Polymer Vol. 37 No. 13, pp. 2589–2595, 1996
Copyright © 1996 Elsevier Science Ltd
Printed in Great Britain. All rights reserved
0032-3861/96/$15.00+0.00

polymer papers

GEMOX: a simple and efficient strategy for the generation of polymer molecular models compatible with X-ray diffraction data

Juan José Navas, Carlos Alemán* and Sebastián Muñoz-Guerra
Departament d'Enginyeria Química, ETS d'Enginyers Industrials de Barcelona, Universitat Politècnica de Catalunya, Diagonal 647, Barcelona E-08028, Spain
(Received 21 September 1995)

A computational method based on a corrected grid search algorithm has been developed to help in the conformational analysis of crystalline polymers by X-ray diffraction. The program was designed upon the difficulty of taking into account all the molecular arrangements that are initially compatible with the experimental diffraction data of polymers, in particular when these consist of chemical repeating units with a wide accessible conformational space. The strategy presented here has been incorporated into a computer program named GEMOX. First, all the models fitting the experimentally determined axial repeat of the chain and having the helical symmetry known for the polymer are rapidly created by computing their atomic coordinates. Secondly, sterically hindered conformations are discarded using force-field energy criteria. For this, a potential energy function containing both non-bonded and torsional terms has been incorporated in the program. The ability of the method to provide a complete survey of the reasonable models for a given system was tested for three well characterized polymers, all of them displaying a complex conformational behaviour. Copyright © 1996 Elsevier Science Ltd.

(Keywords: computer program; molecular models; X-ray diffraction)

INTRODUCTION

Diffraction studies have played a key role in the three-dimensional structure determination of crystalline polymers of both synthetic[1,2] and natural[3–5] origin. Since macromolecules are unable to render monocrystalline samples of the size required for the application of standard X-ray diffraction methods, fibrous specimens have been widely used in the conformational analysis of polymers[6]. Due to the partially ordered nature of polymer fibres, the amount of data obtained by diffraction is usually limited and the molecular structure cannot be accurately determined. Computational procedures carrying out the refinement of the structure against experimental information have been of great help in removing ambiguities and attaining higher precision in the analysis. Nevertheless, a large uncertainty exists in selecting the starting models to be subjected to refinement, a step which is crucial to achieving reliable results.

Two algorithms are usually employed in the refinement of structures against fibre diffraction data: the linked-atom least-squares (LALS) method[7] and the restrained least-squares (RSL) method[8]. Both methods are based on the incorporation of stereochemical information into the refinement procedure in order to reduce the degrees of freedom of the system. Although these methods have been successfully applied in many cases, the reliability of the results is known to depend heavily on the rightness attained in the initial description of the starting model. In order to overcome this difficulty, several strategies have been adopted during the last few years. In polymers with short chemical repeating units (CRUs), a complete and systematic exploration of the conformational space could be easily performed to find all the possible models[9]. Conversely, in compounds with CRUs of medium and large sizes, such a search is rather difficult and scientists tend to resort either to their chemical intuition[10] or to subsidiary methodologies usually based on force-field calculations[11–14]. Of course, the former procedure cannot guarantee the quality of the results since an undetermined number of possible conformations is unavoidably omitted. On the other hand, the application of methods based on force-fields-like molecular mechanics and molecular dynamics entails vast amounts of computer time. The long time period required to perform a large trajectory drastically reduces the number of starting structures analysed in practice. This, together with the probabilistic problems inherent to the dynamical approach, explains the scarce applicability of these method to date.

In this work, we present a low-expense strategy for building up the geometries of all the possible molecular conformations compatible with fibre diffraction data of polymers with medium and large CRUs. The method takes advantage of the 'equivalence principle'[15], according to which the polymer chains adopt regular conformations in the crystalline state, i.e. their conformation is defined by a succession of structurally equivalent units. The procedure combines the application of geometric

---

* To whom correspondence should be addressed

contour conditions with the use of a simple energy analysis in order to select the more favourable models. Since no energy gradient needs to be evaluated, the speed of this model generator to check conformers is several orders of magnitude faster than that of molecular mechanics or molecular dynamics. The methodology developed here has been incorporated into a manageable computer program, named GEMOX (GEneration of polymer MOlecular models compatible with X-ray diffraction data), which broadens its application to a wide variety of users including those who are not computational experts.

METHODS AND RESULTS

Internal geometry of the chemical repeating unit, contour conditions and helix geometry

GEMOX uses as input the internal coordinates of the CRU. The Cartesian coordinates of both backbone and side chain atoms are obtained by a procedure based on the changes of the local coordinate systems using the transformations:

$$\mathbf{G}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{pmatrix} \quad (1)$$

$$\mathbf{G}_2 = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2)$$

where $\mathbf{G}_1$ and $\mathbf{G}_2$ represent the rotation along the dihedral angle ϕ and the bond angle θ, respectively. Once the Cartesian coordinates of every atom contained in the first CRU are determined, contour conditions are applied using the same geometric procedure but taking as starting point the atomic coordinates of the last three atoms of the first CRU. This permits a polymer chain with a regular conformation to be obtained.

Three helical parameters need to be defined for a helical polymer composed of a sequence of identical CRUs: ρ the radius of the helix, h the rise per residue which is a constant axial translation height, and ω the frequency constant of rotation. These can be easily computed from the distance (α) between the CRUs $i-1$ and i, the angle (β) between the CRUs $i-1$, i and $i+1$, and the dihedral angle (γ) between the CRUs $i-1$, i, $i+1$ and $i+2$. In order to determine α, β and γ, the rotation and translation matrices along each CRU are needed[16]. All the molecular models generated by the program can be adequately characterized using this procedure.

Energy calculations

The GEMOX computer program allows the user to accept or reject molecular models using a conformational energy criterion. The program generates polymer chain models using the following procedure: every model holds bond angles and distances within standard values, so dihedral angles are the only internal coordinates that are modified. Once the dihedral angles are set to specific values, the whole CRU is reoriented to fulfil contour conditions at both ends. A single point energy calculation is then performed taking into account the internal energy of the CRU i and its interaction energies with those CRUs between $i-N$ and $i+N$, where N is the number of CRUs that the user wants to consider in the calculation. The conformational energy is computed by adding both the van der Waals and the electrostatic non-bonding contributions to the energy involved in the torsion of the bonds. The electrostatic term tends to favour models with strong electrostatic interactions like hydrogen bonds, whereas with the van der Waals energy function all the molecular models that are sterically forbidden or have contacts that are too close are discarded. Both electrostatic and van der Waals interactions are calculated for all those atoms which are neither directly bonded nor related by a bond angle. A scaling factor of 0.5 has been incorporated when dealing with 1–4 interactions in order to decrease their relative weight.

The set of bonded and non-bonded parameters can be calculated *a priori* according to different methodologies. The PAQMD/AM1 method[17,18] has been demonstrated to be a powerful strategy for computing torsional barriers, whereas for fractional charges the electrostatic method is highly recommended owing to its superiority with respect to other methods[19,20]. The use of electrostatic charges derived from Dewar's semiempirical wavefunctions[21,22] is an efficient method for computing the electrostatic charges of large molecules[23].

Reliable modelling of the electrostatic interactions is crucial for determining the molecular structure. Thus, several works have concluded that the effect of different dielectric constants on conformation is different depending on the molecular system under study[24]. Two different expressions for the dielectric constant were considered in GEMOX: (i) a distance-independent value, $\epsilon = \kappa$, and (ii) a distance-dependent value $\epsilon = \kappa r$, where κ is a constant value selected by the user and r is the distance between two interacting atoms.

Helical symmetry

The helical symmetry of a polymer chain may be expressed as the integer number of residues (u) contained in the integer number of turns (t). The fraction u/t is equivalent to P/h, where h is the unit height and P is the pitch of the helix. For a polymer with a fibre identity distance c the following relations apply:

$$U = N/t \quad (3)$$

$$P = c/t \quad (4)$$

Both helical parameters u and P are allowed to vary within a tolerance factor in the conformational search. Furthermore, in the case of adirectional polymers, like those derived from retro-amino acids, several elements of symmetry can be defined using both the number of monomers required to define a CRU and the symmetry related to the torsional angles sequence. These are: screw axis, glide plane and inversion centre. The program allows both the use of symmetry relations between different dihedral angles and for some dihedral angles to be kept fixed at a certain value.

Program structure

The general flowchart of the program is displayed in *Figure 1*. GEMOX uses as inputs: (i) user information and directives about the search procedure; (ii) the

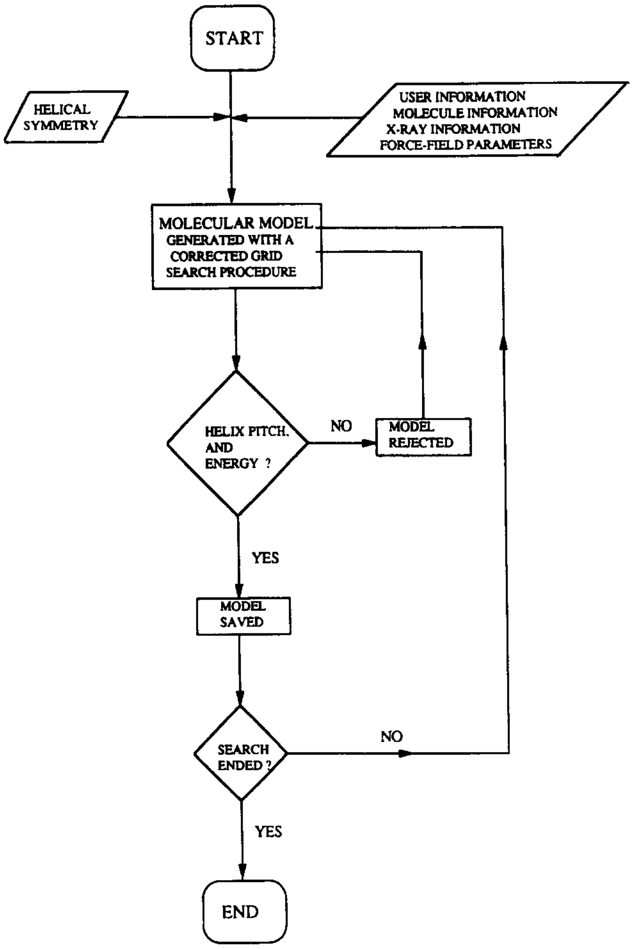

Figure 1 Flowchart of the GEMOX program

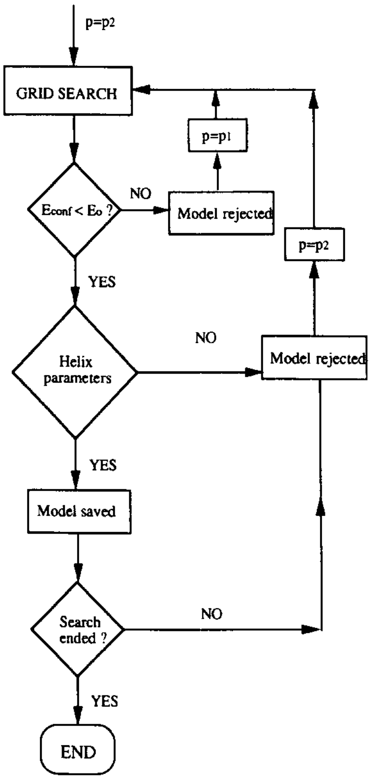

Figure 2 Flowchart of the corrected grid search algorithm used in GEMOX

molecular connectivity, bond lengths and bond angles of the CRU; (iii) the helix pitch derived from X-ray diffraction data; and (iv) the torsional and non-bonded force-field parameters. The program also makes use of the molecular symmetry, which drastically affects the speed of the search procedure. Furthermore, the dihedral angles that are to be kept fixed during the whole process must also be specified at this step.

The program generates all the molecular models compatible with a given helix pitch using a corrected grid search procedure. This implies a systematic variation from 0° to 360° for all the considered dihedral angles. Since the number of molecular conformations thus obtained is usually very large, even when medium size CRUs are considered, acceptance or rejection of a model becomes a complex task. In order to facilitate the selection, two conditions must be fulfilled by all the accepted molecular models. These are: (i) the difference between the experimental and computed values for the helix pitch must be smaller than a threshold value; (ii) the model must be energetically allowed. Thus, for each model the conformational energy is computed. The calculation of the torsional, electrostatic and van der Waals energies for a particular model is an easy task since all the force-field parameters are inputs of the program.

The reliability of the results obtained in a standard grid searching procedure depends largely on the size of the grid step[25]. When a very small grid step is used, good results are obtained but at high computational cost. Conversely a large grid step renders a poor sampling of the conformational space. In order to carry out an efficient search we have employed a modified grid search procedure, in which two grid steps named p_1 and p_2 are adequately combined.

Figure 2 shows the flowchart of the subroutine used in the program for sampling the conformational space. At a first stage the program uses a small grid step (p_2) to generate the molecular models. Since we are only interested in conformations energetically favoured, the energy is calculated for each generated geometry. If the conformational energy is greater than a threshold value (E_0) the program assumes that an unfavoured zone of the conformational space is being explored and the grid step is changed to a large value (p_1). The new grid step is used until an energetically favoured model is found. Then, the program turns back to the small grid step p_2 in order to perform an efficient sampling. For each energetically favoured model the helix pitch is computed using a geometric criterion. If the difference between the experimental and calculated values is smaller than a tolerance factor the model is saved, otherwise the model is rejected. The combination of two grid steps with energy and helical symmetry criteria renders a sampling more convenient than a systematic grid search procedure from both computing time and efficiency points of view. Pilot calculations indicated that a very small change in the dihedral angles may induce a large variation in the helix symmetry. Therefore, the change from p_2 to p_1 must be performed only in energetically unfavoured zones, i.e. those in which the energy criterion has not been satisfied.

Test cases

The structures of poly(L-alanine)[6,26], poly(α-isobutyl-β-L-aspartate)[11] and nylon 2/12[27] were used to test the

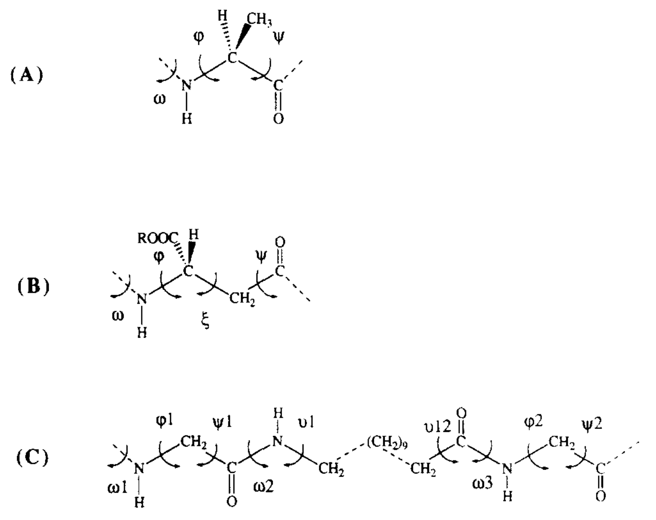

Figure 3 Schematic representation of the chemical repeating units of poly(L-alanine) (A), poly(α-isobutyl-β-L-aspartate) (B) and nylon 2/12 (C)

search procedure, since these three polymers have been refined previously using both the LALS method and energy calculations. A schematic representation of the monomeric units for the three polymers studied, with expression of the notation used in each case for the dihedral angles is displayed in *Figure 3*. Calculations have been performed on a Silicon Graphics RI-4000. In all cases the large grid step was kept fixed at $p_1 = 10°$, whereas p_2 was varied within different ranges depending on the system under study. An energy cutoff of 0.0 kcal mol^{-1} was applied. The tolerance accepted for the helix pitch was 0.1 Å in all cases. Conformational energy maps for poly(L-alanine) and nylon 2/12 were also generated with GEMOX using a grid of 15°.

Poly(L-alanine). Two different crystal forms are known for this polypeptide: the α-helix[6] with 47 residues in 13 turns and the β-sheet[26], which is an almost extended conformation made of 2/1 helices. The repeat chain lengths of these forms are 70.3 Å and 6.89 Å, respectively. Force-field parameters for alanine were taken

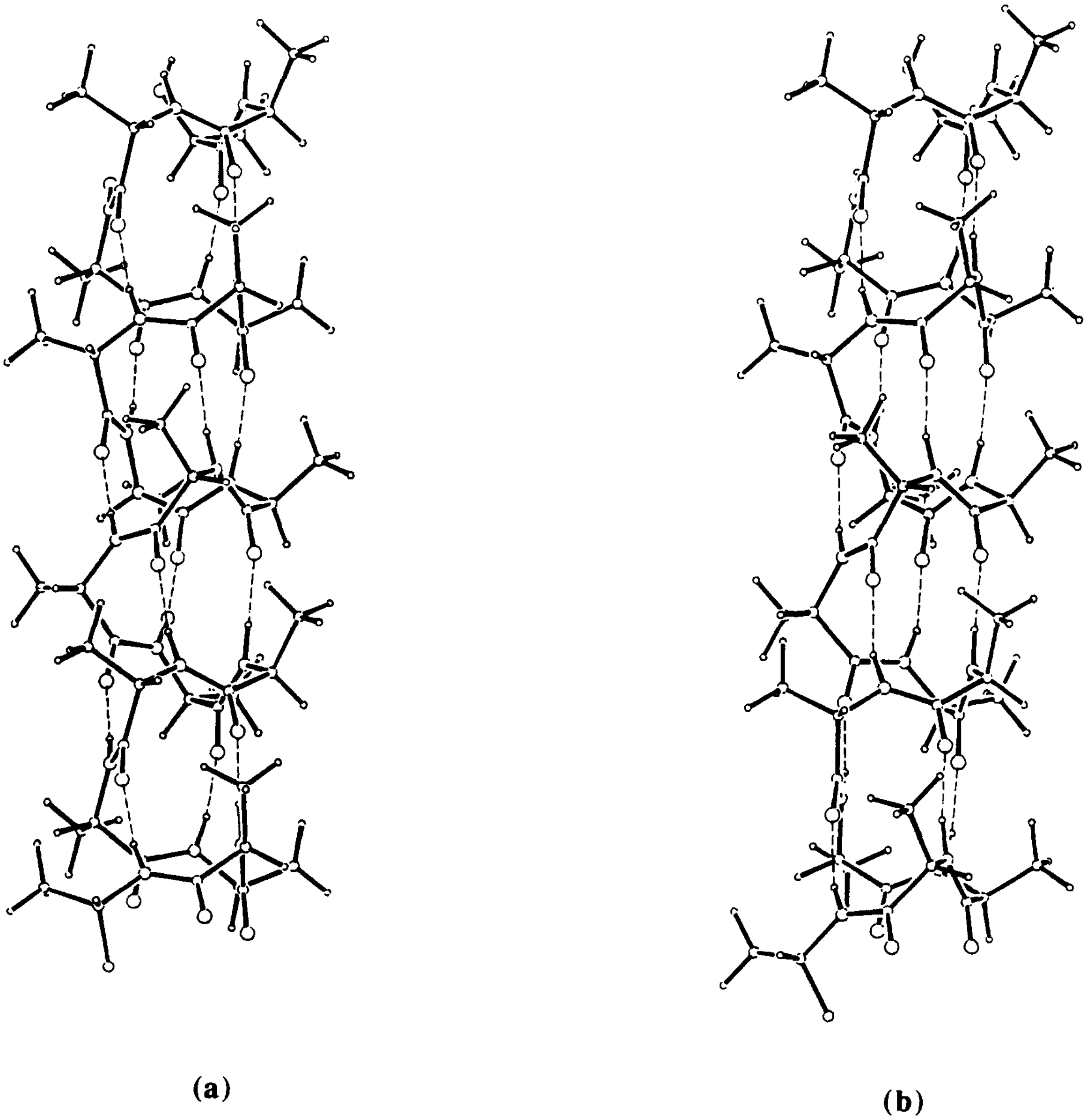
(a)
(b)

Figure 4 Molecular conformations of the standard α-helix (a) and the α-helix modelled by GEMOX (b)

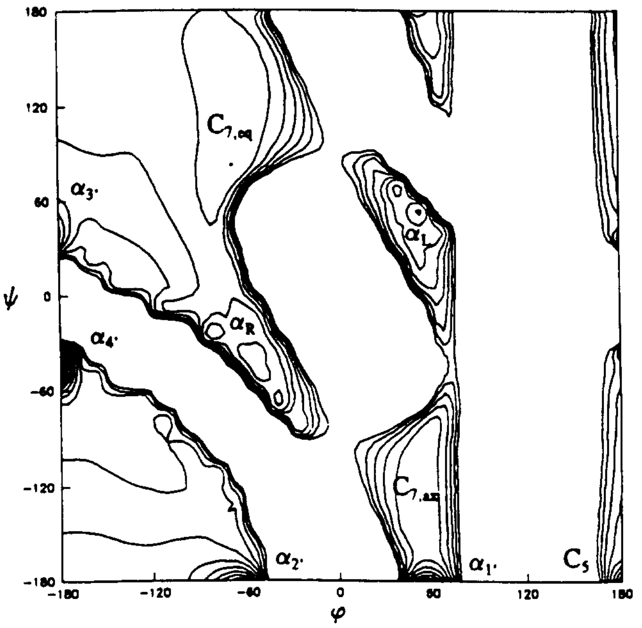

Figure 5 Potential energy surface $E = E(\phi, \psi)$ of poly(L-alanine). Dihedral angles of the minimum energy regions labelled in the map are shown in *Table 1*

Table 1 Low energy regions as calculated by the GEMOX program for poly(L-alanine)

Region	ϕ	ψ	ΔE^a
α_R	−60	−60	0.0
C_5	180	180	2.2
α_1'	60	180	8.6
α_L	60	45	10.1
α_2'	−60	180	14.6
α_4'	180	−60	14.6
α_3'	180	60	19.3
$C_{7,eq}$	−60	60	20.0
$C_{7,ax}$	60	−75	24.4

[a] Torsional angles and relative energies in degrees and kcal mol^{-1} residue^{-1}, respectively

Table 2 Torsional angles[a] obtained by different methods for poly(α-isobutyl-β-L-aspartate)

Form	Method[b]	ω	ϕ	ξ	ψ
Hexagonal	MM	180.0	151.3	−59.1	118.1
	MD	175.0	157.2	−54.1	114.8
	LALS	180.0	151.3	−59.1	118.1
	GEMOX	180.0	159.0	−63.0	120.0
Tetragonal	MM	160.0	144.8	−65.8	136.0
	LALS	180.0	144.2	−86.9	148.5
	GEMOX	180.0	143.0	−85.0	150.0

[a] Torsional angles in degrees
[b] MM, molecular mechanics; MD, molecular dynamics; LALS, linked-atom least-squares

from AMBER libraries[28]. The corrected grid search algorithm found both the α-helix ($\phi, \psi = -57.4°, -47.5°$) and the extended ($\phi, \psi = -138.8°, 134.7°$) conformations for p_2 values ranging from 1° to 7°. The α-helix was the lowest energy model found for all these values of p_2. This must be attributed to the presence of strong intramolecular hydrogen bonding interactions, which are taken into account by the potential energy function of GEMOX. A comparison between the standard α-helix and the helical structure found by GEMOX is shown in *Figure 4*. It is worth noting that the difference between the two conformations is almost negligible. On the other hand, the extended conformation uses neighbouring chains to form hydrogen bonds. Although this structure is included within the set of accepted models, this conformation is slightly unstabilized with respect to the lowest energy model. For p_2 larger than 7° the procedure is similar to a standard grid search algorithm, and therefore the exploration of the conformational space becomes rather crude. The dihedral angles calculated by GEMOX are reproduced within a maximum margin of error of 10°.

In order to prove the reliability of the potential energy function implemented in GEMOX, we computed the conformational energy map for poly(L-alanine). Results are shown in *Figure 5* and *Table 1*. As can be seen, the lowest energy region corresponds to a right-handed α-helix (α_R) with $\phi, \psi \approx -60°, -60°$ which is clearly stabilized by intramolecular hydrogen bonding interactions. The second region of minimum energy corresponds to an extended structure usually referred to as C_5 (intramolecular, five-membered hydrogen bonded ring) in conformational studies of peptides, which is around 2 kcal mol^{-1} residue^{-1} less favoured than the α_R. Other zones energetically accessible are located around the structures named α' by Pople and co-workers[28] and the intramolecular seven-membered hydrogen bonded rings ($C_{7,eq}$ and $C_{7,ax}$). These conformations have been characterized as local energy minima in the potential energy surface of the *N*-acetyl-*N'*-methyl-L-alanylamide at the *ab initio* HF/3-21G and HF/6-31 + G* computational levels[28] confirming the goodness of our results.

Poly(α-isobutyl-β-L-aspartate). This L-aspartic derivative is a stereoregular poly(β-amide) with an alkoxycarbonyl group attached to the β-carbon atom of the repeating unit. This compound adopts two crystal forms, hexagonal and tetragonal, made of 13/4 and 4/1 helices. The repeat chain lengths are 4.95 Å and 19.9 Å respectively[11]. The energy cutoff was increased to 0.5 kcal mol^{-1} in the analysis of the tetragonal form, since helices with a 4/1 symmetry are only possible through some distortion of the peptide bond. (In order

Table 3 Torsional angles[a] obtained by different methods for nylon 2/12[b]

Form	Method[c]	ϕ	ψ	ν_1	ν_2	ν_3	ν_{10}	ν_{11}	ν_{12}
I	LALS	180.0	180.0	180.0	180.0	180.0	180.0	180.0	180.0
	GEMOX	180.0	180.0	180.0	180.0	180.0	180.0	180.0	180.0
II	LALS	78.0	−144.2	107.0	179.0	−178.0	174.0	−165.0	−108.0
	GEMOX	95.0	−142.0	140.0	180.0	180.0	180.0	180.0	−140.0

[a] Torsional angles in degrees
[b] Dihedral angles from ν_4 to ν_9 are considered to be constant and equal to 180°
[c] LALS, linked-atom least-squares

to obtain a central hole of 2.5 Å in the ω-helix a distortion of the peptide bond is required.) GEMOX found the structures of the hexagonal and tetragonal forms for p_2 values in a range from 1 to 3. A comparison of the conformational angles obtained for both forms from LALS, molecular mechanics, molecular dynamics and GEMOX calculations is presented in *Table 2*. Results indicate the capability of this procedure to generate molecular conformations at a reasonable computational cost.

Nylon 2/12. Several members of the nylon $2/n$ family have been investigated by means of X-ray and electron diffraction[27,29–31]. We select nylon 2/12 as a test case since it is the copolyamide $2/n$ with the largest CRU studied so far. Furthermore, its conformational preferences have been examined with the LALS methodology. Two crystal forms were found for nylon 2/12. Form I is made of fully extended chains arranged in hydrogen bonded sheets similar to the α-form of nylons. Form II consists of a hexagonal array of six-fold helices intermolecularly linked in a crystal lattice of $a = 4.79$ Å and $c = 112.8$ Å, with a residue height of 18.8 Å. GEMOX calculations have been performed using the following symmetry restrictions: $\phi_1 = \phi_2$ and $\psi_1 = \psi_2$. The torsional angles ω_i and ν_i were kept in *trans* conformation. The molecular models of form I are found using p_2 values in the range from 1 to 7, whereas form II was located considering p_2 values ranging from 1 to 3. A comparison of the LALS and GEMOX values is shown in *Table 3*. *Figure 6a* shows the potential energy surface $E = E(\phi, \psi)$ for nylon 2/12, which indicates that both crystal forms are energy minima. On the other hand, the map $E = E(\nu_1, \nu_2)$, which is shown in *Figure 6b*, suggests that this potential energy surface is flat, allowing a large number of possible conformations for the methylene groups.

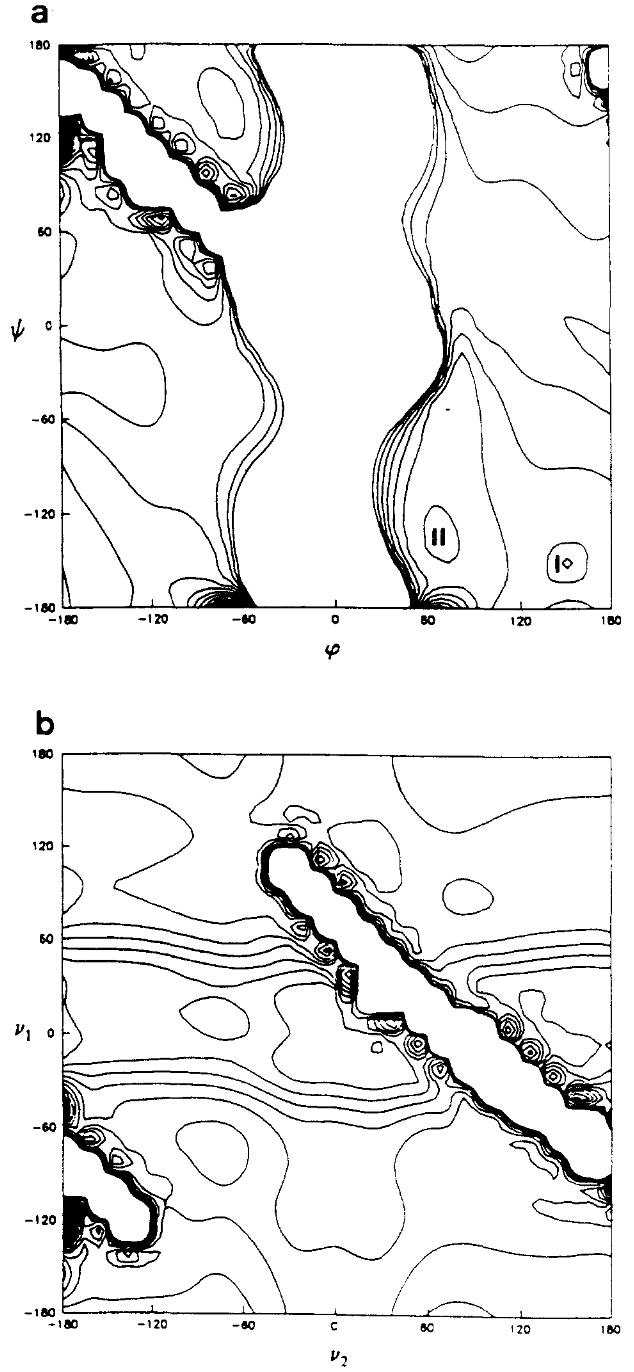

Figure 6 Potential energy surfaces $E = E(\phi, \psi)$ (a) and $E = E(\nu_1, \nu_2)$ (b) of nylon 2/12. Crystal forms of nylon 2/12 are labelled in (a). Both maps were computed considering the dihedral angles fixed at the values found for form I

CONCLUSIONS

GEMOX is a program designed to generate all the molecular models for a polymer which are compatible with a certain set of X-ray diffraction data. It must be stressed that this program has been developed in order to provide people working with fibre diffraction with a tool making the structural analysis of macromolecules easy and rigorous. We do not recommend GEMOX for finding the best molecular model; this should be done preferentially by LALS and/or energy calculations.

The application of GEMOX is justified in the analysis of the conformational space of polymers involving a large CRU, where many feasible molecular models would be unavoidably omitted. The program uses a corrected grid search algorithm in which both experimental diffraction data and conformational energy criteria are taken into account. Test calculations have been performed on three representative systems – poly(L-alanine), poly(α-isobutyl-β-L-aspartate) and nylon 2/12 – which illustrate the usefulness of the program. The results indicate the existence of an important number of molecular models compatible with the diffraction patterns for the three compounds. The best solution for each case, i.e. the conformation resulting from refinement against X-ray data using LALS or energy calculations, is of course included in the set of filtered models.

ACKNOWLEDGEMENTS

This work has been supported by DGICYT (Grant Number PB93-555) and by the CESCA (Centre de Supercomputació de Catalunya).

REFERENCES

1 Tadokoro, H. 'Structure of Crystalline Polymers', Wiley-Interscience, New York, 1979

2 Alexander, L. E. 'X-Ray Diffraction Methods in Polymer Science', Wiley-Interscience, New York, 1969

3 Moorhouse, R., Winter, W. T., Arnott, S. and Bayer, M. E. *J. Mol. Biol.* 1977, **109**, 373

4 Yui, T., Imada, K., Okuyama, K., Obata, Y., Suzuki, K. and Ogawa, K. *Macromolecules* 1994, **27**, 7601

5 Stubbs, G., Namba, K. and Makowski, L. *Biophys. J.* 1986, **49**, 58

6 Fraser, R. D. B. and McRae, T. C. P. 'Conformation in Fibrous Proteins', Academic Press, London, 1973

7 Smith, P. J. C. and Arnott, S. *Acta Cryst.* 1978, **A34**, 3

8 Hendrickson, W. A. 'Methods in Enzymology' (Eds H. W. Wyckoff, C. H. W. Hirs and S. N. Timasheff), Vol. 115, Academic Press, Orlando, 1978, pp. 252–270

9 Alemán, C. and Bella, J. *Biopolymers* 1995, **35**, 257

10 Fernández-Santín, J. M., Muñoz-Guerra, S., Rodríguez-Galán, A., Aymamí, J., Lloveras, J., Subirana, J. A., Giralt, E. and Ptak, M. *Macromolecules* 1987, **20**, 62

11 Bella, J., Alemán, C., Fernández-Santín, J. M., Alegre, C. and Subirana, J. A. *Macromolecules* 1992, **25**, 5225

12 Navas, J. J., Alemán, C., López-Carrasquero, F. and Muñoz-Guerra, S. *Macromolecules* 1995, **28**, 4487

13 Alemán, C., Bella, J. and Pérez, J. J. *Polymer* 1994, **35**, 2596

14 Alemán, C., Franco, L. and Puiggalí, J. *Macromolecules* 1994, **27**, 4298

15 Corradini, P. 'The Stereochemistry of Macromolecules', (Ed. A. D. Ketley), Vol. 3, Marcel-Dekker, New York, 1968

16 Hopfinger, A. J. 'Conformational Properties of Macromolecules', Academic Press, New York, 1973

17 Alemán, C., Canela, E. I., Franco, R. and Orozco, M. *J. Comp. Chem.* 1991, **12**, 664

18 Alemán, C. and Orozco, M. *Biopolymers* 1994, **34**, 941

19 Orozco, M. and Luque, F. J. *J. Comp. Chem.* 1990, **11**, 909

20 Alhambra, C., Luque, F. J. and Orozco, M. *J. Comp. Chem.* 1994, **15**, 12

21 Dewar, M. J. S. and Thiel, W. *J. Am. Chem. Soc.* 1977, **99**, 4899

22 Dewar, M. J. S., Zoebisch, E. G., Healy, E. F. and Stewart, J. J. P. *J. Am. Chem. Soc.* 1985, **107**, 3902

23 Orozco, M. and Luque, F. J. *J. Comp. Aided Mol. Design* 1990, **4**, 411

24 Vega, M. C., Alemán, C., Giralt, E. and Pérez, J. J. *J. Biomol. Struct. Dyn.* 1992, **11**, 429 (and references therein)

25 Crippen, G. 'Distance Geometry and Conformational Calculations', Research Studies Press, Wiley, New York, 1981

26 Elliot, A. 'Poly-α-amino acids' (Ed. G. D. Fasman), Marcel-Dekker, New York, 1967

27 Bermúdez, M., Puiggalí, J. and Muñoz-Guerra, S. *Macromolecules* 1994, **27**, 6325

28 Head-Gordon, T., Head-Gordon, M., Frisch, M. J., Brooks, C. L. and Pople, J. A. *J. Am. Chem. Soc.* 1991, **113**, 5989

29 Puiggalí, J., Muñoz-Guerra, S. and Lotz, B. *Macromolecules* 1986, **19**, 119

30 Puiggalí, J., Muñoz-Guerra, S. and Subirana, J. A. *Polymer* 1987, **28**, 209

31 Puiggalí, J. and Muñoz-Guerra, S. *J. Polym. Sci., Polym. Phys. Edn.* 1987, **25**, 513