Economic simplex optimisation for broad range property prediction: strengths and weaknesses of an automated approach for tailoring of parameters

Thomas J. Müller [a], Sudip Roy [a], and Wei Zhao [a],

[a] *Institut für Theoretische Physikalische Chemie, Technische Universität Darmstadt, Petersenstr.20, 64287 Darmstadt, Germany*

Astrid Maaß [b,*] and Dirk Reith [b]

[b] *Fraunhofer Institute for Scientific Computing and Algorithms, Schloss Birlinghoven, 53754 Sankt Augustin, Germany*

Thomas Müller, Sudip Roy, Wei Zhao, and Astrid Maaß contributed equally to this work.

Abstract

The need for predicting physical compound properties supplementing experimental data is considerable. Nowadays a wide range of classical simulation techniques is available for computing a multitude of such properties with acceptable effort. We here give a field report about our approach, which was to fit an initial model to a single point in the phase diagram. By way of accessing commonly available experimental values we developed a compound specific force-field via simplex optimization. For predicting the desired properties of the novel model we did engage classical equilibrium as well as reverse non-equilibrium molecular dynamics in combination with Monte Carlo methods and report here the performance of these method in detail for the example compound ethylene oxide. We find that the new model describes the experimentally observed behavior of the test compound ethylene oxide (EO) very well in the molecular dynamics section. However, when applying the simplex optimized model to the Monte Carlo section, the limits of transferability become apparent.

Key words: physical property prediction, Molecular dynamics, simplex optimization, YASP, Monte Carlo

* Corresponding author.
Email address: astrid.maass@scai.fraunhofer.de (Astrid Maaß).

1 Introduction

Highly reactive compounds are frequently involved in industrial processes. While reactivity is a desirable feature, it requires extremely careful handling and thus prohibits experiments that are dispensable at any means. Still, in daily practice the need for property data is pressing. Simulation data are useful to fill this gap. This was the motivation of the Industrial Fluid Properties Simulation Collective (IFPSC) for announcing the 4th challenge, when the ethylene oxide (EO) molecule was chosen as a relevant example compound for testing the transferability of methods and force fields to predict a wide range of properties [1].

When participating in this challenge, our main objective was to probe the capability of our simplex optimization procedure for deriving such individually tailored compound models by fitting them to very few basic experimental parameters. According to the contest rules, the results for the created models are compared to those obtained for a previously described reference model [2].

Atomistic soft matter simulations are most commonly used for investigating system properties that depend on intermolecular interactions for which the required large system size prohibits quantum mechanical methods [3]. Simplifying assumptions have to be made that lead to a limited number of analytic terms in order to construct a force field [4]. Compound-specific behavior is incorporated by the individual choice of the force field parameters for each atom instead of considering the electronic structure. A variety of very general force field parameters (AMBER [5], CHARMM [6], GROMOS [7], OPLS [8], UFF [9], others) that describe the behavior of a wide range of compounds globally quite well exists, but these force fields may lack the required accuracy for a specific compound [10]. Individual parameter sets created by quantum mechanical methods generally describe very well the intramolecular properties, but may not sufficiently reproduce *inter*molecular properties due to the inherent limitation of the underlying models [11].

Therefore, our approach to develop an all-atom force-field model of EO that allows extensive simulation studies is to follow a generic, most simple route: starting from the ab initio level and adjusting the model to very few, commonly accessible experimental data.

Thus, as the chosen optimization procedure is quite consumptive, the model is being calibrated with respect to only one point in the phase diagram, i.e. to the requested temperature of 375 K and the associated experimental value for the pressure. The properties we address comprise the densities, heat capacities, isothermal compressibilities, viscosities and thermal conductivities for both the saturated liquid phase and the vapor phase at 375 K, as well as information

about the vapor pressure, the heat of vaporization, and critical properties. As we aim at predicting such a wide spectrum of diverse properties, especially including transport behavior, we opt for an all-atom model that provides the maximum number of internal degrees of freedom, in order to capture friction and exchange of thermal energy comprehensively.

2 Methods and Computational Details

The requested types of properties range from phase equilibria to caloric and dynamic properties. Especially for the latter, the simulation technique of choice is molecular dynamics simulation (MD). For all MD simulations the program package YASP [12] was utilized. Transport properties were determined using reverse non equilibrium molecular dynamics (RNEMD)[13], a comparatively fast converging method, likewise implemented in YASP. When needed, supplementary Monte Carlo simulations were performed with the MCCCS (Monte Carlo for Complex Chemical Systems) Towhee simulation program, version 5.2.1 [14].

Presuming a reasonable molecular geometry, the diverse physical properties mainly depend on the non-bonded interactions between the particles, which are modeled by Lennard-Jones (LJ) and electrostatic interactions in the classical approach. Thus, the LJ parameters (distance of zero-potential σ and potential-well-depth ε) are subject to modification, whereas the atomic charges were kept fixed.

In order to keep not only the computational demand low, but also the optimization problem feasible, we decided to restrict the model to only two atom types involved in LJ-interactions. Considering the numerical proportion of hydrogen to carbon (being 2:1), we neglected the hydrogen atoms consequently, thereby also maintaining comparability to the reference model.

2.1 Models

For starting the optimization procedure, an initial all-atom model representing ethylene-oxide was created with Gaussian 03 [15] using a 6-31G (d) basis set for Hartree-Fock calculations in order to provide a suitable set of atomic charges and appropriate bond and angle geometries. Two sets of charge models were derived from the electrostatic potential (ESP) and fit to the ESP at points selected according to the CHelp [16] and CHelpG [17] schemes. The second charge model emphasizes the negative character of the oxygen atom compared to that of the carbon atoms more, however the overall charge distribution

among the three atom types of EO is quite alike for both models. Finally, we used as partial charges the average atomic charges of both models. As EO is almost rigid due to its three-membered ring structure, throughout all simulations the bondlengths were kept fixed and only angles involving H-atoms were treated as flexible, the force constants being derived from the calculated vibrational spectrum. The LJ-parameters of oxygen and carbon are subject to variation during the optimization.

The reference or round robin model (RR) is labeled as model A by Wielopolski and Smith [2] and treats the methylene groups as united atoms. All employed modeling techniques create the LJ-parameters for mixed interactions according to the rules of Lorentz and Berthelot [18,19]. As implemented in YASP, we apply a reaction field to cope with long range Coulombic forces, whereas Towhee utilizes Ewald summation.

2.2 Simplex optimization

The chosen set of parameters is optimized according to an iterative algorithm, i.e. cycles of MD-simulation, evaluation of target properties, and eventually readjustment of parameters according to the downhill simplex method are performed repeatedly until the abort condition is reached [20], which was the relative quadratic error $((\rho - \rho^{target})/\rho^{target})^2 + ((\Delta H_{vap} - \Delta H_{vap}^{target})/\Delta H_{vap}^{target})^2)$ to fall below the chosen threshold.

The readjustment of parameters is achieved by reflection, expansion or contraction in parameter space [21]. This optimization method is robust and applicable even if an analytical functional form is not available, as is the case here.

The choice of parameters to optimize was based on two considerations. The number of optimization iterations to find a minimum increases exponentially with the number of parameters to optimize in one go. Therefore, we were limited to only few parameters. Test simulations (not shown here) with up to 6 parameters in one optimization have been performed. We decided to keep the values from the quantum calculations fixed and to concentrate on the optimization of other parameters. Besides this technical consideration, we intended to start with a model for EO comparable to the RR model and to refine it stepwise to a new model. Thus, we started experimentally by optimizing σ- and ϵ-values independently and finally went on with the here reported procedure of optimizing σ and ϵ at the same time. When trying to incorporate the optimization of charges the algorithm changed the charges significantly outside of their expected range of values (even inverting the polarity), without improving the overall performance of the force field with respect to the goals

of the challenge. Therefore we report here the results of the optimization of the LJ-parameters for the carbon and oxygen atoms.

For starting the simplex optimization procedure, initial values for the variables have to be provided manually. The values chosen should cover quite a large (but still reasonable) region in parameter space in order to avoid the results being confined in a local minimum all over the time. As the four-dimensional parameter-space needs five starting points for the simplex algorithm, we considered quadruples of values which cover the following spans: $\sigma_{CC} = 0.3 - 0.4$ nm, $\sigma_{OO} = 0.26 - 0.31$ nm, $\varepsilon_{CC} = 70 - 95$ K, $\varepsilon_{OO} = 60 - 80$ K.

Table 1 summarizes all the details of the final models that are used for property prediction. The optimized model differs from the reference model mainly in the presence of hydrogen atoms, i.e. in the distribution of charges, as the LJ parameters for H-atoms are set to zero. Besides that the actual values of σ and ε differ.

2.3 System setup and simulation conditions

2.3.1 Molecular dynamics

In general, uniform settings were used throughout all simulations in order to ensure utmost consistency. However, as the system size affects computation time, even more drastically when repeated iteratively, previous to the optimization process the finite size effect on the density was studied by running liquid phase MD simulations of cubes with 256, 512 and 1024 molecules (data not shown). As a consequence 256 molecules per box were considered sufficient for the equilibrium MD runs of the optimization procedure. Due to different requirements concerning system geometries for other settings (GEMC, NEMD), system sizes of 512 or 768 molecules were used for computing the heat capacities, transport properties or vapor-liquid-coexistence curves, respectively. The computational details for all simulations comprised a non-bonded cut-off of 0.9 nm and a neighbor list cut-off of 1.0 nm in case of the liquid phase systems and a non-bonded cut-off of 5.0 nm in case of the vapor phase systems with a neighbor list cut-off of 5.1 nm. The neighbor list was updated every 10 time steps, the reaction-field-dielectric was 14.5 in both cases. The temperature was coupled to 375 K using a Berendsen thermostat with a coupling constant of 0.2 ps while the pressure was coupled to 1428.5 kPa using a Berendsen manostat with a coupling time of 2 ps and a compressibility of $1*10^{-6}$ 1/kPa. The center of mass drift of the systems was removed every 1000 time steps. The frequency of snapshot collection and the instantly calculated properties was every 200 time steps. The timestep size was 2 fs throughout.

The simplex algorithm implemented in YASP as published by Faller et al.[20]

has been applied to optimize the parameter set. In each optimization cycle at least three different simulations have been carried out:

(1) A pre-equilibration of the starting configuration (which is the output coordinate file of the previous simplex steps with most similar parameter sets; or a standard equilibrated one if none is available) has been performed for 100 000 time steps.
(2) The system was considered sufficiently equilibrated, as soon as there was no significant drift detectable in the pressure of the system, i.e. a linear regression was performed on the recorded pressure values. The product of the slope of the regression line times the total number of datapoint needed to fall below the standard deviation of datapoints. Thus, the final equilibration of the system was achieved by repeating simulation cycles of 100 000 time steps each until the equilibration criterion was met.
(3) Starting with the equilibrated system the finally evaluated production run was performed for another 100 000 time steps.

After this sequence of simulations the simplex algorithm calculated the performance of the current parameter set by comparing the obtained density ρ to the target liquid density ρ^{target} of 747 kg/m^3 and the obtained heat of vaporization ΔH_{vap} to a target heat of vaporization ΔH_{vap}^{target} of 19.8 kJ/mol. The target values were derived from the data given in the ethylene-oxide user-guide [22] and in Walters and Smith [23] and were chosen such that the final model is optimized at the chosen thermodynamic system settings, i.e. 375 K and 1428.5 kPa. With the resulting performance value, the simplex steps were executed, i.e. a new set of parameters was generated and finally a new cycle was started. This procedure was repeated until the performance fell below the threshold of 0.0001 in order to ensure convergence.

For setting up systems under saturated phase conditions, NpT simulations were conducted. We estimated the saturated vapor pressure by use of the definition of the acentric factor as given in [24]: $\log[P_r^{sat}]_{T_r=0.7} = -(\omega+1)$ with P_r^{sat} being the reduced saturation pressure, T_r being the reduced temperature and ω the acentric factor. Applying the experimentally known data ($p_{crit} =$ 7190 kPa, $T_{crit} = 469$ K and $\omega = 0.2114$ [22]) leads to a theoretical vapor phase saturation pressure of 10 % below the experimental value at 328 K. This was transferred to 375 K yielding 1248.5 kPa as reference pressure for vapor phase simulations. Likewise the pressure was increased by 10 % to 1571.4 kPa for the liquid phase simulations in order to ensure stable phase separation. As the compressibility of a liquid is usually very small, such a moderate increase in pressure should not significantly affect the final results.

For computing the transport properties, for all systems (both in liquid and in vapor phase) a $L \times L \times 3L$ box containing 768 EO molecules was built by replicating an equilibrated cubic box containing 256 molecules three times

in z-direction. For both the RR model and the optimized model, RNEMD simulations were performed for the liquid phase at exchange periods of every 60, 200, 400, 800, and 1200 steps for calculating the shear viscosity and at exchange periods of every 100, 200, 400, 800, and 1200 steps for calculating the thermal conductivity. In the vapor phase, exchange periods of every 2, 5, 20, 40, 80, and 120 steps were applied for calculating the shear viscosity. Swap periods of 100, 200, 400, 800, 1200 steps, respectively, were employed for determining the thermal conductivity. In both cases, molecular velocity exchange has been done. All simulations were done for 9 ns and the last 8 ns were used for data analysis.

2.3.2 Monte Carlo

The Gibbs-Ensemble Monte (GEMC) Carlo method was utilized as described in Panagiotopoulos [26] and Martin [10] for recording vapor-liquid coexistence curves. Therefore, 256 + 512 EO molecules of the respective model were distributed among two cubic boxes (for the vapor and the liquid phase), both with periodic boundary conditions. Initial configurations were generated by multiplying a single molecule and placing the copies on a simple cubic lattice to yield cubic boxes. The respective box-lengths were chosen such that the corresponding experimental density was obtained and after subsequent equilibration the phase separation was maintained. In contrast to the reaction field method implemented in YASP, the MC program Towhee [14] utilizes Ewald summation,and for LJ interactions a non-bonded cutoff distance of 1.0 nm and analytical tail correction. Although this is not fully consistent with the conditions used throughout the MD simulations and especially during the optimization, lacking exactly matching methods this proceeding was accepted.

For the rigid RR model only volume moves (p = 0.002), rotational bias two-box molecule transfer moves (p = 0.001) and center of mass translational as well as rotational moves (p = 0.7 each) were considered. Vapor-liquid coexistence data were recorded for temperatures ranging from 230 to 375 K. The systems were equilibrated for 50 000 Monte Carlo cycles, where a cycle consisted of 768 moves. Additional runs of 50 000 cycles were executed and evaluated by splitting the simulation into 20 parts for estimating the uncertainties.

3 Calculation

The list of requested properties given in the challenge announcement is divided into three categories. Out of these we intend to determine the liquid and vapor phase densities, the heat of vaporization, the vapor pressure and the critical temperature and critical density, the heat capacity at constant pressure for

both phases and the associated isothermal compressibilities, as well as the corresponding liquid state and gas state viscosities and thermal conductivities. The utilized techniques are specified below.

3.1 Category 1 properties

The current or finally the average density is a standard output value. According to Equation 1 the density ρ of the simulated system is derived from the volume V of the cubic simulation box with box length l (therefore $V = l^3$) and the total mass in the system m_{sys} which is equal to the n (the number of molecules in the box) divided by the Avogadro constant A times the molecular weight of one ethylene oxide molecule $m_{EO} = 44.05$ g/mol.

$$\rho = \frac{m_{sys}}{V} = \frac{n \cdot m_{EO}}{A \cdot l^3} \tag{1}$$

The heat of vaporization ΔH_{vap} is derived from the total non-bonded energy E_{nb}, the temperature T, the gas constant R, and the number of molecules n according to Equation 2.

$$\Delta H_{vap} = -\frac{E_{nb}}{n} + R \cdot T \tag{2}$$

In order to derive the critical properties (temperature T_{crit} and density ρ_{crit}), the resulting liquid and vapor phase densities were fitted to the law of rectilinear diameters (Equation 3) and the law of order parameter scaling (Equation 4), where the critical exponent β_{crit} was taken as 0.313 [27]:

$$\frac{\rho_{liq} + \rho_{vap}}{2} = \rho_{crit} + C_1(T - T_{crit}) \tag{3}$$

$$\frac{\rho_{liq} - \rho_{vap}}{2} = C_2(T - T_{crit})^{\beta_{crit}} \tag{4}$$

From the latter production runs the resulting thermodynamic pressure of the gaseous box was taken as vapor pressure of the respective model at a given temperature.

3.2 Category 2 properties

The molar heat capacity at constant pressure C_p for the saturated liquid phase and for the vapor phase, respectively, has been calculated with five production

runs at different temperatures (365, 370, 375, 380 and 385 K), but otherwise identical simulation conditions (see above). The total energy $\langle E_{tot} \rangle$ and volume $\langle V \rangle$ were averaged over time. The pressure used in the following formula is the target value of the manostat (p=1571.4 kPa for liquid phase and p=1248.5 kPA for vapor phase). A linear fit of the plot ($\langle E_{tot} \rangle + P\langle V \rangle$) versus T has been used to calculate C_p.

$$C_p = \frac{1}{n_{mol}} \Big[\frac{\partial(\langle E_{tot} \rangle + p\langle V \rangle)}{\partial T} \Big]_{p=1571.4kPa} \tag{5}$$

The isothermal compressibility β is a measure of the relative volume change depending on the pressure P at constant temperature:

$$\beta = -\frac{1}{V} \Big(\frac{\partial V}{\partial P} \Big)_T \tag{6}$$

This equation can be expressed in terms of density (ρ) as follows:

$$\beta = \Big(\frac{\partial \ln(\rho)}{\partial P} \Big)_T \tag{7}$$

As the density of a system is a function of the pressure, the slope of the graph when plotting $ln(\rho)$ against P gives the isothermal compressiblitiy β. In our case, the temperature was fixed to 375 K and simulations were conducted at 10000, 20000, 30000 and 40000 kPa for the liquid phase and 200, 500, 800 and 1000 for the vapor phase, respectively.

3.3 Category 3 properties

For computing the transport properties we apply RNEMD [13] where a steady gradient is established artificially. According to the linear-response theory, the shear viscosity η is the proportionality constant between a shear field and a flux of transverse linear momentum:

$$j_z(P_x) = -\eta \frac{\partial v_x}{\partial z} \tag{8}$$

The shear field, also called the shear rate, is defined as the gradient $\partial v_x / \partial z$: we have chosen the x component of the fluid velocity relative to the z direction. The momentum flux $j_z(p_x)$ is the x component of the momentum p_x, which is transported in the z direction during a given time. In order to calculate the shear viscosity, unphysical momentum transfer needs to be performed within the simulation box.

(1) The simulation box is divided into N (even) slabs labeled from 1 to N. Slab 1 is supposed to have a positive x component for the momentum and the slab in the center of the box (slab M=$N/2$+1) has a negative x component of the momentum.
(2) To achieve the first point, we look for the molecule in slab 1 with the largest negative x component of the momentum in slab 1 and for the molecule with the largest positive x component of the momentum in slab M.
(3) We exchange the x component of the center-of-mass momentum p of the chosen molecules with each other. Since we only have one type of molecules, the mass of the chosen molecules is always the same. Therefore, we can exchange center-of-mass velocities instead. Simultaneously in both chosen molecules, the x-component of the center-of-mass velocity $V_x = \frac{p_x}{\sum_{i=1}^{n} m_i}$ is subtracted from the x-velocity component of every atom in the chosen molecule and added to the atoms in the other molecule.

If we repeat this procedure periodically, after a simulation time t, the total exchanged momentum is $P_{tot} = \sum_{exch}(p_{x,M} - p_{x,1})$ and the momentum flux is

$$j_z(P_x) = \frac{P_{tot}}{2tL_xL_y}, \tag{9}$$

where $p_{x,1}$ and $p_{x,M}$ are the two x components of the exchanged momentum in the slabs 1 and M, respectively, and L_x and L_y are the lengths of the orthorhombic simulation box in the x and y directions. When the system has converged towards a steady state, we can measure the mean velocity in x direction in each slab, determine $\partial v_x/\partial z$, and calculate the value of the shear viscosity using Equation 8.

In our calculation, the strength of the shear field is controlled by the swap period W with which momentum swaps are executed; a larger value of W meaning a weaker shear field.

In analogy, the thermal conductivity λ relates a heat flux j_z to a temperature gradient dT/dz:

$$j_z = -\lambda\left(\frac{dT}{dz}\right) \tag{10}$$

In order to build up a steady temperature gradient induced by continuous heat flux, unphysical heat transfer needs to be performed within the simulation box. Similarly, the method contains three steps.

(1) The simulation box is divided into N (even) slabs labeled from 1 to N. Slab 1, at the beginning of the simulation box, is designated as the hot

slab, and slab M, at the center of the simulation box, is designated as the cold slab.
(2) The Molecule in slab M with the largest modulus of its velocity vector and the molecule in in slab 1 with the smallest modulus of its velocity vector are identified.
(3) The center-of-mass momenta of these two selected molecules are exchanged. This mimics the process of taking heat from slab M to slab 1 in an unphysical way.

Since the energy is conserved, it flows back through the system through a physical transport mechanism which results in a heat flux. After a long enough simulation time a steady temperature gradient will be established.

The heat flux through the system can be deduced from the unphysical heat exchange during the exchange steps.

$$j_z = \frac{1}{2tL_xL_y} \sum \frac{m}{2}(v_{hot}^2 - v_{cold}^2) \tag{11}$$

This allows to calculate the termal conductivity according to equation 10. Again, the strength of the thermal field is controlled by the swap period W.

4 Results

4.1 Performance of models

With the chosen initial points in parameter space, the simplex algorithm yields the model EO_{opt} as summarized in Table 1. For the novel model as well as the reference RR model the physical properties have been determined. The final results are presented in Tables 2 and 3 and compared to experimental reference data.

Table 1

4.1.1 Category 1 properties

We find that the liquid phase density of the RR model in the chosen range of temperatures is about 5-7 % lower than the respective reference density, whereas the vapor phase densities are quite within the expected range. Correspondingly, the vapor pressure at 375 K of 1.468 MPa agrees well with the experimentally observed value of 1.437 MPa. The computed critical temperature of 467.1 K reproduces almost perfectly the reference value of 469.15 K,

whereas the critical density ranges approx. 3 % below the measured value, which is consistent which the general trend of too low densities. Consequently, also the heat of vaporization at 375 K is underestimated by approx. 5 %.

For liquid densities of the optimized model we observe the same, but even more pronounced trend as for the RR model; accordingly the optimized model yields a critical density that is about 8 % lower than the reference. The associated critical temperature of 448.7 K is approximately 20 K below the experimental value. In contrast, the gas phase densities tend to be higher than measured [22] and consequently the thermodynamic pressure exceeds the reference value by 15 %.

The remaining physical properties were obtained by MD simulations. For the RR model the liquid density value at 375 K of 701.3 kg/m^3 matches nicely the value of 702.2 kg/m^3 that results from the Monte Carlo simulations, indicating a good consistency of results obtained by different simulation techniques under liquid phase conditions. For the vapor phase the density obtained by MD is 15 % lower than that achieved in GEMC simulation or by experimental measurements, which is due to the reference pressure that was used for the manostat. This reference pressure was set to 1.248 MPa (instead of 1.468 MPa that were obtained by GEMC).

Concerning the optimized model and in contrast to the previous one, the values for the liquid density as well as for the heat of vaporization match perfectly the experimental values, as these numbers were used as target values during the optimization procedure. The vapor phase density however, is by roughly 50 % much too low, compared to the experimental reference, which traces back to the same reason as above.

4.1.2 Category 2 properties

For the RR model, the heat of vaporization is 15 % below the reference. The same applies for the specific heat capacity, which is consistently too low for the liquid, as well as for the vapor phase systems and thus does not resemble the experimental data. The obtained values for the isothermal compressibility are well reproduced in case of the liquid phase (14 % below reference value), but almost twice as large in case of the vapor phase simulations.

The optimized model performs better in terms of heat of vaporization obviously, and heat capacity: the molar heat capacity for the liquid C_p is 37 % higher than compared to experiment, but the vapor phase heat capacity is only 7% off; the recorded data are presented in Figures 1 and 2. In terms of isothermal compressiblity however, the results consistently exceed by approximately a factor of two the expected values.

Figure 1

Figure 2

4.1.3 Category 3 properties

Shear viscosity and thermal conductivity were studied using RNEMD simulations. Compatible to the observed low density, the liquid phase of the RR model displays a too low viscosity as well, whereas the vapor phase matches nicely the experimental value. The values for the thermal conductivity, in contrast are quite close to the reference values, where the one for the liquid phase was not measured, but calculated as well.

For the optimized model, the results are presented in more detail.

Figure 3

Figure 4

Figure 3 shows the velocity profiles of all systems in liquid phase at different shear rates. Linear behavior was observed for the selected swap periods. We then calculate the shear viscosities at these shear rates and the results are shown in Figure 4. The shear viscosities range from 0.164 – 0.171 mPa·s, displaying almost a constant level, which is indicative for an appropriate magnitude of perturbation and a good signal-to-noise ratio. Therefore the value of 0.171 mPa·s was taken as result and is in agreement with the experimental observation of 0.15 mPa·s. The shear viscosity of the vapor phase was found to be 0.0160 mPa·s which also compares well to 0.012 mPa·s obtained experimentally.

Figure 5

The calculation of the thermal conductivity is in the same spirit as the calculation of shear viscosity. In Figure 5, the temperature profiles of all systems are shown. Linear behavior was found for all perturbation rates tested for the liquid system. The obtained results range from 0.335-0.439 W/(mK), which is considerably higher than the computational reference of 0.12 W/(mK), estimated by the Missenard-method [30]. This is suggested to be a general feature of all-atom models which have more degrees of freedom for transferring kinetic energy than united-atom models such as the RR model which in this case gives a much better value of 0.125 W/(mK). The thermal conductivity of the vapor phase of the optimized model was found to be 0.031 W/(mK) which is in agreement with the experimental observation of 0.02 W/(mK).

5 Discussion

A wide range of simulation techniques ranging from simple MD to RNEMD and MC has been employed to predict very diverse physical properties for EO at conditions that require considerable care and effort when conducting experimental measurements. As a contribution to the fourth IFPSC callenge a new EO-model has been created by simplex optimization with respect to the experimental reference values for the liquid density and the heat capacity.

Generally, the reference model (RR) originally developed by Wielopolski and Smith [2] performs quite well. It is very good in reproducing the critical temperature and the vapor pressure, for the other properties however, the values were generally underestimated. The optimized model naturally performs better in terms of liquid density and heat of vaporization, but also with respect to the heat capacity and the transport properties. Generally, the properties accessible by MD or RNEMD for EO_{opt} were obviously much better predicted than those accessible by GEMC. Hence, the vapor-liquid-coexistence data given in Table 2 display much lower density values over the whole range of chosen temperatures than expected, leading to a much too low critical temperature as well. The associated density, however, is just within tolerable accuracy (8.6 % below the experimental value), only the density-profile, when compared to the experiment, appears to be shifted to lower temperatures. Despite of agreeing very nicely for the RR model, the liquid density values of the optimized model obtained at 375 K by GEMC and MD do not match. This mismatching behaviour was checked thoroughly (data not shown) and unfortunately is persistent, pointing at problems concerning the transferability of our proposed model.

The novel model was optimized explicitly, but also implicitly with respect to very special conditions, therefore not accounting for the subtle algorithmic nuances between YASP and towhee (e.g. energy and pressure corrections, treatment of long range electrostatic interactions, thermostats, barostats). Obviously this makes the model sensitive to changes and thus endangers its transferability from YASP to other programs.

The simplex alogorithm is well known for being a very robust, but not a fast optimization method. Our intention of producing an acceptable model with a (still costly) minimum of computational expense in fact was not entirely satisfied. However, this was not immediately conceivable from the course of the optimization, which is detailed below.

With a starting parameter set of four variables chosen to span the full parameter space, the algorithm took about 40 iterations to achieve an accuracy of 1% for each parameter and about 90 iterations to reach an accuracy of

0.001%. The optimization did abort after 150 iterations while the parameters were only varying in the accuracy of 10^{-4}%. To illustrate the process of the optimization, Figure 6 shows the relative error of the density, heat of vaporization, and of the total error function in dependence of the number of runs performed (the iteration number). We observe high fluctuations at the beginning of the optimization, which decrease to an absolute value of below 0.1% after 40 iterations and decrease further, the more cycles are performed.

Figure 6

The accuracy of the two target properties (density and heat of vaporization) can also be monitored. Figure 7 shows the sequence of results calculated from the iteration runs. The wide fluctuation at the beginning of the optimization corresponds to the error fluctuations in Figure 6. Although all the initial parameter sets produce too high densities and too high values for the heat of vaporization, the iteration converges to a minimum which is located outside the space spanned by the initial parameters.

Figure 7

A similar plot can be produced to show the history of the parameter values. Although all four variables have been optimized at the same time, for the sake of clarity, the two Figures 8 and 9 show only the two ϵ-values and the two σ-values in conjunction. We see here as well that the final parameter set is located outside the parameter space spanned by the initial parameter sets. Convergence is achieved due to the robustness of the method but requires lots of iterations. Nevertheless, computational resources were sufficient to afford this course to obtain parameters that lead to a better behavior of the model with respect to the target properties.

Figure 8

Figure 9

So a broad span of parameter values has been visited during optimization with the adjustments becoming smaller and smaller the closer the final minimum is being encircled. The final σ_{OO}-value is only 1.3% percent below the corresponding value of the RR model, and the σ_{CC} value is 3.2% larger (see Table 1). Given that the assumed bond lengths are slightly shorter, these values appear quite reasonable, as in terms of modeling the LJ-interactions both models are equal. However, the ϵ_{OO} and ϵ_{CC} values for the optimized model are significantly (i.e. 45 and 20 %) lower than those for the RR model. We note that the visited ϵ_{OO}-values fluctuate more than the ϵ_{OO}-values, possibly reflecting the 2:1 ratio of carbon to oxygen atoms, implying that any change in the carbon parameters will have a more pronounced impact on the model-performance than changes in the oxygen parameters. Considering the

smaller partial charges in the optimized model, however, these ϵ-values seemed acceptable, too.

In principle, other parameters could have been chosen for optimization. As mentioned above, test runs including e.g. a reparameterization of partial charges have been performed previously, but did not yield promising models. Due to the nature of the contest, the limitation of time prevented investigating other parameter sets and evaluation functions. On the parameters side, especially the incorporation of the hydrogen atoms with (also optimizable) LJ parameters would introduce additional sterical features to the model. On the evaluation function side, taking into account the dipolar momentum would help to control the partial charges and a term controlling the kinetics (e.g. diffusion coefficients or shear viscosity) of the model is also desired for future investigations.

The above detailed quality of the results however suggests, that at least a second point in the phase diagram or a parallel GEMC simulation during the optimization cycles would have improved the final outcome, especially with respect to transferability. Both options are equally and considerably consumptive, but obviously necessary. Finally when exploring the parameter space with simplex two mathematical aspects have to be remembered: Simplex is a greedy algorithm which looks for local minima. Since the path through the parameter space depends on the starting parameters, the algorithm can get trapped in different local minima. Additionally we have to point out, that optimizing four parameters with only two target values is mathematically an underdetermined system, that consequently has more than one solution. These two issues lead to the fact, that generally simplex finds many parameter sets that fulfil the threshold criterion of sufficient accuracy of the target function. We tested four different starting parameter configurations, which simplex in all cases shifted to the same region in parameters space. The here reported parameter set was the most promising one.

6 Conclusion

Summarizing this work, the optimization of the LJ σ and ε values of oxygen and carbon against the density and the heat of vaporization by the simplex algorithm has led to a new parametrization of the force field for EO. The strengths of this model lie in the evaluation of transport properties calculated with YASP. We obtained a novel parameter set which was located outside the initially estimated parameter space. Therefore the new parameter set may deviate from tabulated ε and σ values of oxygen and carbon which allow a physical interpretation according to LJ-theory; but it represents one solution to the mathematical problem of reproducing the target properties accurately

and is a valid candidate for a new parametrization. We have shown hat using distinctly different simulation techniques, such as molecular dynamics, Monte Carlo, and reverse non-equilibrium molecular dynamics, this new parameter set performed in the tasks of predicting phase equilibrium data, caloric data and transport data mostly within the accuracy requested in the IFPSC challenge. Due to the conditions the model was optimized to, the strength of this model lies in describing the liquid phase properties.

For our future work, however, we will consider several options in order to increase the overall performance of this approach: Although the simplex optimization procedure has proven to be robust and reliable, the method itself could be substituted by alternative, i.e. faster converging algorithms. In any case, as the accuracy of the repeatedly computed test properties (in our case: liquid density and heat of vaporization) is of utmost importance for limiting the number of iteration cycles, thus even longer equilibration and evaluation periods come into account. Furthermore, adding more test properties to the simplex evaluation function will improve the general performance of a model. The self diffusion coefficient (if available) would be a natural choice for that purpose, because this value is easily accessible by simulation and then a dynamic property would be linked to the process of parameterization, which accomplishes the so far more more static set.

7 Acknowledgements

We are grateful to Jadran Vrabec, Gerd Winter and Nico van der Vegt for valuable discussions. We also appreciate the support in fitting data given by Sabine Pott.

References

[1] Industrial Fluid Properties Simulation Collective, http://www.ifpsc.org

[2] P.A. Wielopolski and E.R. Smith, Mol. Phys. 54 (1985)467–478.

[3] A. Leach, Molecular Modelling, second ed., Prentice Hall, 2001.

[4] B.J. Alder and T.E. Wainwright,J. Chem. Phys. 31, No.2 (1959) 459.

[5] W.D. Cornell, P. Cieplak, C.I.Bayly et al, J. Am.Chem.Soc. 117 (1995) 5179–5197.

[6] A.D. MacKerell Jr., D. Bashford, M. Bellott et al, J. Phys. Chem. B 102 (1998) 3586–3616.

[7] W.R.P. Scott, P.H. Hunenberger, I.G. Tironi et al, J. Phys. Chem. A 103 (1999) 3596–3607.

[8] W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, J. Am. Chem. Soc. 118 (1996) 11225 – 11236.

[9] A.K. Rappe, C.J. Casewit, K.S. Colwell et al, J. Am. Chem. Soc. 114 (1992) 10024 – 10035.

[10] M.G. Martin, Fluid Phase Equilibria 248 (2006) 50–55.

[11] R.D. Mountain, J. Phys. Chem. B. 109 (2005) 13352–13355.

[12] F. Müller-Plathe, Comput. Phys. Commun. 78 (1993) 77ff.

[13] P. Bordat, F. Müller-Plathe, J. Chem. Phys. 116 (2002) 3362 – 3369.

[14] MCCCS Towhee, http://towhee.sourceforge.net

[15] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E.Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery Jr, T. Vreven, K.N. Kudin, J.C. Burant

[16] L.E. Chirlian, M.M. Francl, J. Comp. Chem. 8 (1987) 894 ff.

[17] C.M. Brenemann and K.B. Wiberg, J. Comp. Chem. 11 (1990), 361

[18] H.A. Lorentz, Ann. d. Phys. 12 (1881) 1703–1706.

[19] D. Berthelot, Comptes Rendus de l'Academie des Sciences Paris 126 (1889) 1703–1706.

[20] R. Faller, H. Schmitz, O. Biermann, F. Müller-Plathe, J. Comp. Chem. 20 (1999) 1009–1017.

[21] Numerical Recipes: The Art of Scientific Computing, third ed. Cambridge Univresity Press, 2007.

[22] C. Buckles, P. Chipman, M. Cubillas, M. Lakin, D. Slezak, D. Townsend, K. Vogel, M. Wagner, Ethylene oxide user's guide, August (1999), http://www.ethyleneoxide.com

[23] C.J. Walters, J.M. Smith, Chem. Eng. Prog. 48 No. 7 (1952) 337–343.

[24] P. Ghosh, Chem. Eng. Technol.22 (1999)5, 379–399.

[25] P. Atkins, J. De Paula, Physical Chemistry, 8th ed. Oxford University Press, 2006.

[26] A. Panagiotopoulos, Mol. Physics, 61 No.4 (1987) 813–826.

[27] R. Agrawal, E.P. Wallis, Fluid Phase Equilibria 131 (1997) 51–65.

[28] NIST Chemistry WebBook, NIST Standard Reference Database Number 69, June 2005 Release, http://webbook.nist.gov/chemistry

[29] R.A. Fine, F.J. Millero, J. Chem. Phys. 59, No. 10 (1973) 5529 – 5536.

[30] A. Missenard, Comptes Rendus, 260 (1965) 5521 .

Figure Legends

Figure 1: From the temperature dependence of $\langle E_{tot} \rangle + P\langle V \rangle$ the molar heat capacities were derived for the optimized model in liquid phase.

Figure 2: From the temperature dependence of $\langle E_{tot} \rangle + P\langle V \rangle$ the molar heat capacities were derived for the optimized model in vapor phase.

Figure 3: Velocity profiles relative to the slabs of the simulation box for different swap periods as recorded for the optimized model in the saturated liquid phase.

Figure 4: Viscosity in dependence of exchange period.

Figure 5: Temperature profiles relative to the slabs of the simulation box for different swap periods for the optimized model in saturated liquid phase.

Figure 6: Tracking the relative error of the calculated density to the target density, the relative heat of vaporization to the target one, and the total error value throughout the optimization.

Figure 7: Values of ε and σ during the optimization. Consecutively run values are connected by a line to show the evolution of values during optimization.

The values of the five initial parameter sets are shown in light gray triangles while the final parameters are marked by a dark gray circle.

Figure 8: Sequence of the calculated values against which the optimization was performed: density and heat of vaporization. The results of the initial five parameter sets are shown as light gray triangles while the target value (and the results where the optimization forced the system to) is marked by a dark gray circle.

Figure 9: Sequence of the calculated values against which the optimization was performed: density and heat of vaporization. The results of the initial five parameter sets are shown as light gray triangles while the target value (and the results where the optimization forced the system to) is marked by a dark gray circle.

Table 1
Summary of models

	RR	EO_{opt}
Lennard-Jones parameters		
σ_{OO} [nm]	0.2666	0.26308
ε_{OO} [K]	73	55.47
ε_{OO} [kJ/mol]	0.60695	0.46119
σ_{CC} [nm]	0.37143	0.38372
ε_{CC} [K]	90	74.82
ε_{CC} [kJ/mol]	0.7438	0.62209
atomic charges		
q_O	-0.3216	-0.312
q_C	0.1608	-0.084
q_H	-	0.120
bondlengths (fixed)		
r_{CO} [nm]	0.1431	0.1401
r_{CC} [nm]	0.1466	0.1453
r_{CH} [nm]	-	0.1077
bondangles + force constants		
$\alpha(HCH)$[degree], k_α [kJ/mol·rad^2]	-	115.23, 276.33
$\alpha(CCH)$[degree], k_α [kJ/mol·rad^2]	-	119.88, 314.01
$\alpha(OCH)$[degree], k_α [kJ/mol·rad^2]	-	115.24, 293.08

Table 2
Vapor-liquid coexistence data and related properties for both models obtained by GEMC

	RR		EO-opt		experiment [22]	
density [kg/m^3]	liquid	vapor	liquid	vapor	liquid	vapor
230 K	910.6	0.257	870.3	0.287	955	0.253
260 K	875.1	0.747	829.4	0.869	918	0.738
290 K	835.6	2.531	788.4	2.976	878	2.50
330 K	778.2	8.307	724.1	9.71	820	8.29
375 K	702.2	23.897	644.1	28.09	744	24.1
400 K	650.1	37.247	592.7	44.43	694	-
T_{crit}	467.1		448.7		469.15	
ρ_{crit}	301.83		285.4		311	
p_{vapor}^{375K} [MPa]	1.468 ± 0.0032		1.649 ± 0.0169		1.437	

Table 3
Physical properties obtained by MD at a simulation temperature of 375 K. For computing the heat capacity, additional runs at 365, 370, 380 and 385 K were performed. For the liquid phase simulations the reference pressure was set to 1571.4 kPa, in case of vapor phase simulations to 1248.5 kPa.

	RR		EO-opt		experiment[22,23]	
	liquid	vapor	liquid	vapor	liquid	vapor
Category 1 properties						
density [kg/m^3]	701.3	19.8	749.8	11.0	746.7	24.6
ΔH_{vap} [kJ/mol]	17.00		19.87		19.8	
Category 2 properties						
C_p [J/g·K]	0.79	0.86	3.16	1.79	2.30	1.67
β [1/10^{-6}kPa]	2.23	1591	5.23	1800	2.6	819
Category 3 properties						
viscosity [mPa·s]	0.081	0.010	0.171	0.0160	0.151	0.012
	± 0.002	± 0.003	± 0.003	± 0.0004		
th. cond. [W/(mK)]	0.095	0.0125	0.439	0.031	0.12 [30]	0.02
	± 0.002	± 0.0001	± 0.003	± 0.001		

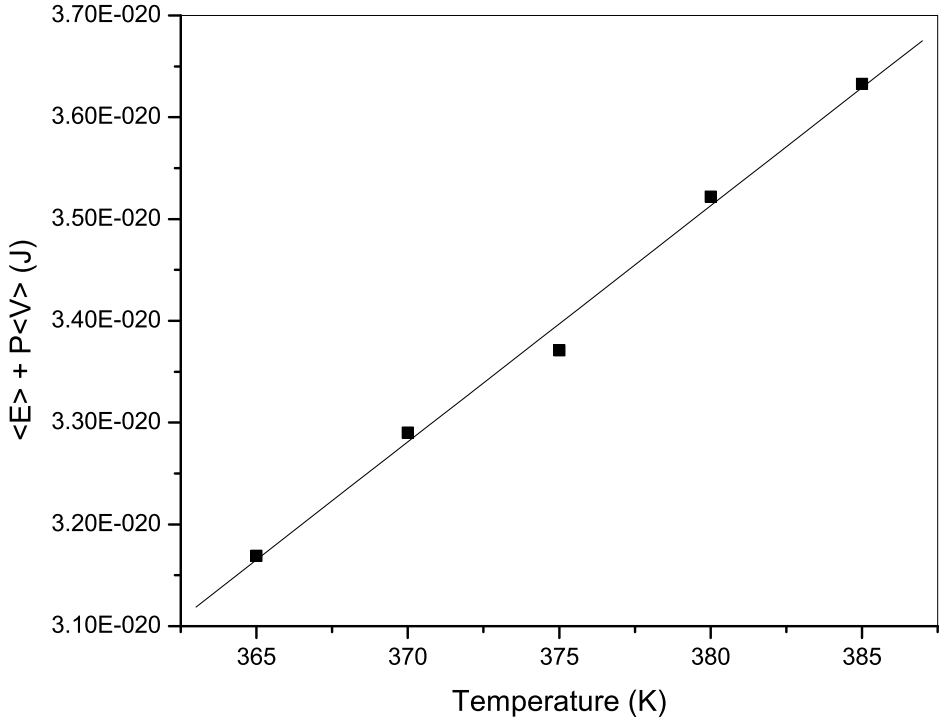
3.70E-020
3.60E-020
3.50E-020
3.40E-020
3.30E-020
3.20E-020
3.10E-020
<E> + P<V> (J)
365
370
375
380
385
Temperature (K)

Fig. 1.

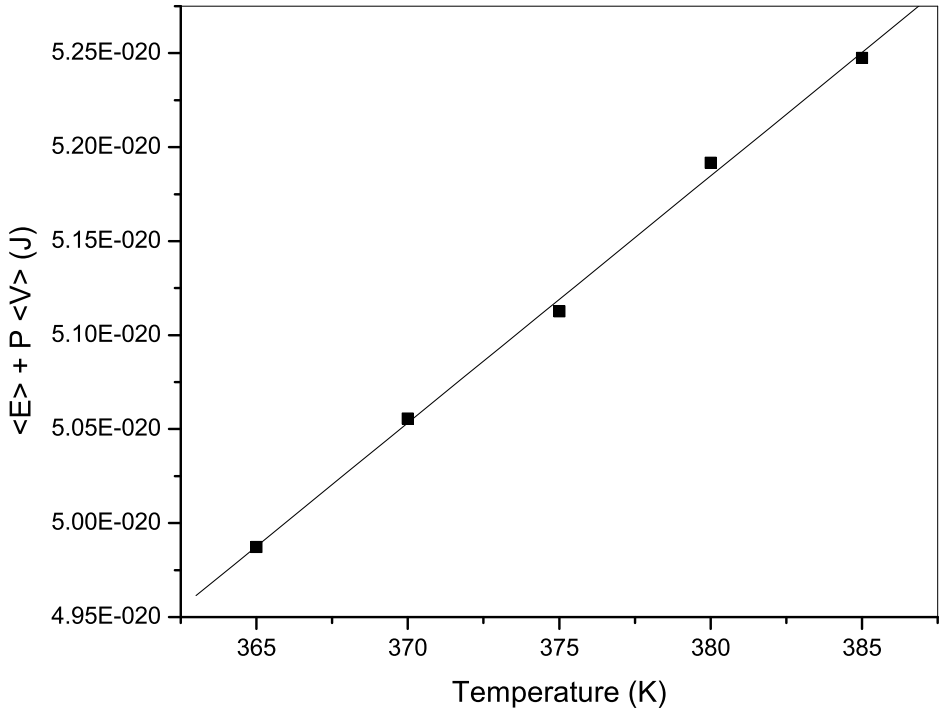

Fig. 2. Müller et al.

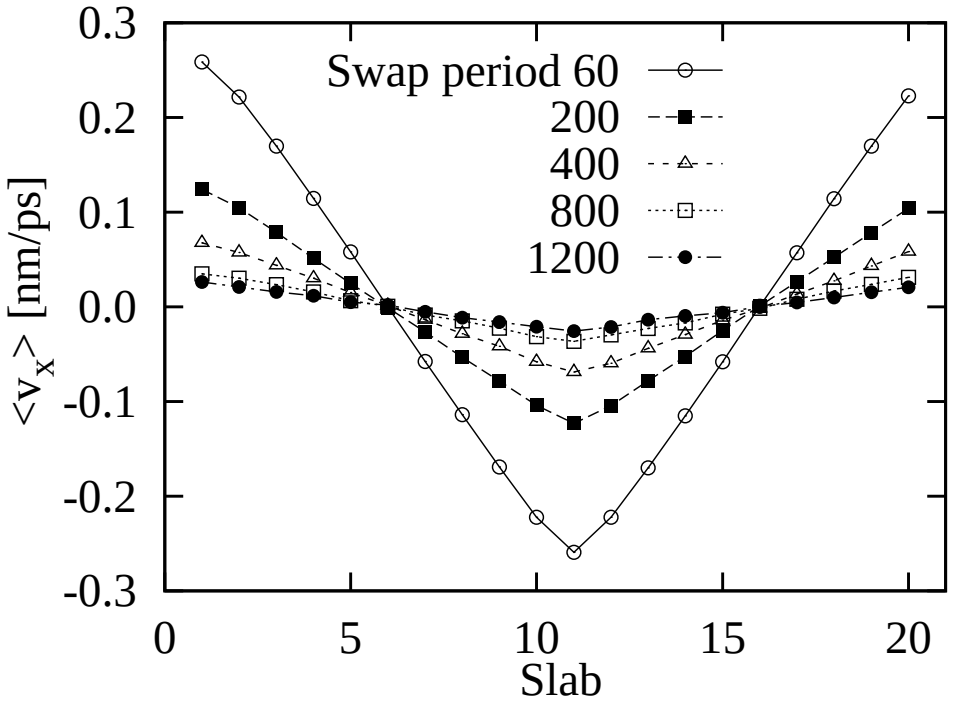

Fig. 3. Müller et al.

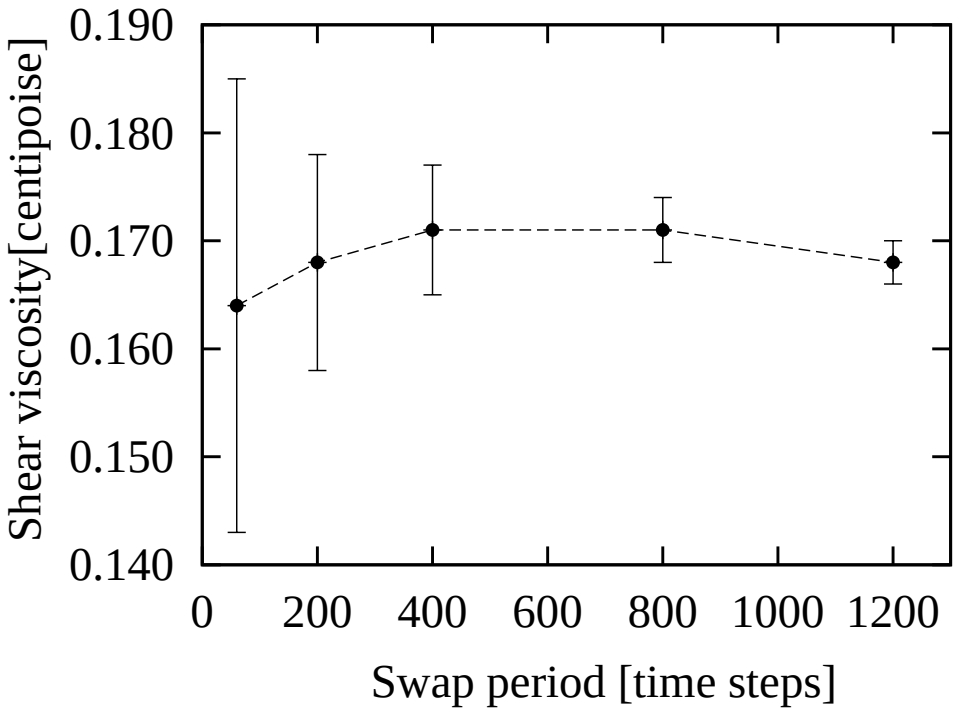

Fig. 4. Müller et al.

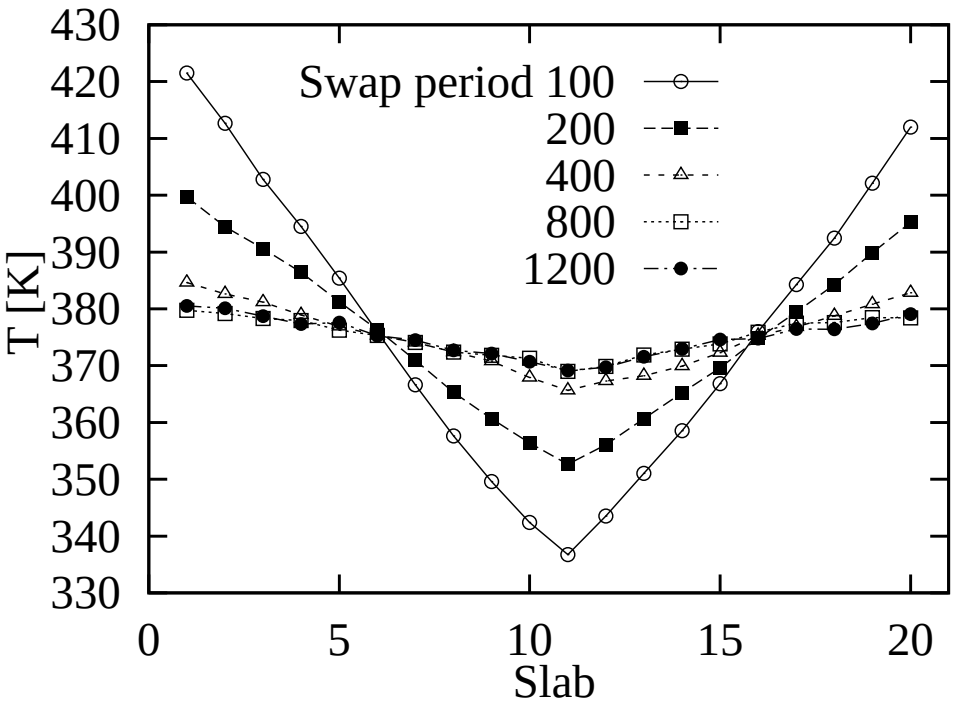

Fig. 5. Müller et al.

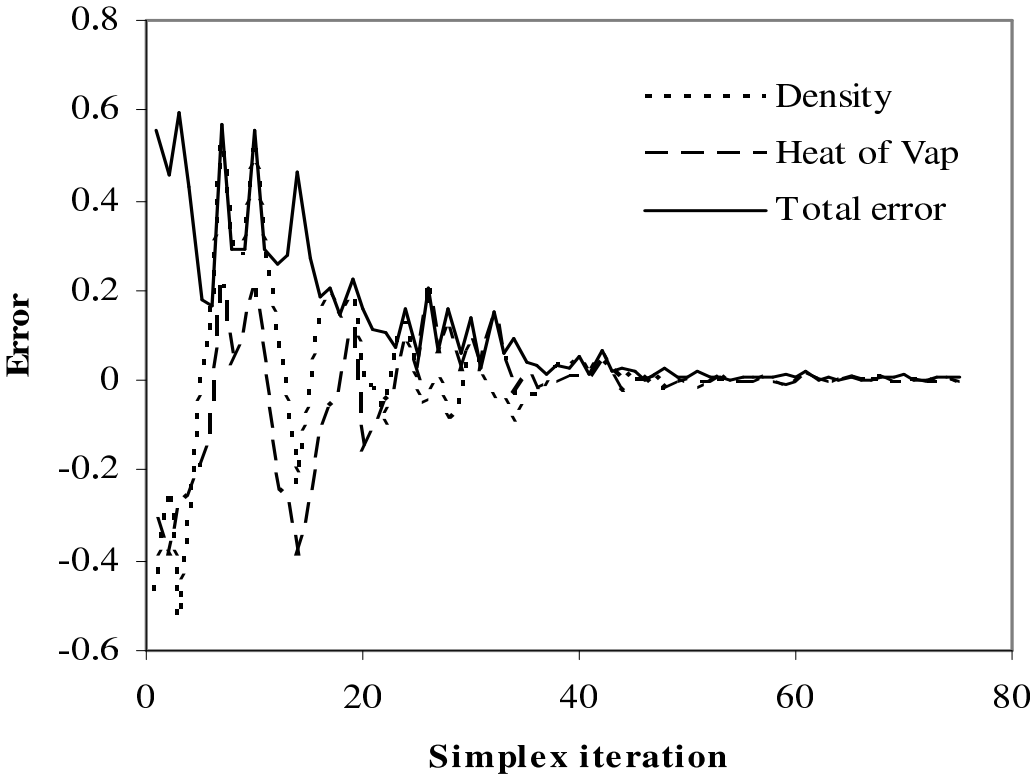
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
Error
0
20
40
60
80
Simplex iteration
Density
Heat of Vap
Total error

Fig. 6.

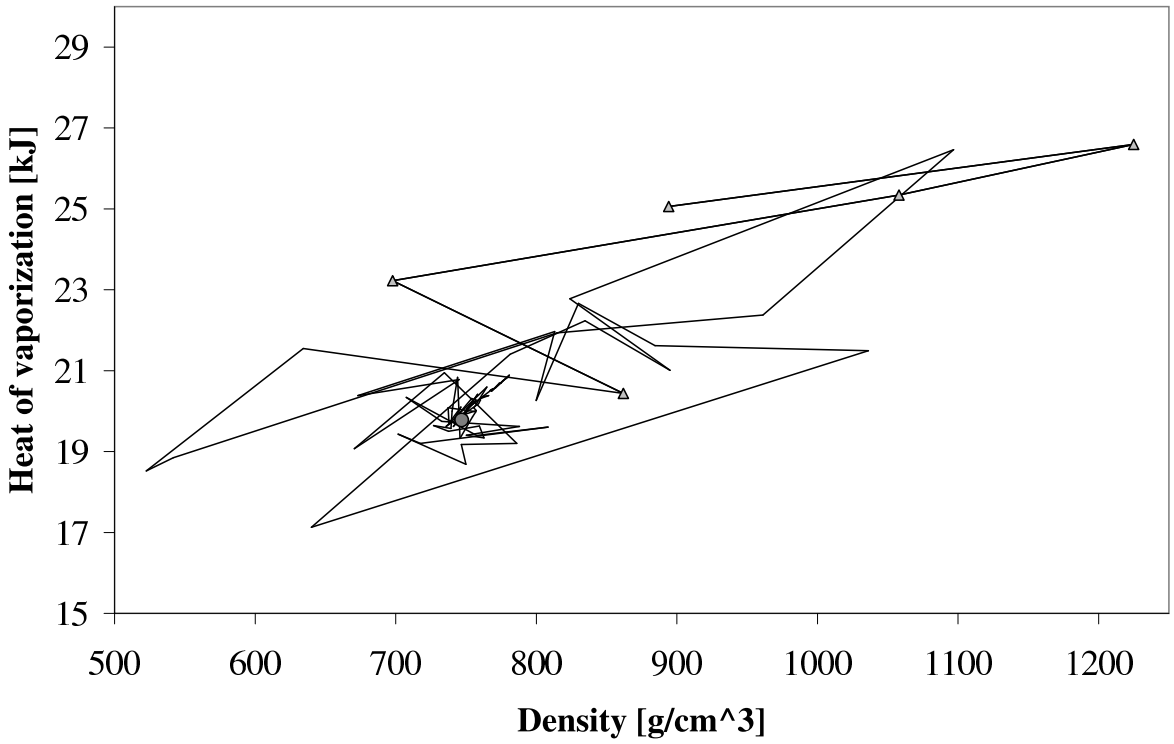

Fig. 7. Müller et al.

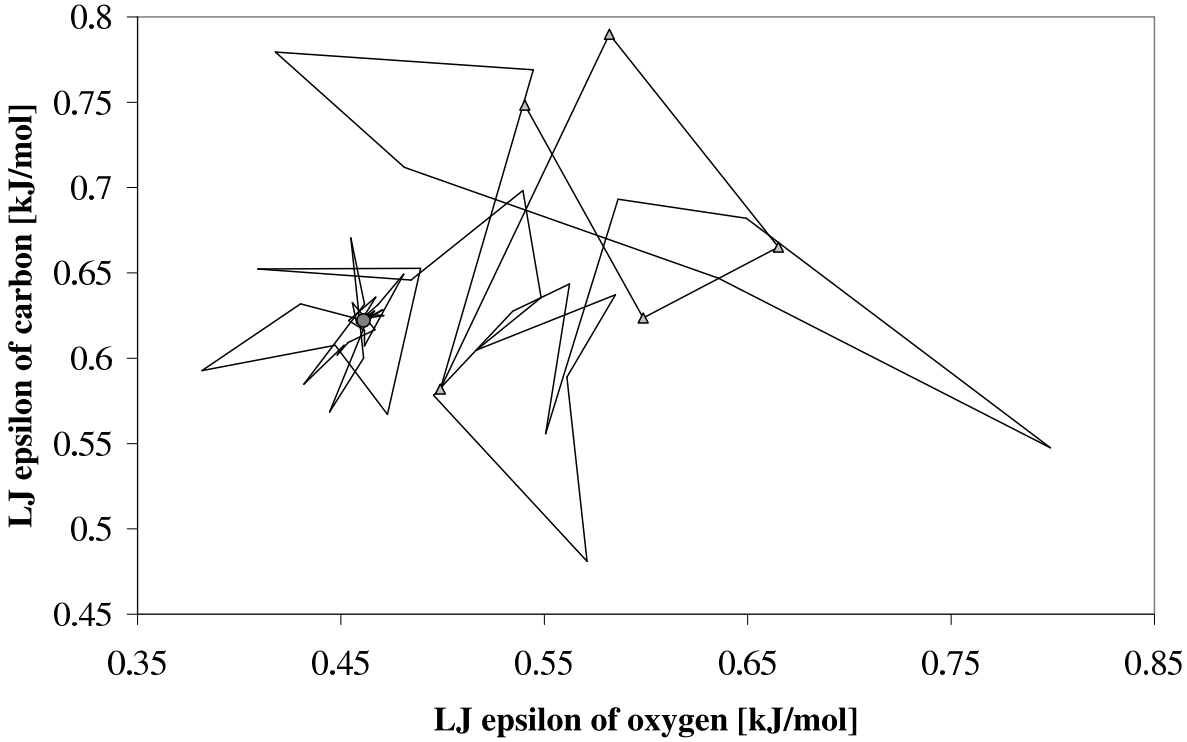

Fig. 8. Müller et al.

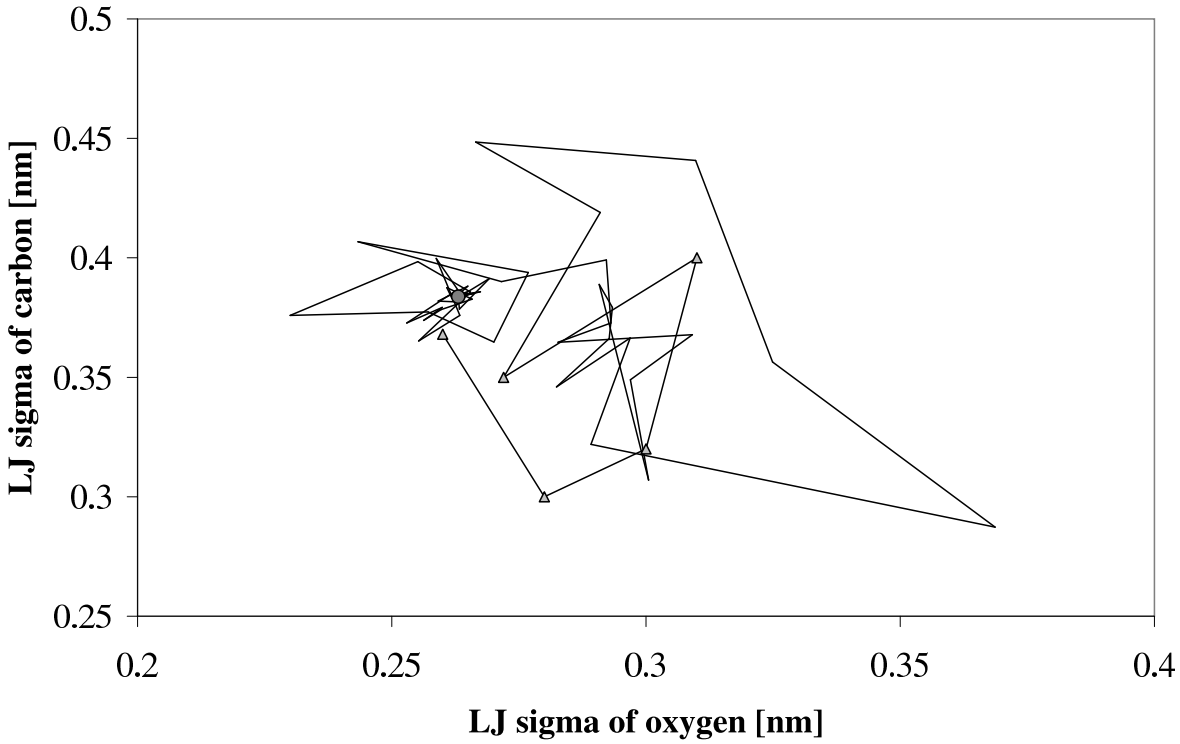

Fig. 9. Müller et al.