Medium to high resolution X-ray structures of DNA and RNA molecules were investigated to find geometric properties useful for automated model building in crystallographic electron density maps. We describe a simple method, starting from a list of electron density “blobs”, to identify backbone phosphates and bases based on properties of the local electron density distribution. We have used this knowledge to propose an algorithm for the automated building of nucleic acid models into electron density maps. The algorithm is based on distances and angles involving C1' and the phosphorus atoms and involves the pseudo-torsion angles η' and θ' that describe the ...P-C1'-P-C1'- ... chain. These quantities show reasonably narrow distributions and an asymmetry that allows the direction of the phosphate backbone to be established.

Keywords: auto-building, nucleic acids, backbone trace

Comparison of biomolecules in 3d is a common task, routinely met in various fields of structural biology and protein crystallography. Well-known examples include the inference on protein function from similarity to structures with known function, prediction of binding sites, choice of sequence-remote models for molecular replacement and others. Due to the large size of protein structures, their comparison often starts with the detection of similarity between their simplified representations, such as secondary structure topology or 3d graphs of secondary structure elements. CCP4 Suite of Programs for Protein Crystallography includes SSM (Secondary Structure Matching), a protein structure aligner, which is built on these principles [1]. SSM is widely used and recognized for speed and efficiency (see, e.g., [2]). However, SSM does not work if secondary structure cannot be calculated, which is often the case in crystallographic applications, when refinement is not complete, or if protein chain appears to be fragmented. An alternative approach to protein structure alignment and comparison is proposed, which does not require the initial structural simplifications and, therefore, is free from SSM shortcomings. Instead, the structures are represented as a manifold of overlapping short fragments, which are clustered by their rotation-translation function of best superposition. The final solution is then chosen as one with the maximal Q-score [1], from the set of top-populated clusters after an additional refinement on Ca-level. The procedure is known to be computationally hard, yet it was developed to match the celebrated SSM performance. Analysis of the new algorithm’s performance, sensitivity and selectivity is presented.

Keywords: Protein Structure, Structural Alignment, Structure Comparison