polymer papers

A comparison of the linked-atom least squares and PS79 systems for determining polymer structures from X-ray fibre diffraction data

R. P. Millane and T. V. Narasaiah
The Whistler Center for Carbohydrate Research, Smith Hall, Purdue University, West Lafayette, Indiana 47907, USA
(Received 18 August 1988; revised 30 January 1989; accepted 1 February 1989)

When determining polymer structures using X-ray fibre diffraction, the diffraction data must be supplemented by stereochemical information. The structure obtained depends on the flexibility of structural models, the types of restraints and constraints used, the relative weighting of different terms, and the balance between the stereochemical and X-ray components. This affects not only final models, but also the important step of rejecting classes of incorrect models. We report a comparison of determinations of a structure using the two most widely used refinement systems (the linked-atom least squares and PS79 systems). Discrimination between classes of model and stereochemistry of the final structures are compared in detail. Implications for polymer structure determination using X-ray fibre diffraction are discussed.

(Keywords: polymer structure; X-ray fibre diffraction; linked-atom least squares; PS79; refinement systems)

INTRODUCTION

Wide angle X-ray diffraction from oriented (and often polycrystalline) specimens can be used to determine molecular and crystal structures of polymers[1-3]. This technique is referred to as X-ray fibre diffraction and is a variation of traditional X-ray crystallographic techniques used to determine the structures of molecules that form single crystals. The molecules or crystallites in polymer specimens (fibres) are randomly rotated relative to each other about their long axes so that the diffraction pattern is cylindrically averaged. The number of diffraction data obtained is therefore substantially reduced compared to that from single crystals. Also, it is usually impossible to solve the phase problem using traditional crystallographic techniques such as isomorphous replacement (except in special circumstances[4]). Structure determination therefore involves generating all possible kinds of model consistent with the chemical sequence of the polymer, and the spacings and symmetry determined from the diffraction pattern. Each model is refined against the diffraction data and stereochemical restraints to obtain optimum models of each kind. A structure is determined unequivocally only if one of the optimized models is significantly superior to the rest on the basis of stereochemical acceptability and agreement between the calculated and observed X-ray amplitudes. Discrimination between competing models is therefore a critical step in fibre diffraction analysis.

The incorporation of stereochemical restraints to supplement sparse diffraction data is necessary to increase the data/parameter ratio to a value for which refinement is meaningful. The linked-atom least squares (LALS)[5,6] and PS79[7] systems are the only generally applicable ones that co-refine structures against X-ray data and stereochemical restraints. The Hendrickson–Konnert system originally designed to refine globular proteins has recently been modified for fibrous structures[8], but this is more suitable for refining larger macromolecular aggregates that have been built into electron density maps[4] and is not discussed here.

Successful refinement and the stereochemistry of final models may depend on features of the refinement system used such as the flexibility of models, the relative weighting of the different stereochemical restraints, and the balance between good stereochemistry and good agreement with the diffraction data. We have assessed the sensitivity of structure determinations to the refinement system used, by comparing determinations of a structure using the LALS and PS79 systems. The structure chosen was ramie cellulose I because this has been examined using a number of different systems. The ability of the two systems to discriminate between different models, and the stereochemistry of the final models were compared.

THE LALS AND PS79 REFINEMENT SYSTEMS

The LALS and PS79 refinement systems both co-refine polymer structures against diffraction data and stereochemical restraints. They therefore have many features in common but there are a number of differences in their implementations.

The LALS system was developed in the 1960s[5] and has since been extended[6,9] to a powerful and flexible system. It has been used to determine over 100 polynucleotide, polysaccharide and polypeptide structures[1]. A linked-atom description of the molecule is used in which interatomic relationships are described in terms of bond lengths, bond angles and conformation (torsion)

0032-3861/89/101763-05$03.00

angles. The quantity Ω given by

$$\Omega = \sum_{k=1}^{N_e} e_k \Delta\theta_k^2 + \sum_{m=1}^{N_x} \omega_m \Delta F_m^2 + \sum_{i=1}^{N_c} k_i \Delta d_i^2 + \sum_n \lambda_n G_n \quad (1)$$
$$= E + X + C + L$$

is minimized by varying a set of chosen parameters consisting of conformation (and possibly bond) angles, packing parameters, and X-ray scale and attenuation factors. Preferred conformational domains are achieved by restraining conformation angles (θ_k) to their respective expected values and they constitute the first term E. The term X involves the differences (ΔF_m) between the observed X-ray amplitudes and those calculated for the model structure. C involves the close, non-bonded interatomic distances, d_i, which are driven beyond normally accepted contact limits. The quantities e_k, ω_m, and k_i are weights that are inversely proportional to the estimated variances of the data. $N = N_x + N_c + N_e$ is the total number of data (which include the restraints). The term L involves constraints which are relationships that are to be satisfied exactly ($G_n = 0$) and the λ_n are Lagrange multipliers. Constraints are used, for example, to ensure connectivity between helical units and to ensure that chemical ring systems are closed. Correct stereochemistry between adjacent helical units is usually achieved during refinement by replicating three equivalent atoms at each end of the unit, and incorporating constraints to ensure that the three atom pairs are related by the appropriate screw symmetry. Minimization of Ω is achieved using full matrix least squares which gives a rather large radius of convergence in parameter space. It is sometimes convenient to apply more constraints than there are independent relationships between parameters, and the normal equations are therefore solved using singular value decomposition to remove the resulting singularities.

The PS79 refinement system was first developed in the 1970s[10–12] and extended to a flexible system[7] that has been used primarily for determining polysaccharide structures, although it can be applied to any kind of polymer. The function minimized is Φ given by

$$\Phi = (1-f)\left\{\sum_k e_k \Delta A_k^2 + \sum_i k_i \Delta d_i^2\right\} + f\left\{\sum_m \omega_m \Delta F_m^2 / \sum_m \omega_m F_{om}^2\right\}^{1/2} \quad (2)$$

where A_k is any bond length, bond angle, or conformation angle, the F_{om} are the observed X-ray amplitudes, f is a constant chosen to balance the stereochemical and X-ray terms, and the remaining symbols are defined as in equation (1). The steric compression can also be minimized using Lennard–Jones potentials. As opposed to the LALS systems, the PS79 system utilizes the 'virtual bond' (the vector joining equivalent positions in a helical repeat unit) method[7] to ensure proper connectivity between helix repeats. The other main difference is that PS79 achieves minimization of Φ using a constrained optimization procedure based on the simplex method[11], rather than least squares.

METHODS

Ramie cellulose I was chosen for the comparison because its structure has been studied using the LALS and PS79 refinement systems[14,15], and it has also been the subject of some controversy concerning chain packing[16]. The LALS and PS79 systems both predict a parallel up chain packing, with all primary hydroxyl groups in the t_g ($\theta_{o6} = \theta$(C4–C5–C6–O6) $\simeq -60°$) domain and with the same hydrogen bonding pattern[15]. Other reports that favour different packings and/or different primary hydroxyl conformations appear to be due to the use of systems that do not properly co-refine structures against X-ray data and stereochemical restraints. We are concerned here with how strongly alternative structures are eliminated, and the details of the final models.

Refinements using both systems were carried out using the X-ray data reported by Woodcock and Sarko[14] so that the results would reflect differences in the refinement systems, rather than in the data. The X-ray data extend to approximately 2.0 Å resolution. There are 49 independent X-ray reflections, 12 of which were too weak to be measured accurately and so were replaced by threshold values. The weak reflections were treated as data in the refinement only where the calculated amplitudes were larger than the threshold values. The refinement of ramie cellulose I using the LALS system has been described[15] although some minor variations were used here to allow a meaningful comparison with the reported refinements using PS79[14].

Different chain packings were compared using refinements in which the sugar rings were flexible (see below).A detailed study of the implications of relaxing the $P2_1$ symmetry showed that there is no justification for accepting a lower symmetry structure for ramie cellulose I, or for the corner and centre chains being conformationally different[15]. The space group $P2_1$ was therefore used with both chains conformationally identical. This requires that the two chains have 2-fold screw symmetry along their molecular axes, are positioned at (0, 0) and (1/2, 1/2) in the a–b plane (*Figure 1*), but that their individual rotations (μ_1 and μ_2) and their relative axial shift (Δw) are variable. The variable parameters were therefore the two bridge conformation angles ($\phi = \theta$(O5–C1–O4–C4) and $\psi = \theta$(C1–O4–C4–C5)), the primary hydroxyl conformation θ_{o6}, the bridge bond angle, the conformation angles and endocyclic bond angles of the glucose ring (restrained to standard values[17]), the packing parameters μ_1, μ_2 and Δw, and the X-ray scale and (isotropic) attenuation factors. This gives a total of 25 parameters and 13 constraints. Refinements were conducted with the three possible packings (parallel up, parallel down and antiparallel) of the chains[15] and compared using refinement statistics as described in the next section.

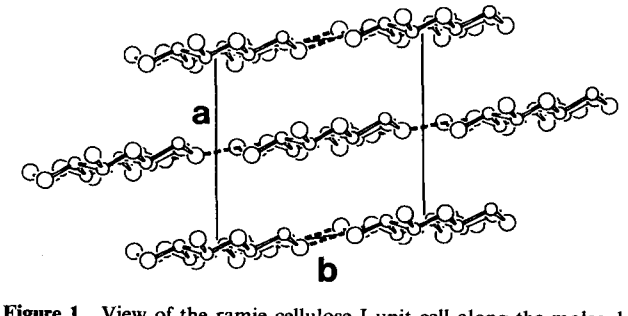

Figure 1 View of the ramie cellulose I unit cell along the molecular axis showing the corner and centre chains that form hydrogen bonded sheets. The O3---O6 intermolecular hydrogen bonds are denoted by broken lines

Results for the PS79 system were taken from the final refinements described by Woodcock and Sarko[14]. They used molecular models without 2-fold screw symmetry, but with the corner and centre chains identical and the centre chain also positioned at (1/2, 1/2) in the a–b plane. Their variable parameters are practically identical to ours except that there are twice as many variable conformation and bond angles because of the two independent sugar rings, and they used an anisotropic (three variable components) attenuation factor.

COMPARISON OF DISCRIMINATIONS

As described in the introduction, discrimination between alternative models is a critical step in the determination of polymer structures from fibre diffraction data. Good discrimination depends on model structures being properly constrained yet sufficiently flexible, and on a proper balance in the refinement between the stereochemical and X-ray discrepancies. The modelling of steric compression should approximate the energy of the structure and penalize structures containing over-short non-bonded contacts in a satisfactory manner.

With the LALS system, different refined models are compared by examining the overall discrepancy Ω. This is done quantitatively using the discrepancy ratio

$$\hat{\Omega} = (\Omega/N)^{1/2}/(\Omega^b/N^b)^{1/2} \quad (3)$$

in Hamilton's test[18] where the superscript b denotes the best model (that with the lowest Ω). The degree to which the X-ray data or steric compression alone support a particular model can be assessed using the ratios

$$\hat{X} = (X/N_x)^{1/2}/(X^b/N_x^b)^{1/2} \quad (4)$$

and

$$\hat{C} = (C/N_c)^{1/2}/(C^b/N_c^b)^{1/2} \quad (5)$$

The absence of over-short non-bonded interatomic distances is also a necessary requirement of acceptable models. The traditional and quadratic X-ray discrepancy indices R and R'' given by

$$R = \sum |F_o - F_c| / \sum F_c \quad (6)$$

and

$$R'' = \{\sum (F_o - F_c)^2 / \sum F_o^2\}^2 \quad (7)$$

where F_o and F_c are the observed and calculated structure amplitudes respectively, are also used to assess how the X-ray data support the different model structures. It is worth noting that R-factors in fibre diffraction analysis are usually lower than those obtained in protein crystallography[19–21]. Other features such as the lengths and geometries of hydrogen bonds, neutralization of charged groups on the polymer by counterions, coordination geometries, and consistency with other physical data are also used when assessing different model structures.

With the PS79 system, the quadratic residual R'' (or R) is primarily used, together with the absence of over-short non-bonded contacts, to compare model structures. The energy due to non-bonded contacts (the sum involving Δd_i^2 in equation (2)) is divided into packing energy and conformation energy terms (denoted here by P and Q respectively) that include the intermolecular and intramolecular contacts respectively. We define $T = P + Q$ which is analogous to the contact term C in equation (1).P, Q and Tcan be used to assess the steric compression of competing models, as well as other features of the models as described above.

Statistics of refinements of the different packing models of ramie cellulose I using the LALS and PS79 systems are listed in *Table 1*. Inspection of the table shows that for the LALS refinement, the parallel down model is clearly inferior to the other two models with respect to both the overall discrepancy and the X-ray agreement, and is not considered further here. For the PS79 system, details of the final refinements of the parallel down model were not available, but this model was rejected early in the analysis due to its poor X-ray agreement[14], consistent with the LALS results.

The parallel up structure is preferred over the antiparallel structure for both refinement systems. For the LALS refinement, both the overall discrepancy and X-ray ratios $\hat{\Omega}$ and $\hat{X}$ allow the antiparallel structure to be rejected at the 0.995 confidence level for a one-dimensional hypothesis[18]. There is little difference

Table 1 Statistics of refinement of the three packing models for ramie cellulose I using the LALS and PS79 systems

	LALS refinement Model				PS79 refinement Model	
Refinement statistic	Parallel up	Anti-parallel	Parallel down	Refinement statistic	Parallel up	Anti-parallel
E	5.9	5.5	5.5	P	24.4	30.2
C	66.7	70.1	67.7	Q	29.1	29.9
$\hat{C}$	1.00	0.98	1.01	T	53.5	60.1
N_c	68	74	67	$\hat{T}$	1.00	1.06
X	19.6	34.7	86.3	R	0.22	0.31
$\hat{X}$	1.00	1.30	2.05	R''	0.19	0.25
N_x	45	47	47	$\hat{R}''$	1.00	1.28
R	0.18	0.26	0.34			
R''	0.17	0.22	0.35			
$\hat{R}''$	1.00	1.35	2.10			
Ω	92	110	159			
$\hat{\Omega}$	1.00	1.06	1.31			
N	130	138	131			

Most of the symbols are defined in the text. $\hat{R}''$ is the ratio of R'' to that for the best model. $\hat{T}$ is the ratio of $T^{1/2}$ to that for the best model

in steric compression in the two packing models because there is minimal interaction between the molecules. The PS79 results also give good discrimination for the X-ray data (>0.995 confidence based on $\hat{R}''$). In this case, the non-bonded energy (T) is lower for the best model, but this is less significant than for the X-ray data. The values of the R-factor ratios $\hat{R}''$ show that the level of discrimination is similar for the two refinement systems.

COMPARISON OF STRUCTURES

Given that the LALS and PS79 refinement systems both predict the same type of model, we now examine the details of the final refined structures. These results are summarized in *Table 2*. The simplest measure of the similarity between two structures is the mean difference in the atomic positions. The value of 0.09 Å indicates excellent agreement for structures determined from fibre diffraction data. The maximum value of 0.21 Å is for one of the hydrogen atoms attached to C6. This is not surprising because this is the most flexible group in the structure. Location of important interactions (such as hydrogen bonds) within and between molecules requires a precision in atomic positions of about 0.1–0.2 Å so the discrepancies observed here are acceptable.

The mean and maximum differences in the ring bond angles are less than one and two standard deviations, respectively, of the variations seen in crystal structures containing pyranose rings[17]. The mean and maximum differences in the ring conformation angles are both less than one standard deviation of the values observed in crystal structures[17]. The glycosidic bond angles are practically identical for the two refinement systems. The glycosidic conformation angles are more flexible but vary by only 3–4° between refinement systems. The primary hydroxyl conformations are still more flexible (although they are restrained by the hydrogen bonds with which they are involved), but deviations of no more than 4° are observed.

Errors in the parameters can be estimated from the covariance matrix during least squares refinement. For the LALS refinement, the average standard deviation of the parameters are 1.2° for the endocyclic bond angles, 1.8° for the ring conformation angles, 2° for the glycosidic conformation angles, 3° for θ_{O6}, 3° for μ and 0.01 Å for Δw. These values are generally consistent with the differences listed in *Table 2*. The rms uncertainty in the atomic coordinates can also be estimated and the average value is 0.2 Å for the LALS refinement. This is somewhat larger than what might be expected compared to the average deviation of 0.09 Å between the two refinement systems. The larger value is primarily due to larger uncertainties in the coordinates parallel to the b cell-edge since there are few contacts (except the O3---O6 hydrogen bond) along this direction and small movements within the hydrogen bonded sheets has little effect on the stereochemistry or calculated X-ray amplitudes.

The average difference in the hydrogen bond lengths between the two refinements is 0.07 Å. Both systems give interchain hydrogen bonds that are consistently longer (by about 0.3 Å) than the intrachain hydrogen bonds, so that this feature appears to be significant. The variation in lengths of the two different intramolecular hydrogen bonds are not the same in the two studies, indicating that these differences are not significant. This seems reasonable since the maximum difference amongst them is only 0.17 Å. Variations in the O3---O5 hydrogen bond length are due primarily to the different glycosidic conformations, whereas the O2---O6 hydrogen bond length is also dependent on the rotation of O6. It is not surprising therefore that there is more variation in the length of the O2---O6 than the O3---O5 hydrogen bonds between the two systems.

The packing parameters that describe the positions of the two chains in the unit cell are also listed in *Table 2*. The rotations of the two chains (measured as the angular coordinate of C6) vary on average by only 1°, and the relative rotations ($\Delta\mu$) by only 2° between refinement systems. The relative translations between the two chains differ by only 0.05 Å.

Table 2 Comparison of stereochemical features (in Å and degrees) of ramie cellulose I crystal structures obtained using the LALS and PS79 systems

Parameter	LALS	PS79
Mean difference in atomic positions	0.09	
Maximum difference in atomic positions	0.21	
Mean difference in endocyclic bond angles	1.1	
Maximum difference in endocyclic bond angles	2.2	
Mean difference in ring conformation angles	0.9	
Maximum difference in ring conformation angles	1.7	
C1–O1–C4 (1)	117.2	117.5
C1–O1–C4 (2)	117.2	117.1
ϕ_1	−89.9	−93.3
ψ_1	−150.3	−146.6
ϕ_2	−89.9	−94.0
ψ_2	−150.3	−146.5
$\theta_{O6}^{(1)}$	−64.5	−61.1
$\theta_{O6}^{(2)}$	−64.5	−59.3
O3---O5(1)	2.58	2.65
O3---O5(2)	2.58	2.66
O2---O6(1)	2.75	2.63
O2---O6(2)	2.75	2.64
O3---O6 (corner chains)	2.97	2.92
O3---O6 (centre chains)	2.89	2.89
μ_1 (corner chain)	96.2	96.3
μ_2 (centre chain)	100.6	98.7
$\Delta\mu$	4.4	2.4
Δw	−2.65	−2.60

(1) and (2) refer to the first and second residues in the chain. $\Delta\mu = \mu_2 - \mu_1$ and Δw is the z-coordinate of the centre chain minus that for the corner chain

CONCLUSIONS

Determination of polymer structures from fibre X-ray data requires a system for co-refining model structures against the X-ray data and stereochemical restraints. The widely used refinement systems that achieve this in an acceptably flexible manner are the LALS and PS79 systems. Good discrimination between alternative models is necessary for successful structure determination. The comparison of the two refinement systems described here shows that they produce very similar levels of discrimination between competing models. It also shows that the two systems produce very similar final models and that important stereochemical features are quantitatively consistent. Overall, this study lends confidence to determinations of precise polymer structures from fibre diffraction data. It is essential, however, that models are properly co-refined against stereochemical restraints and X-ray data, that different

terms are appropriately weighted, and that different models are compared in an objective manner.

ACKNOWLEDGEMENTS

We thank Dr Peter Zugenmaier for his comments on this manuscript, the US National Science Foundation for support (DMB-8606942 to RPM) and Deb Zerth for word processing.

REFERENCES

1 Arnott, S. in 'Fiber Diffraction Methods' (Eds. A. D. French and K. H. Gardner), *ACS Symp. Ser.* **141**, ACS, Washington, DC, 1980, p. 1
2 Millane, R. P. and Arnott, S. *J. Macromol. Sci. Phys.* 1985, **B24**, 193
3 Millane, R. P. in 'Crystallographic Computing 4: Techniques and New Technologies' (Eds. N. W. Isaacs and M. R. Taylor), Oxford University Press, Oxford, 1988, p. 169
4 Namba, K. and Stubbs, G. *Acta Crystallogr.* 1985, **A41**, 252
5 Arnott, S. and Wonacott, A. J. *Polymer* 1966, **7**, 157
6 Smith, P. J. C. and Arnott, S. *Acta Crystallogr.* 1978, **A34**, 3
7 Zugenmaier, P. and Sarko, A. in 'Fiber Diffraction Methods' (Eds. A. D. French and K. H. Gardner), *ACS Symp. Ser.* **141**, ACS Washington, DC, 1980, p. 225
8 Stubbs, G., Namba, K. and Makowski, L. *Biophys. J.* 1986, **49**, 58
9 Millane, R. P., Byler, M. A. and Arnott, S. in 'Supercomputer Applications' (Ed. R. W. Numrich), Plenum, New York, 1985, p. 289
10 Zugenmaier, P. and Sarko, A. *Biopolymers* 1973, **12**, 435
11 Zugenmaier, P. *Biopolymers* 1974, **13**, 1127
12 Zugenmaier, P. and Sarko, A. *Biopolymers* 1976, **15**, 2121
13 Box, M. J. *Comput. J.* 1965, **8**, 42
14 Woodcock, C. and Sarko, A. *Macromolecules* 1980, **13**, 1183
15 Millane, R. P. and Narasaiah, T. V. in 'Cellulose and Wood-Chemistry and Technology' (Ed. C. Schuerch), John Wiley and Sons, New York, 1989, p. 39
16 French, A. D., Roughead, W. A. and Miller, D. P. in 'The Structures of Cellulose: Characterization of the Solid States' (Ed. R. H. Atalla), *ACS Symp. Ser.* **340**, ACS, Washington, DC, 1987, p. 15
17 Arnott, S. and Scott, W. E. *J. Chem. Soc. Perkin II* 1972, 324
18 Hamilton, W. C. *Acta Crystallogr.* 1965, **18**, 502
19 Stubbs, G. *Acta Crystallogr.* 1989, **A45**, 254
20 Millane, R. P. *Acta Crystallogr.* 1989, **A45**, 258
21 Millane, R. P. *Acta Crystallogr.* 1989, in press