

Responses to Reviewer Comments for Manuscript ID: je-2018-012328

Histogram-free reweighting with grand canonical Monte Carlo: Post-simulation optimization of non-bonded potentials for phase equilibria

by

Richard A. Messerly, Mohammad Soroush Barhaghi,
Jeffrey J. Potoff, and Michael R. Shirts

Overview

We are grateful for the feedback we received from each reviewer. We are also encouraged that each reviewer feels the manuscript is ready for publication with only minor revisions. We have included a response for each comment. We believe that the review process has strengthened this manuscript and that it is ready to be published. We appreciate your consideration of the revised manuscript for publication in the *Journal of Chemical & Engineering Data*.

Itemized List of Changes to Manuscript (Page numbers refer to revised manuscript.)

Response to Reviewer 1, Comment 1: pg. 2 (Abstract)
Response to Reviewer 1, Comment 2: pp. 3-4
Response to Reviewer 1, Comment 3: pg. 4
Response to Reviewer 1, Comment 4: pp. 5, 43-44
Response to Reviewer 1, Comment 5: pp. 6-8, S3 (Figure S1)
Response to Reviewer 1, Comment 6: pp. 11, S4 (Table S4)
Response to Reviewer 1, Comment 7: pp. 31, 35
Response to Reviewer 2, Comment 1: pp. 38, 40, 42, Figures 8-9 and Figure S2
Response to Reviewer 2, Comment 2: pg. 10
Response to Reviewer 2, Comment 3: No changes
Response to Reviewer 2, Comment 4: pp. 40, S23-S24, S26-S27 (Figures S4-S5 and S7-S8)
Response to Reviewer 2, Comment 5: pg. 11
Response to Reviewer 2, Comment 6: pg. S37 (Figure S11)
Response to Reviewer 2, Comment 7: pp. 2, 13, 26
Response to Reviewer 3, Comment 1: No changes
Response to Reviewer 3, Comment 2: pg. 42
Response to Reviewer 3, Comment 3: pg. 44
Response to Reviewer 3, Comment 4: No changes

All additions to the manuscript have been highlighted in green

At Revision:

MANUSCRIPT:

- Abbreviate all journal titles in the References as per CASSI.
- Include the date of access in Reference 49.
- Update the publication information in References 51 and 52 when available.

Response #1

We have abbreviated all journal titles and included the date of access for Reference 50 (previously Reference 49). Unfortunately, the publication information is still not available for References 52 and 53 (previously References 51 and 52).

SUPPORTING INFORMATION:

- Re-label "Table SI.I," "Table SI.II," etc. as "Table S1," "Table S2," etc.
- Re-label "Figure SI.1," "Figure SI.2," etc. as "Figure S1," "Figure S2," etc.

Response #1

We have relabeled the tables and figures accordingly. We also relabeled the page numbers and sections as S1, S2, etc.

Reviewer #1

Recommendation: Publish after minor revisions noted.

General:

This interesting work by Messerly et al. introduces a new computational scheme (GCMC-MBAR) to optimize non-bonded force field parameters for phase equilibria by reweighting configurations from GCMC simulations rather than histograms (GCMC-HR). This is a follow-up work of the authors' paper (Ref. 16) in which they applied the same idea to integrate MBAR method with ITIC for predicting vapor-liquid coexistence properties. In this manuscript the authors demonstrate the advantages of GCMC-MBAR over MBAR-ITIC but also point out the existing limitation of the approach and give useful suggestions for the adoption of this method. The key applications of this approach include accurately estimating the prediction of forcefield j using the configurations from GCMC simulations with forcefield i , and optimizing forcefield non-bonded parameters starting from a reference forcefield, which are discussed about in section 3.3 and 3.4 respectively. In general, the authors clearly demonstrate this new and interesting scheme with quite sufficient technical details provided. I suggest publication of this work after the authors address my following comments:

Comment #1

The abstract is too long and much detail is duplicated with the introduction. Please revise and make it concise.

Response #1

We have consolidated the first and second paragraphs of the Abstract to reduce the redundancy with the Introduction.

Comment #2

On page 4 in the first paragraph the authors write "The proposed GCMC-MBAR method is identical to the traditional GCMC-HR approach except that MBAR reweights configurations rather than histograms." It is very natural for the readers to concern about the memory burden and computational cost of this approach since a big amount of additional data needs to be saved from the simulation. The authors address this issue by adopting the basis function in the latter part of the manuscript, but it is better to mention it somewhere early in the paper.

Response #2

We agree with the reviewer that some readers might be concerned about the additional memory burden of storing configuration files. Although we discuss later in the manuscript how basis functions can ease this concern, we have modified the sentence on pages 3-4 to help clarify this issue from the outset:

parameter sets that have not been simulated directly. While storing configuration files is significantly more memory intensive than storing histogram files (and scales as the number of molecules), this additional storage load can be alleviated greatly by utilizing basis functions (see Section 2.4).

However, we would like to emphasize that GCMC-MBAR dramatically reduces the computational cost compared to performing GCMC-HR for numerous parameter sets. For the simple case of computing phase equilibria for a single parameter set, GCMC-MBAR and GCMC-HR have the same computational cost.

Comment #3

On page 4 in the third paragraph the authors talk about the bad performance of MBAR-ITIC when the changes of σ and λ are large. Can you add a short explanation from the cited paper here?

Response #3

Molecular configurations in the liquid phase depend primarily on close-range repulsive interactions (i.e., the size and repulsive exponent parameters). Therefore, two vastly different values of σ or λ will sample from very different regions of configurational space, which leads to poor overlap. We have provided the following explanation on page 4:

large changes in the non-bonded size (σ) and repulsive exponent (λ) parameters. Because cause molecular configurations in condensed phases depend strongly on short-range interactions, this poor “overlap” is primarily observed when $\sigma_{rr} \not\approx \sigma_{ref}$ or $\lambda_{rr} \not\approx \lambda_{ref}$. The

Comment #4

On page 5 in the third paragraph the authors admit that GCMC-MBAR is very similar to HS-GCMC in nature. They are both powerful tools to optimize force field parameters and the authors list several advantages of GCMC-MBAR over HS-GCMC in their opinion. I think this is a very important comparison that the authors need to address more. Does HS-GCMC possess the same capability as GCMC-MBAR and does it suffer from the same problem that GCMC-MBAR has when predicting the liquid phase saturation properties at $\theta_{rr} \approx \theta_{ref}$? It will be very helpful to see some simulation result comparison in the paper between these two methods.

Response #4

Although the objectives of HS-GCMC and GCMC-MBAR are similar, the methods for achieving this objective are actually quite different. HS-GCMC primarily modifies the sampling while GCMC-MBAR only modifies the data analysis. Therefore, it is conceivable to combine the two methods by using HS for sampling and MBAR for post-processing. We have clarified this point on pages 43-44:

The main objective of GCMC-MBAR is similar to that of Hamiltonian scaling (HS), namely, to efficiently estimate phase equilibria properties for many force field parameter

sets (θ) by maximizing the information content extracted from each simulation. Despite some apparent similarities between HS-GCMC and GCMC-MBAR, the means by which these two methods accomplish this objective are quite distinct. Both methods take advantage of the concept of a *mixture distribution* that corresponds to multiple θ .⁶¹ The difference is that GCMC-MBAR generates this mixture distribution by combining configurations (or basis functions) sampled from independent simulations, while HS-GCMC samples directly from this mixture distribution in a single simulation and stores a combined (and scaled) histogram for each θ_{ref} . Thus, HS-GCMC improves efficiency by modifying the sampling, whereas GCMC-MBAR simply modifies the post-simulation data analysis. Because HS-GCMC uses histograms in the analysis, in contrast with GCMC-MBAR, only the parameters that are simulated in the original mixture distribution can be combined. A possible synergy, beyond the scope of this study, could exist between the two methods by employing HS for sampling and MBAR for data analysis.

We agree with the reviewer that it is important to compare HS-GCMC with GCMC-MBAR. GCMC-MBAR possesses some significant capabilities that HS-GCMC does not. Most importantly, GCMC-MBAR can predict phase equilibria properties for arbitrary force field parameter set (θ), whereas HS-GCMC can only estimate phase equilibria properties for force field parameter sets that are sampled directly. In other words, it is actually not possible to compare their performance for predicting liquid phase saturation properties for $\theta_{\text{rr}} \neq \theta_{\text{ref}}$ because HS-GCMC is only applicable for $\theta_{\text{rr}} = \theta_{\text{ref}}$ (where θ_{rr} is one of the parameter sets that constitute θ_{ref}). To clarify this important difference, we have added this statement to page 5:

This final distinction is of utmost importance as GCMC-MBAR is capable of predicting phase equilibria post-simulation for any force field parameter set ($\theta_{\text{rr}} = \theta_{\text{ref}}$ and $\theta_{\text{rr}} \neq \theta_{\text{ref}}$), whereas HS-GCMC can predict phase equilibria only for the parameter sets that are tested at run time.

Furthermore, a rigorous comparison of simulation results for the two methods is unrealistic in the time frame granted for review. We are not aware of any open source Monte Carlo code that includes HS-GCMC. Writing our own HS-GCMC code could require months of testing to ensure reliability. For this reason, we are limited to discussing their algorithmic similarities and differences, rather than performing a benchmark study of their respective efficiencies.

Comment #5

a) On page 6 in the force fields section, authors mention that the bond stretching potential is neglected for all the force fields. They compared their result with those from Mick et al. but Mick also used fixed bond models. This is not valid justification. Have the authors conducted any tests showing that the bond stretching has no influence on the simulation result or is this reported from literatures in the similar systems?

Response #5

The purpose of comparing our results with those of Mick et al. is to validate that GCMC-MBAR and GCMC-HR yield consistent values. Therefore, this comparison would be invalid if we utilized flexible bonds. We have clarified this on pages 6-7:

We employ fixed bond lengths for each force field studied. Note that this is inconsistent with the original NERD force field, which was developed using a harmonic bond potential. The primary reason we utilize fixed bond lengths for the NERD potential is to allow for a valid comparison of our GCMC-MBAR values with the GCMC-HR results of Mick et al.,¹² which were also obtained using fixed bonds. Furthermore, Section S2 of Supporting Information demonstrates that the branched alkane NERD phase equilibria results obtained with fixed bonds (as reported by Mick et al.¹²) agree with the flexible-bond results (as reported by Nath et al.²⁶).

In addition, we have compared the NERD results reported by Mick et al. (fixed bonds) and Nath et al. (flexible bonds). Note that Nath et al. only reported saturated liquid and vapor densities, so we do not compare vapor pressures or enthalpy of vaporization. In all cases, the fixed and flexible results agree to within the combined uncertainties for liquid density. The agreement for vapor density is not as good, but most compounds agree to within the combined uncertainty. Note that Mick et al. utilized GCMC-HR while Nath et al. used GEMC. Therefore, it is also possible that the different methods, simulation codes, and data analysis are the cause of these subtle deviations in vapor phase properties. This figure has been included on page S3 in Supporting Information:

S2 Fixed vs. flexible bonds

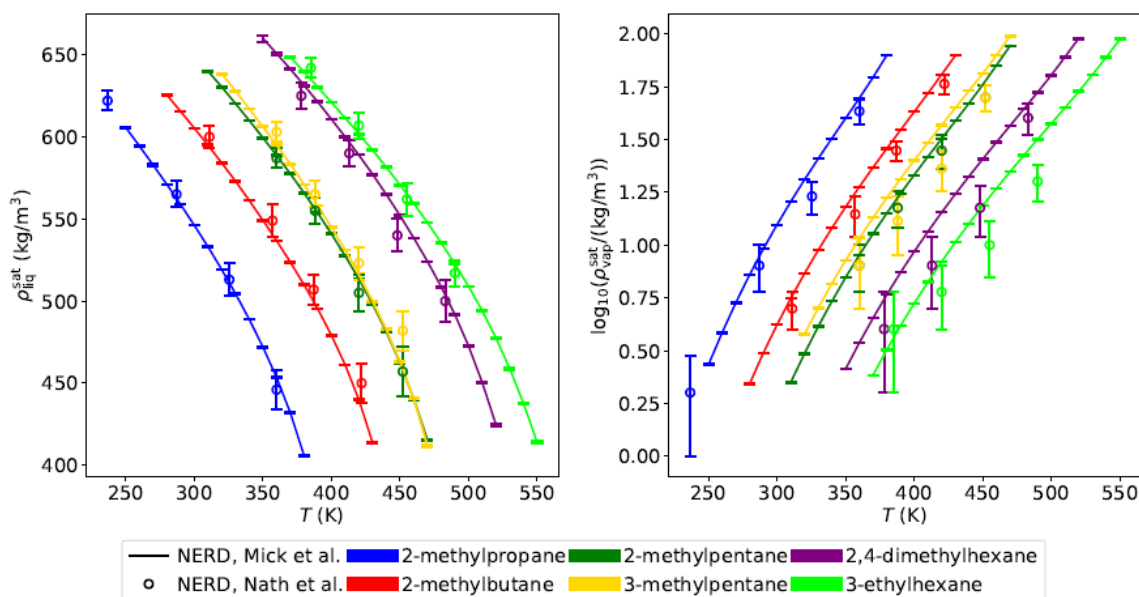


Figure S1: Comparison of saturated liquid (left panel) and vapor (right panel) densities for fixed (Mick et al.) and flexible (Nath et al.) bonds. Mick et al. used GCMC-HR while Nath et al. utilized GEMC. Note that Nath et al. did not report tabulated values for vapor pressure or enthalpy of vaporization.

b) The same question comes to electrostatic interactions. Please show evidence that it is valid to ignore these interactions.

Electrostatics are not included in the MiPPE, TraPPE, and NERD united-atom force fields for the hydrocarbons examined in this study. Again, a valid comparison of our GCMC-MBAR results with the GCMC-HR results of Mick et al. and Soroush Barhaghi et al. necessitates that we not modify the underlying force field. The following sentence was added on page 8 to clarify this point:

repulsive exponent. **Note that Coulombic (electrostatic) interactions are not computed because the MiPPE, TraPPE, and NERD force fields do not include charges for any of the compounds studied.**

Furthermore, Soroush Barhaghi et al. demonstrated that the inclusion of electrostatics in the optimization of the MiPPE parameters for alkynes did not improve the agreement with experimental phase properties. Therefore, it is unlikely that much improvement would be achieved by including electrostatics in our optimization of the new MiPPE cyclohexane parameters.

Comment #6

On page 10 in the last paragraph, the authors state that they use CBMC moves to enhance the insertion acceptance rate. It would be nice to show the actual acceptance ratio from sample simulations in the supporting information.

Response #6

We appreciate this suggestion from the reviewer. In accordance, we have provided the following table on page S4 in Supporting Information:

S3 CBMC acceptance rates

Table S4: Percentage acceptance of CBMC moves for cyclohexane with MiPPE force field. Averages were computed from 20 replicate simulations with $L_{\text{box}} = 3.0$ nm.

T (K)	μ (K)	Acceptance (%)
450	-4370	68.8
500	-4370	75.1
550	-4370	39.5
500	-4135	5.50
460	-4025	2.20
410	-3890	1.11
360	-3790	0.19

We also reference this table on page 11 in the manuscript:

site trials differ slightly. Section S3 provides an example of the CBMC acceptance rates for the different simulation state points.

Comment #7

In section 3.3, the authors show the effective snapshot results for GCMC-MBAR and use them to explain the bad performance of the method in predicting liquid phase properties. They also compare with MBAR-ITIC but the comparison is not very clear to me. It would be nice to have some data from MBAR-ITIC listed in the comparison. Also the authors listed two of their hypothesis that GCMC-MBAR would experience better overlap than MBAR-ITIC when $\theta_{rr} \approx \theta_{ref}$. Are these hypotheses proved to be correct after the comparison? Does GCMC-MBAR still do relatively better even for liquid phase since the two hypothesis the authors raise should hold true for liquid phase?

Response #7

Our results support both of our initial hypotheses.

However, it is actually not possible to include some data from MBAR-ITIC because MBAR-ITIC does not compute K_{eff} at saturated liquid or vapor conditions. Recall that ITIC is a vastly different approach than GCMC for computing phase equilibria. ITIC performs a series of NVT simulations along a supercritical temperature and along compressed liquid isochores. The lowest temperature along the liquid isochore is a “near-saturation” liquid state point though. As such, it is the most important state point for determining the reliability of MBAR-ITIC at predicting the saturated liquid phase, i.e., MBAR-ITIC will be unreliable for $\rho_{\text{liq}}^{\text{sat}}$ and $U_{\text{liq}}^{\text{sat}}$ if $K_{\text{eff}} < 50$ at this near-saturated liquid state point. Therefore, the best comparison that can be made with the GCMC-MBAR saturated liquid K_{eff} values are the MBAR-ITIC K_{eff} values at the lowest temperature isochore state point.

For the scenario where $\theta_{rr} \approx \theta_{ref}$, i.e., TraPPE \rightarrow MiPPE ($\lambda_{ref} = 12$ and $\lambda_{rr} = 16$) Messerly et al. reported for MBAR-ITIC that $K_{\text{eff}} \approx 1$ at this near-saturated liquid state point. In other words, only a single configuration

contributes to the MBAR estimate because there is essentially no overlap between the two force fields. By contrast, $K_{eff} > 1$ for the saturated liquid with GCMC-MBAR. Therefore, Figures 5 and 7 do support our initial hypothesis that GCMC-MBAR will experience better overlap for the saturated liquid compared to MBAR-ITIC..

Furthermore, although the GCMC-MBAR saturated liquid K_{eff} values are still typically less than 50 for TraPPE -> MiPPE, the phase equilibria estimates are significantly more reliable than those obtained with MBAR-ITIC with similar K_{eff} values. This supports our second hypothesis that, in contrast to MBAR-ITIC, GCMC-MBAR does not fail catastrophically in the case of poor overlap ($K_{eff} < 50$).

We added the following statements to emphasize the improvement of GCMC-MBAR relative to MBAR-ITIC.

Page 31:

able. In comparison, MBAR-ITIC experiences even worse overlap at near-saturated liquid conditions ($\bar{K}_{snaps,liq}^{eff} \approx 1$) for TraPPE \Rightarrow MiPPE-gen ($\lambda_{rr} \neq \lambda_{ref}$).¹⁶

Page 35:

When $\lambda_{rr} = \lambda_{ref}$, MBAR-ITIC is reliable ($\bar{K}_{snaps,liq}^{eff} > 50$) for $\sigma_{rr} = \sigma_{ref} \pm 0.0025$ nm,¹⁶ while Figure 7 (top left panel) suggests that GCMC-MBAR is reliable over a much wider range ($\sigma_{rr} \approx \sigma_{ref} \pm 0.01$ nm). Although the overlap is significantly better for GCMC-MBAR compared to MBAR-ITIC for $\lambda_{rr} \neq \lambda_{ref}$, Figure 7 demonstrates that GCMC-MBAR and MBAR-ITIC follow a similar trend, namely, the high $\bar{K}_{snaps,liq}^{eff}$ region corresponds to $\sigma_{rr} < \sigma_{ref}$ when $\lambda_{rr} > \lambda_{ref}$ and $\epsilon_{rr} > \epsilon_{ref}$. This poses a challenge for parameterization because the optimal σ_{CH_2} value is fairly constant while the optimal ϵ_{CH_2} value tends to increase with respect to λ_{CH_2} .

Page 35:

tic would be important when implementing gradient descent optimization schemes. By contrast, MBAR-ITIC yields sporadic contours for $\lambda_{rr} \neq \lambda_{ref}$ that are unreliable for optimization.¹⁶

Reviewer #2

Recommendation: Publish after minor revisions noted.

General:

The manuscript concerns molecular simulations of phase equilibria and optimization of force field parameters. The authors apply MBAR to GCMC simulations for calculating phase equilibria and use the property of MBAR that reweighting can be conducted for intensive state variables (T and chemical potentials) as well as for Hamiltonian parameters (theta), on equal ground. This strong-point of MBAR is used for optimization of force field parameters.

This study is an important contribution to the field and should be published after some minor revision, as detailed below. Very fine work, indeed.

Comment #1

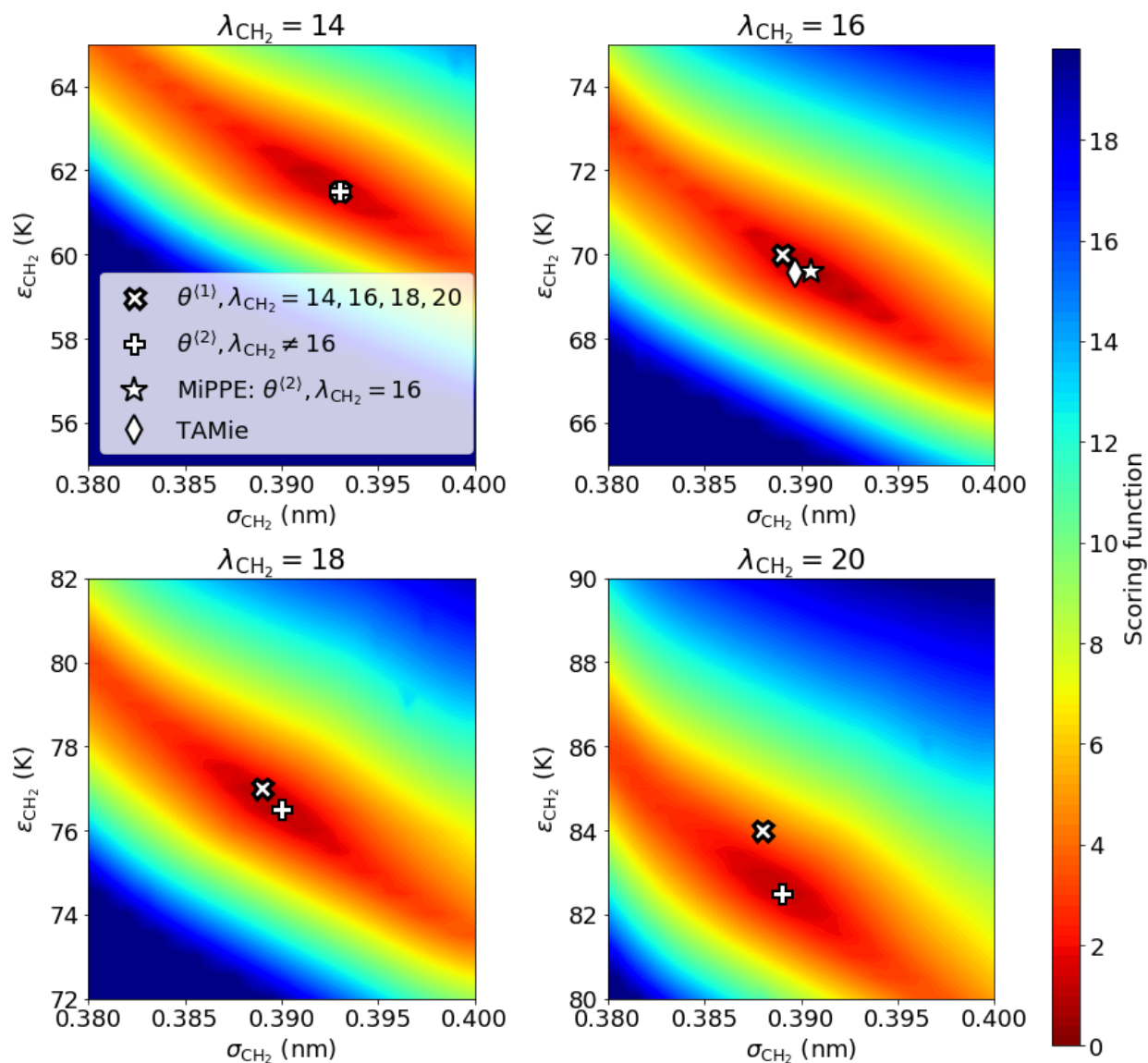
section 3.4: this case study ignores a paper where a Mie force field (also $\lambda=16$) was earlier proposed for cyclohexane, with (I guess) essentially the same results. Differences should appear only because the objective function is slightly different here. The conclusion of this work is that the here proposed force field is „the most accurate“. Omitting the previous study is (at least) awkward, because the authors seem to be aware of the line of developments from these authors, as the reference list shows. Is there a reason for omitting the study mentioned above from the list of „most reliable force fields from literature. Ref 22,24,27,55,56,57“ ?

Response #1

We sincerely apologize for this omission. Despite our assumed familiarity with the TAMie series of publications, we failed to include the TAMie Mie parameters for cyclohexane.

In fact, the newly optimized MiPPE parameters for cyclohexane agree very closely with the TAMie parameters. Because we did not consider the TAMie parameters prior to this reviewer's comment, the good agreement between the MiPPE and TAMie parameters (inadvertently) provides supporting evidence that our GCMC-MBAR results are reliable.

To better capture the literature, we now depict the TAMie parameters in Figure 8:



And refer the reader to this addition on page 38:

mization for each λ_{CH_2} . Also depicted in the $\lambda_{\text{CH}_2} = 16$ panel are the TAMie cyclohexane parameters.⁵⁶ Note that the first iteration optimal parameter sets (white "X"s) are similar

We also added the following statements to discuss the close agreement between MiPPE and TAMie.

Page 38:

ering even integer values. This optimal $\lambda_{\text{CH}_2} = 16$ value also agrees with the TAMie force field. The overall optimal parameter set ($\theta^{(2)}$, $\lambda_{\text{CH}_2} = 16$) is included in Table 1 as the MiPPE cyclohexane parameters. ϵ_{CH_2} and σ_{CH_2} are reported with three and four digits, respectively, consistent with other MiPPE parameters and to provide a qualitative measure of uncertainty. Note the close agreement between the MiPPE and TAMie parameters⁵⁶ ($\epsilon_{\text{CH}_2}/k_{\text{B}} = 69.568 \text{ K}$, $\sigma_{\text{CH}_2} = 0.38967 \text{ nm}$, and $\lambda_{\text{CH}_2} = 16$), which were optimized with a slightly different objective function and experimental data set.

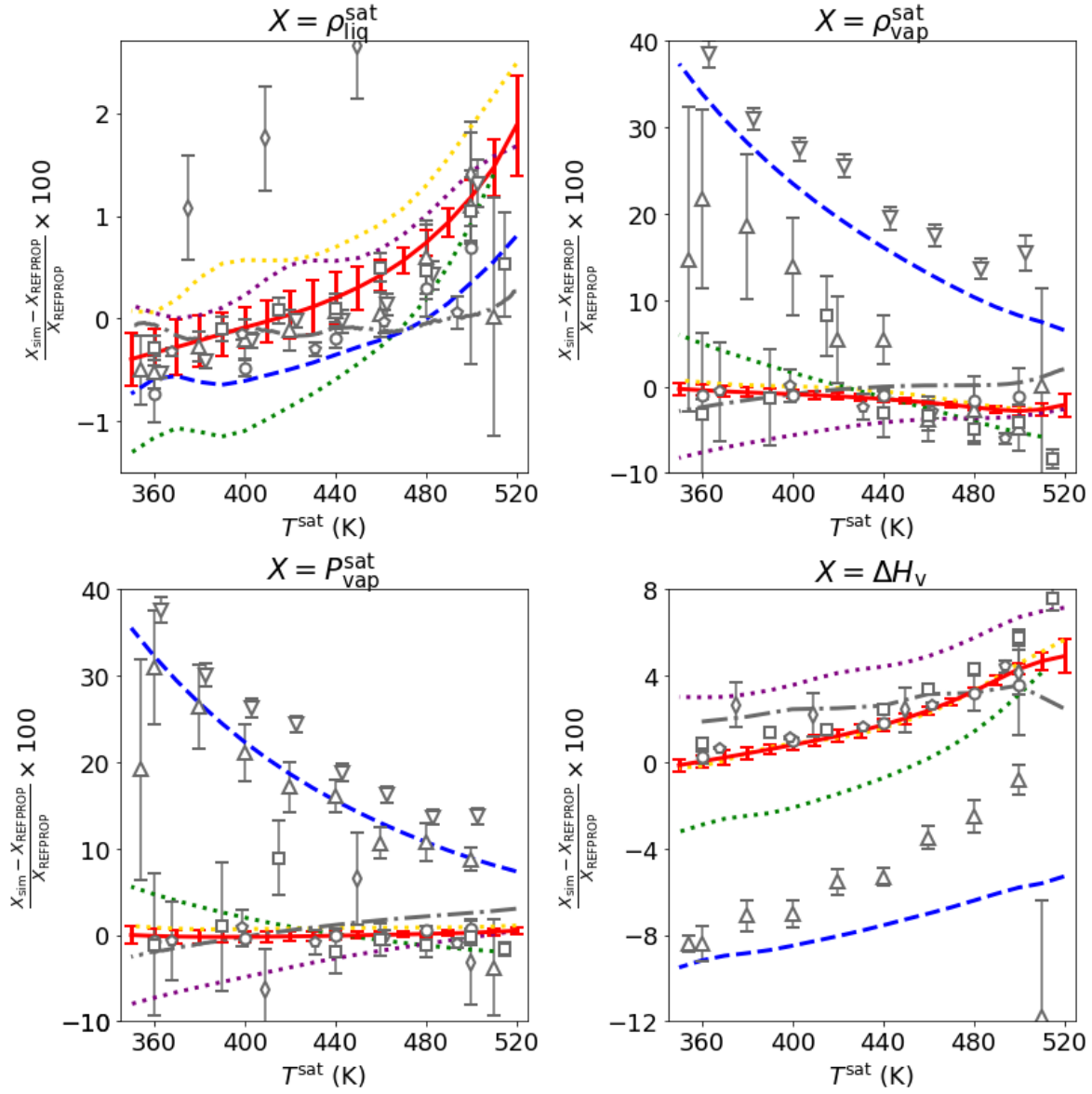
Page 42:

ishing the accuracy for $\rho_{\text{liq}}^{\text{sat}}$ and ΔH_{v} . In fact, the first iteration $\lambda_{\text{CH}_2} = 16$ parameter set achieves similar deviations as the TAMie force field. The second (final) iteration provides further improvement in each property compared to the first iteration. In comparison with literature force fields, MiPPE (and TAMie) are arguably the most accurate at predicting

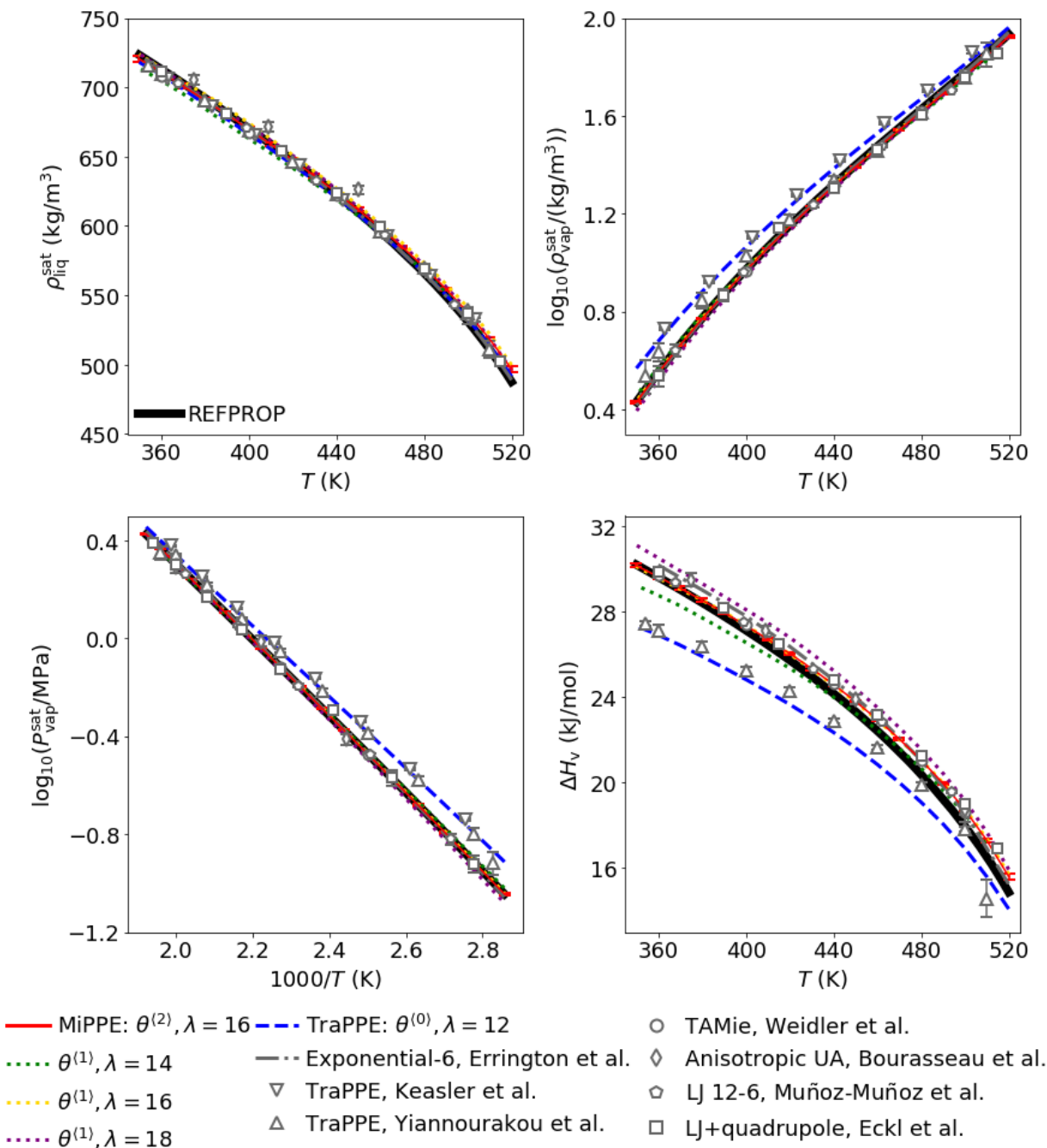
We also refer the reader to the supporting information for a detailed comparison of MiPPE and TAMie (see comment #4) on page 40:

Information, including a detailed comparison between the MiPPE and TAMie force fields.

We now include the reported TAMie values (Weidler et al., grey circles) in Figure 9:



And in Figure S2 of Supporting Information:



Comment #2

page 9 reads: „The system volume is constant with respect to μ and T for a given compound“. This statement is confusing. Things would be clear if the statement simply read: „The system volume is constant for a given compound“. Do the authors combine simulations for several volumes (for various (μ, T) -conditions)?? It would be unclear to me how this is realized technically (and would thus require clear explanations). And it would be questionable conceptually, because finite size scaling effects would be obscured.

Response #2

To improve clarity, we have modified this sentence on page 10 as suggested:

The system volume is constant for a given compound. The cubic box side length is 3

Comment #3

comment about equation 13: I am glad that the sign of the μ^*N term is now corrected to be negative. I assumed the positive sign to be a misprint in the first MBAR paper. But when the wrong positive sign reappeared in subsequent publications I started to have doubt about whether some intricacy existed about the sign.

Response #3

The reviewer is right that the negative sign found in equation 13 is correct. An upcoming review of MBAR will help clarify the sign error found in previous MBAR publications.

Comment #4

Eq. 17: internal energies for vapor phases are suffering strong finite size scaling effects in GCMC simulations with low number / vanishing number of molecules. Among all quantities considered in this work, it is probably the least size-robust quantity. The enthalpy of vaporization determined from eq. 17 therefore suffers from substantial finite size effects. Robust values for enthalpies of vaporization that more closely resemble infinite size systems are determined from the Clausius-Clapeyron equation (as stated by Weidler and Gross, 2016, see above). The derivative $d\ln(p_{\text{sat}})/d(\beta)$ is a native quantity in the spirit of histogram reweighing.

Response #4

Although we agree with the reviewer that it is important to consider the impact of finite-size effects, we have verified that finite-size effects are negligible for ΔH_v at low temperatures. This validation was performed by repeating the analysis for 20 replicate simulations with a larger box size (3.5 nm box length instead of 3.0 nm). This box size (volume of 42.875 nm³) is consistent with that utilized by Weidler and Gross.

The results suggest that finite-size effects are only statistically significant for ΔH_v (and $\rho_{\text{vap}}^{\text{sat}}$) near the critical point. However, this deviation of 1% to 2% near the critical point is unrelated to the reason presented by the reviewer, i.e., a vapor phase suffering from low number / vanishing number of molecules. The behavior near the critical point is well-known and not a consequence of Eq 17 or the data analysis. The following paragraph has been added to page 40 to discuss the finite-size test:

Furthermore, because we have performed the MiPPE parameterization with a relatively small box size (27 nm³), it is important to test for the existence of finite-size effects. By performing additional simulations with a larger box size (42.875 nm³), we conclude that finite-size effects are only significant (larger than the combined uncertainties) for $\rho_{\text{vap}}^{\text{sat}}$ and ΔH_v near the critical point. Specifically, we estimate that finite-size effects for both $\rho_{\text{vap}}^{\text{sat}}$ and ΔH_v are between 1% and 2% for $T^{\text{sat}} > 480$ K (see Section S6.5 in Supporting Information).

The corresponding figures found on pages S26 and S27 of Supporting Information are also presented below:

S6.5 Finite-size effects

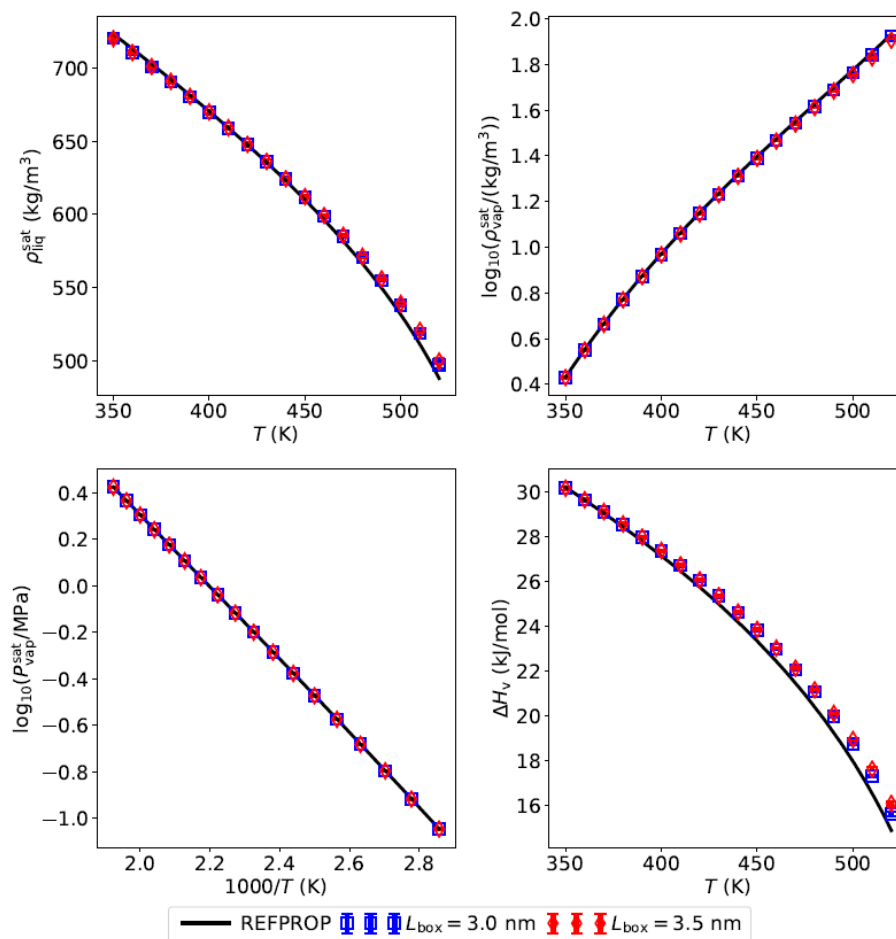


Figure S7: Comparison of phase equilibria using simulations with 3.0 nm and 3.5 nm box lengths for cyclohexane with MiPPE force field. Estimated values and uncertainties are obtained from 20 independent replicate simulations. See Figure S8 for quantitative assessment of finite-size effects.

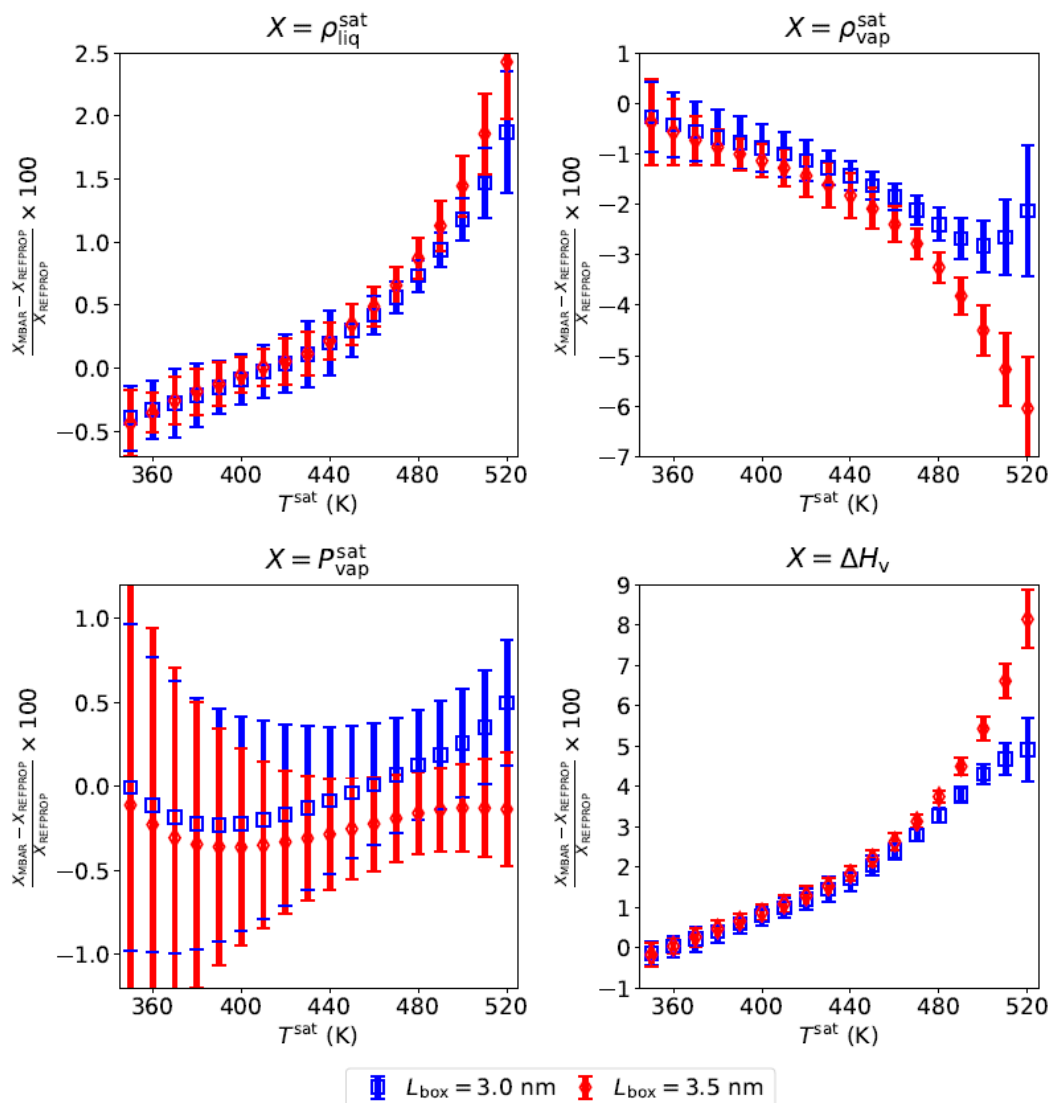


Figure S8: Comparison of percent deviations using simulations with 3.0 nm and 3.5 nm box lengths for cyclohexane with MiPPE force field. Agreement is typically within the combined uncertainties for low to moderate temperatures. Finite-size effects are most prevalent in $\rho_{\text{vap}}^{\text{sat}}$ and ΔH_v near the critical temperature ($T^{\text{sat}} > 480$ K). Estimated values and uncertainties are obtained from 20 independent replicate simulations.

In addition, because Sections 3.1, 3.2, and 3.3 compare GCMC-MBAR results with GCMC-HR results from the previous publications of Mick et al. and Soroush Barhaghi et al., it is important that we utilize the same approach for estimating ΔH_v , namely, Eq 17. Although we could utilize the $\frac{d \ln(p_{\text{sat}})}{d \beta}$ in our cyclohexane optimization, we are concerned that utilizing two different methods would create confusion and would be inconsistent with how MiPPE was parameterized in previous studies.

Furthermore, we have performed additional simulations with the TAMie force field for cyclohexane and computed the phase equilibria (labeled “TAMie, This work” in the figures below). Although some disagreement (beyond the statistical uncertainties) exists between our TAMie ΔH_v values (obtained with Eq 17) and those reported by Weidler et al. (computed with $\frac{d \ln(p_{\text{sat}})}{d\beta}$), the deviations in ΔH_v are typically less than 0.5%.

Note that slightly larger deviations (around 1%) are observed between our TAMie saturation densities and vapor pressures and those of Weidler et al. As demonstrated previously, these systematic deviations are not related to finite-size effects. Such discrepancies in phase equilibria properties are not completely unexpected when comparing simulation results across different platforms. As the results of Weidler et al. were not obtained with an open-source code, we are left only to speculate as to the source of these small (albeit non-negligible) discrepancies.

These figures have been added to the Supporting Information on pages S23 and S24:

S6.3 TAMie comparison

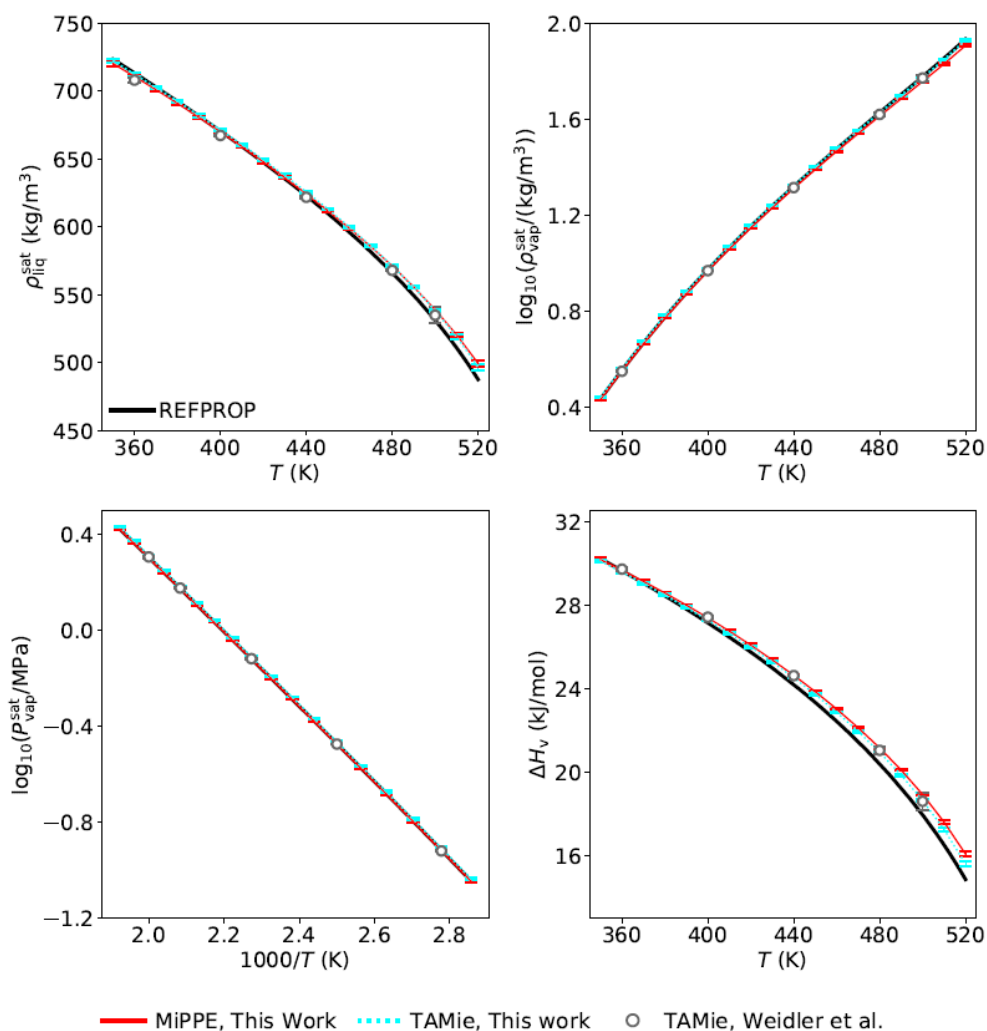


Figure S4: Comparison of MiPPE and TAMie (from this work and Weidler et al.) cyclohexane phase equilibria. All simulations from this work utilized 3.5 nm box length. TAMie simulations used a 1.4 nm cut-off.

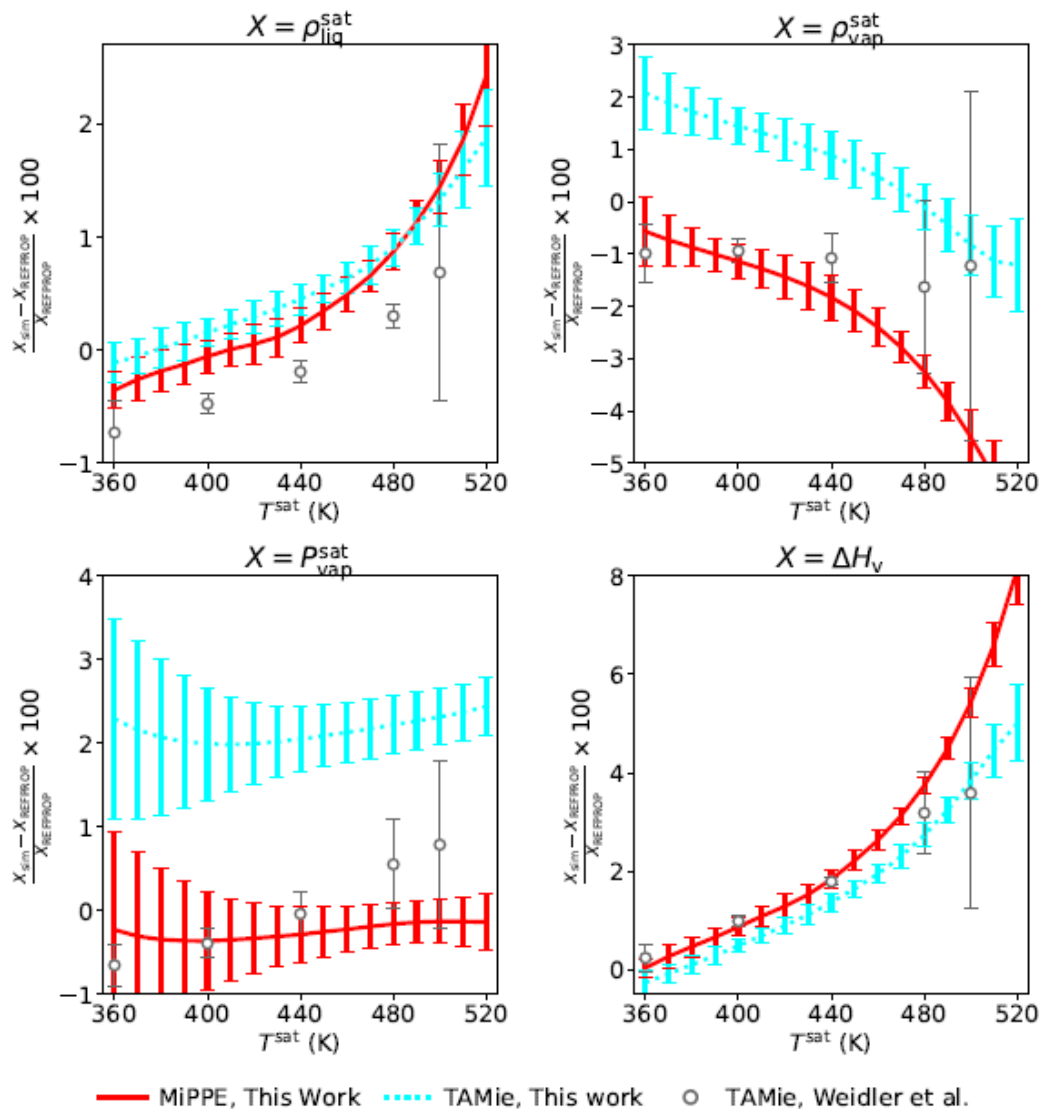


Figure S5: Comparison of MiPPE and TAMie (from this work and Weidler et al.) deviations relative to REFPROP cyclohexane values. All simulations from this work utilized 3.5 nm box length. TAMie simulations used a 1.4 nm cut-off. Note that discrepancies between TAMie results from this work and Weidler et al. are larger than the combined uncertainties (although they are typically between 0.5 and 2%). Such systematic discrepancies in phase equilibria are not completely unexpected when comparing simulation results across different platforms. Efforts to elucidate the source of these small (albeit non-negligible) discrepancies were unsuccessful.

Comment #5

page 10: the authors speak about molecule swap moves. What is meant, I guess, is insertion and deletion moves. However, „swap moves“ for most readers will suggest identity swap moves often performed for efficiently sampling the N-space of mixtures. I suggest modifying the terminology.

Response #5

We have replaced the term “swap moves” with “insertion and deletion moves.” The text on page 11 now reads:

the simulation temperature while molecule **insertion and deletion** moves ensure that μ is

Comment #6

why is the pressure defined along a slope extrapolation? In HR-GCMC it is simply determined from the probability density function at $N=0$, where it the value relates to the grand canonical partition function.

Response #6

We are utilizing the same method as Mick et al. and Soroush Barhaghi et al. to allow for a consistent comparison with these previous publications. The reviewer is correct that some studies (e.g., Weidler et al.) determine the ideal gas limit by computing the probability density at $N=0$. However, both methods are equally valid. Furthermore, in our experience, the difference between the pressures computed with these two methods is negligible.

To clarify, although our TAMie Psat results disagree by 1-2% with those reported by Weidler et al. (see Comment #4), we do not attribute this to the method used to compute the absolute pressure. The discrepancy in vapor pressure is consistent with the disagreement in vapor density. In other words, the compressibility factor in the vapor phase is consistent between our results and those of Weidler et al. (see figure below). Therefore, the ideal gas reference pressure is not the source of the discrepancy, rather the deviation in vapor density leads to the deviation in vapor pressure. The following plot has been included as Figure S11 on page S37 in supporting information:

S7.3 TAMie validation

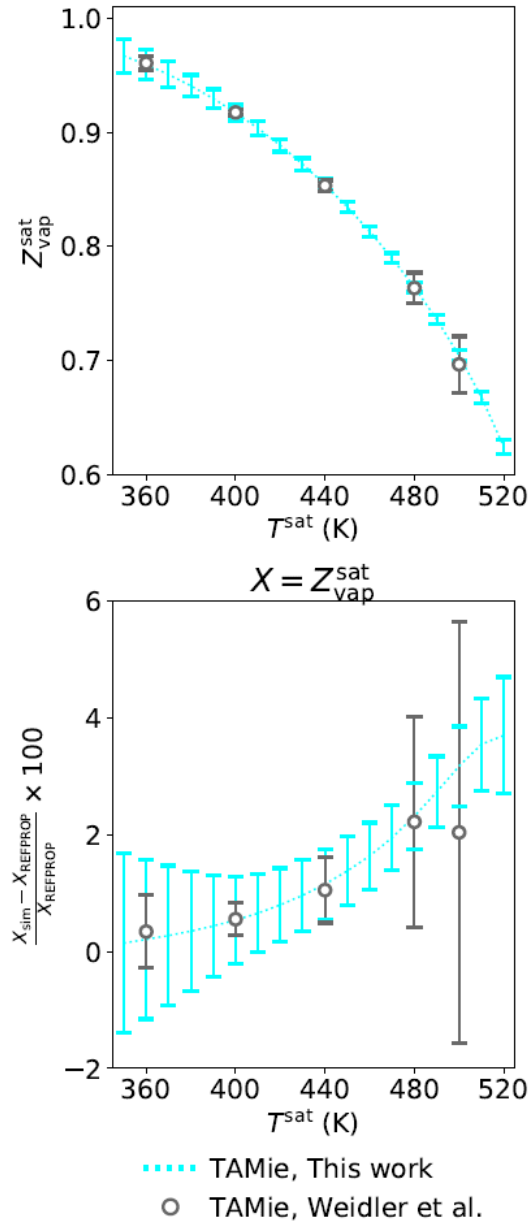


Figure S11: Comparison of compressibility factor in saturated vapor phase ($Z_{\text{vap}}^{\text{sat}}$) for TAMie (from this work and Weidler et al.). Simulations from this work utilized 3.5 nm box length with a 1.4 nm cut-off. Error bars apply standard propagation of error that assumes independence of $\rho_{\text{vap}}^{\text{sat}}$ and $P_{\text{vap}}^{\text{sat}}$.

Comment #7

Writing:

page 2: it is more appropriate to speak of „large number of simulations“ instead of „large amount of simulations“. Two appearances. (I think - I am not a native speaker).

Response

We agree with the reviewer's grammatical correction. We have removed both instances of “large amount of simulations.”

page 13, bottom: I propose to clarify that $K_{\text{snaps},i}$ denotes the number of uncorrelated configuration samples of simulation i .

Response

We have added “uncorrelated configuration” to the description of $K_{\text{snaps},i}$ on page 13:

where R is the number of runs, $K_{\text{snaps},i}(N, U)$ is the number of (uncorrelated) configuration “snapshots” in the i^{th} run (with β_i and μ_i) that have N molecules and U within the

Eq. 23: The explanation for the derivative terms (Eq.25) is given at the end of the section (page 23). I would think many readers would appreciate the explanation be moved forward to eq. 25.

Response

We tried to move Eq 25 and/or the explanation of derivatives closer to Eq 23, but the discussion of weights would then be out of place as it would precede the paragraph where the weights are explained. For this reason, we did not modify the manuscript.

page 26: Weigler should read Weidler.

Response

We appreciate the reviewer's attention to detail. This typo has been corrected.

Reviewer #3

Recommendation: Publish after minor revisions noted.

General:

The authors have demonstrated use of histogram-free reweighting for optimizing the van der Waals component of the non-bonded interactions for the simulation of vapor-liquid phase equilibria. This is very good work, and the paper is clearly written. I recommend publication of this work subject to minor revisions. The authors may want to consider the following points:

Comment #1

In figure 2, the authors perform small local variations to identify a slightly shifted local minima for the different molecules. Have the authors considered performing larger variations in the scaling factor for the well-depth to check if there is an even lower minima in the scoring (error) function? It may be useful to also show the sensitivity of these plots to reasonable variations in the scoring function. (Let's say vary the distribution of weights between liquid density and vapor pressure by about 10%).

Response #1

We appreciate the reviewer's suggestion. Figure 2 presents results from a 1-D optimization of a highly pre-tuned force field. Consistent with the results of Weidler et al., the magnitude of the scaling parameter (ψ) is approximately unity. We have also explored a wider range of ψ values, but the errors grow monotonically for $\psi > 1$ and $\psi < 1$. Figures 6 and 8 also demonstrate that, over a wide range of ϵ values, a unique optimal value of ϵ exists for a given σ value. Therefore, we already have strong evidence that we have obtained the global optimal ψ values.

To provide a fair assessment of the transferability of the original MiPPE parameter set, the scoring function (and the underlying experimental dataset) must be consistent between the MiPPE and the individualized MiPPE (iMiPPE). Varying the weights in the scoring function would inevitably impact the optimal scaling parameter value and, therefore, most compounds would result in $\psi \neq 1$. This would unfairly characterize the MiPPE parameters as being non-transferable.

For these reasons, no changes were made to the manuscript.

Comment #2

The authors present a very clear example of cyclohexane to demonstrate the utility of this approach in parametrizing new force fields. Can the authors comment on the complexity of the optimization process when more than two parameters need to be optimized (for example, a molecule comprising of more than one type of interaction site)? Further, while this may be beyond the scope of this work, it may be useful to comment on the ease/difficulty of using this approach when vapor-liquid equilibria of mixtures are considered as part of the scoring function.

Response #2

We agree with the reviewer that it would be beneficial to discuss how to perform the optimization for a compound with more than one united-atom site. The primary reason we performed a single-site optimization is that the 2-D heat maps are easy to visualize. For a higher dimensional optimization, a grid search based scheme would likely no longer be feasible.

More importantly, the GCMC-MBAR analysis would be unchanged. If employing basis functions, however, additional basis functions would need to be computed for each of the ii and ij pair interactions. The addition of basis functions does not significantly increase computational costs and the memory load is still much less than

storing configurations. To assist future application of this approach, the following paragraph has been added to page 42:

While a higher dimensional parameterization would necessitate a more sophisticated optimization scheme (Step 3b), the GCMC-MBAR analysis (Step 3a) would be unchanged. If implementing basis functions with GCMC-MBAR (Step 2), however, recomputing the energy for θ_{tr} requires an attractive and repulsive basis function for all ii (same) and ij (cross) pair interaction sets. For example, propane would require 6 basis functions (an attractive and repulsive basis function for the $\text{CH}_3\text{-CH}_3$, $\text{CH}_3\text{-CH}_2$, and $\text{CH}_2\text{-CH}_2$ interactions). Although the memory requirement scales linearly as the number of basis functions, the storage load should still be manageable and significantly less compared to storing configurations. Furthermore, although generating these additional basis functions also requires a larger number of “rerun” calculations (recall Section 2.4), the increase in computational cost is negligible compared to performing direct GCMC simulations with each proposed parameter set.

We agree with the reviewer that including mixtures in the scoring function is both of interest but also outside of the scope of the present work. We do not wish to address this topic as it merits an entire paper for an adequate discussion.

Comment #3

For polar molecules, do the authors anticipate that the GCMC-MBAR approach will be capable of parametrizing the electrostatic charges with a similar efficiency as it can be done for the non-polar interactions?

Response #3

GCMC-MBAR is applicable to parameterizing electrostatics. For example, previous work (Ref. 46) has demonstrated the capability of MBAR to estimate solvation free energies when varying both LJ and electrostatic parameters, e.g., the point charge parameter (q).

In general, with sufficient overlap (high number of effective snapshots), MBAR will provide accurate estimates regardless of the parameter type. However, we would only be able to speculate as to the size of the reliability region for electrostatic parameters. The following paragraph has been added on page 44:

The present study has focused on the van der Waals Mie λ -6 non-bonded parameters (ϵ , σ , and λ). However, GCMC-MBAR can also be applied to parameterize electrostatic non-bonded parameters, e.g., point charges (q).⁴⁶ While the number of effective snapshots should still provide a reasonable estimate of reliability for $q_{\text{tr}} \neq q_{\text{ref}}$, we recommend that future work test the range of q_{tr} over which GCMC-MBAR is reliable.

Comment #4

Typographical corrections:

Page 4, line 3: "scheme"

Page 15, line 20: "probably"

Response #4

The word on page 4, line 3 is actually "schema." Definition – a representation of a plan or theory in the form of an outline or model

The word on page 15, line 20 is actually "provably." Definition – in a way that is capable of being proved.

No changes were made.