Out of these, the last term is least understood and very often it is neglected. We shall also do so. Similarly, if only short range interactions are considered, which form the dominant factor determining the conformation in globular proteins[6], then V_{hb} also can be neglected. Thus, the total potential energy can be written in the form

$$V(\phi, \psi) = V_1(\phi, \psi) + V_2(\phi, \psi) \quad (3)$$

where

$$V_1 = V_{nb} + V_{es} \quad (4)$$

and this term is dependent on the nonbonded interatomic distances r_{ij}, and

$$V_2(\phi, \psi) = V_{tor}(\phi, \psi) = V_t(\phi) + V_t(\psi) \quad (5)$$

and this term is dependent only on the values of the dihedral angles (ϕ, ψ). Of these, the total potential energy in equation (3), namely $V(\phi, \psi)$, can be determined for each amino acid residue from the observed distribution $P(\phi, \psi)$, in the (ϕ, ψ)-plane, as obtained from the (ϕ, ψ)-data in a large number of proteins, using the Boltzmann relation[7, 8]:

$$P(\phi, \psi) \alpha \exp - V(\phi, \psi)/RT. \quad (6)$$

On the other hand, $V_1(\phi, \psi)$ can be calculated, for each amino acid residue, using the known potential energy functions V_{nb} and V_{es} for which the parameters

---

*To whom reprint requests should be addressed.

employed are well established. We have used, for, this purpose, the 6-exp nonbonded interaction potential function with the constants given by Chandrasekaran and Balasubramanian[9], except for those involving hydrogens, for which a softened potential is used, by reducing the van der Waals radius of hydrogen by 0·1 Å[10, 11]. The electrostatic energy V_{es} was calculated using the monopole approximation and an effective dielectric constant[12].

Knowing V and V_t, the effective torsional potential energy, $V_s(\phi, \psi)$, was deduced using eq. (3). This was then analysed as the sum of two components $V_t(\phi)$ and $V_t(\psi)$, as in eq. (4) and each was represented as a sum of Fourier components of the form

$$V_t(\theta) = V_{\theta 1}\cos(\theta - \delta_{\theta 1}) + V_{\theta 2}\cos(2\theta - \delta_{\theta 2}) + V_{\theta 3}\cos(3\theta - \delta_{\theta 3}) \quad (7)$$

in which $V_t(\phi)$ and $V_t(\psi)$ are obtained by substituting for θ, ϕ and ψ respectively.

To start with, $V_t(\phi)$ was put equal to zero, as indicated by Kolaskar *et al.*[4] and the experimental data for $V_{exp}(\psi)$ were obtained by summing the experimental data of $P_{exp}(\phi, \psi)$ over the variable ϕ and thus obtaining $P_{exp}(\psi)$, andthen using a formula similar to eq. (6) to convert $P_{exp}(\psi)$ into $V_{exp}(\psi)$. Similarly, $V_t(\psi)$ values were obtained from theoretical studies on dipeptide units while $V_t(\phi)$ values were got using eq. (3). The Fourier coefficients $V_{\psi 1}$, $V_{\psi 2}$ and $V_{\psi 3}$ as well as the phase angles, $\delta_{\psi 1}$, $\delta_{\psi 2}$ and $\delta_{\psi 3}$, were obtained from these data of $V_t(\psi)$ *versus* ψ by standard procedures. These values were refined by calculating back

$$V_{th}(\psi) = V_t(\psi) + \sum_{k=1}^{8} V_{\psi k}\cos(k\psi - \delta_{\psi k}) \quad (8)$$

and converting these into theoretical probability data *viz.*, $P_{th}(\psi)$. Then, $V_{\psi k}$ and $\delta_{\psi k}$ were refined by finding the best fit between $P_{th}(\psi)$ and $P_{exp}(\psi)$, as judged by the value of the R-factor

$$R = \frac{\Sigma\left|P_{th} - P_{exp}\right|}{\Sigma P_{exp}} \quad (9)$$

Having thus refined the ψ-potential, this was used to obtain $P_{th}(\phi)$ from $V_{th}(\phi, \psi)$ and $V_t(\phi)$, in exactly the same manner as was done for the ψ-potential. Then $V_{\phi k}$ and $\delta_{\phi k}$ ($k = 1, 2, 3$) were determined, first roughtly by Fourier analysis, and then refined by making use of the R-value. This completes one cycle of refinement.

In the same way, three cycles of refinement were carried out, first by summing over ϕ and analysing $V_t(\psi)$, and then by summing over ψ and analysing $V_t(\phi)$, each time. The accuracy attainable for $V_{\phi k}$ and $V_{\psi k}$ was only of the order of 0·5 kcal/mole and for $\delta_{\phi k}$ and $\delta_{\psi k}$ was the order 30°.

Results and Discussion

The functions, $V_t(\phi)$ and $V_t(\psi)$, thus obtained after three cycles of refinement for a dipeptide unit containing an Ala side chain, are

$$V_t(\phi) = -1{\cdot}0\cos(\phi + 60°) - 0{\cdot}5\cos(2\phi + 30°) \quad (10)$$

and

$$V_t(\psi) = -0{\cdot}5\cos(\psi + 60°) - 1{\cdot}0\cos(2\psi + 30°) - 0{\cdot}5\cos(3\psi + 30°). \quad (11)$$

Applying these torsional potential functions, the value of $V_{tot}(\phi, \psi)$ was calculated, without including hydrogen bond energy, for an Ala side chain. For comparison with the observed data of $P_{exp}(\phi, \psi)$, the calculations were made for three values of $\tau(C^{\alpha})$ namely, 105°, 110° and 115°, and the total potential energy was obtained as a "weighted" average of these three. The distribution of $V_{th}(\phi, \psi)$ thus obained is shown in Fig. 1 by continuous lines. The

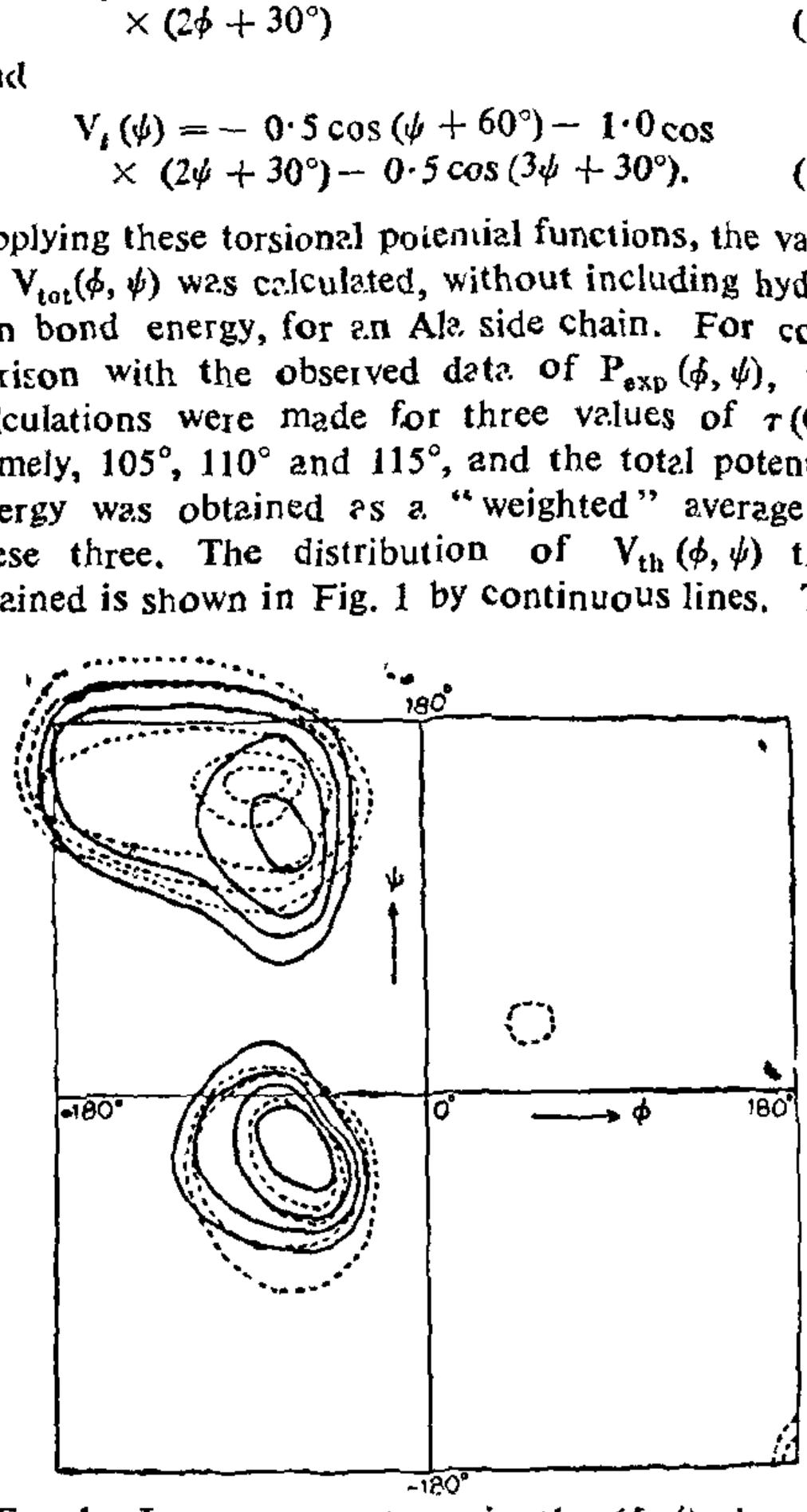

FIG. 1. Isoenergy contours in the (ϕ, ψ)-plane for an alanine dipeptide unit at intervals of 0·5 kcal/mole. The continuous curves are drawn using values of $V_{tot}(\phi, \psi)$ obained from calculations using the functions $V_t(\phi)$ and $V_t(\psi)$, as given by eq. (10) and eq. (11). The dotted curves are drawn by using the $P_{exp}(\phi, \psi)$ data of Ala residues from 22 globular proteins.

isoenergy contours are drawn at intervals of 0·5 kcal/mole, above the *absolute minimum*. In the same figure, the dotted curves correspond to the isoenergy contours which were obtained from the values of $P_{exp}(\phi, \psi)$, calculated for Ala residues, from the distribution in the (ϕ, ψ)-plane observed in the crystal structure data of the globular proteins mentioned above. It must be noted tht we have considered only those Ala residues which were found to occur

outside the α-helical region and have avoided the influence of long range interactions to a very good extent. Hence, this map, obtained from the observed dipeptide conformations in protein structure data, can be directly compared with that obtained from calculations on a dipeptide unit having an Ala side chain. A glance at Fig. 1 would indicate the very good agreement between theory and experimental data. However, we would like to caution that the experimental isoenergy curves are approximate as we had to draw them using only data from about 300 examples.

Similar calculations were carried out on dipeptides having Val, Leu, Ile, Ser, and Phe side chains, and the functions $V_t(\phi)$ and $V_t(\psi)$ were obtained in each case making use of the experimental crystal structure data from globular proteins and the calculated values of $V_1(\phi, \psi)$, taking into account the interactions with the side chains and the torsional potential for the side chains.

These studies have given potential functions having only marginal differences in the values of the parameters for the functions $V_t(\phi)$ and $V_t(\psi)$. Considering all these functions obtained for the different residues, a function which fits best the experimental protein data on these residues was deduced. This is given by

$$V_t(\phi) = -1{\cdot}0\cos(\phi + 60°) \tag{12}$$

and

$$V_t(\psi) = -0{\cdot}5\cos(\psi + 60°) - 1{\cdot}0\cos \times (2\psi + 30°) - 0{\cdot}5\cos(3\psi + 30°) \tag{13}$$

These functions were then used to calculate the total "averaged" potential energy distribution in the (ϕ, ψ)-plane, for the different dipeptide units having side chains Ala, Val, Leu, Ile, Ser and Phe. The potential energy values were properly normalised and multiplied by a statistical weight factor proportional to the number of examples of each type. The potential energy map thus obtained is shown in Fig. 2 (*a*) and its isoenergy contours may be compared with those in Fig. 2 (*b*), which are obtained by using the experimental data $P_{exp}(\phi, \psi)$ for the same six amino acid residues. It can be seen from Fig. 2 that the experimental curves and theoretical "averaged" energy contours agree very well. This amount of agreement has never been obtained earlier between theory and experiment.

It should be mentioned that the potential functions (12) and (13) refer to the usual L-amino acid residues. For a D-residue, the corresponding equations will be similar to (12) and (13), with the same $V_{\phi k}$ and $V_{\psi k}$, but with $\delta_{\phi k}$ and $\delta_{\psi k}$ of the same magnitudes as in (12) and (13), but of opposite signs.

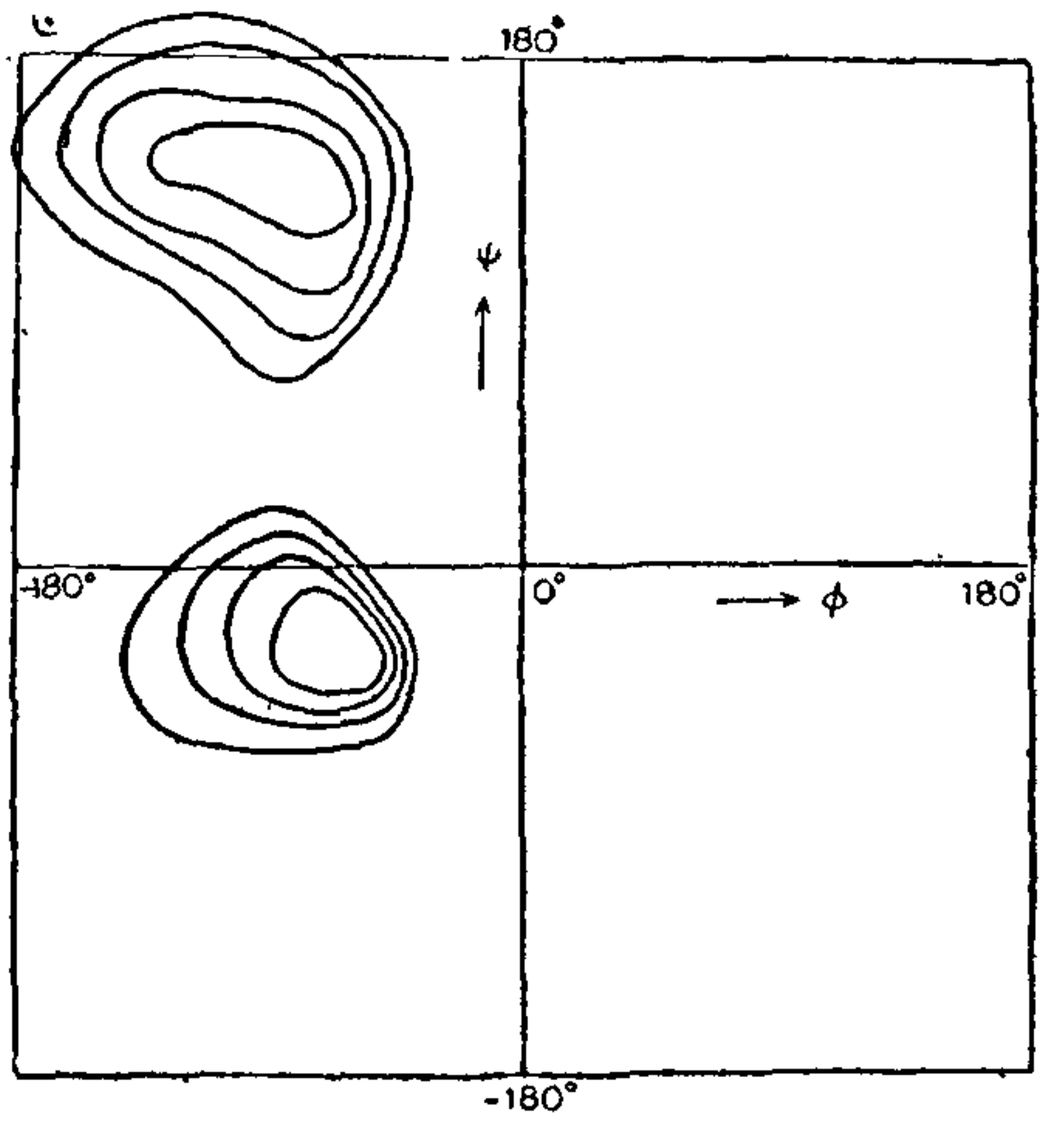

FIG. 2 (*a*). Isoenergy contours of $V_{tot}(\phi, \psi)$ drawn, at intervals of 0·5 kcal/mole, for a dipeptide unit, after properly averaging over different side chains (*see* text). The fnctions $V_t(\phi)$ and $V_t(\psi)$ as given in eq. (12) and eq. (13) were used for the total potential energy calculations.

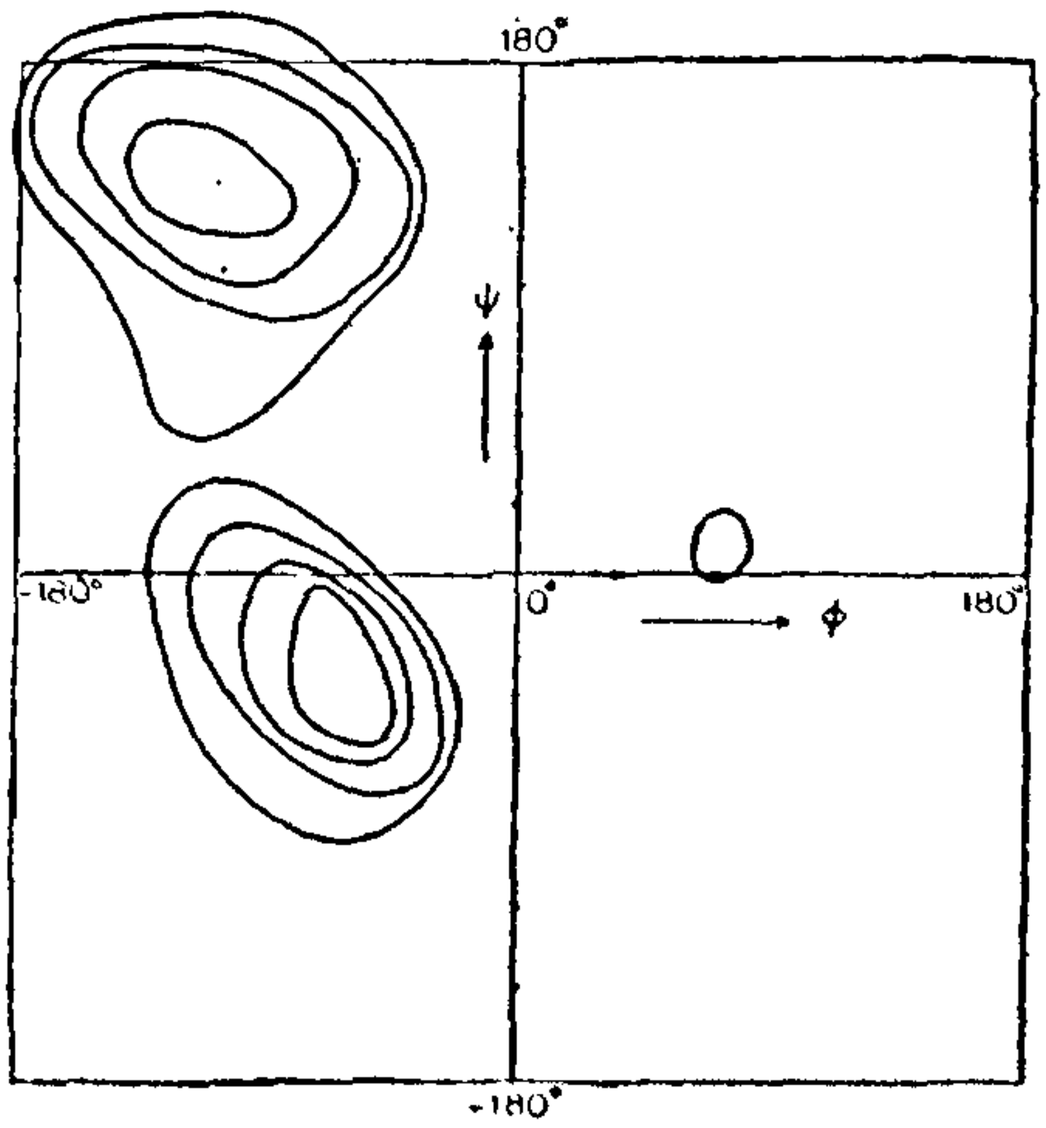

FIG. 2 (*b*). Isoenergy contours, at intervals of 0·5 kcal/mole, drawn using the (ϕ, ψ) data for Ala, Val Leu, Ile, Phe and Ser residues from twentytwo proteins. Note the very good agreement between Fig. 2 (*a*) and Fig. 2 (*b*).

The results presented here clearly indicate that the data on macromolecules, which are accumulating very rapidly, can be used to derive semiempirical potential energy functions, which can be used no

only in the study of macromolecules, but also for small molecules. The details of the results presented here are expected to be published elsewhere.

ACKNOWLEDGEMENT

We thank Prof. G. N. Ramachandran for suggesting this problem.

We wish to acknowledge the financial support from the Department of Science and Technology (India).

1. Ramachandran, G. N., In "Conformation of biological molecules and polymers", (Bergman, E. D. and Pullman, B., Eds.), *Israel Acad. of Sci. and Humanities*, Jarusalem, 1973, pp. 1–11.
2. —, Sarathy, K. P. and Kolaskar, A. S., *Z. fur. Kristall.*, 1973, **138**, 299.
3. Kolaskar, A. S., *Ph.D. Thesis*, Indian Institute of Science, Bangalore 1976.
4. —, Sasisekharan, V. and Sarathy, K. P., *Theoret. Chim. Acta (Berl.)*, 1975, **38**, 109.
5. —, Sarathy, K. P. and Sasisekharan, V., *Curr. Sci.*, 1975, **44**, 35.
6. Chou, P. Y. and Fasman, G. D., *Biochemistry*, 1974, **13**, 211.
7. Flory, P. J., *Statistical Mechanics of Chain Molecules*, Wiley-Interscience, New York, 1969.
8. Ramachandran, G. N., Unpublished, 1977.
9. Chandrasekharan, R. and Balasubramanian, R., *Biochem. Biophys. Acta*, 1969, **188**, 1.
10. Ramachandran, G. N., *Jour. Ind. Inst. Sci.*, 1975, **56**, 375.
11. Hagler, A. T. and Lapiccirella, A., *Biopolymers*, 1976, **15**, 1167.
12. Ramachandran, G. N. and Sasisekharan, V., *Advan. Protein Chem.*, 1968, **23**, 283.

ELECTROPHORESIS OF ^{32}P-LABELLED OLIGONUCLEOTIDES ON THIN-LAYER DEAE-CELLULOSE FOR RAPID SEPARATION IN FINGERPRINTING

P. JAGADEESWARAN AND JOSEPH D. CHERAYIL

Department of Biochemistry, Indian Institute of Science, Bangalore 560 012, *India*

THE two-dimensional fractionation procedure for radioactive oligonucleotides, developed by Sanger *et al.*[1] is now widely used in fingerprinting and sequencing of nucleic acids. In this procedure separation in the first dimension is by electrophoresis on a cellulose acetate membrane strip at high voltage and in the second dimension on a DEAE-cellulose paper. Brownlee and Sanger[2] have employed chromatography on thin-layer DEAE-cellulose for fractionation of large oligonucleotides, especially deoxy-oligonucleotides. Grohmann and Sinsheimer[3] have noted certain advantages in carrying out the first dimension electrophoresis on thin-layer cellulose instead of cellulose acetate. We have devised a simple procedure for carrying out electrophoresis in the second dimension on a thin-layer DEAE-cellulose plate. In this method voltages of the order of 50–60 per cm can be applied without appreciable heating effect and the fractionation completed in a few hours.

EXPERIMENTAL

Mycobacterium smegmatis, used in these studies, was from Microbiology and Cell Biology Laboratory of this Institute. DEAE-cellulose (0·85 meq/g) and cellulose powder (microcrystalline) were purchased from Centron Research Laboratory, Bombay. RNase T_1 and RNase A were from Sigma Chemical Company, St. Louis. Carrier free ^{32}P-orthophosphate was from Bhabha Atomic Research Centre, Bombay. All other reagents were of analytical grade.

Preparation of ^{32}P-labelled oligonucleotides

^{32}P-labelled total RNA was isolated from *Mycobacterium smegmatis* grown with radioactive phosphate in low phosphate medium[4,5]. 5S RNA was separated by polyacrylamide gel-electrophoresis and digested with RNase T_1 or RNase A according to the procedure of Sanger and Brownlee[6].

Preparation of thin-layer DEAE-cellulose plates

DEAE-cellulose and microcrystalline cellulose in the required proportions were thoroughly mixed with water (6 ml per g of mixture) in a Waring Blendor and the slurry was applied to glass plates using a Desaga spreader with thickness setting at 0·5 mm. The plates were dried at 60–80° C.

Determination of base composition

The radioactive spots, as revealed by autoradiography, were scraped from the thin layer plates, eluted with 30% triethylammonium carbonate[2] and digested with 0·2 M NaOH. The mononucleotides were separated by electrophoresis at pH 3·5 on Whatman No. 1 paper using cold nucleotides as markers, and counted in a liquid scintillation counter.