Chapter 18.7. The TNT refinement package

D. E. Tronrud and L. F. Ten Eyck

18.7.1. Scope and function of the package

TNT (Tronrud et al., 1987) is a computer program package that optimizes the parameters of a molecular model given a set of observations and indicates the location of errors that it cannot correct. Its authors presume the principal set of observations to be the structure factors observed in a single-crystal diffraction experiment. To complement such a data set, which for most macromolecules has limitations, stereochemical restraints such as standard bond lengths and angles are also used as observations.

A molecule is parameterized as a set of atoms, each with a position in space, an isotropic $B$ factor and an occupancy. The complete model also includes an overall scale factor, which converts the arbitrary units of the measured structure factors to e Å$^{-3}$, and a two-parameter model of the electron density of the bulk solvent.

Because a TNT model of a macromolecule does not allow anisotropic $B$ factors, TNT cannot be used to finish the refinement of any structure that diffracts to high enough resolution to justify the use of these parameters. If one has a crystal that diffracts to 1.4 Å or better, the final model should probably include these parameters and TNT cannot be used. One may still use TNT in the early stages of such a refinement because one usually begins with only isotropic $B$'s.

At the other extreme of resolution, TNT begins to break down with data sets limited to only about 3.5 Å data. This breakdown occurs for two reasons. First, at 3.5 Å resolution, the maps can no longer resolve β-sheet strands or α-helices. The refinement of a model against data of such low resolution requires strong restraints on dihedral angles and hydrogen bonds – tasks for which TNT is not well suited. Second, the errors in an initial model constructed with only 3.5 Å data are usually of such a magnitude and quality that the function minimizer in TNT cannot correct them.

18.7.2. Historical context

The design of TNT began in the late 1970s, and the first publishable models were generated by TNT in 1981 (Holmes & Matthews, 1981). Its design was greatly influenced by observations of the strength and weaknesses of programs then available.

The first refinement of a protein model was performed by Jensen and co-workers at the University of Washington (Watenpaugh et al., 1973). This structure refinement was atypical because of the availability of high-resolution data. The techniques of pre-least-squares small-molecule refinement were simply applied to this much larger model. Since many of the calculations were performed manually, no comprehensive software package was created for distribution.

It was quickly realized that for macromolecular refinement to become common, the calculations had to be fully automated and ideal stereochemistry had to be enforced. In the late 1970s, four programs became available, all of which automated the refine- ment calculations, but each implemented the enforcement of stereochemistry in different ways. They were PROLSQ (Hendrickson & Konnert, 1980), EREF (Jack & Levitt, 1978), CORELS (Sussman et al., 1977) and FFTSF (Agarwal, 1978). PROLSQ was, ultimately, the most popular.

At one end of the spectrum lay FFTSF. This program optimized its models to the diffraction data while completely ignoring ideal geometry. Following a number of iterations of optimizing the fit of the model to the structure factors, the geometry was idealized by running a separate program. At the other extreme was CORELS. It optimized its models to the diffraction data while allowing no deviations from ideal stereochemistry. The model was allowed to change only through the rotation of single bonds and the movement of rigid groups. Both approaches were frustrating to a certain extent. With FFTSF it was a struggle to find a model that agreed with all observations at once. With CORELS it was difficult to get the model to fit the density, because small and, apparently, insignificant deviations from ideality often added up after many residues to large and significant displacements, and these were forbidden. Neither approach to stereochemistry seemed very convenient, although CORELS was used for early-stage refinement for many years because of its exceptional radius of convergence.

Both PROLSQ and EREF enforced ideal stereochemistry and agreement with the diffraction data simultaneously. This strategy proved very convenient and generated models that satisfied their users. The two programs differed significantly in the form in which they required the ideal values to be entered. PROLSQ required that the ideal values for both bond lengths and bond angles be entered as distances, e.g. an angle was defined by the distance between the two extreme atoms. EREF required that the standard value for an angle simply be entered as the number of degrees. Since EREF stored its library of standard values in the same terms as those with which people were familiar, it was much easier to enter the values.

These two programs differed in another way as well. PROLSQ stored ideal values for the stereochemistry of each type of residue (e.g. alanine, glycine etc.), while EREF parameterized the library in terms of atom types. For example, the angle formed by three atoms, the first a keto oxygen, the second a carbonyl carbon and the third an amide nitrogen, would have a particular ideal value regardless of where these three atoms occurred. In this matter, PROLSQ was more similar to the thought patterns of crystallographers.

18.7.3. Design principles

TNT was designed with three fundamental principles in mind. Each principle has a number of consequences that shaped the ultimate form of the package.

18.7.3.1. Refinement should be simple to run

The user should not be burdened with the choice of input parameters that they may not be qualified to choose. They also should not be forced to construct an input file that is obscure and difficult to understand. It is hard now to remember what most