

Nautilus software for automated nucleic acid building

K Cowtan

The 'nautilus' software is a new tool for automated building of RNA/DNA from electron density. It uses similar ideas to the 'buccaneer' software (Cowtan, 2008) for protein model building, but with a different and highly efficient target function for identifying nucleotide features.

The software will locate likely nucleotide features including sugars and phosphates, grow these into chains, merge overlapping chains, match the built chains to the sequence, and build the bases. The resulting structure is refined using 'refmac', and the calculation is iterated to obtain a more complete structure.

The software may be used to build nucleotide structure in experimentally phased maps, molecular replacement maps, or to add the nucleotide components to protein complexes.

The calculation consists of the following steps:

- Locate likely phosphate and sugar features in the electron density.
- Grow the features into chains.
- Merge overlapping chains.
- Join chains whose 3' and 5' ends are in close proximity.
- Rebuild the chains using fragments from the Richardson's database of well determined nucleotide structures.
- Match the chains against the known sequence.
- Build the bases and other peripheral atoms.

The calculation is notable for its speed, typically taking no more than a few tens of seconds. When combined with refinement in refmac (itself a fast package) (Murshudov et al, 1999), more than 90% of the time is spent in the refinement step. This speed is achieved through use a highly optimised 'fingerprint' for detecting structural features from the electron density values at a few highly informative points, which must have extreme density values if the feature is present in a given orientation. A fast rotation and translation search can be carried out using a method similar to the that employed by the ESSENS software (Kleywegt & Jones, 1997). The fingerprint for a sugar group is shown in figure 1.

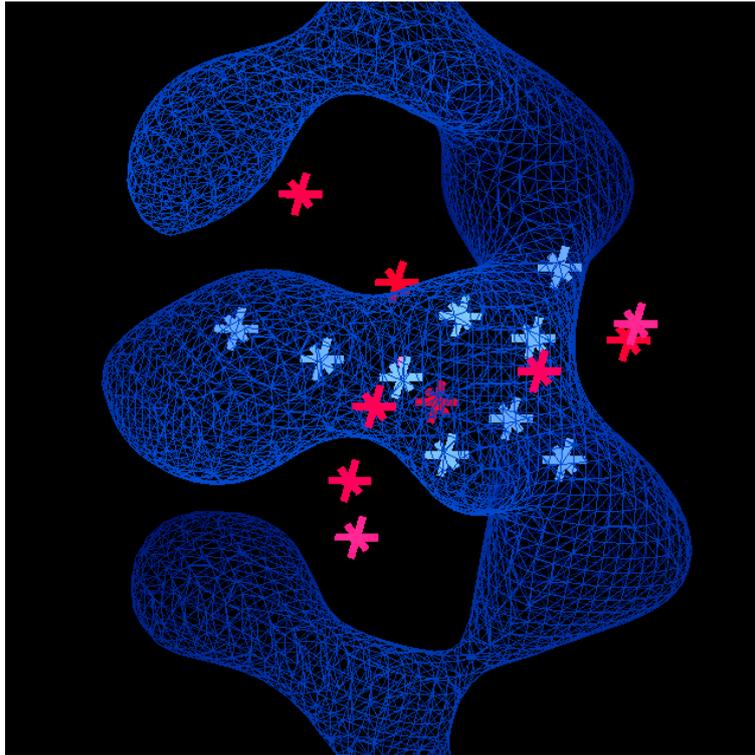


Figure 1: Nautilus sugar fingerprint. Blue crosses are locations where high density is required, red crosses are locations where low density is required.

Similar fingerprints are used to identify phosphates and to distinguish between different base types.

Version 0.3 of Nautilus is included in CCP4 version 6.3.0. The pipeline may be run through the CCP4i graphical user interface, or from the command line as either a build/refine pipeline or for a fast build only. In future an interactive version of some of the functionality will be available in the 'Coot' model building software (Emsley et al, 2010).

This article may be cited freely.

References

- Kleywegt, G. J. & Jones, T. A. (1997). Acta Cryst. D53, 179-185.
- Murshudov G. N., Lebedev A., Vagin A. A., Wilson K. S. and Dodson E. J. (1999). Acta Cryst. section D55, 247-255.
- Cowtan, K. (2008). Acta Cryst. D64, 83-89.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). Acta Cryst. D66, 486-501.