NucleoFind: A Deep-Learning Network for Interpreting Nucleic Acid Electron Density

J. Dialpuri[1], J. Agirre[1], K. Cowtan[1], P. Bond[1]

[1] *York Structural Biology Laboratory, Department of Chemistry, University of York, York, UK*

jordan.dialpuri@york.ac.uk

Nucleic acid electron density interpretation after molecular replacement remains a difficult problem for computer programs to deal with. Programs tend to rely on time-consuming and computationally exhaustive searches to recognise characteristic features. We present NucleoFind, a deep-learning-based approach to interpreting and segmenting electron density. Using an electron density map from X-ray crystallography, the positions of the phosphate group, sugar ring and nitrogenous base group can be predicted with high accuracy. On average, 84 % of phosphate atoms, 86 % of sugar atoms and 90 % of base atoms are located by the models in maps produced following protein molecular replacement. The wealth of context these predicted maps provide can then be used to automatically build more accurate and complete nucleic acid molecular models in a short amount of time.

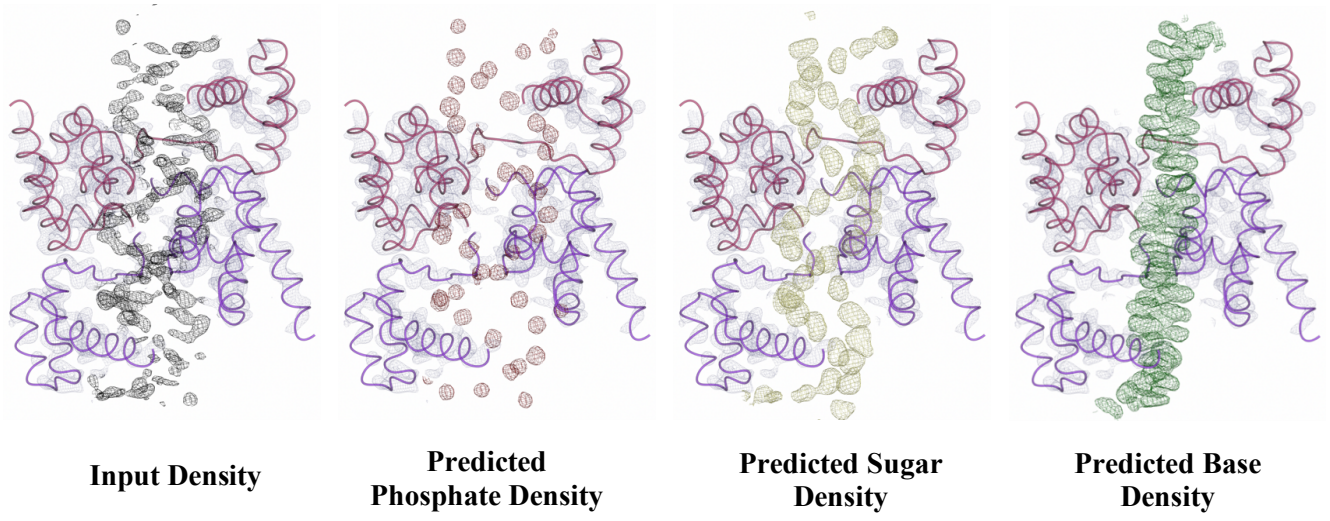

Figure 1. Output of all three deep-learning models corresponding to phosphate group, sugar group and base group predictions. To generate the input density, molecular replacement was performed on a POU DNA binding domain, PDB Code: 3L1P [1]. The input density, shown in black at 1.5 σ can be seen as the characteristic DNA duplex, however, the density is noisy and discontinuous which often causes automated model-building software packages to struggle with locating features. NucleoFind can highlight the phosphate, sugar and base positions well from the input density, highlighting the usefulness of the program as post molecular replacement tool.

[1] Esch, D., Vahokoski, J., Groves, M. et al. A unique Oct4 interface is crucial for reprogramming to pluripotency. Nat Cell Biol 15, 295–301 (2013).