JCTC Journal of Chemical Theory and Computation

LETTER

pubs.acs.org/JCTC

Singular Value Decomposition for Analyzing Temperature- and Pressure-Dependent Radial Distribution Functions: Decomposition into Grund RDFs (GRDFs)

Philipp J. di Dio,* Martin Brehm, and Barbara Kirchner

Wilhelm-Ostwald-Institut für Physikalische und Theoretische Chemie, Universität Leipzig, Linnéstr. 2, D-04103 Leipzig, Germany

ABSTRACT: Singular value decomposition paves the way for systematic investigations of temperature- and pressure-dependent radial distribution functions. The decomposition into (weighted) Grund radial distribution functions (GRDF) shows that the temperature-dependent water structure can easily be understood by only three contributions: a major temperature-independent contribution from the first GRDF, a major temperature-dependent contribution from the second GRDF, and a minor temperature-dependent fine structure contribution from the third GRDF.

We present a new and unbiased way of analyzing temperature as well as pressure- or density-dependent radial distribution functions (RDFs). For many decades, radial distribution functions of liquids, especially water, were investigated experimentally as well as theoretically to gain information about the structure of liquids and, in special cases, their temperature as well as pressure dependencies.[1,2] For instance, the two-state model of water[2] is based on such investigations. Unfortunately, up to now, the analyses of the temperature- and pressure-dependent RDFs were restricted to the changing forms of the RDFs and the actual measured phase point.

Our approach of analyzing temperature- and pressure-dependent radial distribution functions is to use the singular value decomposition (SVD) of a matrix. The matrix **A** is formed by the temperature- and/or pressure-dependent RDFs (columns a_i of **A**) and then is decomposed with the SVD to determine the linear dependency of the RDFs. We found that five O−O RDFs (obtained from ab initio molecular dynamics simulations of water at 300, 350, 400, 600, and 1000 K) are, in a first approximation, a linear combination of only two (main) contributions:

$$g_{\mathrm{O-O}}(r,T) \approx v_1(T)\cdot w_1(r) + v_2(T)\cdot w_2(r) \tag{1}$$

The analysis of our set of RDFs shows that $v_1(T)$ is almost constant, and the whole temperature dependency is condensed in $v_2(T)\cdot w_2(r)$. A possible application could lie in coarse-grained force fields, which often suffer from a lack of temperature transferability of the iterative Boltzmann inversion.[3] The iterative Boltzmann inversion iteratively constructs potentials $V_i(r)$ with

$$V_{i+1}(r) = V_i(r) + k_{\mathrm{B}}T \ln \frac{g^i(r)}{g(r)} \tag{2}$$

which reproduce $g(r)$ at one single temperature. We will therefore present a simple method of how RDFs can be decomposed as in eq 1, so that in a further step the fitting procedure (eq 2) can be expanded to temperature-dependent RDFs over a large temperature range. Other applications of the SVD in signal processing[4] and even in biology[5] have been reported.

The aim of the singular value decomposition is the following. Supposing we have a real matrix $\mathbf{A} = (a_1, a_2, ..., a_n) \in \mathbb{R}^{m\times n}$ with $m,n \in \mathbb{N}$ and $m \geq n$, i.e., $a_i = (a_{1i}, ..., a_{mi})^{\mathrm{T}} \in \mathbb{R}^m$ for all $i = 1, ..., n$. The identification of **A** as a compact operator $\mathbf{A}: \mathbb{R}^n \rightarrow \mathbb{R}^m$ between the two Hilbert spaces $\mathbb{R}^m$ and $\mathbb{R}^n$ with the standard scalar product $\langle \cdot \mid \cdot \rangle$ leads to the application of a well-known theorem from functional analysis.[6] It states in our case that every matrix **A** is a product of three matrices, $\mathbf{U} \in \mathbb{R}^{m\times n}$, $\Sigma \in \mathbb{R}^{n\times n}$, and the transpose of $\mathbf{V} \in \mathbb{R}^{n\times n}$:

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^{\mathrm{T}} \tag{3}$$

The matrices $\mathbf{U} = (u_1, u_2, ..., u_n)$ and $\mathbf{V} = (v_1, v_2, ..., v_n)$ are orthogonal, and Σ is a diagonal matrix:

$$\Sigma = \mathrm{diag}(s_1, ..., s_n) = \begin{pmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & s_n \end{pmatrix} \tag{4}$$

with $s_1 \geq s_2 \geq ... \geq s_n \geq 0$. This in only one out of several equivalent formulations, but the most useful for our purpose. The s_i are called singular values of **A**. In this decomposition, the singular values of the matrix **A** are unique. However, the matrices **U** and **V** are not unique. If there are some degenerate singular values, for instance $s_1 = s_2$, then the vectors u_1, u_2, v_1, and v_2 are unique up to an orthogonal transformation. In the case of no degeneracy, i.e., $s_1 > s_2 > ... > s_n \geq 0$, all vectors are unique up to their sign, i.e., an orthogonal transformation on a one-dimensional subspace. However, the nondegeneracy is mostly the case, especially in applications.

From a geometrical point of view, **U**Σ as well as **A** map the $(n-1)$-dimensional unit sphere

$$S^{n-1} := \{x = (x_1, ..., x_n)^{\mathrm{T}} \in \mathbb{R}^n \mid ||x||_2^2 = x_1^2 + ... + x_n^2 = 1\} \tag{5}$$

Received: May 18, 2011
Published: August 31, 2011

dx.doi.org/10.1021/ct2003385 | J. Chem. Theory Comput. 2011, 7, 3035–3039

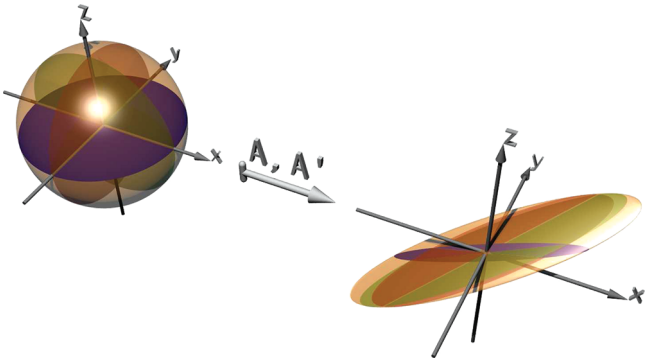

Figure 1. Demonstration and geometrical interpretation of the singular value decomposition (SVD) of two matrices $\mathbf{A} = (a_1, a_2)$ and $\mathbf{A}' = (a_1, a_2, a_3)$ with $a_1 = (0.8,0.9,0.6)^T$, $a_2 = (0.4,0.7,0.2)^T$, and $a_3 = (0.1,0.6,0.0)^T$ as an example of two maps. On the left, the blue disk represents the unit sphere S^1 in the $x-y$ plane, and the orange sphere is the unit sphere S^2. On the right, both unit spheres are deformed by the maps $\mathbf{A}$ and $\mathbf{A}'$, respectively, with the singular values $s_1 \approx 1.55770$, $s_2 \approx 0.20571$, $s'_1 \approx 1.64630$, $s'_2 \approx 0.39792$, and $s'_3 \approx 0.03664$.

to an ellipsoid

$$E_{\mathbf{A}}^{n-1} = \{x_1 s_1 u_1 + \ldots + x_n s_n u_n | x_1^2 + \ldots + x_n^2 = 1\} = \{x_1 a_1 + \ldots + x_n a_n | x_1^2 + \ldots + x_n^2 = 1\} \subset \mathbb{R}^m \quad (6)$$

This is demonstrated by two matrices $\mathbf{A}$ and $\mathbf{A}'$ in Figure 1 with $m = 3$ and $n = 2$ or 3. In Figure 1, the unit sphere S^1 (blue disk, left) is mapped to a one-dimensional ellipsoid in $\mathbb{R}^3$ (blue disk, right), and S^2 is mapped to a two-dimensional ellipsoid (both orange). As seen from this simple illustration, we find that the vectors u_i of $\mathbf{U}$ are the semiaxes of the ellipsoid $E_{\mathbf{A}}^{n-1}$ and the singular values s_i are their length. From the definition of $E_{\mathbf{A}}^{n-1}$, we find the relations

$$s_1 = \max_{x \in E_{\mathbf{A}}^{n-1}} ||x||_2 \text{ and } s_n = \min_{x \in E_{\mathbf{A}}^{n-1}} ||x||_2 \quad (7)$$

i.e., the greatest and the smallest distance from the origin.

If we compare both matrices $\mathbf{A}$ and $\mathbf{A}'$, only the vector a_{n+1} is "added" to $\mathbf{A}$ ($n = 2$), i.e., $\mathbf{A}' = (A, a_{n+1}) = (a_1, a_2, \ldots, a_{n+1})$. The SVD of $\mathbf{A}'$ gives three other matrices $\mathbf{U}'$, Σ', and $\mathbf{V}'$ with $\mathbf{A}' = \mathbf{U}'\Sigma'(\mathbf{V}')^T$. In general, all vectors of the primed matrices are different from the nonprimed vectors. For instance, see Figure 1, where the singular values s_1 and s_2 of the blue one-dimensional ellipsoid on the right differ from the three singular values s'_1, s'_2, and s'_3 of the orange two-dimensional ellipsoid. However, we have the relation

$$E_{\mathbf{A}}^{n-1} \subseteq E_{\mathbf{A}'}^{n} \quad (8)$$

which leads to the relation between the largest singular value s_1 of $\mathbf{A}$ and s'_1 of $\mathbf{A}'$:

$$s'_1 = \max_{x \in E_{\mathbf{A}'}^{n}} ||x||_2 \geq \max_{x \in E_{\mathbf{A}}^{n-1}} ||x||_2 = s_1 \quad (9)$$

as well as the smallest singular value s_n of $\mathbf{A}$ and s'_{n+1} of $\mathbf{A}'$:

$$s'_{n+1} = \min_{x \in E_{\mathbf{A}'}^{n}} ||x||_2 \leq \min_{x \in E_{\mathbf{A}}^{n-1}} ||x||_2 = s_n \quad (10)$$

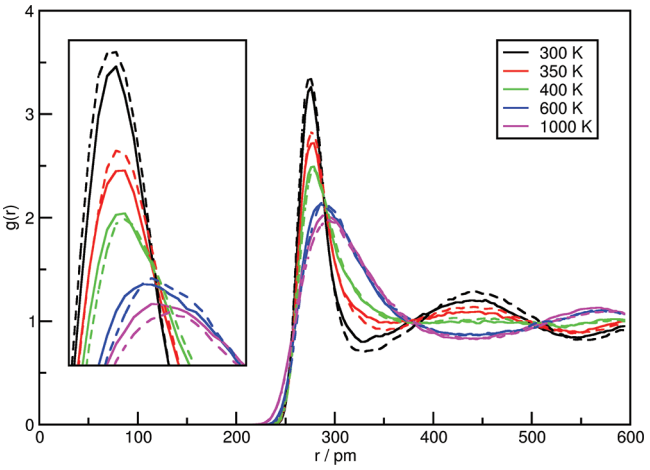

Figure 2. Two sets of temperature-dependent O−O radial distribution functions of water from a previous study[7] used as examples in the present study. The solid line corresponds to the ruthenium(VI) ester simulation and the dashed line to the ruthenium(VIII) ester simulation.

Table 1. Singular Values s_i from the Decomposition of the Raw RDFs in Figure 3 for both Water Simulations

	simulation sets	
	Ru(VI) ester	Ru(VIII) ester
s_1	29.12231	29.07127
s_2	4.17852	4.87964
s_3	0.41898	0.61054
s_4	0.18050	0.19100
s_5	0.13057	0.10789

Both are useful relations when additional data are used for the analysis.

In order to apply the singular value decomposition to temperature- and pressure-dependent radial distribution functions, the RDFs led to the construction the matrix $\mathbf{A}$. Each RDF is a single vector, i.e., column $a_i = g(r,T_i)$ of $\mathbf{A}$. The matrix has the form $\mathbf{A} = (g(r_i,T_j))_{i=1,\ldots,m;j=1,\ldots,n}$ ($m \geq n$). To demonstrate this, we use the oxygen−oxygen RDFs of water from a previous temperature-dependent ab initio molecular dynamics simulation[7] where $m = 199$ and $n = 5$, see Figure 2. Two series of five ab initio molecular dynamics simulations (CP2k,[8] Nosé−Hoover chain thermostat,[9] BLYP-D[10]) were performed with two different ruthenium esters together with 60 water molecules with the temperature ranging from 300 to 1000 K at a constant volume (1203.8 and 1223 pm box length, i.e., ρ = 1.212 and 1.170 g cm^{-3}; the ruthenium esters increase the density). The simulation time for each temperature was at least 20 ps. The trajectories were analyzed with TRAVIS.[11] For the singular value decomposition, we used the algorithm published by Golub and Reinsch in 1970.[12] Additionally, the singular values s_i are sorted in descending order, and the signs of the vectors $u_i = (u_{1i}, \ldots, u_{mi})^T$ and v_i were changed to satisfy

$$\sum_{j=1}^{m} u_{ji} \geq 0 \quad (11)$$

for all $i = 1, \ldots, n$. This was done in order to obtain maximal positive u_i.

dx.doi.org/10.1021/ct2003385 |J. Chem. Theory Comput. 2011, 7, 3035–3039

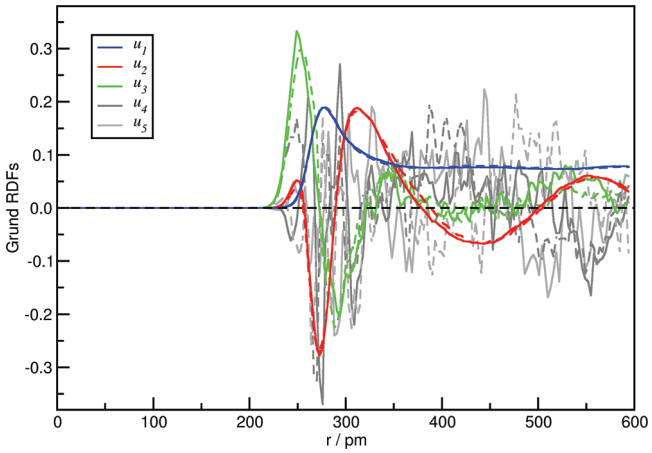

Figure 3. Resulting Grund RDFs u_1-u_5 from the singular value decomposition of the raw RDFs from Figure 2 with the corresponding singular values s_i in Table 1.

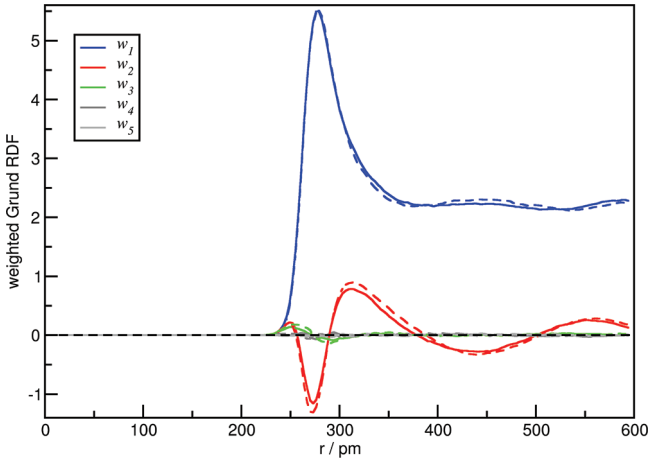

Figure 4. Weighted Grund RDFs w_1-w_5 from the singular value decomposition of the raw RDFs from Figure 2.

Applying the SVD to our RDFs, we find the singular values s_i in Table 1. Comparing both simulations, their corresponding singular values are similar. In both cases, the first two singular values are well separated from the last three, which are at least 1 order of magnitude smaller. Because **U** and **V** are orthogonal, their vectors u_i and v_i are normalized, and the singular values s_i represent the weighting factors in the linear combinations gaining the raw RDFs. Because of the small values of s_3, s_4, and s_5 compared to s_1 and s_2, the raw RDFs in Figure 2 are essentially linear combinations of only two main contributions. The lower singular values only represent noise, which can be filtered out by our approach. Furthermore, from eq 9, we find that with an increasing number of RDFs (spectra), the largest singular value never decreases, while from eq 10, we find that the smallest singular value never increases. Therefore, with an increasing number of RDFs or other data sets, the signal to noise ratio becomes better in contrast to traditional interpretations, where an increased number of spectra only leads to problems in the graphical representation and the noise can only be reduced by better statistics, i.e., longer simulation time for each spectra.

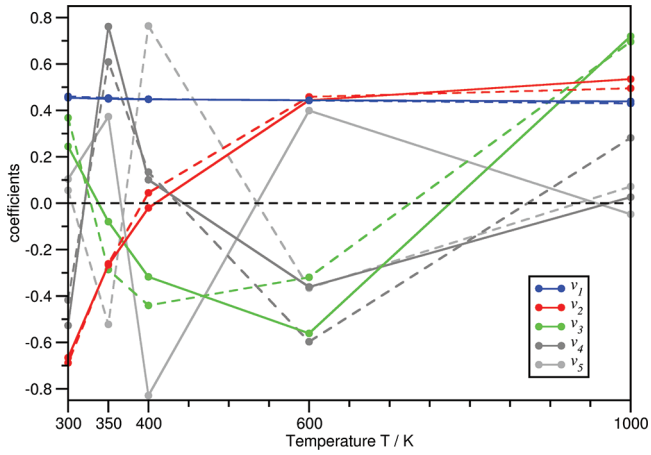

Figure 5. Temperature-dependent development of the coefficients v_1-v_5 in **V** for both simulation sets (Ru(VI) simulation, solid lines; Ru(VIII) simulation, dashed lines).

The Grund RDFs are the columns u_i of **U**, see Figure 3, while the weighted GRDFs (Figure 4) are the columns of $\mathbf{W} = \mathbf{U} \cdot \Sigma$, i.e., $w_i = s_i \cdot u_i$, and they represent the different contributions in a clearer way. In both representations, the first three GRDFs with the highest singular values are colored, while the GRDFs with the smaller singular values (minor GRDFs) are gray. In our application to the RDFs at five temperatures, we observe that the main GRDF (highest singular value of approximately 29) is very smooth and represents the water structure where no second hydration shell occurs. The second GRDF represents the main contributions to the temperature change at a constant density of the raw RDFs. Essentially, all information about the temperature dependence of all five raw RDFs are condensed into one (weighted) GRDF (the second one) and its temperature-dependent coefficients. The third GRDF seems to be a fine structure for the temperature dependence affecting only the first hydration shell up to 325 pm, see Figure 4. At larger distances, it contains much noise. Obviously, this needs further investigation because of the insufficient statistics. In our application, the last two GRDFs are considered statistical noise.

The temperature-dependent coefficients of the GRDFs (columns v_i of **V**) are shown in Figure 5. The coefficients v_1 of the main GRDF (blue) change slightly over the whole temperature range from 300 to 1000 K. Therefore, in our application, the main GRDF seems to be a constant contribution to all water structures. A systematic change of the coefficients is found for the second GRDF, see v_2 (red curve) in Figure 5. With increasing temperature, the coefficients rise from ∼− 0.7 at 300 K up to ∼0.6 at 1000 K. This behavior is in accordance with the previous interpretation that the main temperature changes are condensed in the second GRDF. The second GRDF has negative values in the region of the second hydration shell. Values of v_2 smaller than zero characterize a phase with a second hydration shell for the water molecules. But with increasing temperature, v_2 increases and characterizes a vanishing second hydration shell. In the general case, the coefficients and GRDFs are temperature- as well as density-dependent, as seen from the comparison in Figure 5. For a better interpolation, additional temperatures are needed. In general, interpolation of the coefficients at temperatures

dx.doi.org/10.1021/ct2003385 |J. Chem. Theory Comput. 2011, 7, 3035–3039

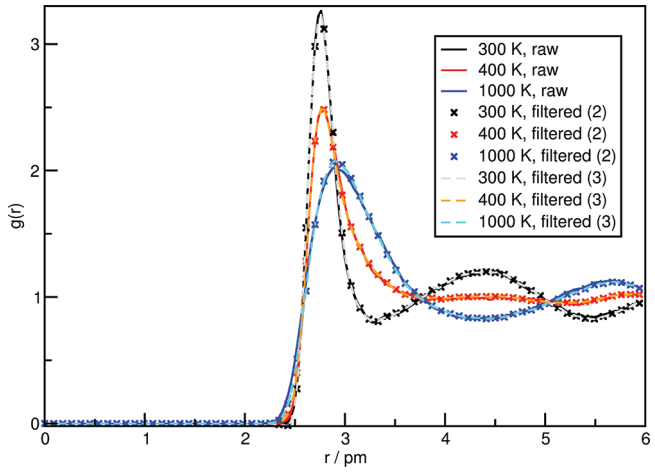

Figure 6. Reconstruction of three raw RDFs (Figure 2) at 300, 400, and 1000 K using only the contributions from the two or three first GRDFs and without the minor contributions from the remaining GRDFs (noise).

between the simulated temperatures is feasible and leads to the possibility of predicting all RDFs in a given temperature range and therefore predicting all averages depending on $g(r)$:[13]

$$\langle A \rangle = 2\pi N\rho \int_0^\infty a(r)\, g(r)\, r^2\, \mathrm{d}r = \sum_i v_i(T) \cdot 2\pi N\rho \int_0^\infty a(r)\, w_i(r)\, r^2\, \mathrm{d}r = \sum_i v_i(T) \cdot A_i \quad (12)$$

as well as their temperature derivatives. Additionally, the irregular temperature development of v_3, v_4, and especially v_5 is in accordance with its interpretation as noise from the statistics. The reconstruction of the temperature-dependent RDFs without the noise from u_3, u_4, and u_5 is shown in Figure 6. The reconstructed RDFs are not changed significantly in comparison to the raw RDFs in Figure 2, where the noise (u_4, u_5) is left out. At 300 and 400 K, the raw and filtered RDF coincide over the full range, while at 1000 K the missing contribution from the fine structure (u_3) slightly affects the first hydration shell. The contribution from the fine structure u_3 restores the first hydration shell at 1000 K, and the raw and reconstructed RDFs coincide over the whole range. As seen from the RDFs in Figure 2, our approach is not restricted by describing only the temperature changes accurately, because similar changes of the RDFs (or even other functions like autocorrelation functions etc.) are also found for many other continuous dependencies, and SVD can therefore be applied.

In summary, we presented a simple, systematic, and unified approach for analyzing temperature- and pressure-dependent radial distribution functions and structures from the singular value decomposition. Its generality is also suited for the analysis of other dependencies and functions. Furthermore, it could be useful for the comparison of molecular dynamics methods like force field techniques with different force fields and ab initio techniques with different DFT functionals, especially by explaining temperature- and pressure-induced structural changes. The main advantages of our approach are that an increasing number of RDFs systematically improves the results, as seen from eqs 9 and 10, and that the comparison and the analysis of a large number of RDFs (spectra) are easier because all RDFs are reduced to some main GRDFs and the minor GRDFs can be left out.

The main GRDFs contain the main contributions and are easier to handle. The minor GRDFs are found to contain the statistical noise only, which can be filtered out by our approach. We investigated five raw oxygen−oxygen RDFs of water from ab initio molecular dynamics simulations and found that all RDFs are essentially a linear combination of only two GRDFs. The temperature-dependent problem reduced therefore from five RDFs (spectra) to only two GRDFs (spectra) and their temperature-dependent coefficients. The remaining three degrees of freedom are found to be statistical noise, which was filtered out without significantly changing the resulting raw RDFs. In AIMD simulations, the simulation time is very limited. Unfortunately, this results in unsatisfactory statistics, a major drawback for extensive temperature studies with AIMD simulations. This drawback is overcome by our approach because the number of spectra and, therefore, the overall simulation time (not only the simulation time at each temperature) improves the results and the statistics, and intermediate temperatures can be interpolated. Therefore, our approach reveals new ways of investigating and interpreting the temperature and pressure dependence of molecular systems.

AUTHOR INFORMATION

Corresponding Author
*E-mail: didio@uni-leipzig.de.

ACKNOWLEDGMENT

This work was supported by the DFG, in particular by the project KI 768/6-1, graduate school BuildMoNa, and the ESF. Computer time from the RZ Leipzig was gratefully acknowledged.

REFERENCES

(1) (a) Frauenfelder, H.; Petsko, G. A.; Tsernoglou, D. *Nature* **1979**, *280*, 558–563. (b) Bertagnolli, H. *Angew. Chem., Int. Ed.* **1992**, *31*, 1577–1578. (c) Corongiu, G.; Clementi, E. *J. Chem. Phys.* **1992**, *97*, 2030–2038. (d) Laasonen, K.; Sprik, M.; Parrinello, M.; Car, R. *J. Chem. Phys.* **1993**, *99*, 9080–9089. (e) Mancera, R. L.; Buckingham, A. D. *Chem. Phys. Lett.* **1995**, *234*, 296–303. (f) Mancera, R. L.; Buckingham, A. D. *J. Phys. Chem.* **1995**, *99*, 14632–14640. (g) Sprik, M.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **1996**, *105*, 1142–1152. (h) Yoshii, N.; Yoshie, H.; Miura, S.; Okazaki, S. *J. Chem. Phys.* **1998**, *109*, 4873–4884. (i) Robinson, G. W.; Cho, C. H.; Urquidi, J. *J. Chem. Phys.* **1999**, *111*, 698–702. (j) Sorenson, J. M.; Hura, G.; Glaeser, R. M.; Head-Gordon, T. *J. Chem. Phys.* **2000**, *113*, 9149–9161. (k) Handgraaf, J.-W.; van Erp, T. S.; Meijer, E. J. *Chem. Phys. Lett.* **2003**, *367*, 617–624. (l) Hura, G.; Russo, D.; Glaeser, R. M.; Head-Gordon, T.; Krack, M.; Parrinello, M. *Phys. Chem. Chem. Phys.* **2003**, *5*, 1981–1991. (m) van Erp, T. S.; Meijer, E. J. *J. Chem. Phys.* **2003**, *118*, 8831–8840. (n) Mantz, Y. A.; Chen, B.; Martyna, G. J. *Chem. Phys. Lett.* **2005**, *405*, 294–299. (o) Lee, H.-S.; Tuckerman, M. E. *J. Chem. Phys.* **2006**, *125*, 154507:1–14. (p) Lehmann, S. B. C.; Spickermann, C.; Kirchner, B. *J. Chem. Theory Comput.* **2009**, *5*, 1640–1649. (q) Lehmann, S. B. C.; Spickermann, C.; Kirchner, B. *J. Chem. Theory Comput.* **2009**, *5*, 1650–1656. (r) Guse, C.; Simionescu, A.; Schünemann, B.; Hentschke, R. *J. Phys.: Condens. Matter* **2010**, *22*, 325105:1–14.

(2) (a) Urquidi, J.; Singh, S.; Cho, C. H.; Robinson, G. W. *Phys. Rev. Lett.* **1999**, *83*, 2348–2350. (b) Head-Gordon, T.; Hura, G. *Chem. Rev.* **2002**, *102*, 2651–2670.

(3) Carbone, P.; Varzaneh, H. A. K.; Chen, X.; Müller-Plathe, F. *J. Chem. Phys.* **2008**, *128*, 064904:1–11.

(4) (a) Biglieri, E.; Yao, K. *Signal Process* **1989**, *18*, 277–289. (b) Edfors, O.; Sandell, M.; van de Beek, J.-J.; Wilson, S.; Borjesson, P. *IEEE Trans. Commun.* **1998**, *46*, 931–939.

(5) (a) Alter, O.; Brown, P. O.; Botstein, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10101–10106. (b) Troyanskaya, O.; Cantor, M.; Sherlock, G.; Brown, P.; Hastie, T.; Tibshirani, R.; Botstein, D.; Altman, R. B. *Bioinformatics* **2001**, *17*, 520–525.

(6) (a) Reed, M.; Simon, B. *Methods of Modern Mathematical Physics: Functional Analysis*; Academic Press: San Diego, CA, 1980; Vol. 1, Theorem VI.17. (b) Werner, D. *Funktionalanalysis*, 6th ed.; Springer-Verlag: Berlin, Germany, 2007; Satz VI.3.6.

(7) di Dio, P. J.; Brehm, M.; Kirchner, B. Temperature Dependent Hydration of Ruthenium(VI)-dioxo-2,5-dioxolane and Ruthenium-(VIII)-trioxo-2,5-dioxolane: An Ab Initio Molecular Dynamics Study. Submitted.

(8) (a) Lippert, G.; Hutter, J.; Parrinello, M. *Theor. Chem. Acc.* **1999**, *103*, 124–140. (b) CP2k developers group under the terms of the GNU General Public License. See: CP2P Developers Home Page. http://cp2k.berlios.de/ (accessed Jul 26, 2011).

(9) (a) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519. (b) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697. (c) Martyna, G. J.; Klein, M. L.; Tuckerman, M. E. *J. Chem. Phys.* **1992**, *97*, 2635–2643.

(10) (a) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100. (b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789. (c) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463–1473. (d) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787–1799.

(11) Brehm, M.; Kirchner, B. J. *Chem. Inf. Model.* **2011**, *51*, 2007–2023.

(12) Golub, G. H.; Reinsch, C. *Numer. Math.* **1970**, *14*, 403–420.

(13) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, Great Britain, 1987.