



AUTOMATIC BUILDING OF PROTEIN AND NUCLEIC ACID STRUCTURES

F. Pavelčík^{1,2}

¹Department of Chemical Drugs, FaF VFU, Brno

²Department of Inorganic Chemistry, PRIF UK, Bratislava

Email: pavelcik@vfu.cz, pavelcik@fns.uniba.sk

Keywords:

Model building, RNA, Proteins, Conformation families, Electron density

Conformation family

A concept of conformation family is useful for classification of proteins and nucleic acid structures, for prediction of protein structures, various model building programs, and as a tool for protein structure verification. There is no clear definition of conformation family in the literature. Clusters of similar structures usually represent conformation family. In our concept the conformation family is a region of a conformation space highly populated with experimental conformations. The smoothed conformation density in this region should have a local or global maximum. The conformation space is an infinite periodic torsion angle space. The conformation families were determined by direct multidimensional mapping (2-D, 4-D, and 6-D).

Conformation families were determined for di-peptides, tri-peptides, tetra-peptides and penta-peptides. These are related to model building fragments called AlphaD, AlphaT, AlphaQ, and AlphaP. A method of Pavelcik & Vanco [1] was used. All PDB structures (Febr. 2007) with resolution better than 1.5 Å, and 90% homology criterion were selected for analysis. The number of calculated torsion angles was almost 500 000. The grid of mapping was 16. The search probe was variable: R=RD 0.5N; RD is empirically found radius for 2-D search, N is dimension of the conformation space. Penta-peptide conformations (8-D) were generated by a combination of two tri-peptide conformation families. Less populated families were removed. The number of conformation families for di-peptides is 6, for tri-peptides 24-26, and 130-140 for tetra-peptides. The families were classified according to positions in the Ramachandran map and studied for the purpose of automatic protein model building. Similar approaches were used also for a verification of conformation families of mononucleotides (phosphate-to-phosphate type).

Model building

A method for automatic building of bio-macromolecular structures has been developed. Individual molecular frag-

ments (AlphaT and AlphaQ) are located in an electron density map by a phased rotation conformation and translation function, as implemented in the program NUT [2]

The protein fragment connecting procedure was tested on a green fluorescent protein. The crystal structure with PDB code 1EMB was selected for the testing. The PDB file of the structure was created with a program HEL and the structure building was analyzed in details with respect to sequence, secondary structures, loops and cis peptide bonds. The results are compared with model building techniques based on mono-peptides and di-peptides. The novel method is a promising tool for building low resolution protein structures.

The same method was applied to building of DNA and RNA structures. The fragments are RNAbone and DNAbone, which represent mononucleotide of phosphate-sugar-phosphate type. Fragments are flexible and all backbone torsion angles can be varied during the search. For computation reasons the search is restricted by a table of allowed conformations (conformation families). The RNAbone/DNAbone is suitable for intermediate resolution (2.0 - 3.0 Å). Individual fragments are connected into polynucleotide chains by a program HEL. In the case of RNAbone/DNAbone also side-chains can be built. The PDB file is result of connecting. The procedures were tested on RNA/DNA structures ranging from small nucleotide (1QYL) to ribosome (1FFK, 1J5E). About 70 % to 90 % of the structure can be built, depending on resolution, fragment used and phase quality. A rigid double helical fragment NAhelix of 90 atoms is used to locate stretches of regular A-RNA (DNA) structures. The position and orientation of the fragment can be refined (but not conformation). About 50-70 % of the structure can be located in typical RNA structures. NAhelix is suitable for low-resolutions (3-4 Å).

References

1. F. Pavelcik; J. Vanco, *J. Appl. Cryst.*, **39**, (2006), 315-319.
2. F. Pavelcik., *J. Appl. Cryst.*, **39**, (2006), 483-486.