On the contribution of hydrogen atoms to X-ray scattering

Afonine, P.V. [1,#] and Adams, P.D. [1,2]

[1]*Lawrence Berkeley National Laboratory, One Cyclotron Road, MS64R0121, Berkeley, CA 94720 USA*
[2]*Department of Bioengineering, University of California Berkeley, Berkeley, CA, 94720, USA.*

[#]Correspondence email: PAfonine@lbl.gov

Approximately half of all atoms in bio-macromolecular structures are hydrogens. While X-ray diffraction data rarely allows direct determination of their positions, with a few exceptions the geometry of hydrogen atoms can be inferred from the positions of other atoms. Even though a hydrogen atom is a weak X-ray scatterer its contribution to the total scattering is not negligible (for a review see Afonine *et al.*, 2010). Until recently it was customary to ignore hydrogen atoms throughout the process of crystallographic X-ray structure determination. However, it has been demonstrated (Chen *et al.*, 2010; Headd *et al.*, 2009; Davis *et al.*, 2007; Word *et al.*, 1999) that using hydrogens in structure determination typically improves model geometry and highlights problems otherwise difficult to detect. In this article we illustrate the contribution of hydrogen atoms to calculated X-ray structure factors and *R*-factors.

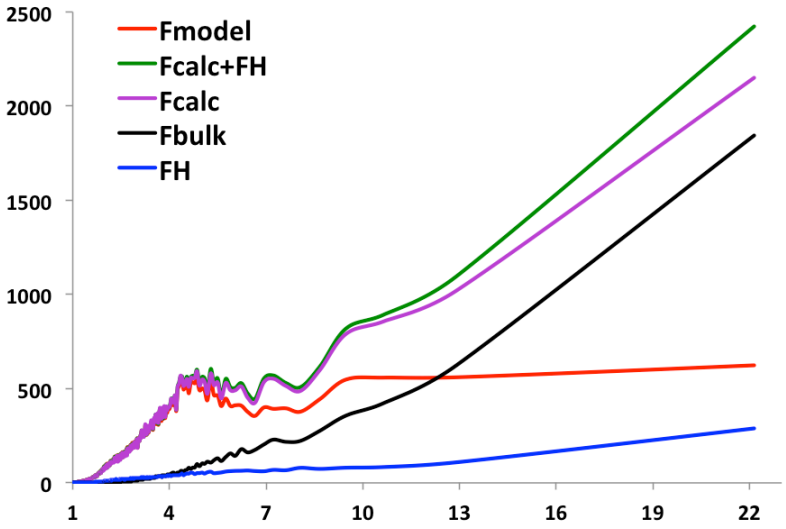

Figure 1. Resolution bin averaged values of each term in formula 1 calculated for a structure with code 1F8T using all theoretically possible reflections up to 1Å resolution.

Defining the total model structure factor (F_{model}) as the scaled sum of structure factors calculated from non-hydrogen atoms (F_{calc}), hydrogen atoms (F_H) and bulk-solvent (F_{bulk}) (Afonine *et al.*, 2005; Cooper *et al.*, 2010)

$$\mathbf{F}_{\text{model}} = k(\mathbf{F}_{\text{calc}} + \mathbf{F}_{\text{H}} + \mathbf{F}_{\text{bulk}}) \qquad (1)$$

allows us to illustrate the individual contributions in (1). Figure 1 shows resolution bin averaged values of each term of (1) calculated for a structure taken from the Protein Data Bank (PDB; code 1F8T) using all theoretically possible reflections to 1Å resolution. The bulk-solvent contribution was calculated as $F_{bulk} = k_{sol}\exp(-B_{sol}s^2/4)F_{mask}$ with $k_{sol} = 0.35 e/Å^3$ and $B_{sol} = 50 Å^2$ (for details see Afonine *et al.*, 2005), and k=1 since all the calculated terms in (1) are in absolute scale. As expected, the contribution from the hydrogen atoms (F_H; blue line) it not negligible compared to the total structure factor (F_{model}; red line). At low resolution F_{model} is significantly smaller than the components ($F_{calc} + F_H$) or F_{bulk} because even though the bulk-solvent contribution is large at this resolution, the bulk scatterers are out of phase with the protein scatterers and therefore the bulk-solvent contribution is out of phase with the other terms (Podjarny & Urzhumtsev, 1997). Our observation is that the plots in figure 1 are characteristic and do not vary significantly from structure to structure.

To illustrate the impact of hydrogen atoms on the crystallographic *R*-factor and how this depends on data resolution we selected approximately 250 structures from PDB. The structures were selected such that each of six resolution ranges (bins): 0-1, 1-1.5, 1.5-2, 2-2.5, 2.5-3 and 3-3.5Å contained approximately the same number of structures. Additional selection criteria aimed to select the best available structures and included 99% complete data across the whole resolution range, no twinning, *R*-factors lower than average, and minimal geometry violations (clashscore, C_β

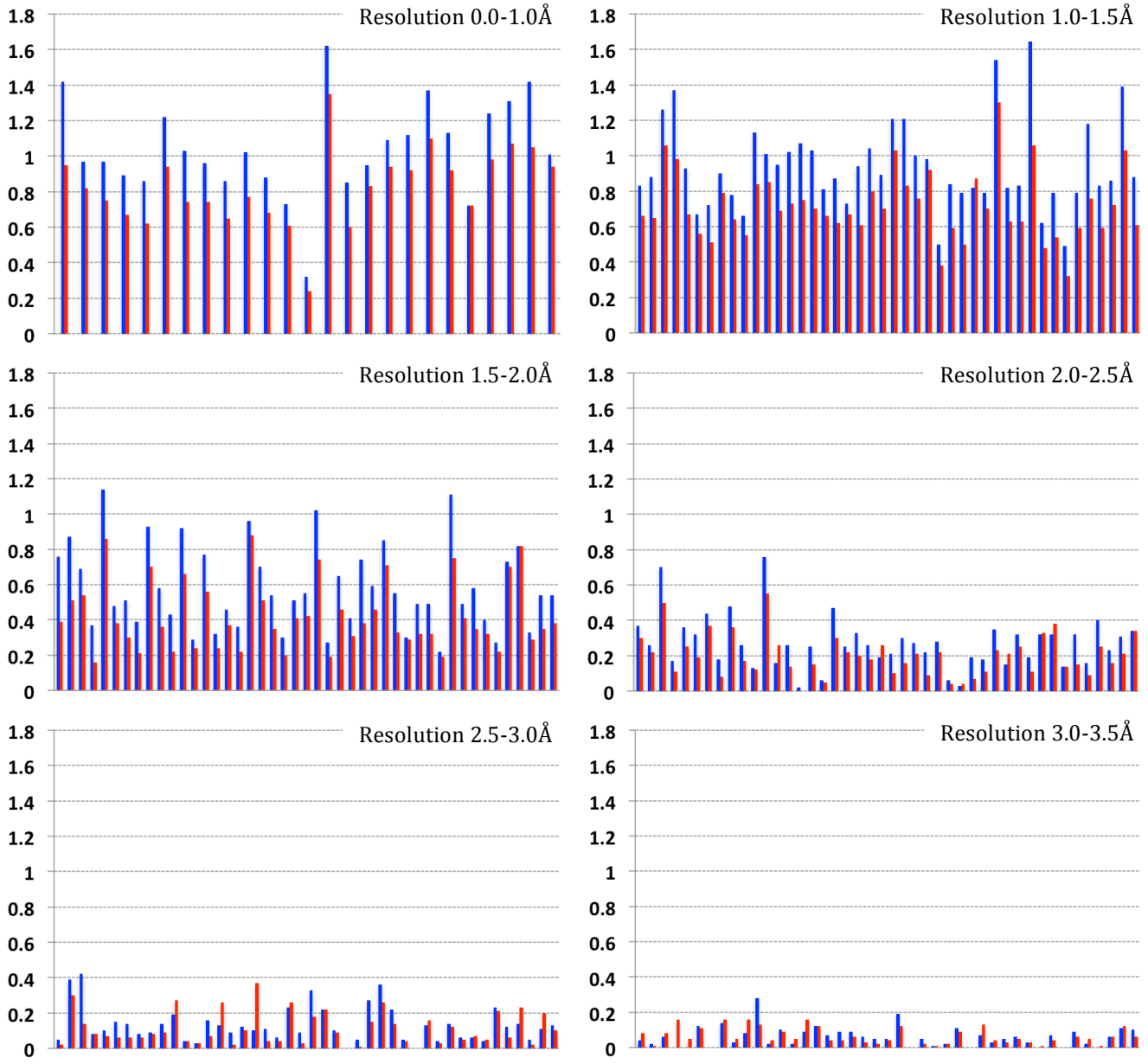

Figure 2. See text for details. R_{noH}-R_H(blue) and R_{noH}-R_{H_unique} (red). PDB codes of structures used in calculations in the order of their appearance in the plots; (0,1): *1mnz, 1nwz, 1r2m, 2ddx, 2e4t, 2fvy, 2gg2, 2gud, 2h5c, 2h5d, 2ov0, 2pne, 2ppp, 2rh2, 2zq7, 3agn, 3ago, 3cnj, 3f7l, 3gyi, 3gyj, 3l8w, 3mi4, 3noq, 3pyp*; (1,1.5): *1uww, 2eht, 2ivj, 2j23, 2j3p, 2jhq, 2jin, 2jju, 2o7i, 2qfe, 2r8o, 2rb2, 2rbo, 2rbp, 2rby, 2rc2, 2uxw, 2w47, 2xeu, 2z3h, 2zw0, 3a3v, 3ahs, 3b36, 3b3a, 3bc9, 3bne, 3ce1, 3deo, 3dxl, 3f9b, 3fm6, 3fwk, 3i3f, 3ioh, 3ivv, 3jsc, 3kwu, 3laa, 3m1z, 3m6b, 3mea, 3mnb, 3moy*; (1.5,2): *1lka, 1so3, 1t7t, 1vh2, 1xkg, 1y57, 1yb1, 1ze3, 2bik, 2d20, 2ei9, 2fp2, 2fxs, 2h9b, 2p5q, 2qvm, 2qwm, 2uya, 2ves, 2vun, 2wgb, 2wgp, 2wm3, 2wn2, 2x9w, 3a23, 3c67, 3cpg, 3d3i, 3ddw, 3dms, 3f3s, 3fj4, 3flu, 3fuy, 3g28, 3hbn, 3hbu, 3hfk, 3hgm, 3krr, 3kvc, 3l3v, 3m5v, 3m8u*; (2,2.5): *1o51, 1t73, 1w6w, 1x31, 2dbi, 2dso, 2fb0, 2hha, 2q36, 2qtb, 2uv2, 2vx0, 2vz6, 2w5o, 2wmr, 2wo3, 2wul, 3a7r, 3bbd, 3bbe, 3bkq, 3byi, 3dd3, 3dv5, 3e2k, 3e32, 3e87, 3g2f, 3gg3, 3haz, 3hd5, 3hk8, 3hy6, 3hzu, 3i3d, 3ib3, 3ic5, 3igu, 3iqa, 3l48, 3lnt, 3ls8, 3mtc, 3n51*; (2.5,3): *1e3h, 1h2n, 1o8c, 1yah, 1yc0, 1zkf, 2c44, 2ecf, 2f89, 2iun, 2izv, 2jas, 2qkx, 2rd5, 2vd5, 2vqa, 2vwk, 2w2c, 2wb7, 2zjy, 2zs9, 3a29, 3a2j, 3bic, 3c9l, 3csn, 3d2z, 3d8a, 3dd1, 3ffb, 3gn4, 3gxe, 3i4e, 3ibg, 3ihl, 3ir6, 3k5d, 3khj, 3llm, 3ly2, 3m4p, 3mg2, 3mle, 3n3k*; (3,3.5): *1b9x, 1l8h, 1n21, 1q9c, 1sjp, 1t7z, 1vg2, 1vg7, 1wdl, 1x03, 2a81, 2dgl, 2ffl, 2fqq, 2jjd, 2o0i, 2qqv, 2uy9, 2v8a, 2wdr, 2wdv, 2wfn, 2wuy, 2wy6, 2x8c, 2zrc, 2zrk, 3a2i, 3bbp, 3br1, 3dbc, 3dbd, 3dbe, 3dbf, 3ef7, 3ej1, 3fbn, 3gzp, 3hd7, 3hy5, 3ibp, 3krx, 3l2j*.

deviations, Ramachandran plot and rotamer outliers). For each structure we then computed three *R*-factors corresponding to a structure without hydrogen atoms (R_{noH}), a structure with all hydrogens added to expected positions (R_H) using the Reduce program (Word *et al.*, 1999) as

implemented in *phenix.reduce*, and a structure with hydrogens added to uniquely defined positions only (no hydrogens with rotational degrees of freedom; R_{H_unique}). Since *phenix.reduce* adds hydrogens to nuclear positions we re-optimized the X-H bond lengths (X is the heavy atom the hydrogen, H, is bonded to) such that the new hydrogen positions correspond to the electron cloud distance (for details see Afonine *et al.*, 2010). Since hydrogen atoms were added to already well refined structures we had to scale their contribution F_H to account for the fact that the refined ADPs of non-hydrogen atoms may have been inflated to account for absent hydrogens. This effect has been observed previously when anisotropic ADPs can model deformation density at ultra-high resolution (Afonine *et al.*, 2004). The scaling of F_H consisted of multiplying it by a resolution dependent factor $k_h \exp(-B_h s^2/4)$ with two refinable parameters k_h and B_h. Also, this scaling of F_H is intended to account for the effect of hydrogen atom abstraction (when applicable) described by Meents *et al.* (2009).

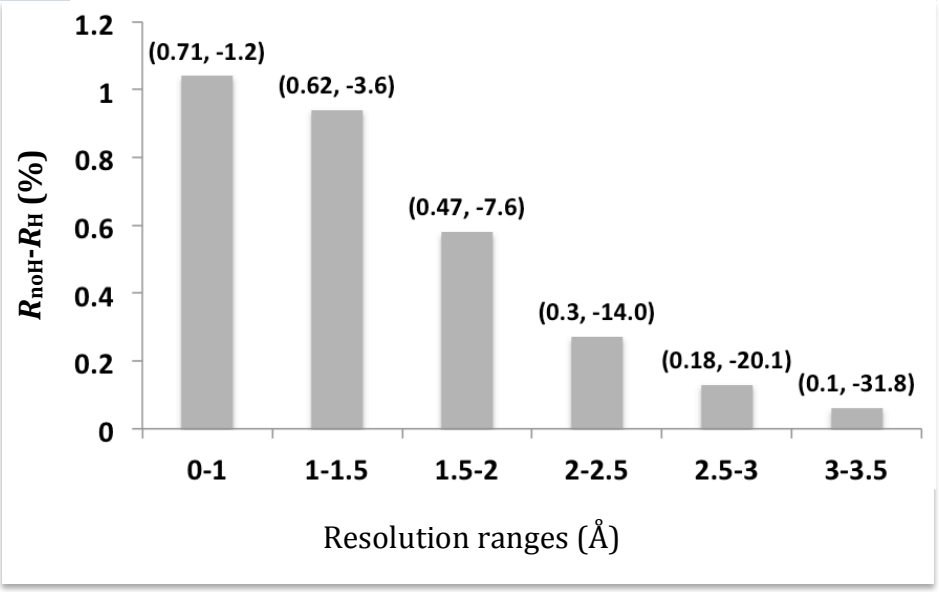

Figure 3. Averaged R_{noH}-R_H values shown for six resolution ranges. Each bar caption shows corresponding bin-averaged pairs of (k_h , B_h).

Figure 2 shows six plots corresponding to six selected resolution bins. Each plot presents two series of bars representing the *R*-factor differences R_{noH}-R_H (blue) and R_{noH}- R_{H_unique} (red) for each structure. It is clear that contribution of hydrogen atoms is non-zero across all six selected resolution ranges, and it ranges from an average of approximately 1% at highest resolution to about 0.04% at lowest resolution. In all cases adding hydrogen atoms improved the *R*-factors. Not including hydrogens with rotational degree of freedom almost always diminishes the *R*-factor improvement at resolutions up to 2.5Å and has a mixed effect at resolutions worse than 2.5Å. The decrease of *R*-factor improvement (R_{noH}-R_H) at lower resolution may have at least two explanations: 1) the positional error of non-hydrogen atoms is higher at lower resolution which in turn means a higher positional error for the hydrogen atoms and therefore less improvement in *R*-factor, and 2) as mentioned above the inflated ADPs of non-hydrogen atoms may already have partly compensated for the absence of hydrogens.

Figure 3 shows averaged R_{noH}-R_H values shown for six resolution ranges along with corresponding bin-averaged values of k_h and B_h. These values may be different for models where hydrogen atoms were used throughout the process of structure determination and refinement, as this may make the contribution of hydrogen atoms more distinct.

In summary, the contribution of hydrogen atoms to X-ray scattering is not negligible, hydrogens do contribute to the total model structure factor and we have illustrated how this affects the *R*-factors. The effect on the *R*-factor diminishes with resolution and could be the result of using well refined structures in our tests or/and at lower resolution the predicted positions of hydrogen atoms are less accurate. It is therefore possible that the effect on *R*-factor may be more significant if hydrogens were used throughout the process of structure determination and refinement. Finally, good quality structures at high resolution permit the inclusion of hydrogens possessing rotational degree of freedom. However this is not the case for lower resolution structures, and is likely also not the case for partially refined models.

References

Afonine, P.V., Mustyakimov, M., Grosse-Kunstleve, R.W., Moriarty, N.W., Langan, P. & Adams, P.D. (2010). *Acta Cryst*. D**66**, 1153-1163.

Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. (2005). *Acta Cryst.* D**61**, 850-855.

Chen, V.B., Arendall III, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. & Richardson, D.C. (2010). *Acta Cryst*. D**66**, 12-21.

Cooper, R.I., Thompson, A.L. and Watkin, D.J. (2010). *J. Appl. Cryst.* **43**, 1100-1107.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall III, W.B., Snoeyink, J., Richardson, J.S. and Richardson, D.C. (2007). *Nucleic Acids Res.* **35**, W375-W383.

Headd, J.J., Immormino, R.M., Keedy, D.A., Emsley, P., Richardson, D.C. and Richardson, J.S. (2009). *Journal of Structural and Functional Genomics*. **10**, 83-93.

Meents, A., Dittrich, B. and Gutmann S. (2009). *J. Synchrotron Rad.* **16**, 183-190.

Podjarny, A. D. & Urzhumtsev, A.G. (1997). *Methods Enzymol*. **276**, 641-658.

Word, J.M., Lovell, S.C., Richardson, J.S. and Richardson, D.C. (1999). *JMB*. **285**, 1735-47.