short communications

New developments of the *TWIN* algorithm for phase extension and refinement in disordered supramolecular structures

K. Bethanis,[a] **P. Tzamalis,**[a] **A. Hountas,**[a]* **G. Tsoucaris,**[b] **A. Kokkinou**[a] **and D. Mentzafos**[a]

[a]Physics Laboratory, Agricultural University of Athens, 75 Iera Odos, Votanikos, Athens 118-55, Greece, and [b]Laboratoire de Physique, Tour B, Centre Pharmaceutique, Université Paris Sud, 92290 Châtenay-Malabry, France. Correspondence e-mail: gphy2xoa@aua.gr

A new development of the *TWIN* algorithm is described and used for phase extension/refinement in supramolecular complexes. A small number of phased reflections at low resolution are sufficient for the quasi-automated determination of all atomic coordinates, including disordered atoms.

1. Introduction

In the present short note, we examine a problem often encountered in supramolecular (SM) structures where an inherent disorder impairs the determination of the positions of the sites of all atoms. Wide experience has shown that this complete determination, if finally successful at all, requires lengthy and cumbersome calculations. In these problems, however, a small number of reflections are often approximately phased *a priori* by (insufficiently) isomorphous replacement, but only at low (2.5 Å) resolution. Thus, this problem is not an *ab initio* one but a *phase extension/refinement* one, specific for the disorder case; yet a quasi-automatic algorithm is desirable. Note that the recent developments in direct methods enable one now to successfully attack *ab initio* structure determination for 1000–2000 atoms *via* the algorithms *Shake-and-Bake* (Weeks *et al.*, 1994) and *half-baked* (Sheldrick & Gould, 1995).

The twin algorithm has been recently upgraded and used for protein phase extension from a small set of 200 reflections at low resolution (LR) to a large one of 10000 reflections at high resolution (HR) (Bethanis *et al.*, 2000). Here we have adapted the method to handle SM structures where the guest molecules, the water molecules (filling the space between and within the host molecules) and some of the host atoms are disordered. The convergence of the procedure is highly speeded by using alternately the *TWIN* and the *SHELXL97* (Sheldrick, 1997) algorithms.

A main theoretical innovation of the *TWIN* algorithm consists of the use of an *auxiliary* function

$$\psi(r) \underset{\text{FT}}{\longleftrightarrow} \Psi(H)$$

endowed with a clear *quantum-mechanical connotation as a wave function*. In fact, the algorithm searches for a function $\psi(r)$ such that

$$\rho(r) \underset{\text{FT}}{\longleftrightarrow} E_H$$

obeys the experimental data E_H^{obs}. We write the fundamental relations linking the usual E set with the 'wave function Ψ set' (Hountas & Tsoucaris, 1995):

$$E_H = \sum_K \Psi_K (\Psi_{K-H})^* \underset{\text{FT}}{\longleftrightarrow} \rho(r) = |\psi(r)|^2 \qquad (1)$$

$$\Psi_K = \sum_H E_H \Psi_{K-H}. \qquad (2)$$

The alternation of calculations by (1) and (2) is continued until the cut-off value of the symmetry test reaches a satisfactory value. The *TWIN* procedure is entirely automated thanks to the adaptation of the pertinent parameters and level of tests allowing the optimum utilization of the theoretical background. In particular, the use of the Ψ set provides a new test based upon the symmetry of the space group. This symmetry, although assumed known, **is not introduced** in the initial Ψ set. Thus, this initial Ψ set has the general form of $P1$ symmetry, with the Friedel condition $(\Psi_K)^* = \Psi_{-K}$ and the corresponding phases are set randomly. However, as soon as the phase-determination procedure is developing, the Ψ set is gradually modified and, at the same time, it gradually acquires the correct symmetry. This progressive symmetry acquisition is used as a diagnostic test for the correctness of the phases generated by the *TWIN* algorithm.

A new area of application appears with the SM problems where a rough estimation of the position and orientation of the host molecule is a rather standard crystallographic problem, while the determination of the atomic positions of the guest and water molecules is often a difficult problem on account of their low values of occupancy factors associated with large thermal movements and statistical disorders. This short note describes the necessary developments of the algorithm and its successful application to the structure of a cyclodextrin (CD) host–guest compound: β-CD–indole-3-butyric acid complex (102 independent non-hydrogen atoms). The available data extend to 1.05 Å resolution and comprise 3146 unique observed reflections in $C2$ space group.

2. Transfer of approximate and incomplete isomorphous structural information into reciprocal space

The values of the initial coordinates borrowed from a known very roughly isomorphous structure are, in general, corrupted by an error that can be considerable. However, an envelope of the host CD molecule could be obtained from these coordinates, containing the

necessary information to start the procedure. This hypothesis led us to calculate as usual the structure factors from the initial coordinates of the CD molecule, but to retain only the LR ones, rejecting the HR E's, which would mainly contain noise. Thus, we are able to handle the poor information of the initial coordinate set, which consists of the approximate positions of the sole CD host atoms without water and guest molecules. In this way, the large initial error associated with atomic positions is now reflected in the LR of accepted E's to be introduced as initial information. In the present case, the actual number of symmetry-independent initial phases is only 186 at 2.5 Å resolution.

3. The use of the SR wave function

In the *TWIN* algorithm, the Ψ set is arbitrarily defined by the user and it can be extended as well *beyond the observed data resolution*. In other words, the Ψ set can be endowed 'by construction' from the very beginning with the resolution considered by the user as appropriate for a particular problem. We have tentatively used here a Ψ set of 0.8 Å resolution. This generates 4000 Ψ's (Friedel independent) within the shell $0.8 < d < 1.05$ Å, in addition to the 5191 Friedel independent Ψ set associated with Miller indices within the observed sphere at 1.05 Å resolution. Thus, the determination of the phases of E's is under the control of a total number of 9191 independent Ψ's.

We propose to associate the term '*super-resolution wave function*' with the above Ψ set, which comprises a significant part located beyond the observed resolution sphere. By introducing this Ψ set into the fundamental equation (1), we are now able to enter HR contributors $\Psi_K(\Psi_{K-H})^*$ into the E set (K and/or $K - H$ are located beyond the sphere). This allows the transfer of HR information into the phases of the observed structure factors. Then the whole new E set generates in turn new values of the Ψ set (including the SR wave function) through the regression equation (2), and so on. In the present work, the E set is the observed one with $d > 1.05$ Å. Note, however, that the same SR Ψ set could be used to generate phases of E's beyond the observed sphere as well. The latter, known as 'super-resolution E', have been tentatively calculated by other methods (Collins, 1982).

A further adaptation concerns the minimum value of E's used throughout the procedure. The nature of the problem, focusing on rather small contributions of the water and guest molecules, suggested setting a very low cut-off value for the structure factors, both those observed and those calculated with (1): E^{obs}_min $= 0.32$;
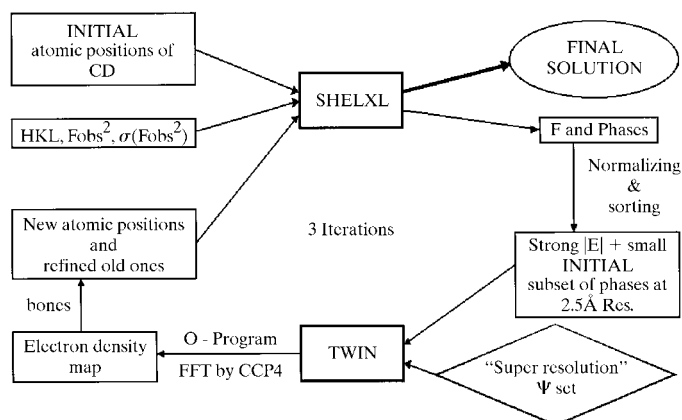
E^{calc}_min (AMIN in Bethanis *et al.*, 2000) varies from 1.6, at the beginning of the procedure, to 0.3 at the end.

4. Combining the *TWIN* and *SHELX* algorithms

At the end of the first run of the above version of the *TWIN* algorithm, the new phases revealed four additional atoms (water molecules and guest atoms); the positions of the CD atoms have been refined as well. The new list of coordinates is introduced in the *SHELXL* algorithm, which provides new and/or improved coordinates; these in turn led to new values of phases for the 186 reflections at 2.5 Å resolution to be introduced in a new run of the *TWIN* algorithm. Three cycles of alternate *TWIN* and *SHELXL* algorithms have been performed and have finally shown 17 water molecule sites and all 15 atoms of the guest molecule, with a final symmetry mean phase error S_MPE $= 5°$. Fig. 1 shows the flow chart of the procedure within the three iterations. Fig. 2 shows (*a*) the initial and (*b*) the final electron-density maps. The final R factor (anisotropic refinement) is 0.12, which is about normal for this type of supramolecular compound on account of the usually occurring disorder. Note also that the occupancy factor of the guest molecule is about 0.5. The electron-density map of the guest molecule is given in Fig. 3. The final result has shown that the use of a *super-resolution wave function* Ψ set (9119 Friedel independent Ψ's) that is considerably larger than the observable E set (5191 reflections) has been very efficient for the determination of all atomic positions of the host, guest and water molecules.
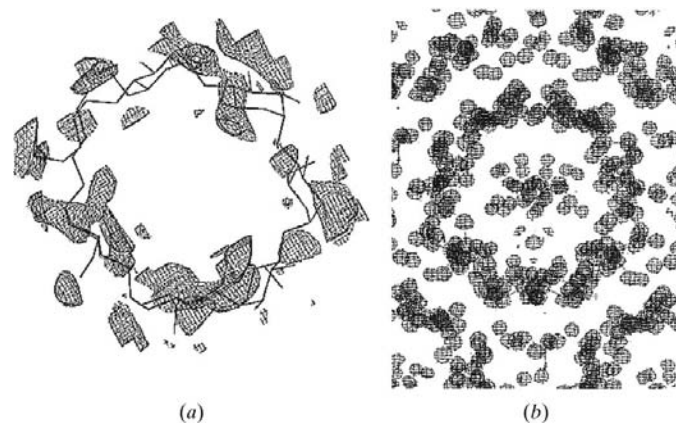

Figure 2
(*a*) Electron density corresponding to the initial set of 186 reflections at 2.5 Å resolution. (*b*) Final electron density plotted with observed moduli at 1.05 Å; the phases are calculated by the *SR wave function* extended set at 0.8 Å.
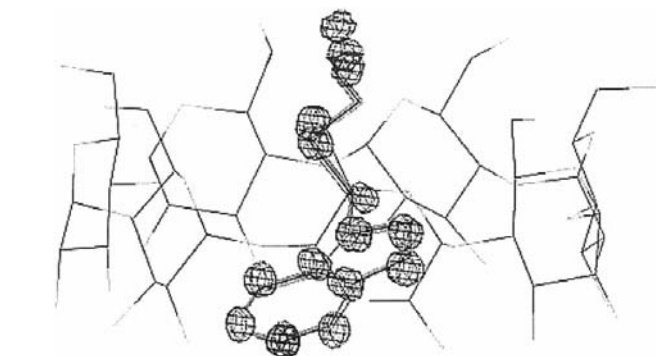

Figure 1
Flow chart of the whole procedure.

Figure 3
Electron-density map of the guest molecule at the end of the procedure.

short communications

Summarizing, a novel adaptation in the *TWIN* algorithm is proposed for host–guest compounds. It is characterized by two main innovations:

(i) The number of initial phases is severely limited by the resolution criterion: this achieves a transfer of 'lack of information' from direct to reciprocal space.

(ii) The 'wave function SR' technique can usefully produce about as many Ψ's (Fourier coefficients of the wave function) outside the observed sphere as those associated within the observed sphere.

In addition, new values for several parameters (cut-off values) have been necessary to handle SM problems. Finally, the convergence towards the correct solution is enhanced by the performance of a small number of sequences of the *TWIN* and the *SHELXL* algorithms.

References

Bethanis, K., Tzamalis, P., Hountas, A., Mishnev, A. & Tsoucaris, G. (2000). *Acta Cryst.* A**56**, 105–111.
Collins, D. M. (1982). *Nature (London)*, **298**, 49–51.
Hountas, A. & Tsoucaris, G. (1995). *Acta Cryst.* A**51**, 754–763.
Sheldrick, G. M. (1997). *SHELXL97: Program for the Refinement of Crystal Structures.* University of Göttingen, Germany.
Sheldrick, G. M. & Gould, R. O. (1995). *Acta Cryst.* B**51**, 423–431.
Weeks, C. M., De Titta, G. T., Hauptman, H. A., Thuman, P. & Miller, R. (1994). *Acta Cryst.* A**50**, 210–220.