Supplementary Information

Water structure in solution and crystal molecular dynamics simulations compared to protein crystal structures

Octav Caldararu,[1] Majda Misini Ignjatović,[1] Esko Oksanen[2] and Ulf Ryde *[1]

[1] Department of Theoretical Chemistry, Lund University, Chemical Centre, P. O. Box 124, SE-221 00 Lund, Sweden

[2] Instruments Division, European Spallation Source Consortium ESS ERIC, P. O. Box 176, SE-221 00 Lund, Sweden

Correspondence to Ulf Ryde, E-mail: Ulf.Ryde@teokem.lu.se, Tel: +46 – 46 2224502, Fax: +46 – 46 2228648

2020-02-14

Table S1. Average residence time (ps) of water molecules in the crystallographic water sites of R- and S-galectin-3C in the four simulations. Residence times were calculated by counting the number of continuous frames in which a certain water molecule is within 2.5 Å of the crystallographic water sites in the MD simulations. The residence time is averaged over all crystallographic water sites and over all ten simulations.

Simulation	Residence time
Crystal R	84
Solution R	55
Crystal S	82
Solution S	52

Table S2. Recall of crystallographic water molecules in the 100 ns crystal or solution MD simulations of R– and S–galectin-3C against the 100 K (cryo) crystal structure from the grid-based global clustering. The number of crystallographic waters that have at least one MD water cluster within 1.0, 1.5, 2.0, 2.5 or 3.0 Å is given and the percentage of the total number of crystallographic waters is given in parentheses.

MD	1.0	1.5	2.0	2.5	3.0
R–galectin-3C					
Crystal	38 (18%)	114 (54%)	177 (84%)	190 (90%)	205 (97%)
Solution	36 (17%)	91 (43%)	168 (80%)	191 (91%)	200 (95%)
S–galectin-3C					
Crystal	62 (29%)	112 (53%)	149 (71%)	192 (91%)	195 (93%)
Solution	56 (27%)	117 (55%)	157 (74%)	187 (89%)	200 (95%)

Table S3. Recall of crystallographic water molecules in the 100 ns MD simulations of R– and S–galectin-3C (in the crystal or in solution) against the 100 K crystal structure from the non-grid-based global clustering. The number of crystallographic waters that have at least one MD water cluster within 1.0, 1.5, 2.0, 2.5 or 3.0 Å is given and the percentage of the total number of crystallographic waters is given in parentheses.

MD	1.0	1.5	2.0	2.5	3.0
R–galectin-3C					
Crystal	53 (25%)	123 (58%)	156 (74%)	177 (84%)	197 (93%)
Solution	45 (21%)	118 (56%)	160 (76%)	177 (84%)	190 (90%)
S–galectin-3C					
Crystal	57 (27%)	113 (53%)	146 (68%)	187 (87%)	205 (96%)
Solution	49 (23%)	111 (52%)	150 (70%)	190 (89%)	200 (93%)

Figure S1. Box volume in the 1 ns NPT equilibration of the crystal MD simulation of R-galectin-3C.

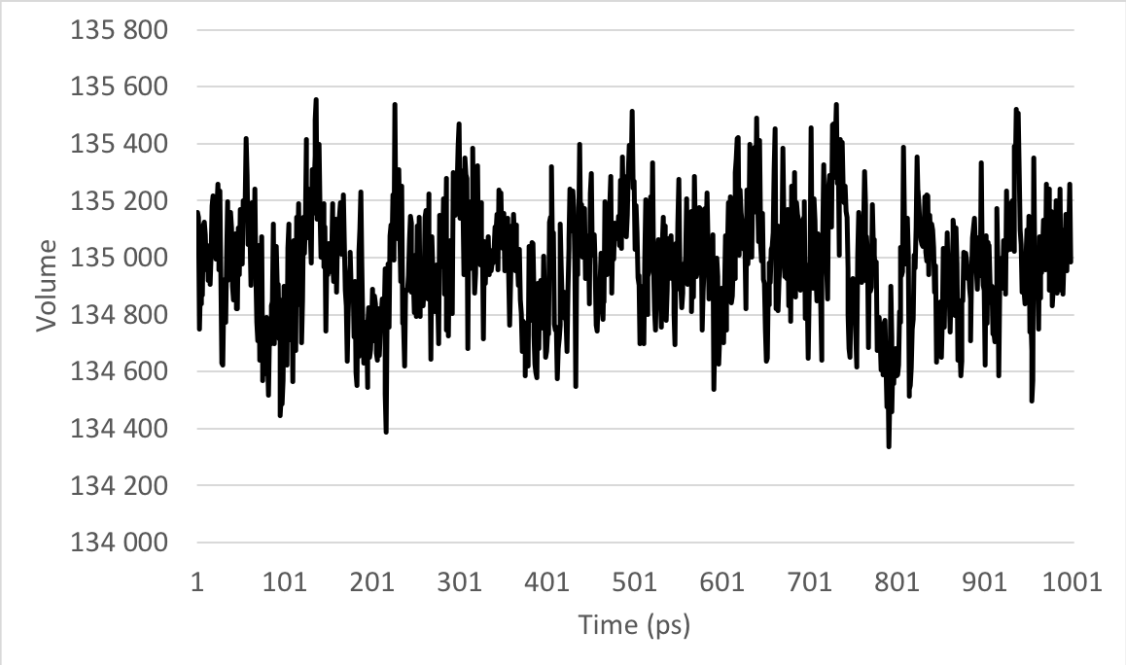
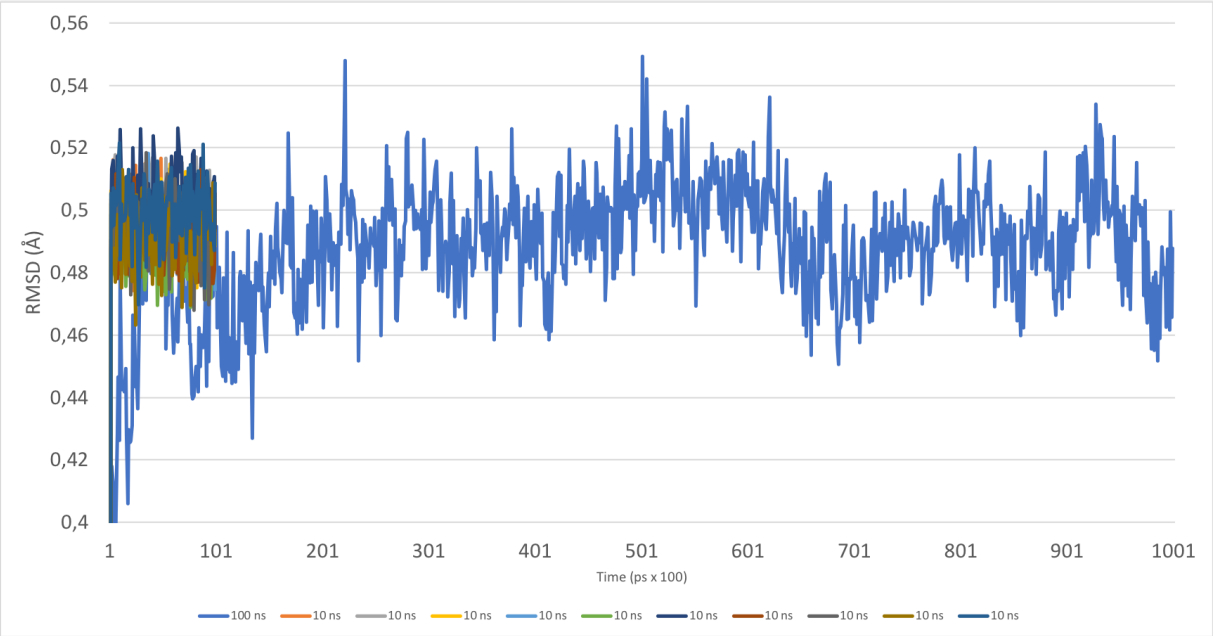
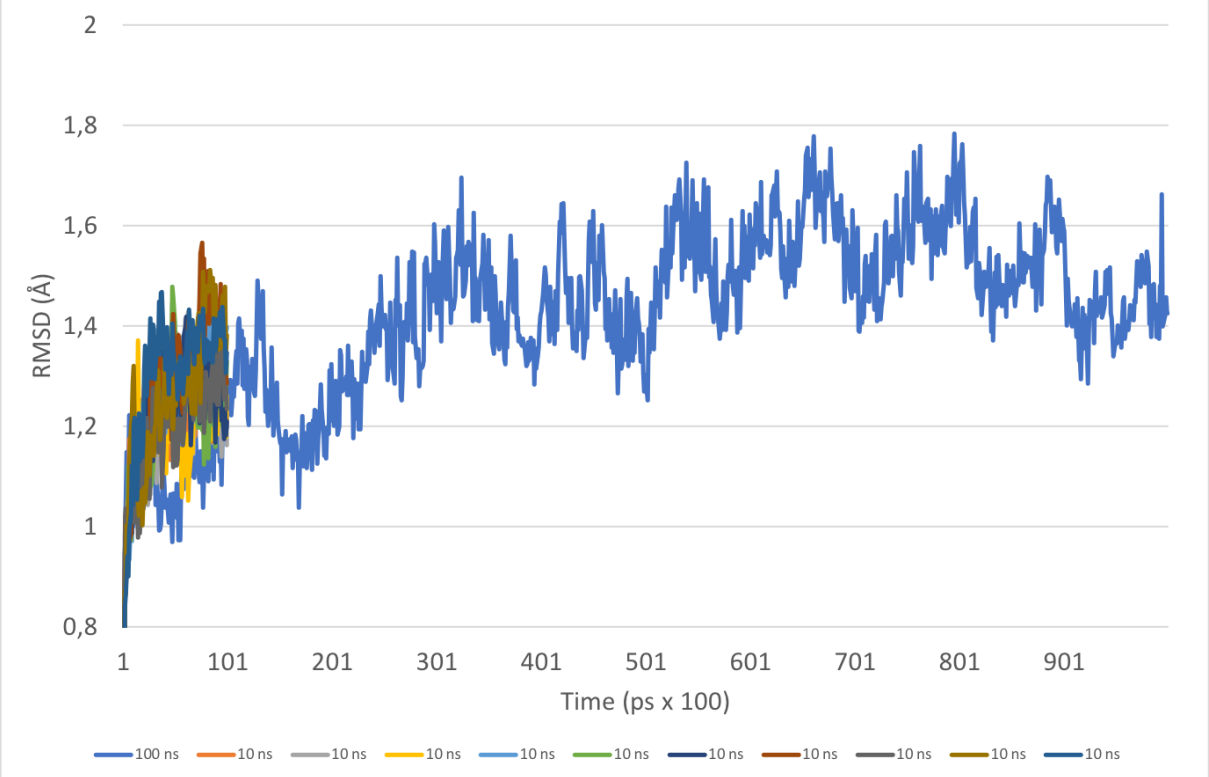


Figure S2. Protein heavy-atom RMSD in the 10 × 10 ns simulations and in the 100 ns simulations.

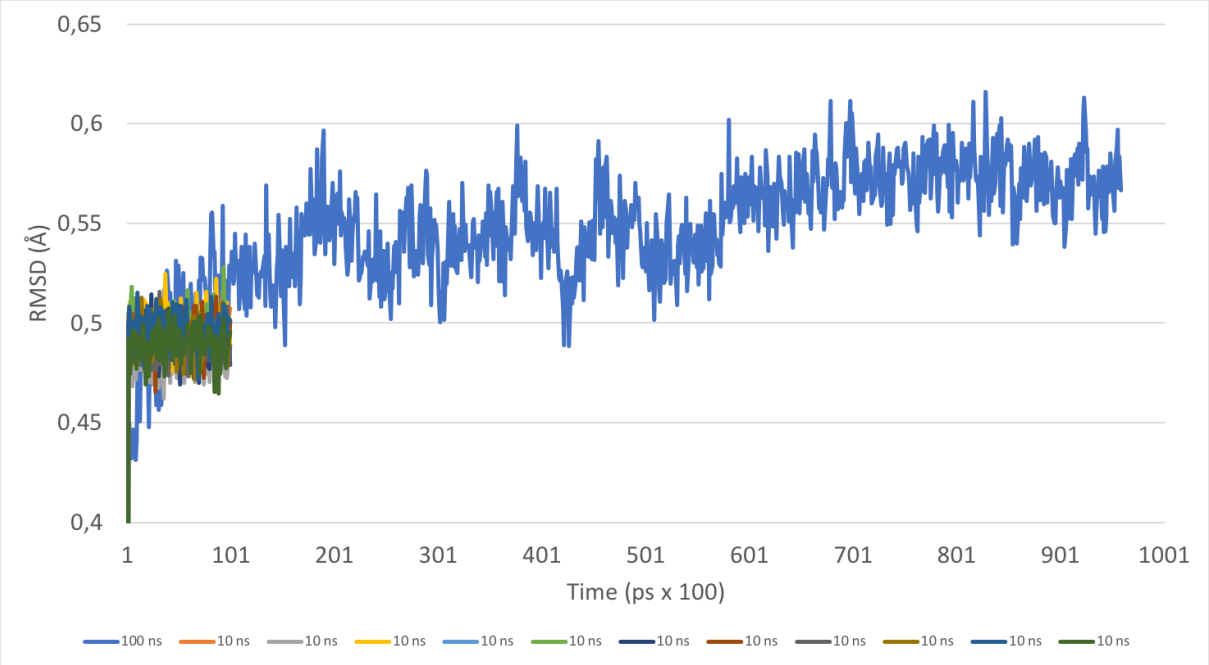
(a) Crystal MD simulation of R-galectin-3C



(b) Solution MD simulation of R-galectin-3C



(c) Crystal MD simulation of S-galectin-3C



(d) Solution MD simulation of S-galectin-3C

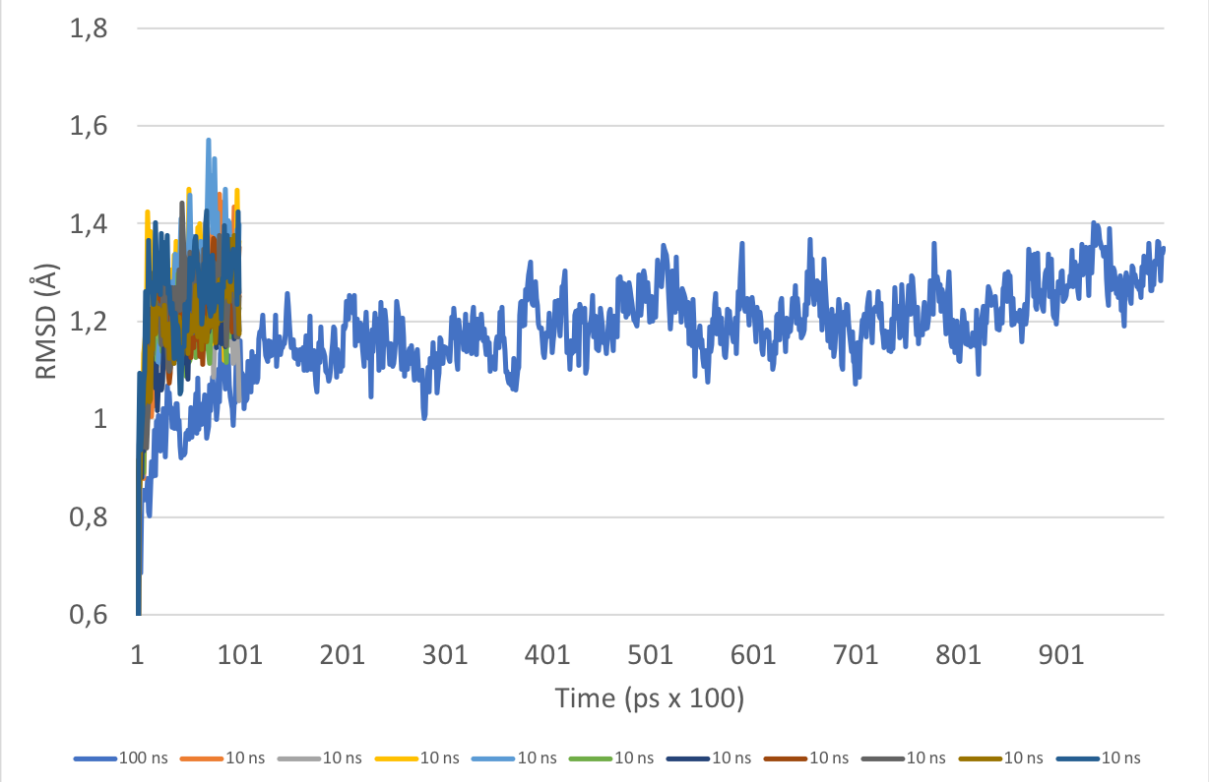
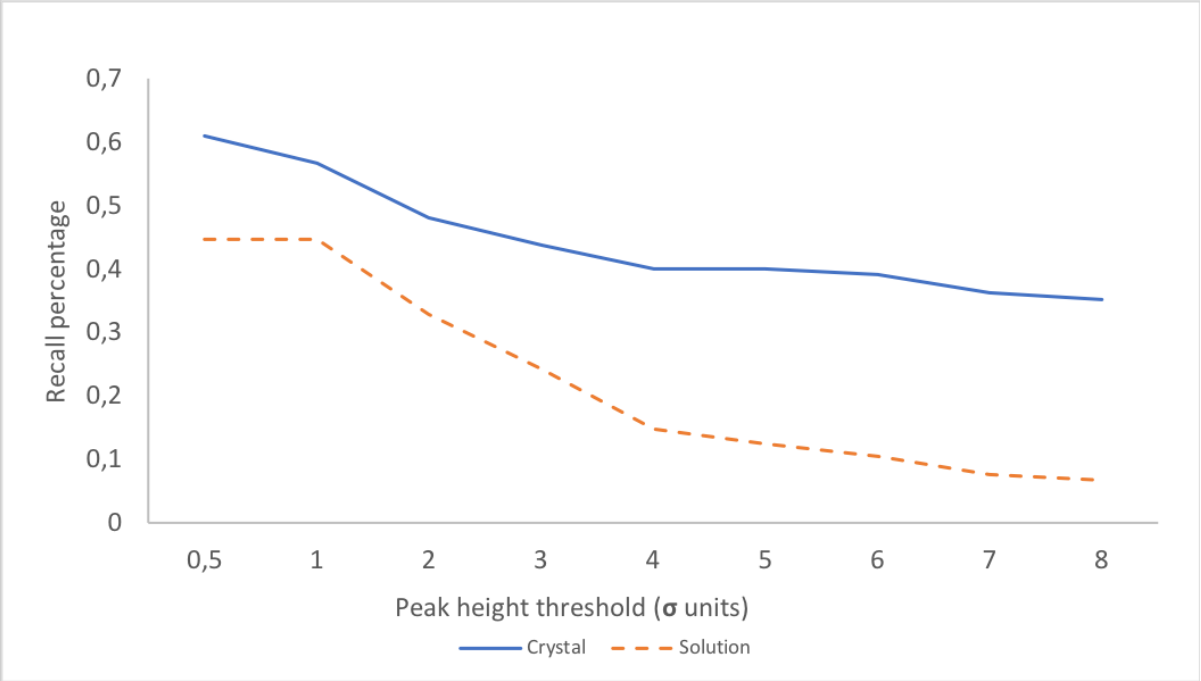


Figure S3. Dependence of the (a) recall and (b) prediction statistics on the density threshold. The statistics were computed using a distance cutoff distance of 1.5 Å and using only the crystallographic water molecules in the cryo-temperature structure.

(a) recall statistics



(b) prediction statistics