Supporting Information: Critical Analysis of the Accuracy of Models Predicting or Extracting Liquid Structure Information.

Marc Van Houteghem, An Ghysels, Toon Verstraelen, Ward Poelmans, Michel Waroquier, and Veronique Van Speybroeck*

Center for Molecular Modeling, QCMM Alliance Ghent-Brussels, Ghent University, Technologiepark 903, B-9052 Zwijnaarde, Belgium

E-mail: veronique.vanspeybroeck@ugent.be

*To whom correspondence should be addressed

Table S.1: Lengths of the MD production runs.

Table S.2: Velocity correlation times of the centers of mass of the solvents.

Table S.3: Details CP-computations. Thet total number of used pairs are masked above 10.0 kJ/mol and below 0.1 kJ/mol.

Fig. S.1: COM RDFs from the NPT/DFT-D3 MD simulations.

Fig. S.2: CP-values from CP2K as a function of the CP-model correction for MeOH, MeCN, TCM, THF, EtOH. CP-values are masked above 10.0 kJ/mol and below 0.1 kJ/mol.

Table S.4: Sampling covariance matrices of the optimized (A_i, τ) parameters and the condition numbers λ of the normalized sampling covariance matrices for MeOH, TCM, MeCN, THF, EtOH. $[A]$ = kJ/mol, $[\tau]$ = /.

Fig. S.3: Intermolecular radial distribution functions of MeOH. Experimental data from Yamaguchi *et al.*, *Mol. Phys.*, 96(8):1159, 1999; Mol. Phys., 97(4):603, 1999.

Fig. S.4: Intermolecular radial distribution functions of TCM. Experimental data from Pothoczki *et al.*, *J. Phys.: Condens. Matter*, 22(40):404211, 2010.

Fig. S.5: Intermolecular radial distribution functions of MeCN.

Fig. S.6: Intermolecular radial distribution functions of THF. Experimental data from Bowron *et al. J. Am. Chem. Soc.*, 128(15):5119, 1996.

Fig. S.7: Intermolecular radial distribution functions of EtOH.

Fig. S.8: Partial intermolecular structure factors of THF.

Fig. S.9 Total (inter + intra) radial distribution functions of THF.

Fig. S.10: Partial structure factors of the total RDFs of THF.

FIG. S.11.: Comparison of Fourier-based approach (applying Eqs. (2), (3), (4) on g) in the left column, with a direct weighted summation of all $g_{\alpha\beta}$ in the right column (Eq. (5)) for THF.

Molecular Dynamics - some more details:

Table S.1: Lengths of the MD production runs.

solvent	NVT/DFT-D3 [ps]	NPT/DFT-D3 [ps]	NPT/DFT-D3 + CP [ps]
MeOH	8.1	12.5	19.9
TCM	8.1	15.0	16.0
MeCN	8.1	15.0	16.6
THF	8.1	15.0	16.6
EtOH	8.1	14.0	16.3

Table S.2: Velocity correlation times of the centers of mass of the solvents.

solvent	$\tau_{\text{corr}}^{\text{vel,com}}$ [fs]
MeOH	81
TCM	270
MeCN	119
THF	147
EtOH	97

BSSE and CP corrections - some more details:

Table S.3: Details CP-computations. Thet total number of used pairs are masked above 10.0 kJ/mol and below 0.1 kJ/mol.

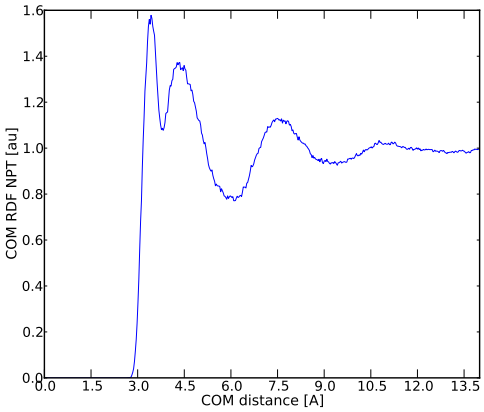
solvent	$d_{\text{cut-off}}$ [Å]	total # of pairs	total # of used pairs	total # of frames	cell [Å]
MeOH	6.2	15714	15111	16	11.86
TCM	7.9	13476	10687	19	15.91
MeCN	6.4	15749	11418	19	14.99
THF	7.5	8127	7982	19	16.46
EtOH	7.0	16821	15819	18	15.74

Parameter calibration for the CP correction terms:

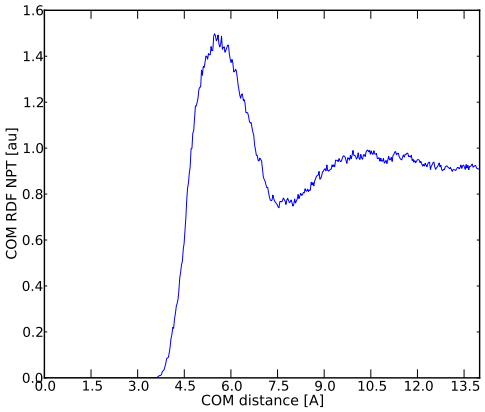
The considered cost function is the root-mean-square deviation (RMSD) between the natural logarithms of the model $\Delta E_{\text{CP}}^{\text{mod}}$ and the training $\Delta E_{\text{CP}}^{\text{ref}}$ corrections.

The natural logarithms are used to make the cost function sensitive to relative deviations between model and training values rather than absolute values, i.e. we want to achieve an

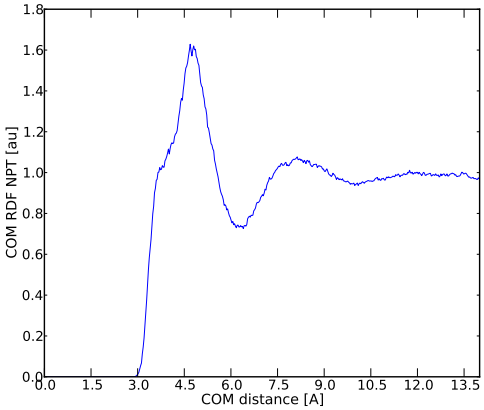
FIG. S.1.: COM RDFs from the NPT/DFT-D3 MD simulations.



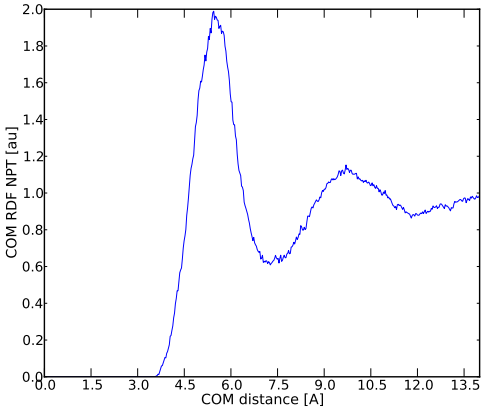
MeOH



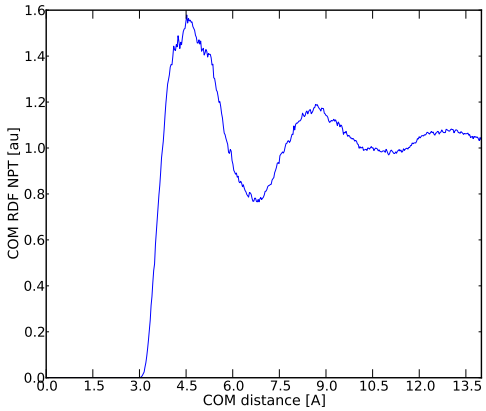
TCM



MeCN



THF



EtOH

equal sensitivity in all energy regions:

$$\ln\left(\Delta E_{\mathrm{CP}}^{\mathrm{mod}}\right) - \ln\left(\Delta E_{\mathrm{CP}}^{\mathrm{ref}}\right) = \ln\left(1 + \frac{\Delta E_{\mathrm{CP}}^{\mathrm{mod}} - \Delta E_{\mathrm{CP}}^{\mathrm{ref}}}{\Delta E_{\mathrm{CP}}^{\mathrm{ref}}}\right). \tag{1}$$

Negative values for the (A_i,τ) parameters are undesirable as we want the CP-model to be repulsive in nature. However, a (partially) negative set of optimized parameters is not inherently prevented by the CMA-strategy. Therefore, ($\ln A_i, \ln \tau$) are varied as parameters in the algorithm. To estimate the influence of the stochastic factor of CMA on the final result, the CMA-optimization procedure is conducted 200 times. This way, an average set of parameters can be calculated, with an accompanied statistical error on this average.

For each solvent, the *sampling covariance matrix* $\mathrm{Cov}(i,j)$ is calculated from the 200 parameters estimates. They are reported in Table S.4. This is useful for two reasons:

(i) The diagonal element $\mathrm{Cov}(i,i)$ is the variance σ_i^2, and quantifies the influence of the stochastic behaviour of the CMA-algorithm. The square root of each diagonal element gives the statistical standard error on the average of the parameter (0.00279, 0.067, 0.121, 0.663 for τ, $A_{\text{H-H}}$, $A_{\text{C-C}}$, $A_{\text{O-O}}$ respectively; see parameter set MeOH in Table II main manuscript).

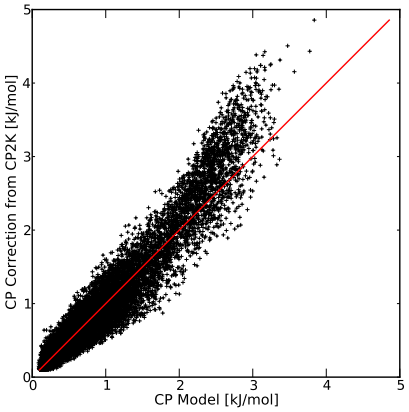
(ii) The condition number of the normalized sampling covariance matrix:

$$\frac{\mathrm{Cov}(i,j)}{\sigma_i \sigma_j} \tag{2}$$

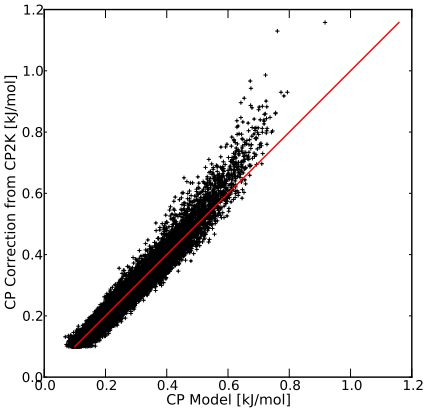
is a measure of how ill-conditioned the parameter estimation is. In practice, a high condition number implies that the calibrated parameters may not be transferable to other solvents. In the case of methanol, the condition number is relatively large (6.1E4). A high condition number limits the accuracy to which a solution can be obtained.

The sampling covariance matrices and associated condition number for MeCN, TCM, EtOH, THF are of the same order of magnitude, and are given in Table S.4.

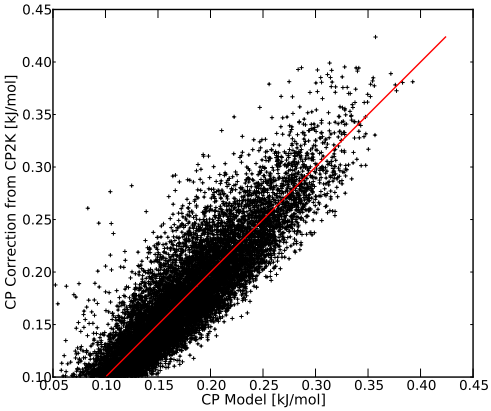
FIG. S.2: CP-values from CP2K as a function of the CP-model correction for MeOH, MeCN, TCM, THF, EtOH. CP-values are masked above 10.0 kJ/mol and below 0.1 kJ/mol.



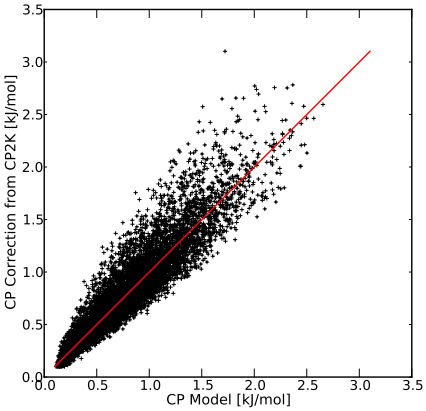
MeOH



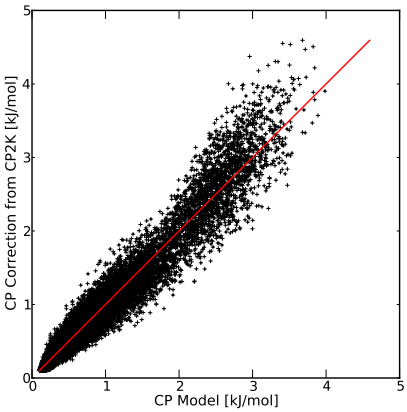
TCM



MeCN



THF



EtOH

Table S.4: Sampling covariance matrices of the optimized (A_i, τ) parameters and the condition numbers λ of the normalized sampling covariance matrices for MeOH, TCM, MeCN, THF, EtOH. $[A]$ = kJ/mol, $[\tau]$ = /.

	Cov(i,j)	λ
MeOH	$\begin{array}{c} \\ \tau \\ A_{\mathrm{H}} \\ A_{\mathrm{C}} \\ A_{\mathrm{O}} \end{array} \begin{array}{c} \begin{array}{cccc} \tau & A_{\mathrm{H}} & A_{\mathrm{C}} & A_{\mathrm{O}} \end{array} \\ \begin{pmatrix} 7.768 \times 10^{-6} & 0.000143 & \text{-}000325 & \text{-}0.00184 \\ 0.000143 & 0.00447 & \text{-}0.00700 & \text{-}0.0353 \\ \text{-}0.000325 & \text{-}0.00700 & 0.0147 & 0.0785 \\ \text{-}0.00184 & \text{-}0.0353 & 0.0785 & 0.439 \end{pmatrix} \end{array}$	60 897
TCM	$\begin{array}{c} \\ \tau \\ A_{\mathrm{H}} \\ A_{\mathrm{C}} \\ A_{\mathrm{Cl}} \end{array} \begin{array}{c} \begin{array}{cccc} \tau & A_{\mathrm{H}} & A_{\mathrm{C}} & A_{\mathrm{Cl}} \end{array} \\ \begin{pmatrix} 7.144 \times 10^{-7} & -3.963 \times 10^{-6} & -2.780 \times 10^{-6} & -2.437 \times 10^{-5} \\ -3.963 \times 10^{-6} & 9.590 \times 10^{-5} & \text{-}0.000262 & 0.000251 \\ -2.780 \times 10^{-6} & \text{-}0.000262 & 0.00127 & \text{-}0.000464 \\ \text{-}2.437 \times 10^{-5} & 0.000251 & \text{-}0.000464 & 0.00109 \end{pmatrix} \end{array}$	2653
MeCN	$\begin{array}{c} \\ \tau \\ A_{\mathrm{H}} \\ A_{\mathrm{C}} \\ A_{\mathrm{N}} \end{array} \begin{array}{c} \begin{array}{cccc} \tau & A_{\mathrm{H}} & A_{\mathrm{C}} & A_{\mathrm{N}} \end{array} \\ \begin{pmatrix} 9.009 \times 10^{-6} & -3.247 \times 10^{-5} & -4.247 \times 10^{-6} & -5.123 \times 10^{-5} \\ -3.247 \times 10^{-5} & 0.000136 & 1.150 \times 10^{-5} & 0.000209 \\ -4.247 \times 10^{-6} & 1.150 \times 10^{-5} & 2.838 \times 10^{-6} & 1.933 \times 10^{-5} \\ -5.123 \times 10^{-5} & 0.000209 & 1.933 \times 10^{-5} & 0.000323 \end{pmatrix} \end{array}$	18524
THF	$\begin{array}{c} \\ \tau \\ A_{\mathrm{H}} \\ A_{\mathrm{C}} \\ A_{\mathrm{O}} \end{array} \begin{array}{c} \begin{array}{cccc} \tau & A_{\mathrm{H}} & A_{\mathrm{C}} & A_{\mathrm{O}} \end{array} \\ \begin{pmatrix} 1.359 \times 10^{-5} & 0.000456 & \text{-}0.000684 & \text{-}0.000532 \\ 0.000456 & 0.0161 & \text{-}0.0241 & \text{-}0.0174 \\ \text{-}0.000684 & \text{-}0.0241 & 0.0369 & 0.0262 \\ \text{-}0.000532 & \text{-}0.0174 & 0.0262 & 0.021 \end{pmatrix} \end{array}$	8129

Continued on next page

Table S.4 – *Continued from previous page*

	Cov(i,j)				λ
		τ	A_{H}	A_{C}	A_{O}	
EtOH	τ	2.036×10^{-6}	0.000111	-0.000213	-0.000553	48520
	A_{H}	0.000111	0.00653	-0.0120	-0.0306	
	A_{C}	-0.000213	-0.0120	0.0227	0.0584	
	A_{O}	-0.000553	-0.0306	0.0584	0.151	

Radial Distribution Functions (RDFs) - Structure factors (SFs) : additional data

Atom types for the solvents MeOH, TCM, MeCN, THF, EtOH.

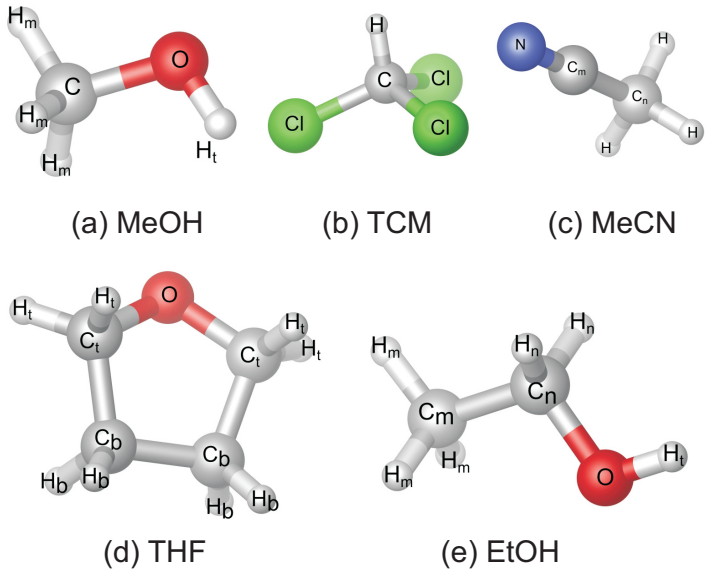
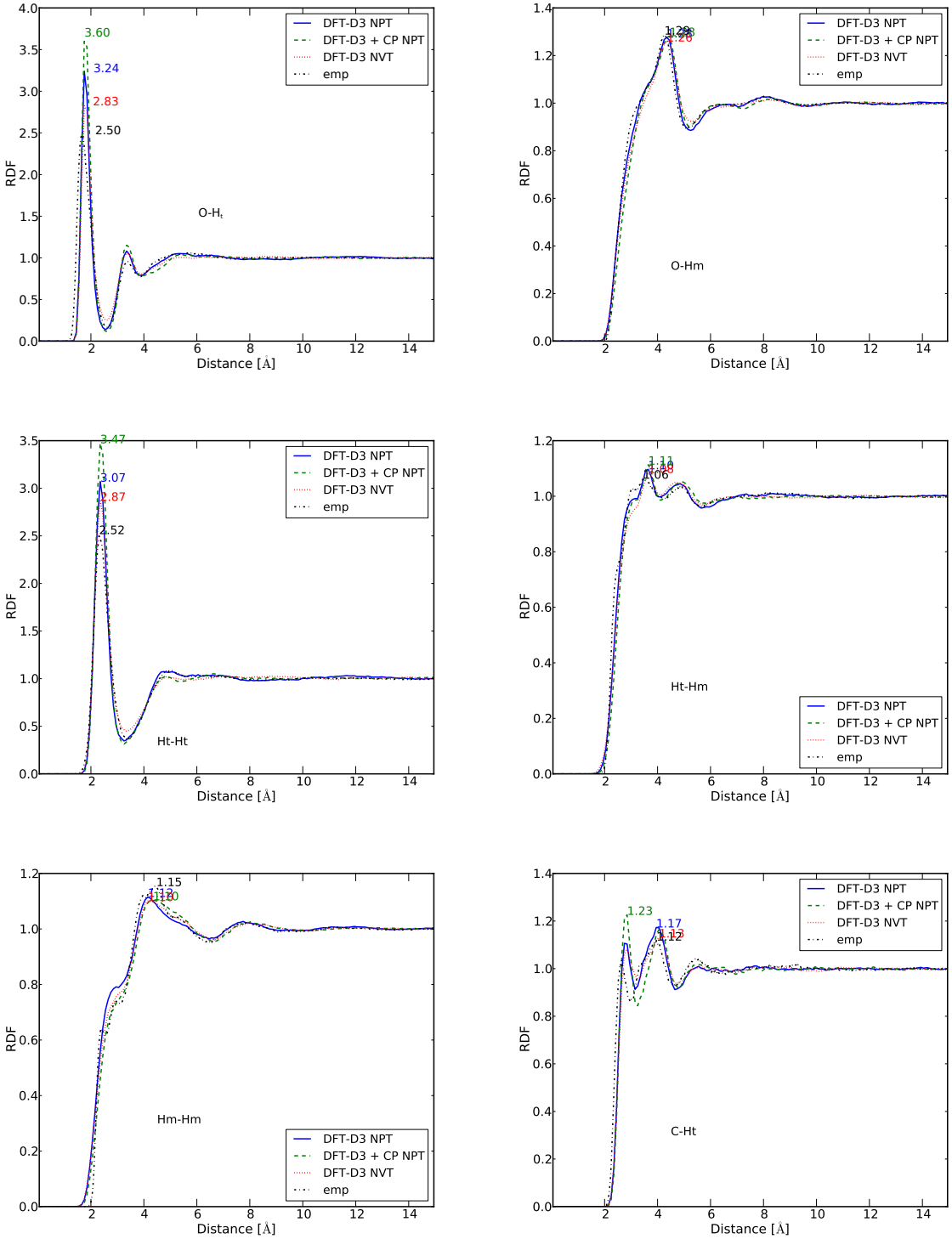


FIG. S.3.: Intermolecular radial distribution functions of MeOH. Experimental data from Yamaguchi *et al.*, *Mol. Phys.*, 96(8):1159, 1999; Mol. Phys., 97(4):603, 1999.



Continued on next page

FIG. S.3 – *Continued from previous page*

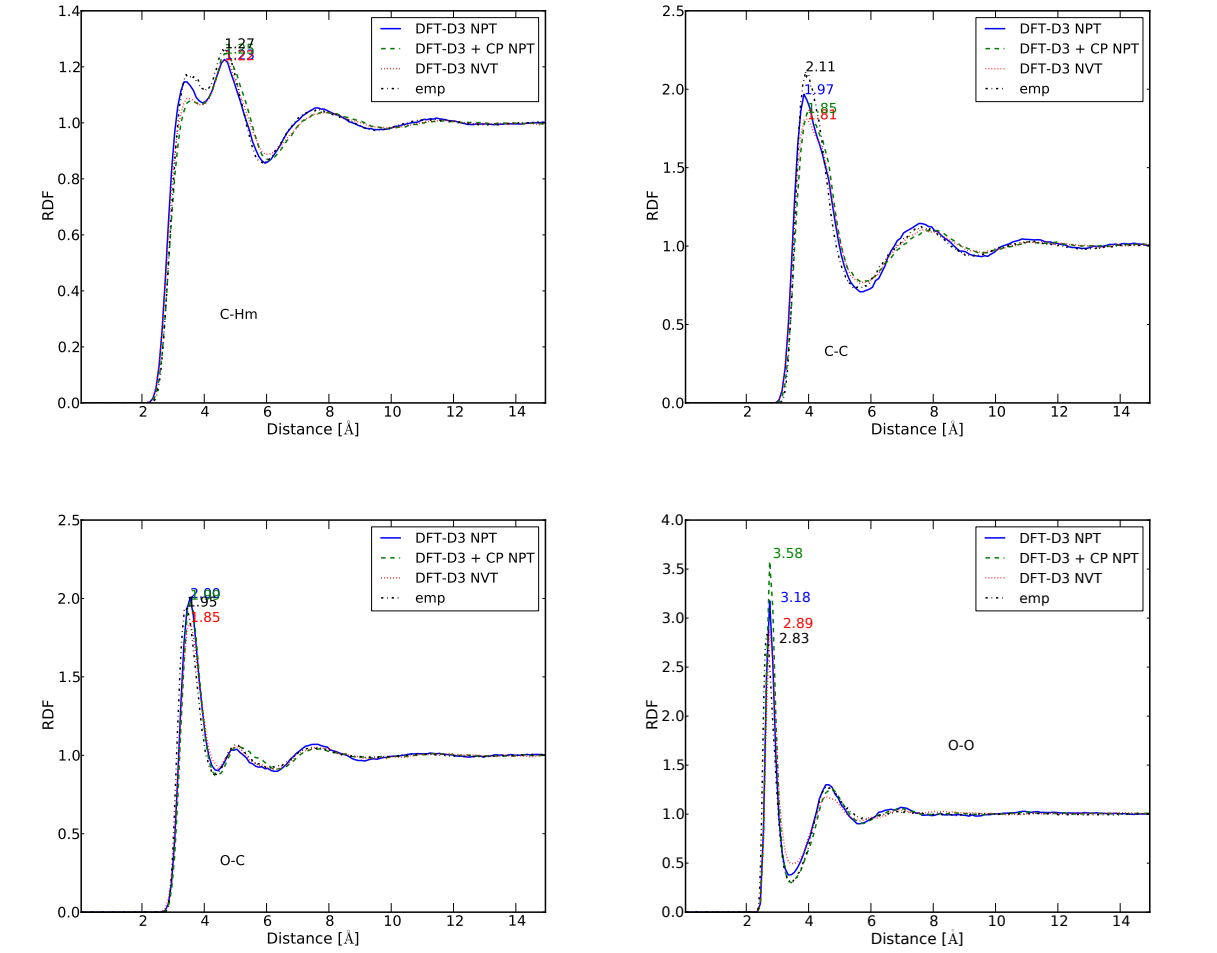


FIG. S.4.: Intermolecular radial distribution functions of TCM. Experimental data from Pothoczki *et al.*, *J. Phys.: Condens. Matter*, 22(40):404211, 2010.

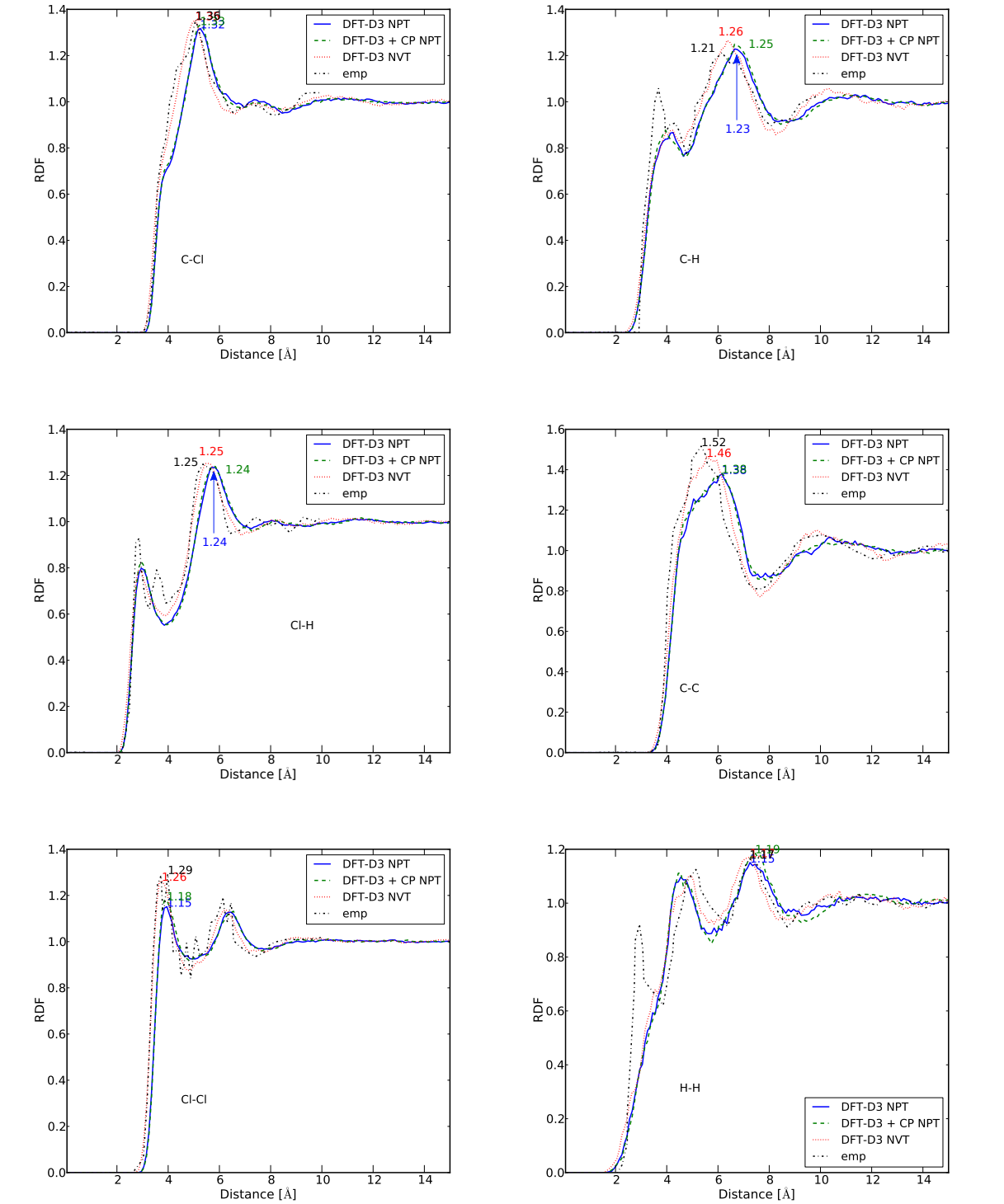
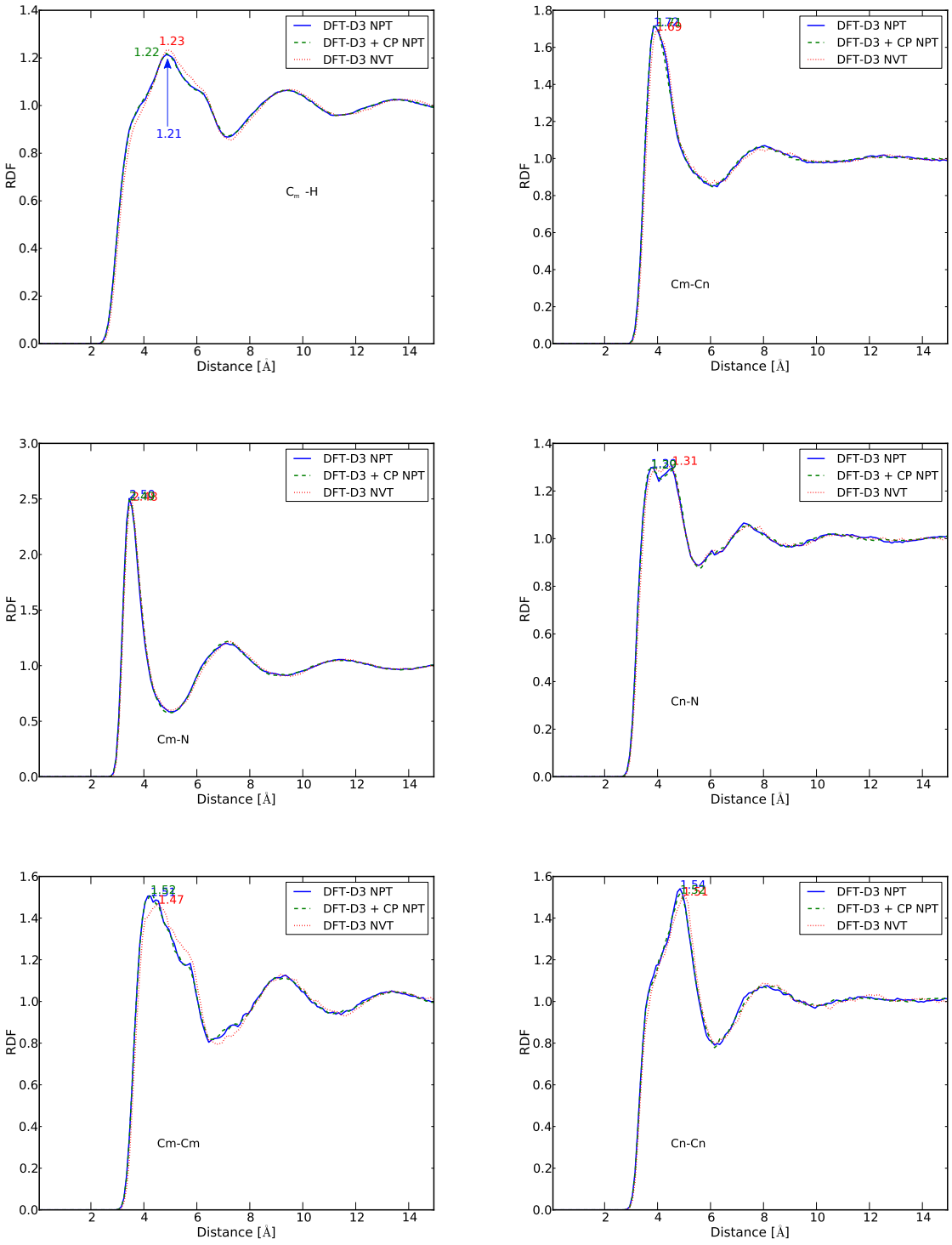
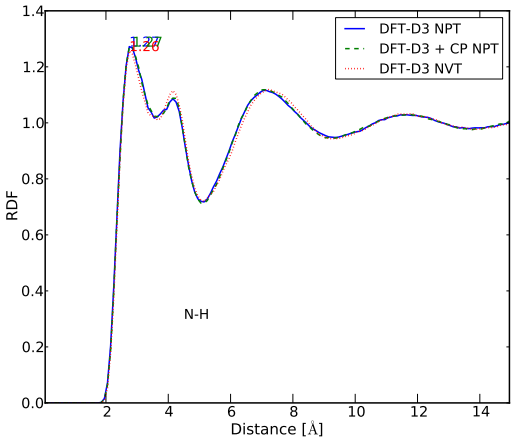


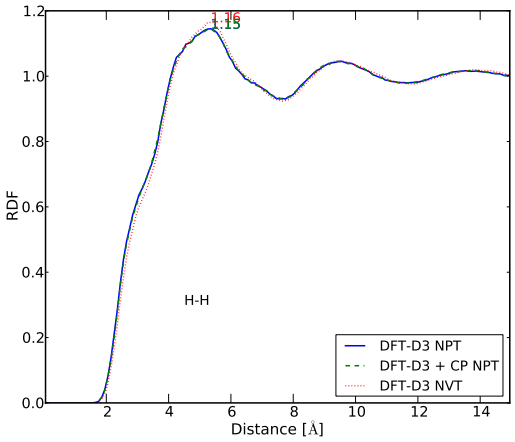
FIG. S.5.: Intermolecular radial distribution functions of MeCN.

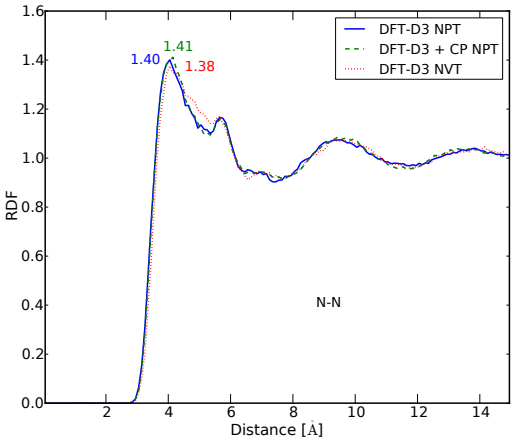


Continued on next page

FIG. S.5 – *Continued from previous page*







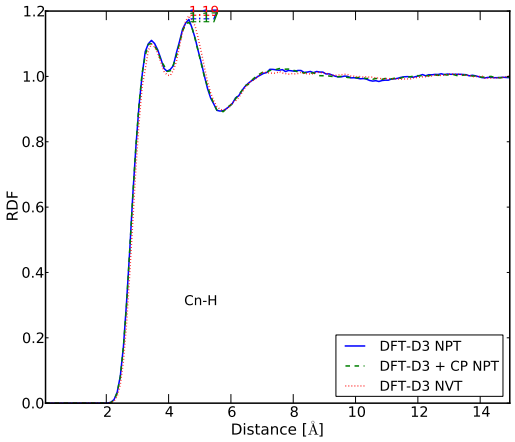
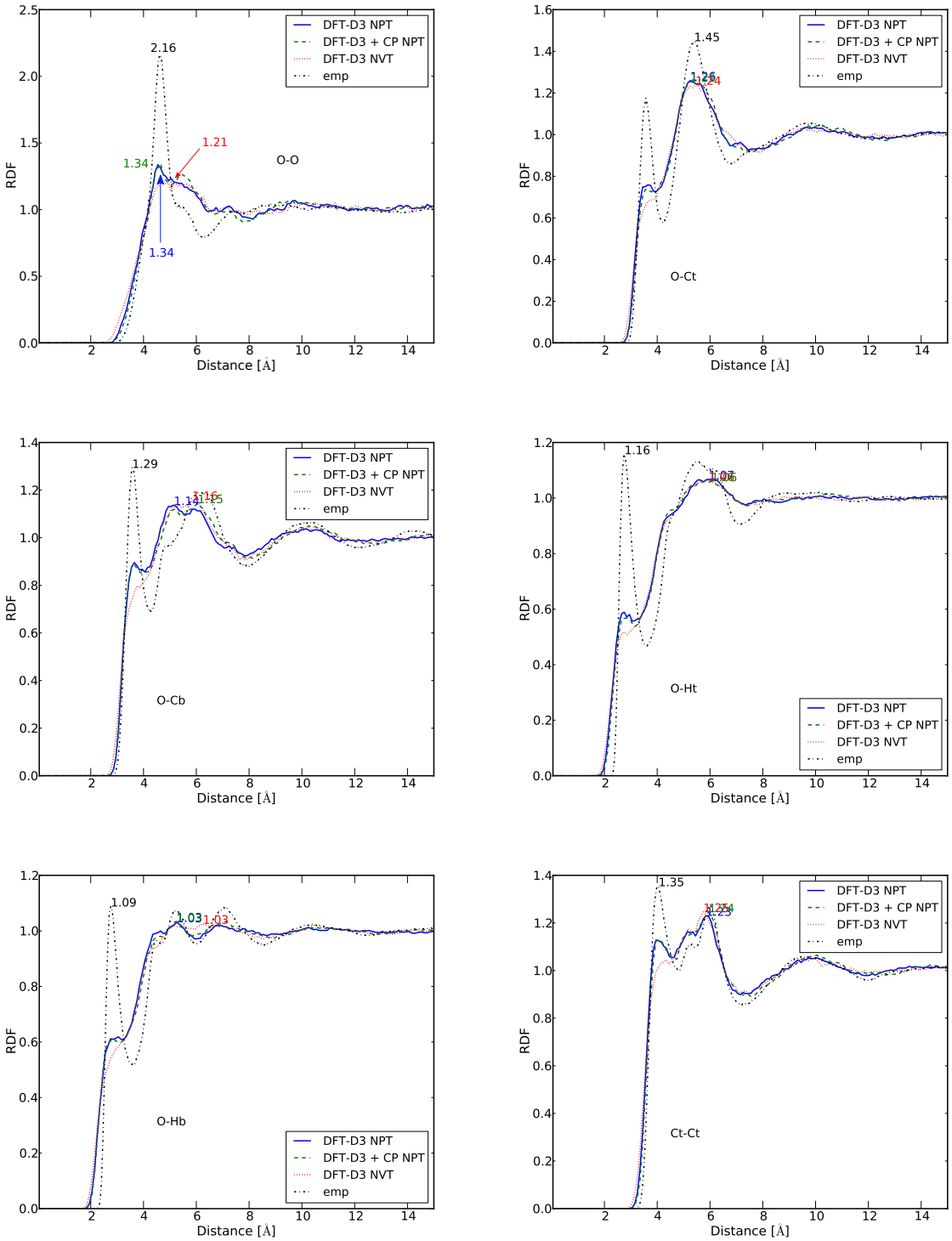
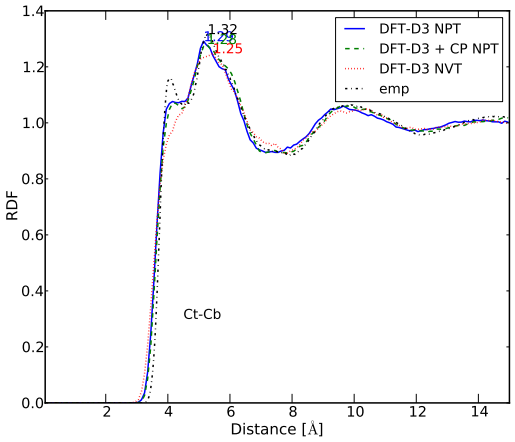


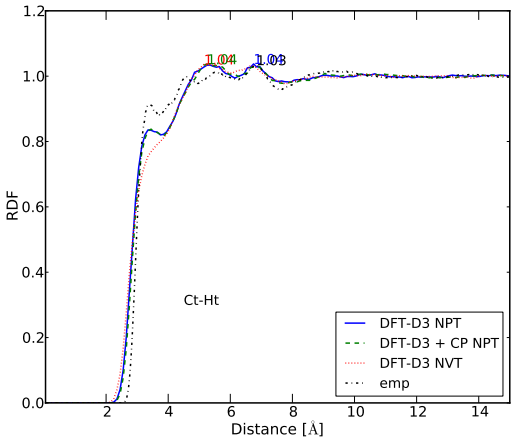
FIG. S.6.: Intermolecular radial distribution functions of THF. Experimental data from Bowron *et al. J. Am. Chem. Soc.*, 128(15):5119, 1996.

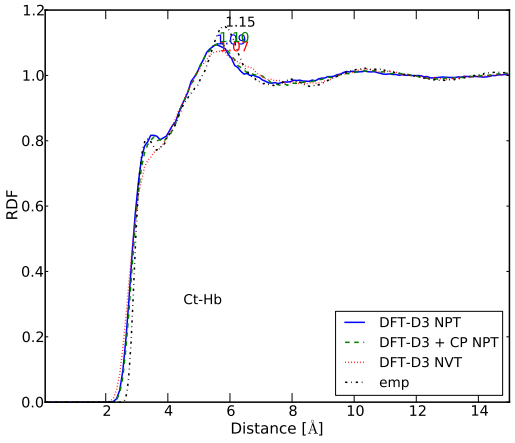


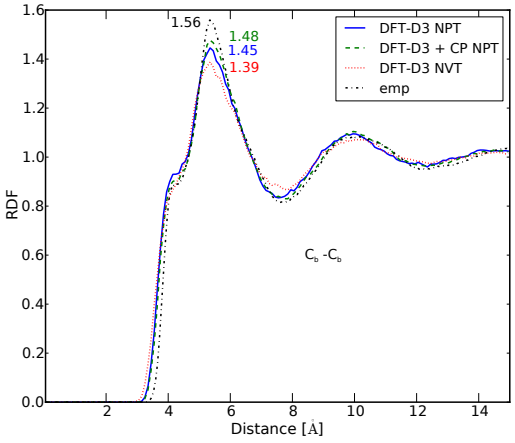
Continued on next page

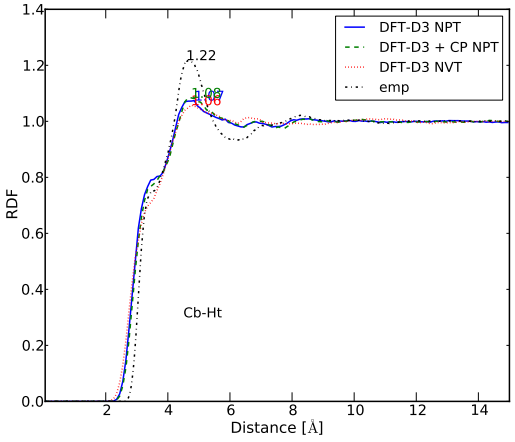
FIG. S.6 – *Continued from previous page*

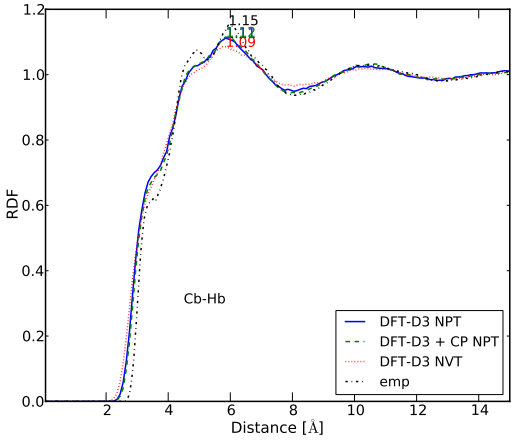






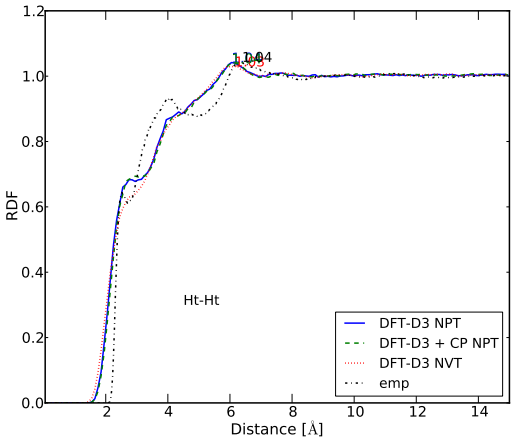


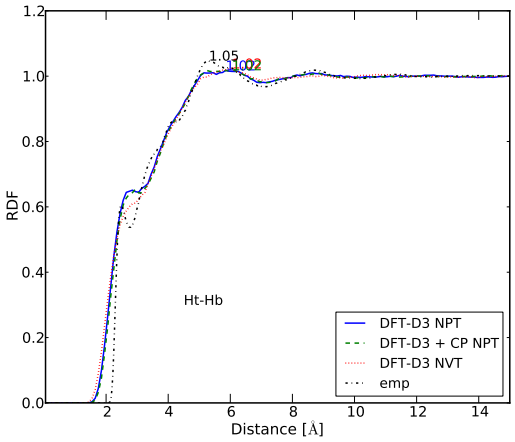




Continued on next page

FIG. S.6 – *Continued from previous page*





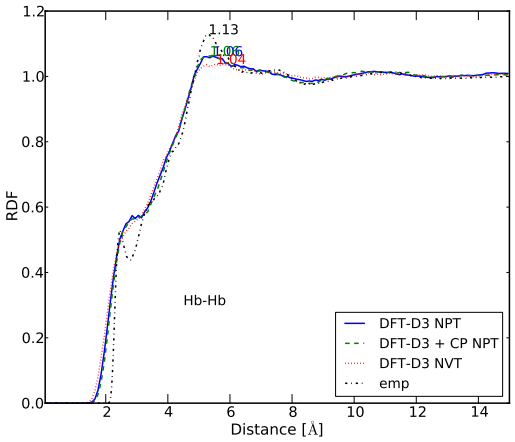
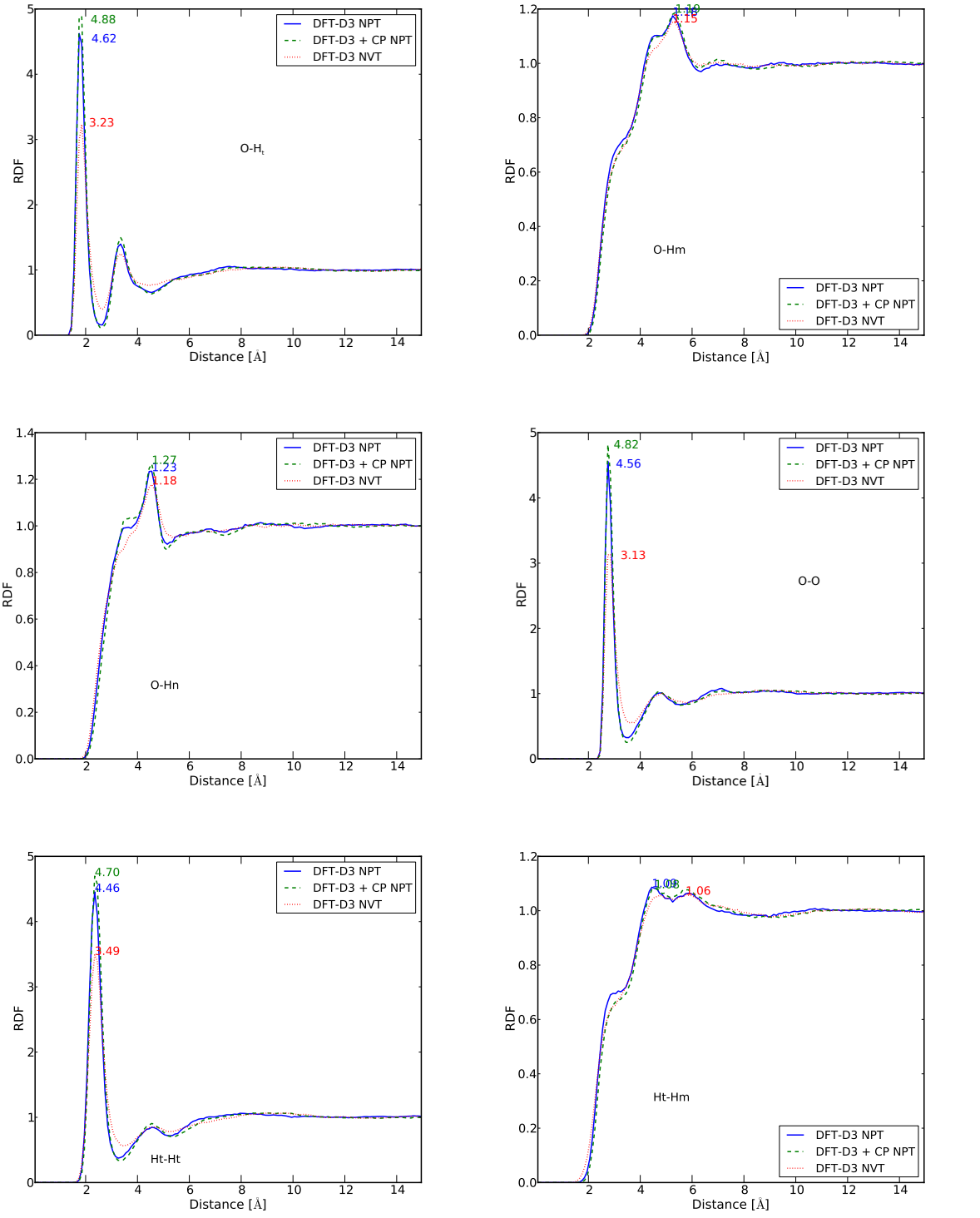
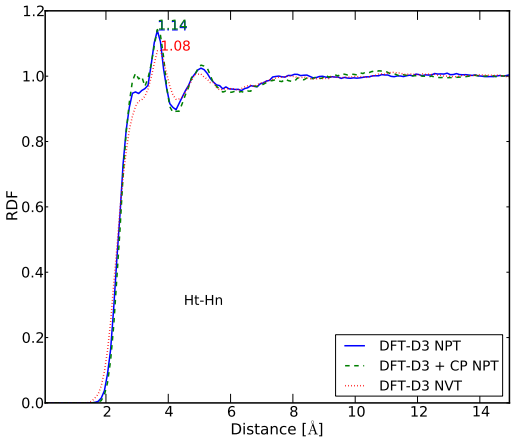


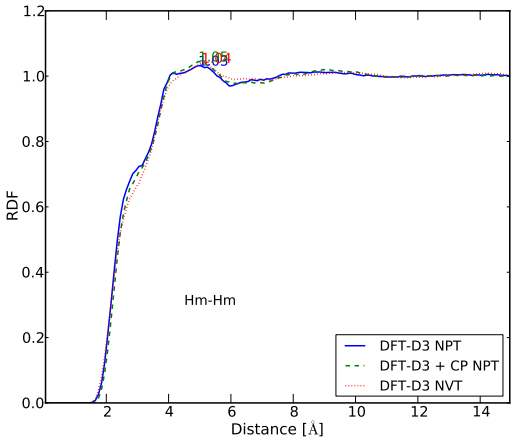
FIG. S.7.: Intermolecular radial distribution functions of EtOH.

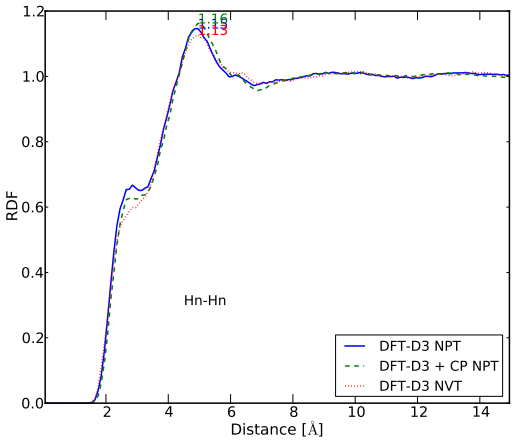


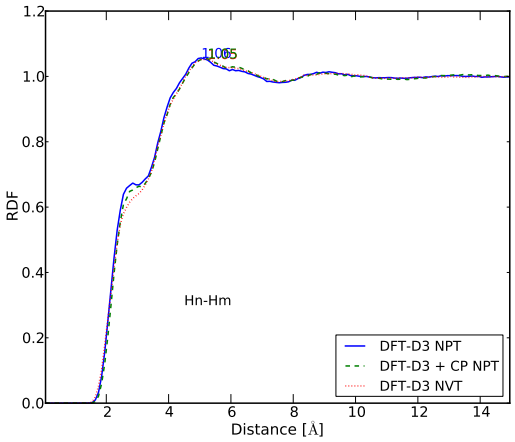
Continued on next page

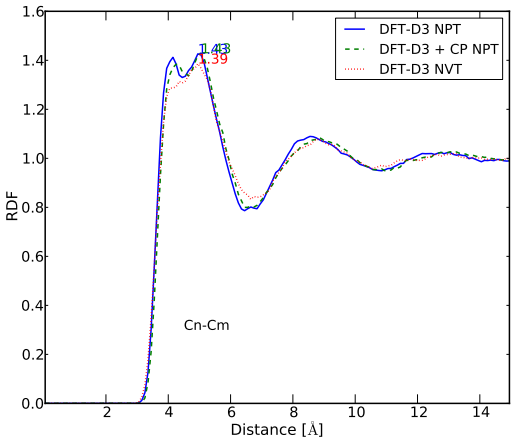
FIG. S.7: – *Continued from previous page*

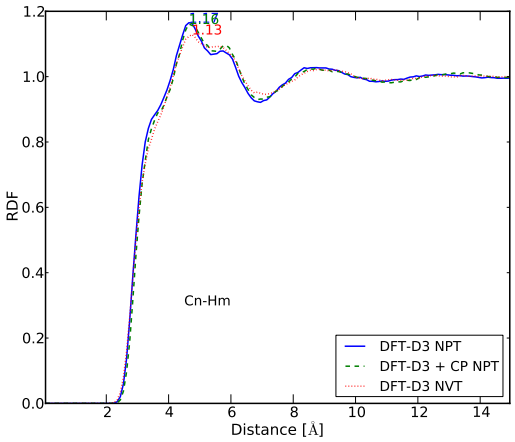






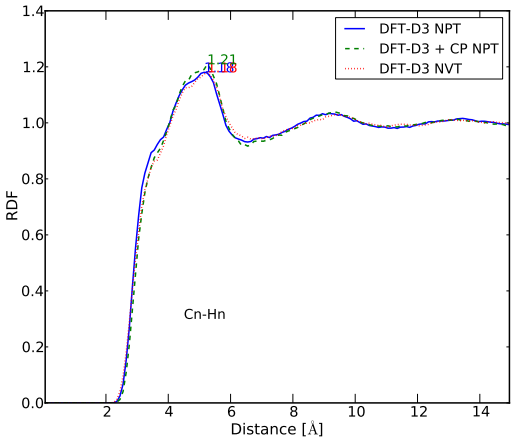


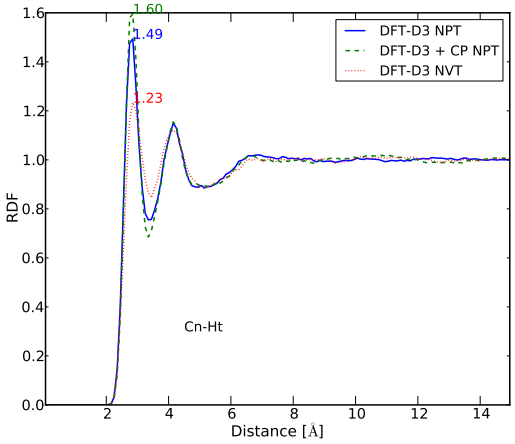


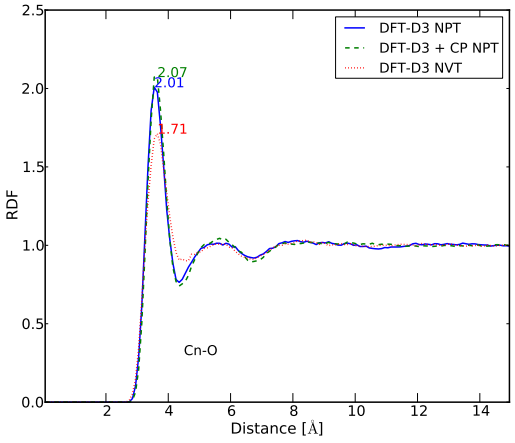


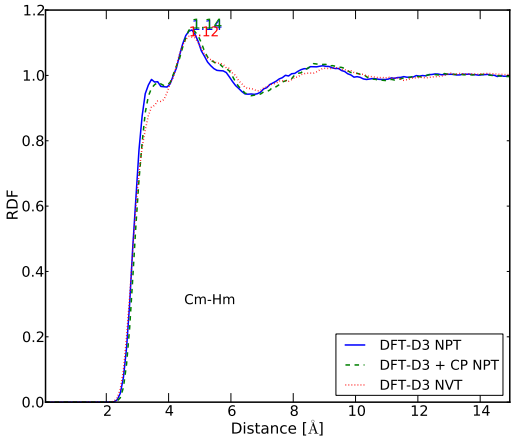
Continued on next page

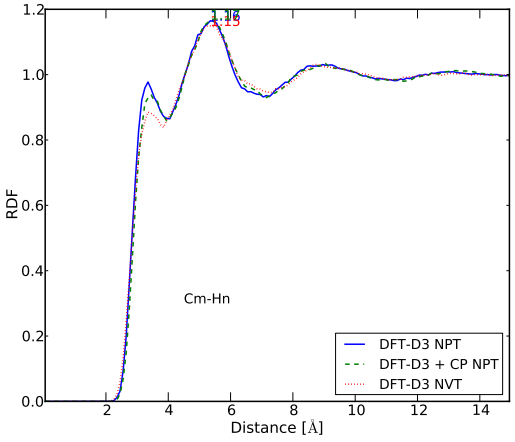
FIG. S.7: – *Continued from previous page*

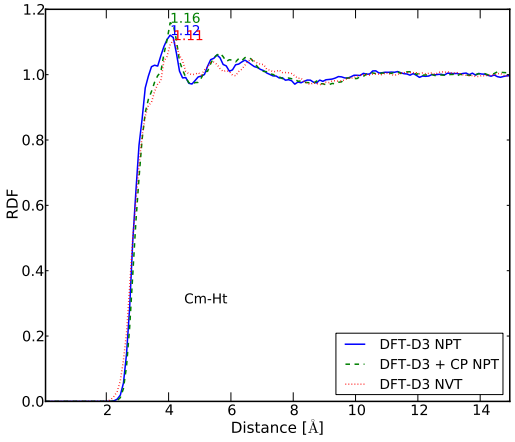






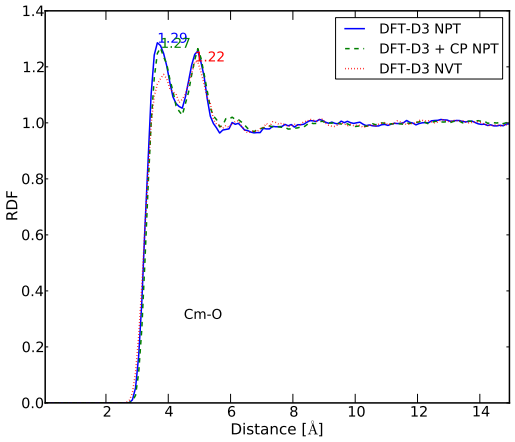


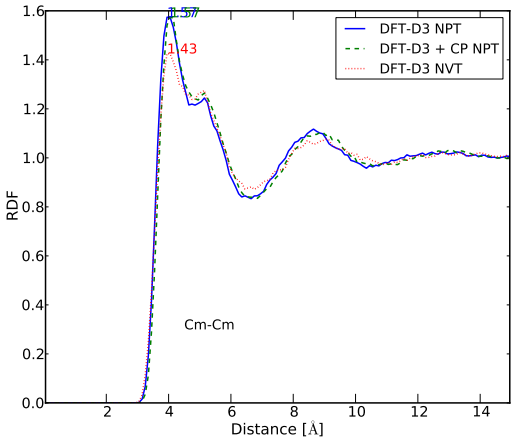


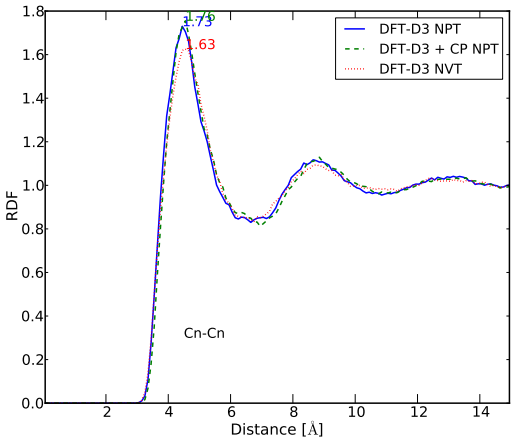


Continued on next page

FIG. S.7: – *Continued from previous page*

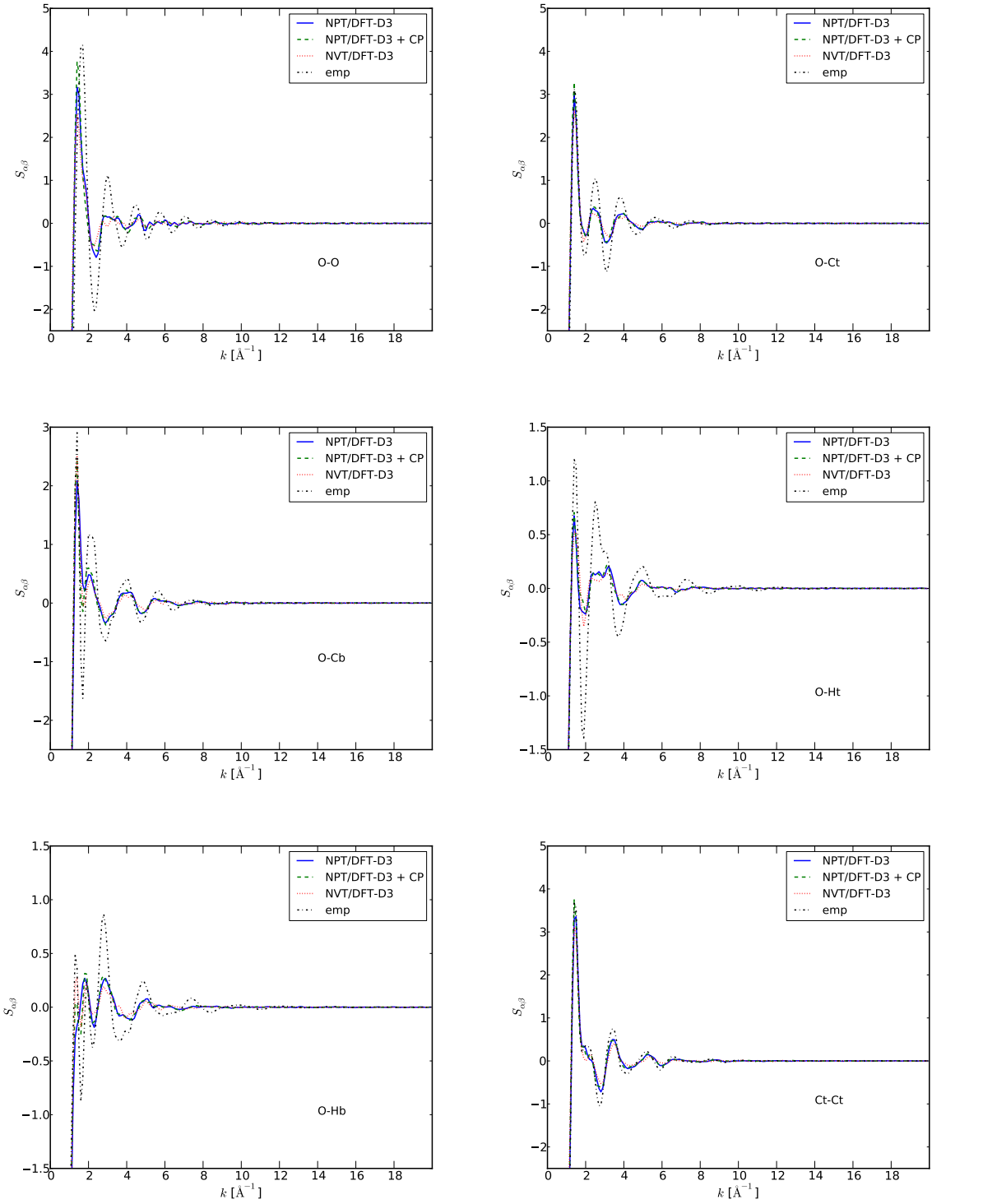






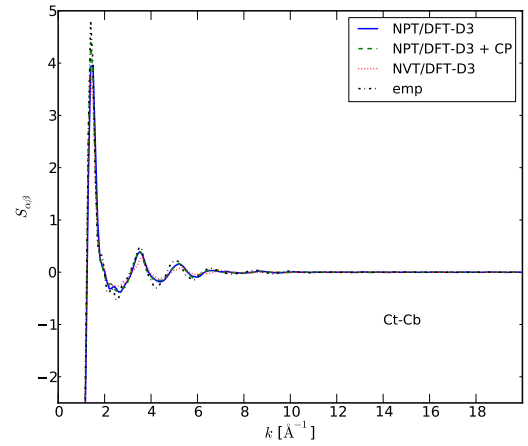
Structure factors THF - additional data :

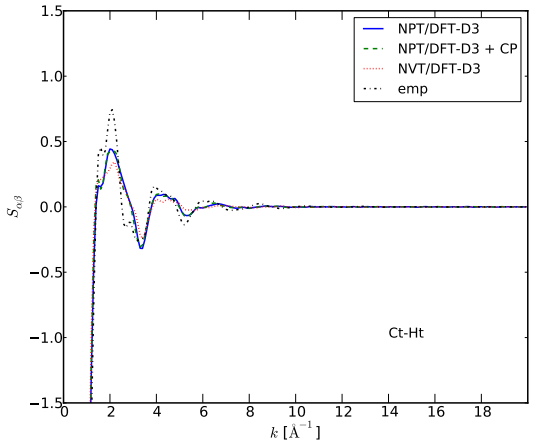
FIG. S.8.: Partial intermolecular structure factors of THF.

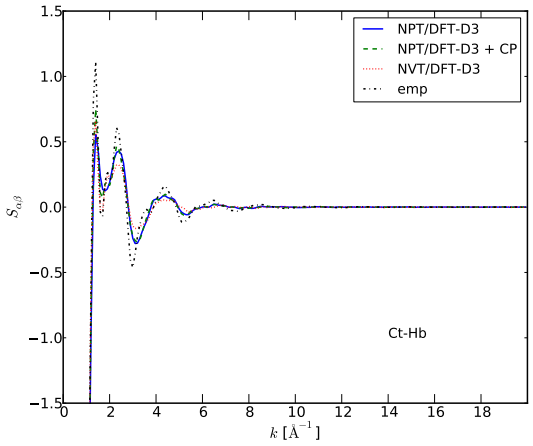


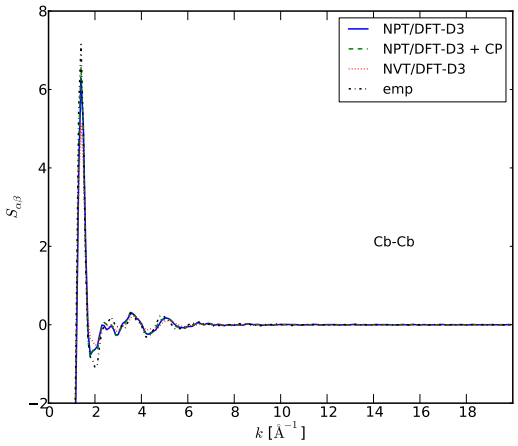
Continued on next page

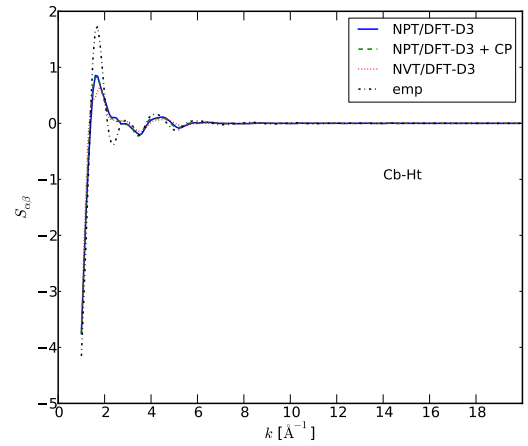
FIG. S.8 – *Continued from previous page*

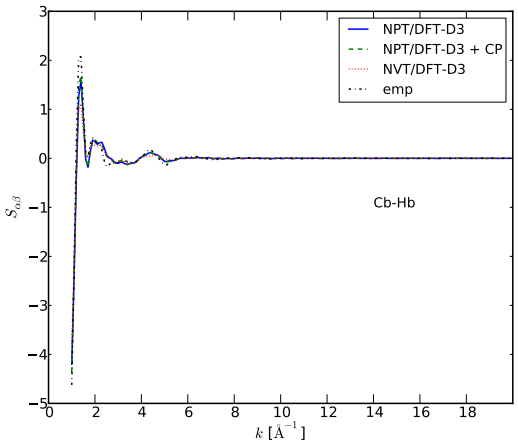






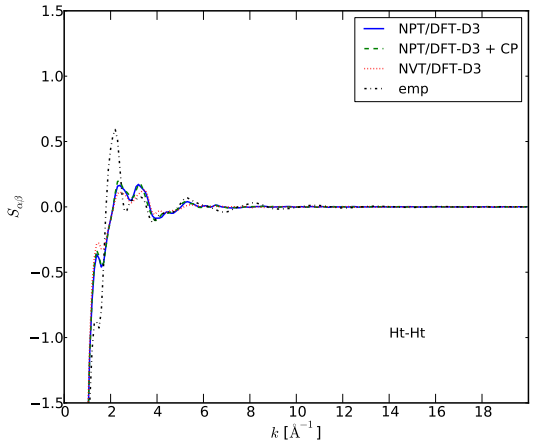


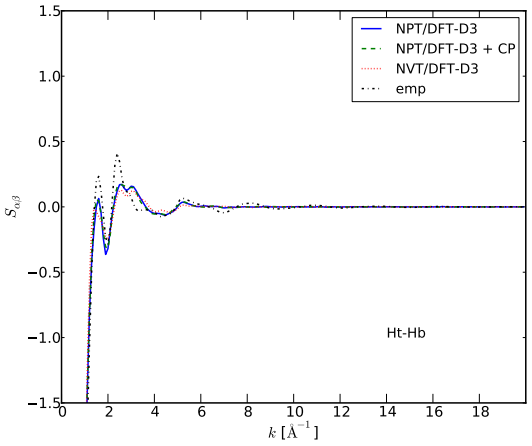




Continued on next page

FIG. S.8 – *Continued from previous page*





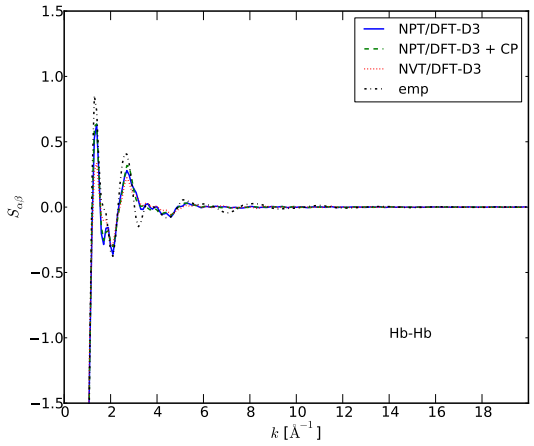
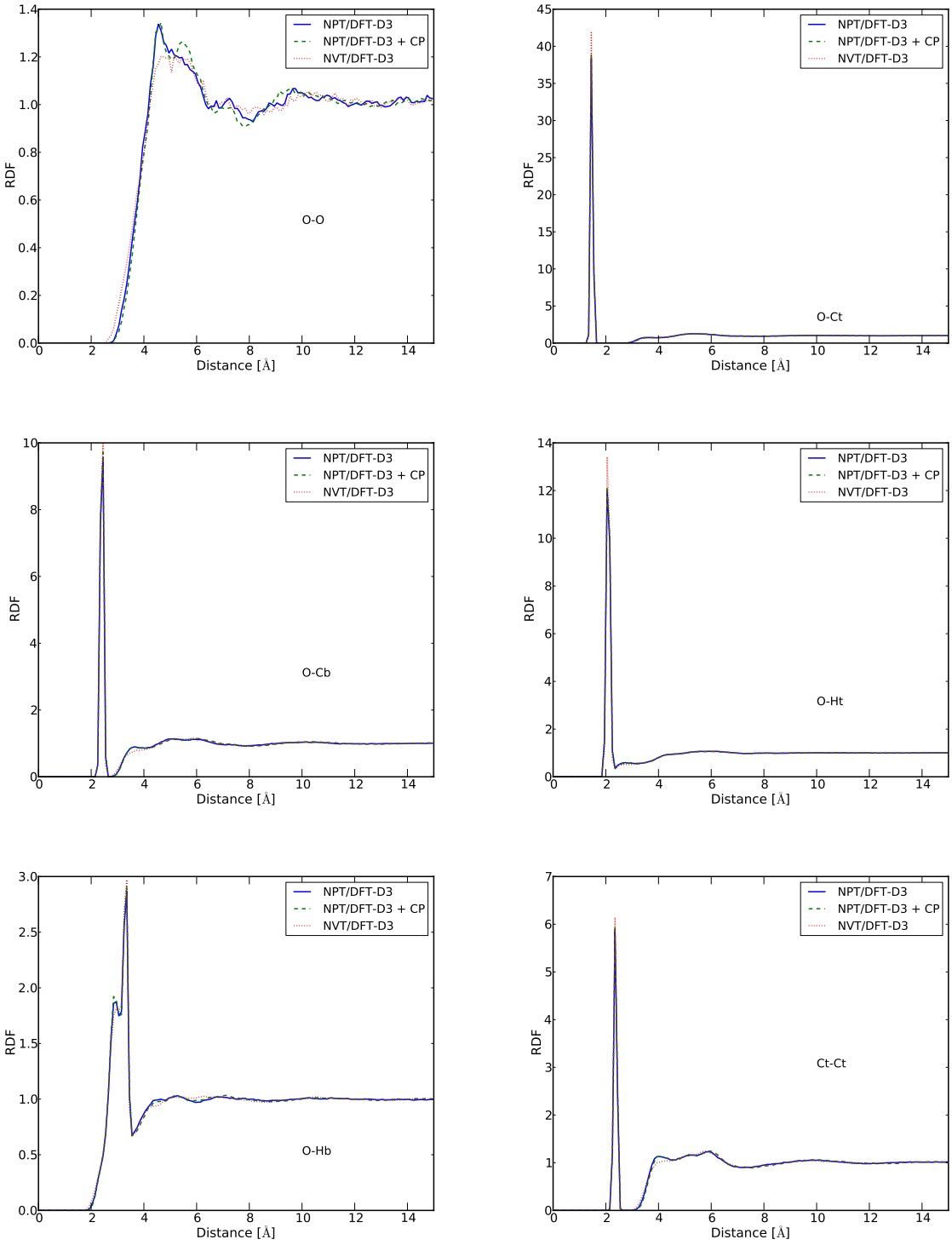
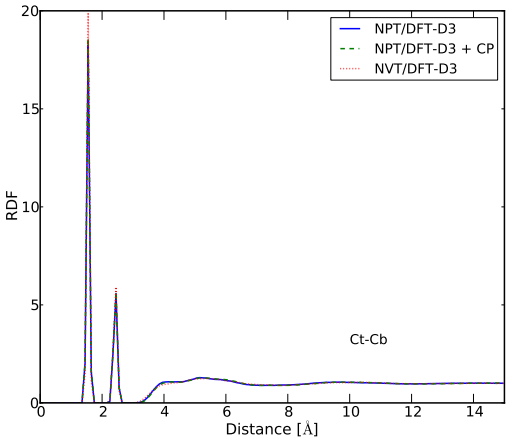


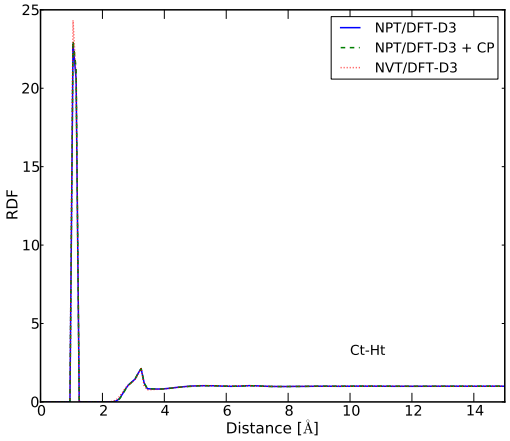
FIG. S.9.: Total (inter + intra) radial distribution functions of THF.

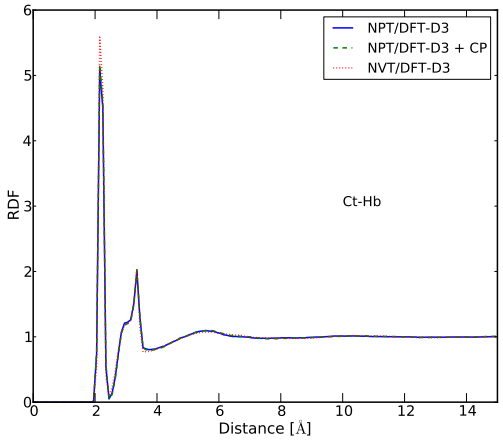


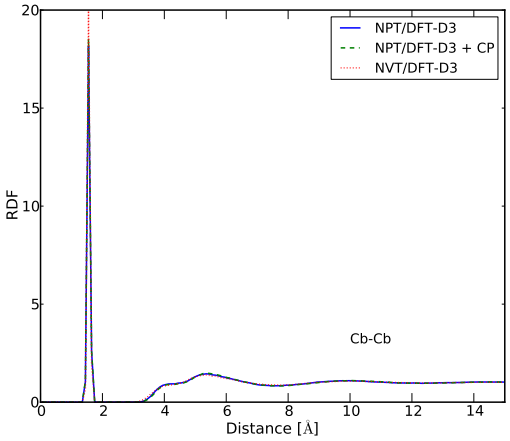
Continued on next page

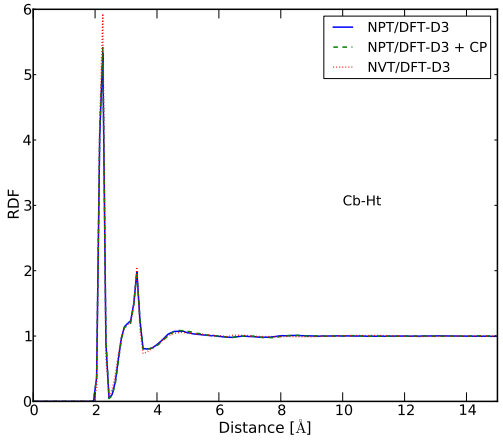
FIG. S.9 – *Continued from previous page*

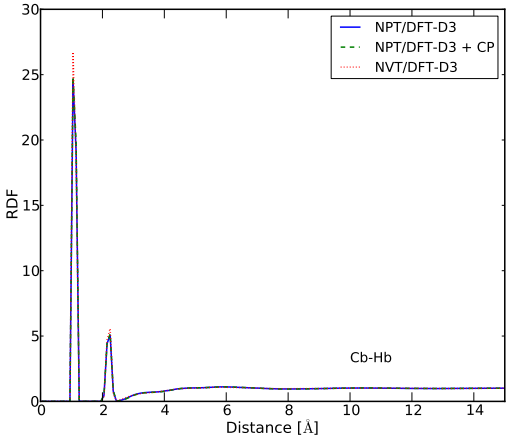






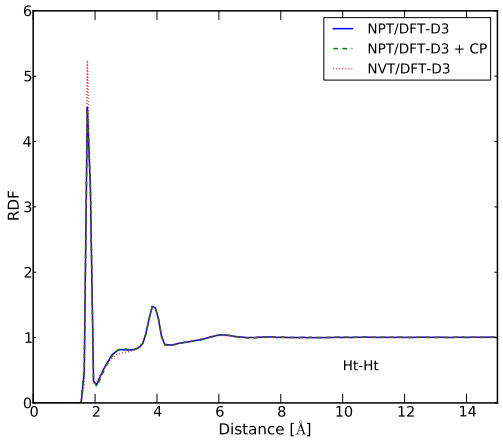


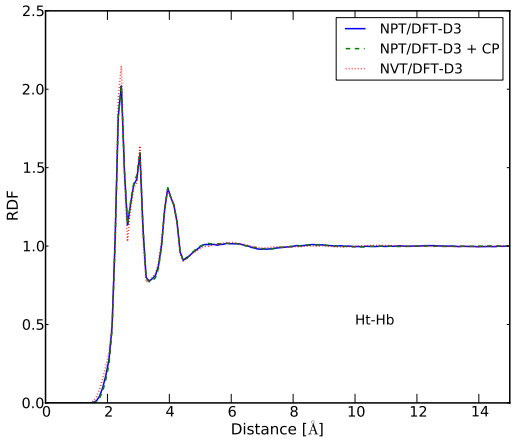




Continued on next page

FIG. S.9 – *Continued from previous page*





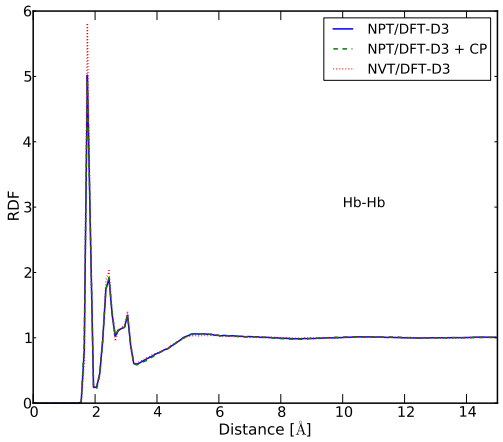
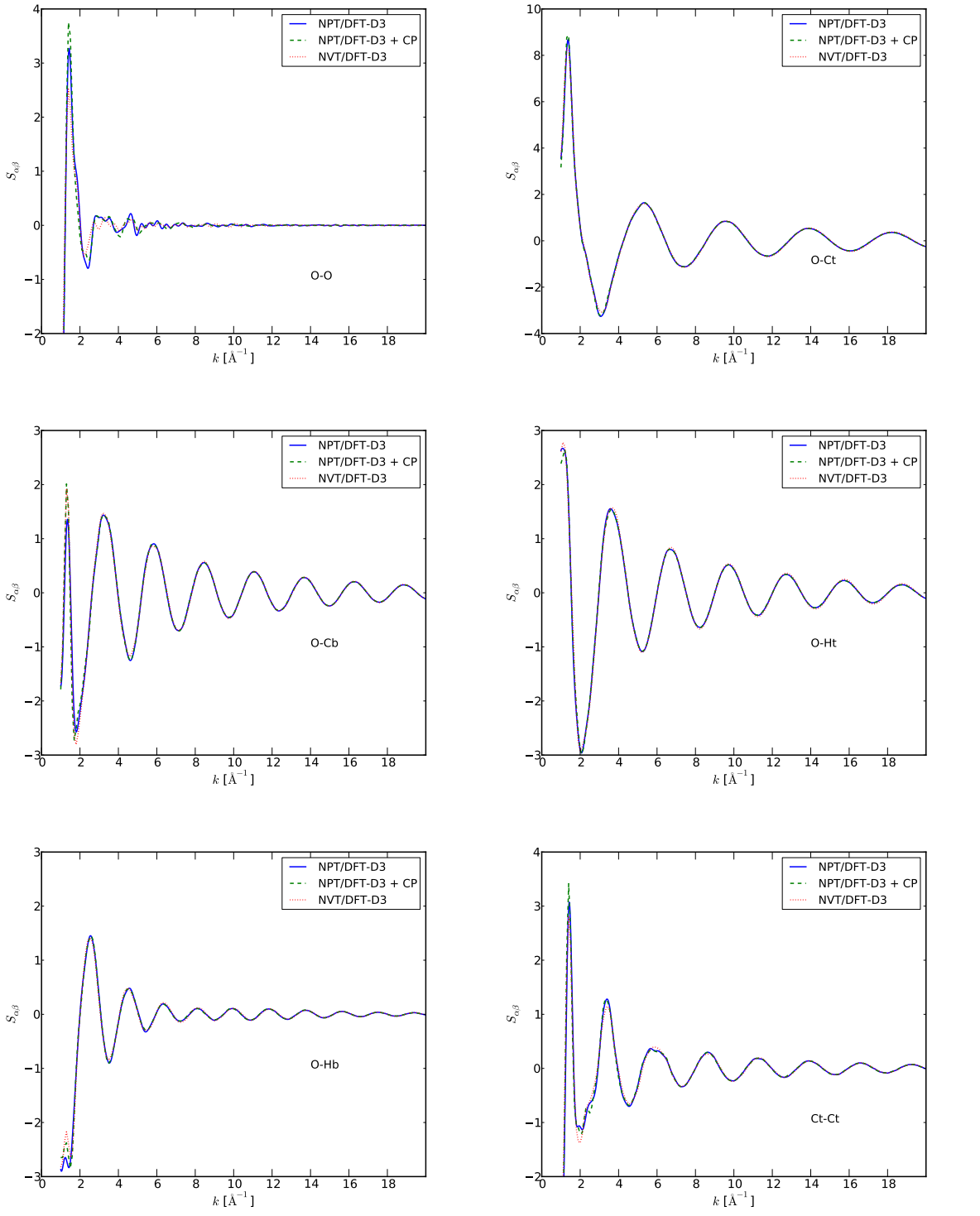
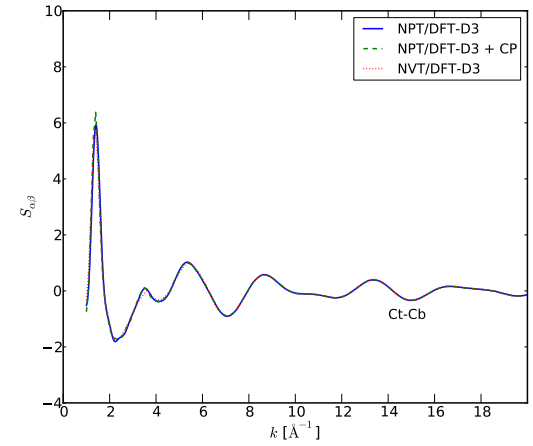


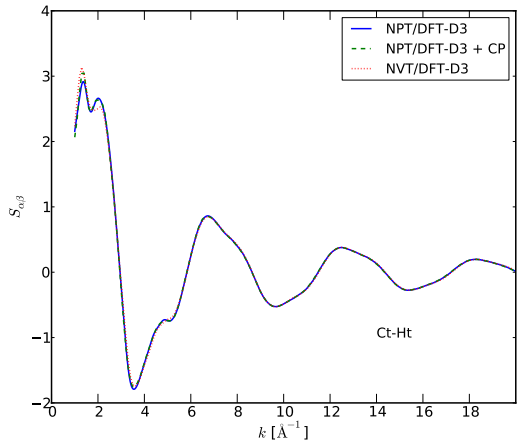
FIG. S.10.: Partial structure factors of the total RDFs of THF.

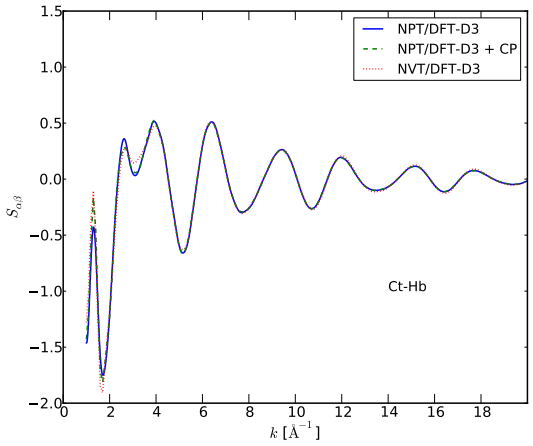


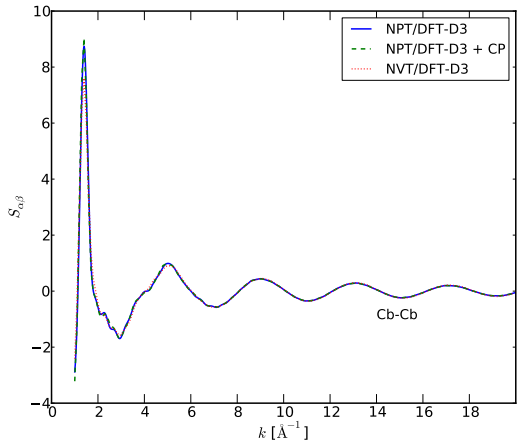
Continued on next page

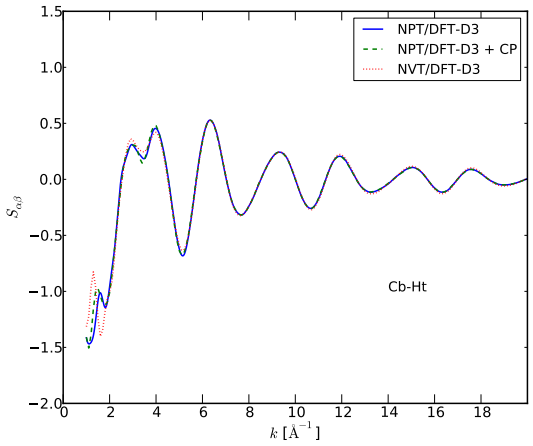
FIG. S.10 – *Continued from previous page*

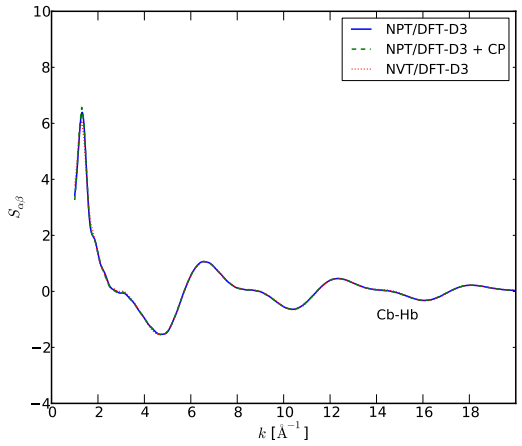






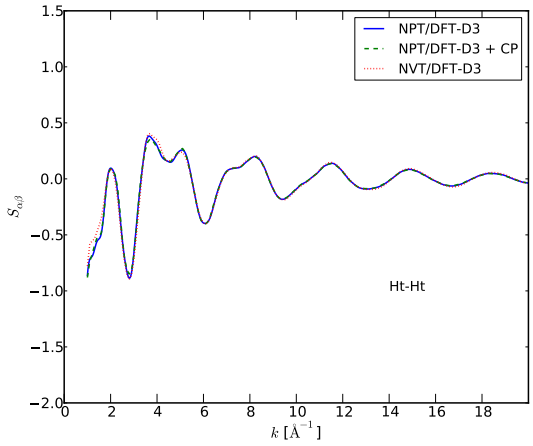


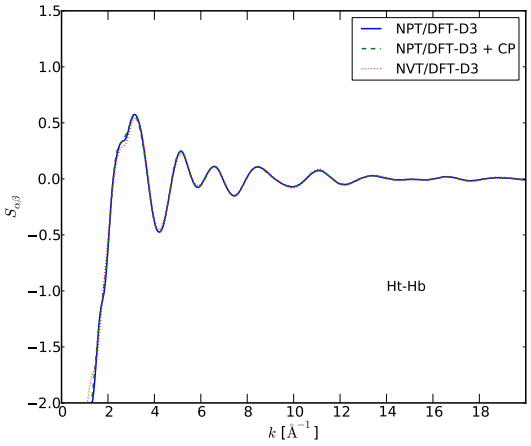




Continued on next page

FIG. S.10 – *Continued from previous page*





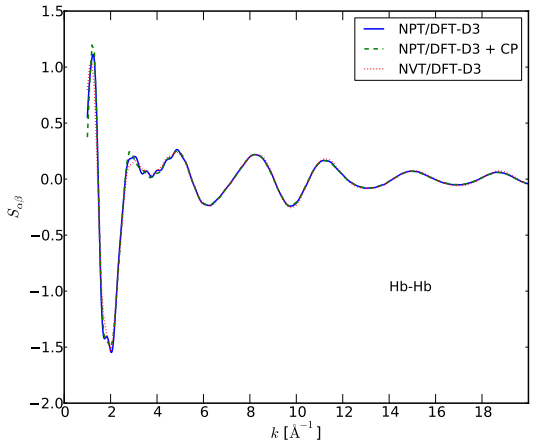
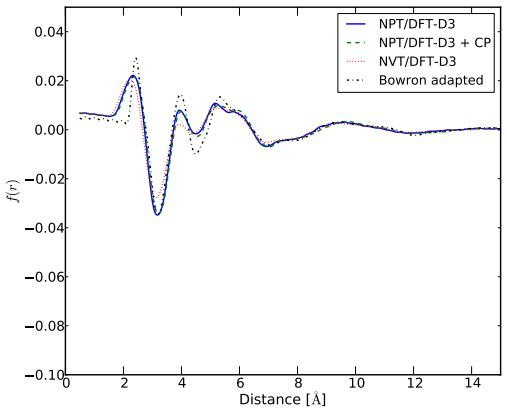
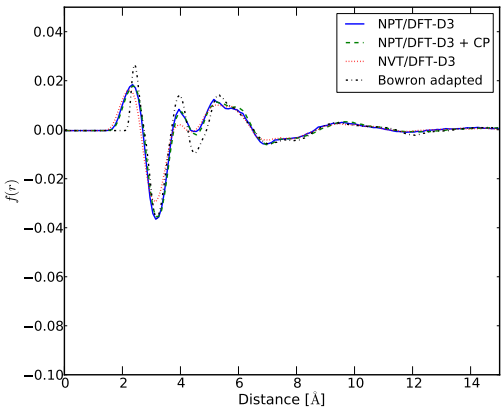


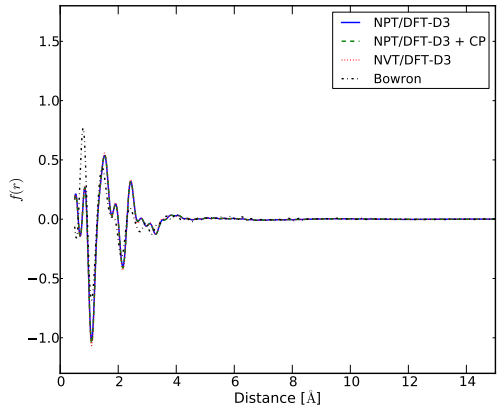
FIG. S.11.: Comparison of Fourier-based approach (applying Eqs. (2), (3), (4) on g) in the left column, with a direct weighted summation of all $g_{\alpha\beta}$ in the right column (Eq. (5)) for THF.



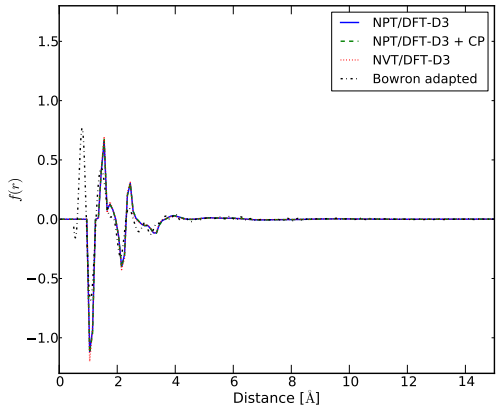
Total intermolecular SF with Eqs. (2), (3), (4)



Total intermolecular SF with Eq. (5)



Total inter- and intramolecular SF with Eqs. (2), (3), (4)



Total inter- and intramolecular SF with Eq. (5)