BIOLOGICAL
CRYSTALLOGRAPHY

Volume 71 (2015)

Supporting information for article:

The solvent component of macromolecular crystals

Christian X. Weichenberger, Pavel V. Afonine, Katherine Kantardjieff and Bernhard Rupp

Since the *MATTPROB* approach is based on statistical analysis of existing data, we have put emphasis on the type of bio-macromolecule that is most prevalent in the PDB. The Matthews probability calculator has been constructed on a non-redundant data set consisting of 60,218 protein structures, 998 nucleic acid structures and 2,414 structures of protein/nucleic acid complexes. Resolution *independent* probability predictions are available in the MATTPROB program also for protein/nucleic acid complexes and nucleic acid structures, but a *resolution-dependent prediction is not reliably possible* in these cases due to the low number of available structures.

A reviewer suggested that the manuscript could contain more information on solvent content prediction of nucleic acids. We decided to provide the corresponding figures for RNA/DNA – along with the appropriate caveats – as supplemental material. References to these supplemental figures are made in the respective protein figure legends of the main manuscript.

We have re-plotted Figure 2 with 998 models of nucleic acid structures. A dependency of V_S on resolution as observed in protein structures cannot be derived from the distribution shown in Supplemental Figure 1, which can be reasoned by the currently low number of available data points. The future will show if an increase of DNA/RNA structures deposited with PDB eventually establishes a similar trend as observed with protein crystal structures.

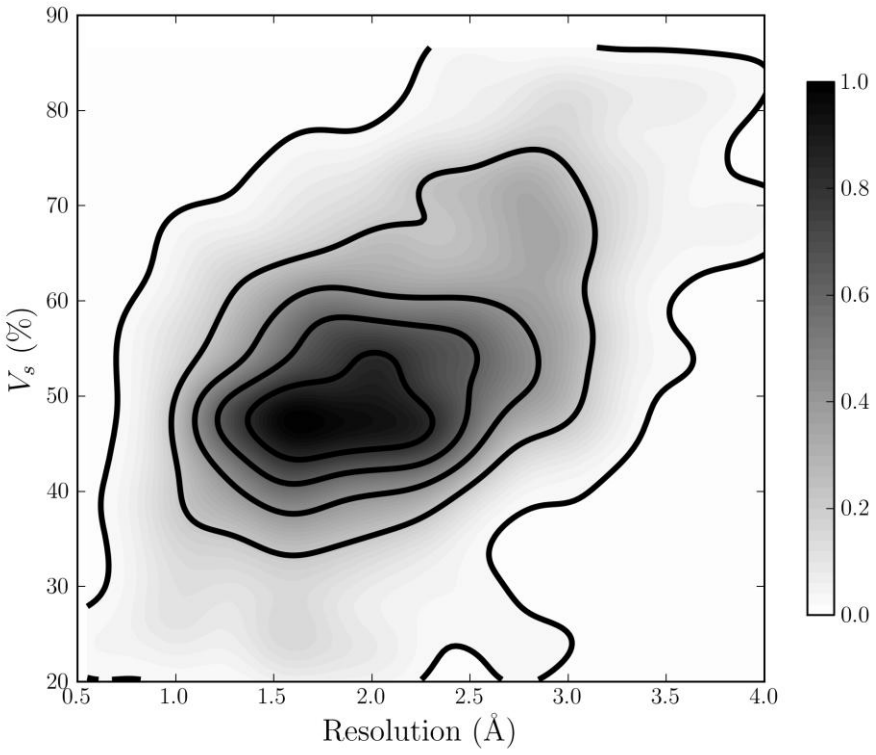

Figure S1 Similar plot as Figure 2 in the main manuscript but with 998 DNA and RNA structures. Half of the nucleic acid structures have been determined at resolutions between 1.6Å and 2.5Å. The density for this range shows little dependence of V_S on the resolution, rather a uniform solvent content of slightly below 50%.

We have also plotted Figure 6 with data for nucleic acid structures. The nature of these macromolecules is rather different compared to protein molecules. Nucleotides have a larger number of atoms per residue; the crystallized chains are rather short; and in crystal structures they arrange often linearly, lacking a hydrophobic core. Given these structural differences and the low number of deposited DNA/RNA structures we caution against a corresponding 'rule of thumb' and prefer to omit this figure from the main manuscript. The suggested RNA/DNA plot corresponding to main Figure 6 is shown in Supplemental Figure 2.

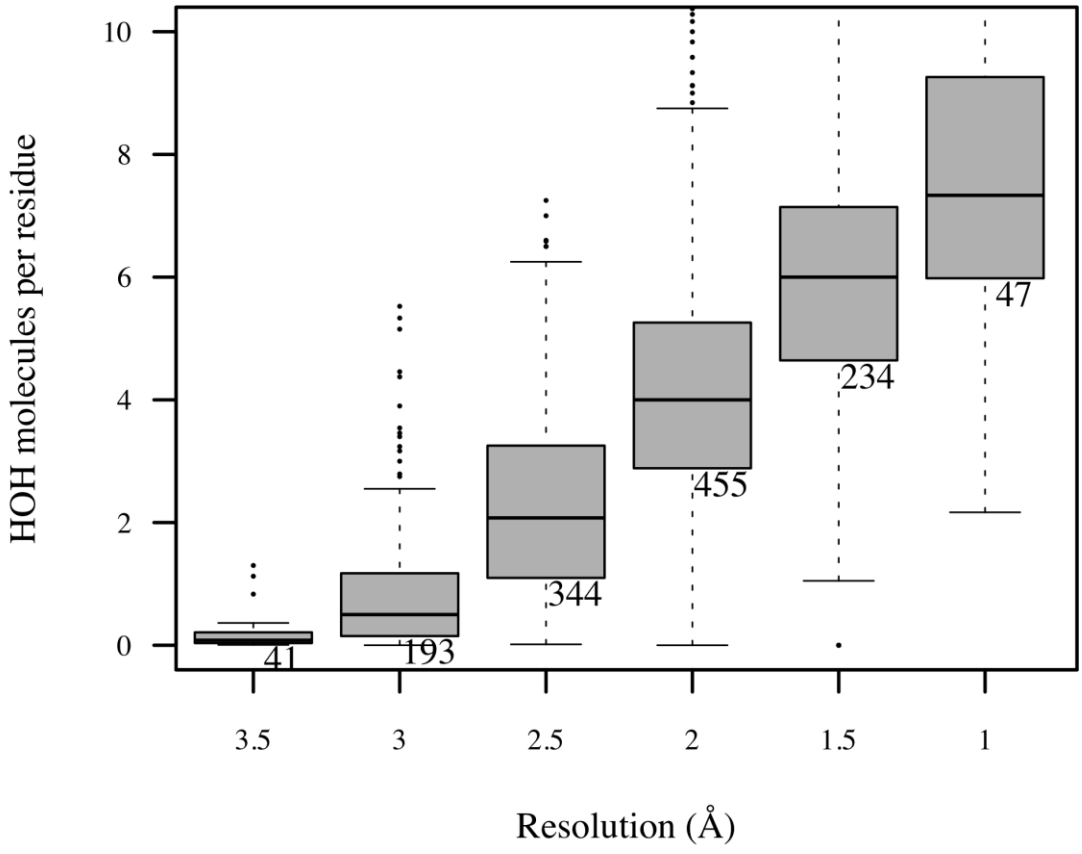

Figure S2 Manuscript Figure 6, for PDB DNA/RNA structures. Due to the low number of deposited structures, we have increased the binning interval to 0.5Å. Box plots are labeled with the number of DNA/RNA structures of the respective bin. The much higher variability of water molecules per residue is visible on the scale of the y-axis. For example, PDB entry 2dpb contains 236/24=9.83 waters per residue, whereas PDB entry 1z3f contains only 27/12=2.25 waters per residue. However, both structures have been refined to 1.5Å.