

SCIENTIFIC REPORTS



OPEN

Data driven inference for the repulsive exponent of the Lennard-Jones potential in molecular dynamics simulations

Lina Kulakova¹, Georgios Arampatzis¹, Panagiotis Angelikopoulos^{1,3}, Panagiotis Hadjidoukas¹, Costas Papadimitriou² & Petros Koumoutsakos¹

The Lennard-Jones (LJ) potential is a cornerstone of Molecular Dynamics (MD) simulations and among the most widely used computational kernels in science. The LJ potential models atomistic attraction and repulsion with century old prescribed parameters ($q=6$, $p=12$, respectively), originally related by a factor of two for simplicity of calculations. We propose the inference of the repulsion exponent through Hierarchical Bayesian uncertainty quantification. We use experimental data of the radial distribution function and dimer interaction energies from quantum mechanics simulations. We find that the repulsion exponent $p \approx 6.5$ provides an excellent fit for the experimental data of liquid argon, for a range of thermodynamic conditions, as well as for saturated argon vapour. Calibration using the quantum simulation data did not provide a good fit in these cases. However, values $p \approx 12.7$ obtained by dimer quantum simulations are preferred for the argon gas while lower values are promoted by experimental data. These results show that the proposed LJ 6- p potential applies to a wider range of thermodynamic conditions, than the classical LJ 6-12 potential. We suggest that calibration of the repulsive exponent in the LJ potential widens the range of applicability and accuracy of MD simulations.

The Lennard-Jones (LJ) potential is one of the centerpieces in Molecular Dynamics (MD) simulations, the key computational method for studying atomistic phenomena across Chemistry, Physics, Biology and Mechanics. Despite the widespread use of MD simulations, an often overlooked fact is that the classic LJ potential involves a century old and rather ad-hoc prescribed repulsion exponent. In this study we demonstrate that this parameter needs to be modified in order to enhance the predictive capabilities of MD simulations.

The structure of the LJ potential depends on the inter-atomic distance (r) and consists of two parts: an attractive term $-r^{-q}$ that models the Van der Waals forces and a repulsive term r^{-p} that models the Pauli repulsion. While the exponent $q = 6$ has a theoretical justification¹, the $p = 12$ exponent was chosen for computational efficiency as the square of the attractive term. In addition, two scaling parameters σ and ε control the shape of the potential. The ε and σ parameters have been the subject of numerous calibration studies²⁻⁵ and more recently the subject of Bayesian inference techniques^{6,7}. Bayesian Uncertainty Quantification (UQ) employs experimental data and provides a probability distribution of the parameters. The parameter uncertainty can then be propagated by the model in order to obtain robust predictions on a quantity of interest^{7,8}. In cases where the data sets correspond to different inputs for the system, e.g. different thermodynamic conditions, the use of Hierarchical Bayesian (HB) methods provides a stable method for UQ^{9,10}.

Here we employ a HB method to infer the parameters (ε , σ , p) of LJ 6- p force-field. In the past, several values of the exponent p of the LJ 6- p potential, ranging from 10 to 20, have been considered¹¹. The authors calibrated using pressure and viscosity data for various thermodynamic conditions and concluded that the exponent 12 is the best choice.

We perform the HB inference for the LJ 6-12 and LJ 6- p parameters of argon based on experimental RDFs of liquid argon and saturated argon vapor for six different temperature and pressure pairs, as well as on one dataset

¹Computational Science and Engineering Laboratory, Clausiusstrasse 33, ETH Zürich, CH-8092, Switzerland.

²Department of Mechanical Engineering, University of Thessaly, Pedion Areos, GR-38334 Volos, Greece. ³Present address: D.E.Shaw Research LLC, New York, NY 10036, USA. Correspondence and requests for materials should be addressed to P.K. (email: petros@ethz.ch)

	Ref. ³	Ref. ²	Ref. ⁴	Ref. ⁵
ε	0.2385	0.2824	0.2381	0.2498
σ	3.4000	3.3605	3.4050	3.3450
T	94.4	86.64–168.86	137.77	88–127
ρ	1.374	0.435–1.479	0.156, 0.972	0.283–3.897
Phase	liquid	gas, liquid, solid	gas + liquid	gas + liquid
Data	RDF	P, E	TP, ρ, L	P, B

Table 1. LJ parameters for argon used in literature. The last row shows the data used for fitting. Notations: T (temperature), P (pressure), ρ (density), B (second virial coefficient), E (energy), TP (gas-liquid transition pressure), L (latent heat of evaporation).

from quantum calculations, representing gaseous argon. We present a rigorous model selection process for the LJ 6-12 vs LJ 6- p potentials for each of the cases and perform robust posterior predictions for the diffusion coefficient and density.

We find that the most likely values for the LJ repulsive exponent for liquid argon are $p \approx 6.5$, strongly differing from the value of $p = 12$ that is being used, while the gaseous argon is simulated best with the exponent $p \approx 12.7$, much closer to the conventional one. We remark that our results have been obtained in the case of a simple system. However, we consider that they offer significant evidence that the repulsive exponent should be reconsidered when the parameters of the LJ potential are being fitted to data.

Results

In our work, we perform several inferences of the parameters of the classical and modified LJ potentials for argon at different thermodynamic conditions. Here and further, ϑ denotes a set of parameters to be calibrated.

We first calibrate the parameters ε , σ of the classical LJ 6-12 potential, as well as the LJ model error parameter σ_n , which accounts for the inadequacy of the LJ description, even with the best possible parameters, as compared to the physical reality. The parameter set ϑ is thus equal to $\{\varepsilon, \sigma, \sigma_n\}$. We use RDF as data and denote this inference as $B_{12,R}$. Subsequently, we include the exponent p of the repulsion term into the parameter set (inference $B_{p,R}$) and perform model selection for the LJ 6-12 and LJ 6- p force fields.

We then perform a non-hierarchical and a hierarchical Bayesian inference for each of the potentials using the methodology from ref.¹⁰. In the non-hierarchical inference the parameters for each data set are being inferred completely independently with a common prior information, encoded in $p(\vartheta|\mathcal{M})$, see Supplementary Fig. S1c. In HB inference the parameters corresponding to each data set are being connected to a hyper-parameter ψ , see Supplementary Fig. S1b. In this approach the prior information of the model parameters is data dependent, $p(\vartheta|\psi, \mathbf{d})$, where \mathbf{d} denotes the set of all available data, allowing information to flow between the different data sets leading to more robust and accurate predictions for the model parameters. The aim of this procedure is to infer the hyper-parameters ψ governing the variability of the parameters ϑ between the different sets of data and to subsequently update the distributions of the parameters ϑ (see Supplementary Material S1 for details). These inferences are denoted $HB_{12,R}$ and $HB_{p,R}$.

The experimentally measured RDFs are taken from ref.¹². The RDFs are computed for 6 temperature/pressure pairs (T, P). We denote the pairs as follows: $L_1 = (84.4, 0.8)$, $L_2 = (91.8, 1.8)$, $L_3 = (126.7, 18.3)$, $L_4 = (144.1, 37.7)$, $L_5 = (149.3, 46.8)$, $V = (149.3, 43.8)$, where L stands for “liquid” and V stands for “vapor”. The corresponding datasets (RDFs) are denoted as R_{L_i} for liquid and R_V for vapor.

Finally, we perform the inference with LJ 6- p using quantum dimer energy calculations from ref.¹³ as data, which corresponds to the gaseous argon. We then compare the obtained parameter distribution with those computed for the liquid argon and saturated argon vapour from the RDF data. The quantum dimer dataset is denoted as Q and the corresponding inference is denoted as $B_{p,Q}$.

Calibration of LJ 6-12. We present results of parameter calibration for ε , σ , σ_n , while p is fixed to 12. We use a wide enough uniform prior for each of the parameters and each of the datasets R_{L_i}, R_V ($\vartheta \in [0.05, 3] \times [3, 4] \times [10^{-6}, 1]$). We observe that the values which were obtained in the calibration process are close to those found in literature (Table 1). In Fig. 1 the most probable values (MPV) of the parameters along with 5–95% quantiles is presented (light red). Notice that results are only presented for the four out of six datasets as the LJ 6–12 potential failed to simulate the liquid argon for conditions L_1 and L_2 . By “failed” we mean that there were no parameters for LJ 6-12 within the prior bounds which would produce liquid argon, the MD simulation was trapped in either gas or solid phase.

A large difference in the values of ε for liquid and vapor is observed, which implies that one cannot perform the simulations using the same parameters for the two phases. We define the uncertainty in a parameter as the ratio of the 5–95% quantile spread to the MPV. The uncertainty in ε varies from 14% to 20% depending on the dataset, while the value of σ is identified more precisely with uncertainty of 2–6%. This difference is attributed to the type of data used in the inference process: the location of the RDF peak, which gives the most significant contribution to the sum of squared errors (SSE) in the log-likelihood, is more sensitive to σ . On the other hand, ε affects the height of the RDF peak which has a smaller contribution to the log-likelihood. Despite the generally small uncertainty in σ , we observe it to be quite big for L_5 , as compared to the other liquid cases. The reason for this lies in the shape of the RDF for L_5 . Fig. 2 shows that the RDF for L_5 , unlike for the other liquid cases, is very

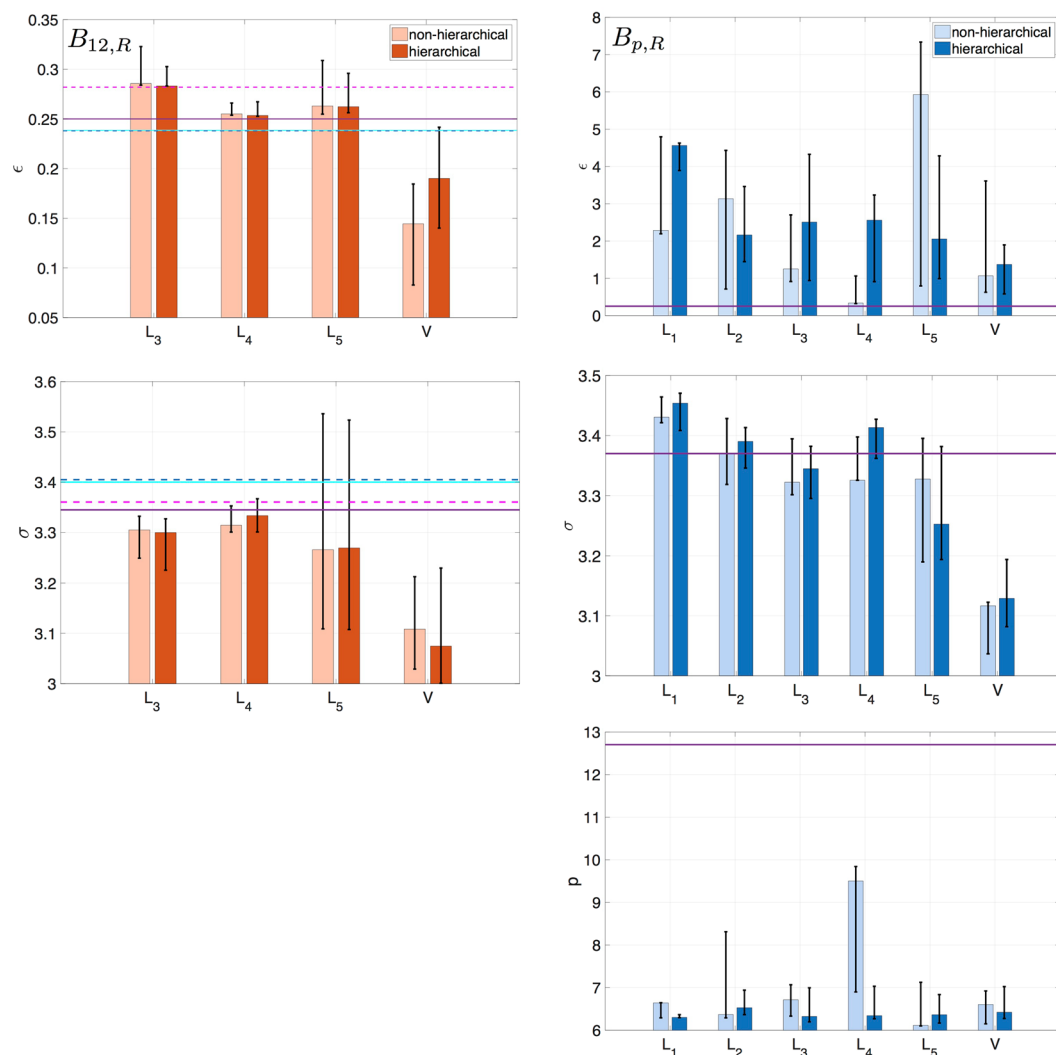


Figure 1. Posterior parameter values: MPVs along with 5–95% quantiles obtained in $B_{12,R}$ (light red), $HB_{12,R}$ (dark red), $B_{p,R}$ (light blue), $HB_{p,R}$ (dark blue). Horizontal lines for LJ 6–12 indicate the reference values: ref.² (magenta dashed line), ref.⁵ (purple solid line), ref.³ (cyan solid line), ref.⁴ (blue dashed line). Horizontal line for LJ 6– p indicates the MPV for $B_{p,Q}$.

Name	Symbol	Unit
Temperature	T	K
Pressure	P	atm
Distance	r	Å
LJ well depth	ε	kcal/mol
LJ well location	σ	Å
LJ repulsion exponent	p	—
RDF model error	σ_n	—
Density	ρ	g/cm ³
Diffusion coefficient	D	cm ² /s

Table 2. Units used in this work.

flat. The RDF peak location is controlled by σ , and since the peak is not well defined for L_5 , the sampling algorithm allows for a wide variety of σ values.

Next, we infer the LJ parameters using the HB approach. We select the prior $p(\vartheta_i|\psi)$ by using Bayesian model selection (see Supplementary Material S2 for details).

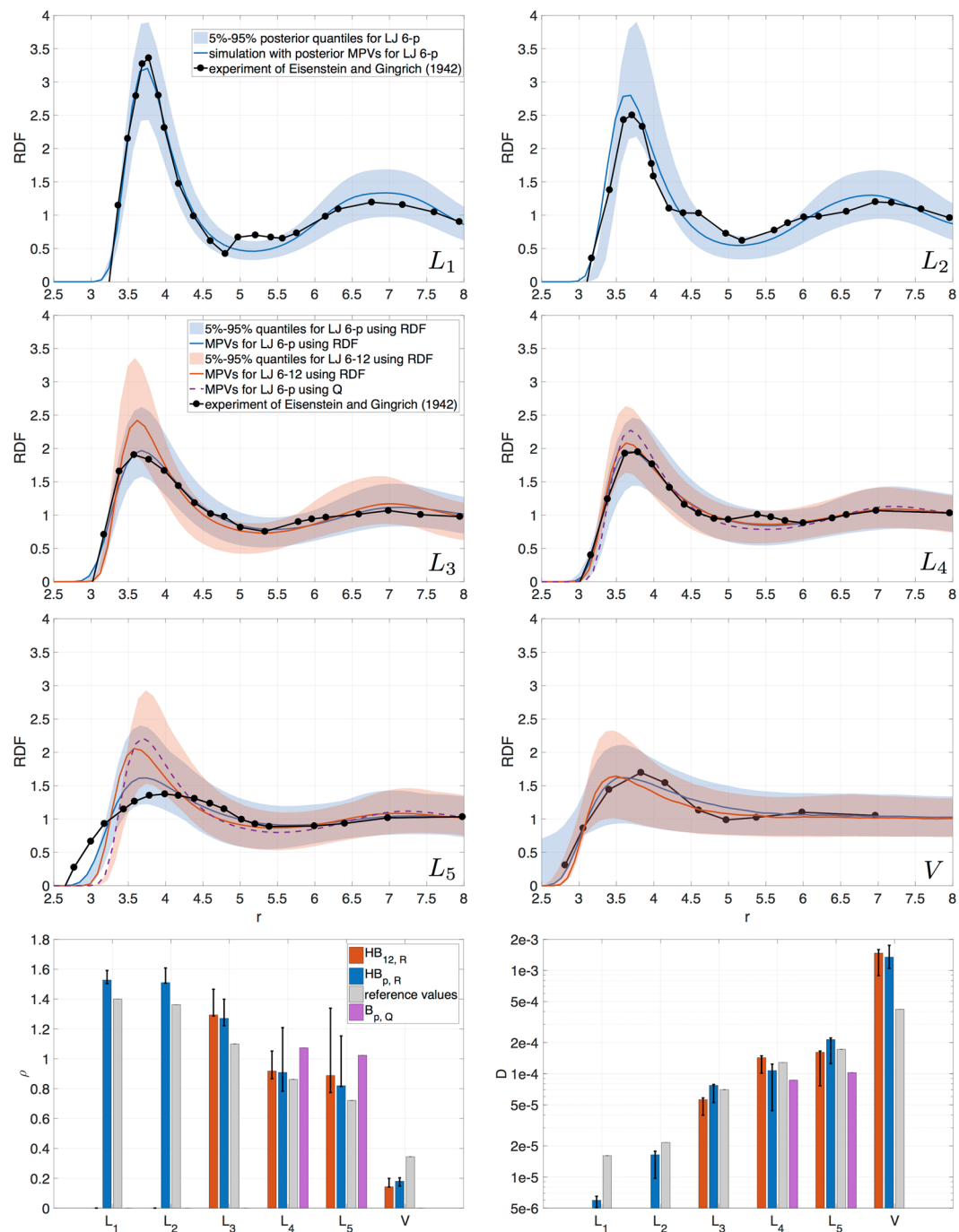


Figure 2. Robust posterior predictions: MPVs along with 5–95% quantiles obtained in $HB_{p,R}$ (blue), $HB_{12,R}$ (red), MPV obtained in $B_{p,Q}$ (purple). Black line with dots: experimental data for RDF. Grey bars: experimental data for ρ , analytically computed values for D .

The values of the LJ parameters are presented in Fig. 1 (dark red). The MPVs and the quantiles of the parameters are almost the same as in the $B_{12,R}$, which means that for each dataset $d_i \in \{R_{L_1}, \dots, R_{L_5}, R_V\}$ no information about the parameters can be extracted from the other datasets.

The full set of the MPVs and distribution quantiles for each dataset R_{L_i}, R_V is given in Table 3, while the full posterior distributions are shown in Supplementary Fig. S3.

Calibration of LJ 6- p . *Dataset R.* Here we include the LJ exponent p into the parameter set ϑ . As in the LJ 6-12 case, we choose a uniform prior with wide enough bounds ($[0.05, 10] \times [3, 4] \times [6.01, 15] \times [10^{-6}, 1]$). Note that with LJ 6- p the sampling algorithm dictated much wider bounds for ε compared to the LJ 6-12 case. As will be seen later, this is due to a strong correlation between ε and p . We observe again the non-transferability of

		$b(\varepsilon)$	$q(\varepsilon)$	$b(\sigma)$	$q(\sigma)$	$b(\sigma_n)$	$q(\sigma_n)$
$B_{12,R}$	L_3	0.286	[0.284, 0.323]	3.305	[3.250, 3.332]	0.168	[0.158, 0.314]
	L_4	0.255	[0.254, 0.266]	3.314	[3.301, 3.353]	0.089	[0.080, 0.140]
	L_5	0.263	[0.255, 0.309]	3.266	[3.110, 3.536]	0.317	[0.292, 0.586]
	V	0.144	[0.083, 0.184]	3.109	[3.029, 3.213]	0.147	[0.119, 0.304]
$HB_{12,R}$	L_3	0.283	[0.283, 0.303]	3.300	[3.226, 3.327]	0.177	[0.156, 0.310]
	L_4	0.253	[0.254, 0.267]	3.333	[3.301, 3.367]	0.073	[0.073, 0.147]
	L_5	0.262	[0.256, 0.296]	3.269	[3.108, 3.523]	0.337	[0.301, 0.579]
	V	0.190	[0.140, 0.242]	3.075	[3.001, 3.229]	0.180	[0.192, 0.385]

Table 3. Posterior values of each parameter $\vartheta \in \{\varepsilon, \sigma, \sigma_n\}$ of LJ 6-12: MPV $b(\vartheta)$ and 5–95% quantiles $q(\vartheta)$. The values are rounded from 16-digit precision.

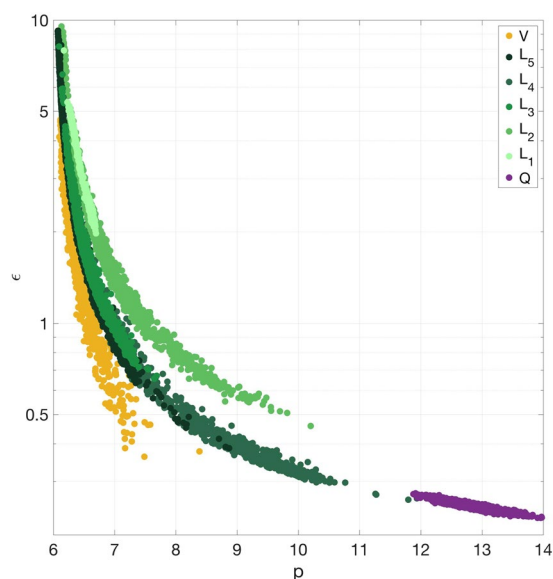


Figure 3. Posterior samples of $B_{p,R}$ projected onto (p, ε) subspace: yellow circles correspond to V , green circles correspond to L_i in the temperature increasing order from the lightest to the darkest color, purple circles correspond to Q . Note that the ε scale is logarithmic.

the LJ parameters from liquid to vapor simulations: the values of σ lie in disjoint domains for L_i and V (Fig. 1). Being a more flexible potential, LJ 6- p can simulate a wider range of thermodynamic conditions, including L_1 and L_2 , which result in the values of LJ parameters similar to those obtained for the other three liquid conditions. We observe that the 95% quantile of p , as well as its MPV, is for four out of six RDF datasets below 7.5 and for all the datasets below 10, which is much smaller than the conventional 12. This can be explained by the fact that the repulsion energies predicted by the standard 6-12 LJ potential are very high for the liquids. The configurations with such energies happen with probability close to zero, and the MD simulation is not able to sample them. As in the LJ 6-12 case, the parameter ε exhibits significant variation within each dataset R_{L_i}, R_V (uncertainty 110–216%, computed the same way as for LJ 6-12), while p and σ are well-defined with the uncertainty of 5–30% and 1–6%, respectively. In addition, ε differs substantially among the RDF datasets, but always in accordance with p : the higher the p , the lower the ε (see Fig. 3). We observe that the uncertainty in σ is once again bigger for L_5 , as compared to the other liquid cases, which is consistent with the LJ 6-12 inference.

Similarly with the inference of the LJ 6-12 parameters, we proceed by calibrating the parameters using the HB approach. Details for the selection of the prior can be found in Supplementary Material S2. The results of the inference are given in Fig. 1. We observe that the uncertainty in ε gets significantly reduced for conditions L_1, L_2, L_5 and V indicating that the inference benefited from the information contained in the two remaining datasets R_{L_3} and R_{L_4} with narrow posterior distributions of ε (Supplementary Figs S2 and S4). On the other hand, the uncertainty in ε for L_3 and L_4 increases adjusting to the wide ranges in the other four cases. A similar situation can be seen for p , where narrow distributions for L_1, L_3, L_5, V shift the posterior values for L_2 and L_4 . The RDF is, as noticed before, very sensitive to the changes in σ , which controls the location of the LJ potential well, and therefore σ is well determined for each of the datasets R_{L_i}, R_V and extracts almost no information from the other data sets.

		$b(\varepsilon)$	$q(\varepsilon)$	$b(\sigma)$	$q(\sigma)$	$b(p)$	$q(p)$	$b(\sigma_n)$	$q(\sigma_n)$
$B_{p,R}$	L_1	2.286	[2.193, 4.794]	3.431	[3.422, 3.464]	6.644	[6.292, 6.647]	0.157	[0.185, 0.302]
	L_2	3.134	[0.712, 4.416]	3.369	[3.319, 3.428]	6.370	[6.293, 8.313]	0.222	[0.215, 0.518]
	L_3	1.250	[0.914, 2.700]	3.322	[3.301, 3.395]	6.715	[6.332, 7.068]	0.098	[0.080, 0.159]
	L_4	0.337	[0.322, 1.060]	3.325	[3.325, 3.398]	9.501	[6.900, 9.842]	0.057	[0.066, 0.153]
	L_5	5.928	[0.794, 7.318]	3.328	[3.190, 3.395]	6.116	[6.102, 7.126]	0.164	[0.168, 0.326]
	V	1.065	[0.625, 3.611]	3.117	[3.037, 3.122]	6.604	[6.151, 6.923]	0.100	[0.099, 0.190]
$HB_{p,R}$	L_1	4.561	[3.890, 4.626]	3.454	[3.408, 3.471]	6.302	[6.296, 6.366]	0.422	[0.450, 0.602]
	L_2	2.081	[1.333, 3.515]	3.387	[3.335, 3.424]	6.565	[6.340, 7.051]	0.211	[0.178, 0.351]
	L_3	2.506	[0.941, 4.325]	3.345	[3.295, 3.382]	6.324	[6.198, 6.996]	0.093	[0.081, 0.186]
	L_4	2.588	[0.892, 3.676]	3.403	[3.358, 3.432]	6.339	[6.226, 7.063]	0.082	[0.075, 0.146]
	L_5	2.055	[0.992, 4.281]	3.252	[3.194, 3.382]	6.364	[6.169, 6.837]	0.183	[0.159, 0.316]
	V	1.371	[0.582, 1.896]	3.129	[3.082, 3.194]	6.422	[6.277, 7.025]	0.111	[0.103, 0.233]
$B_{p,Q}$		0.252	[0.239, 0.261]	3.370	[3.367, 3.375]	12.703	[12.333, 13.309]	0.006	[0.006, 0.010]

Table 4. Posterior values of each parameter $\vartheta \in \{\varepsilon, \sigma, p, \sigma_n\}$ of LJ 6- p : MPV $b(\vartheta)$ and 5–95% quantiles $q(\vartheta)$. The values are rounded from 16-digit precision.

	$E_{12,R}$	$E_{p,R}$	$e^{E_{p,R} - E_{12,R}}$
L_1	—	−7.05	—
L_2	—	−14.8	—
L_3	−9.72	2.81	2.74×10^5
L_4	5.10	5.18	1.09
L_5	−15.8	−8.76	1.18×10^3
V	−3.83	−4.94	3.31×10^{-1}

Table 5. Log-evidences $E_{12,R}$ ($E_{p,R}$) for $B_{12,R}$ ($B_{p,R}$).

Dataset Q. We examine the suitability of the repulsion exponent 12 by performing a calibration using as data the calculated quantum dimer scans of argon. These data describe the behavior of the gaseous argon. We infer the LJ 6- p parameters by fitting the LJ potential to the binding energy of the quantum dimer (Fig. 1). The resulting value of p is much closer to the conventional 12 (Table 4), suggesting that for the gaseous argon, unlike for the liquid one, LJ 6-12 is a reasonable choice.

The full set of MPVs and distribution quantiles of the LJ parameters for R_{L_i} , R_V and Q is given in Table 4 and the full posterior distributions are plotted in Supplementary Figs S2, S4 and S5.

LJ 6-12 vs LJ 6- p : Comparison. *Model selection.* We choose between the LJ 6-12 and LJ 6- p potentials by applying the Bayes selection criterion. We observe that LJ 6- p is significantly better than the LJ 6-12 for L_3 and L_5 (Table 5). Recalling that LJ 6-12 is not able to produce a liquid for L_1 and L_2 , we conclude that LJ 6- p is preferred for four RDF datasets out of six. In the case of L_4 the two potentials provide results that are indistinguishable by the Bayesian model selection. The only dataset on which the LJ 6-12 potential produces better results (3 times more probable than LJ 6- p) is V , the vapor case. That brings us to the conclusion that LJ 6- p is either much better or not worse than LJ 6-12 for all the liquid cases considered. For the vapor case, the LJ 6- p is over parametrized, as compared to LJ 6-12.

LJ potentials. Studying the reasons for LJ 6- p being more plausible than LJ 6-12, we take a closer look at the inferred shapes of the potentials. We observe a very stable correlation in the (p, ε) subspace (Fig. 3) for all the datasets used. This result is expected as p regulates the strength of the repulsion and ε alters the strength of both repulsion and attraction simultaneously. The difference between the R_i and Q datasets shows up in the region of the subspace which gets populated. The quantum dimer-based calibration prefers high values of p , which correspond to the tails of the distributions inferred using the RDF data. We performed a calibration with L_3 and narrow prior bounds ($p \in [12, 14]$) to see whether this is indeed a tail of the full posterior distribution (Supplementary Fig. S6). The narrow posterior values are below 3.95, while the values of the full posterior start from 4.18, which explains why the tails of the full distributions for L_i and V have a negligible number of samples in the region $p \in [12, 14]$ preferred by the Q -based inference. As the parameters ε and p are highly correlated, one could expect that the inference will be able to recover values of ε for LJ 6-12 such that the resulting potential is close to the inferred LJ 6- p . However, the effect that p and ε have on the LJ potential is not entirely the same. As ε acts as a scaling factor for the whole potential, it is not able to make the potential less deep and at the same time flat enough to avoid switching to the gas phase (compare simulations with MPVs for L_5 , V in Fig. 4). The same reasoning can

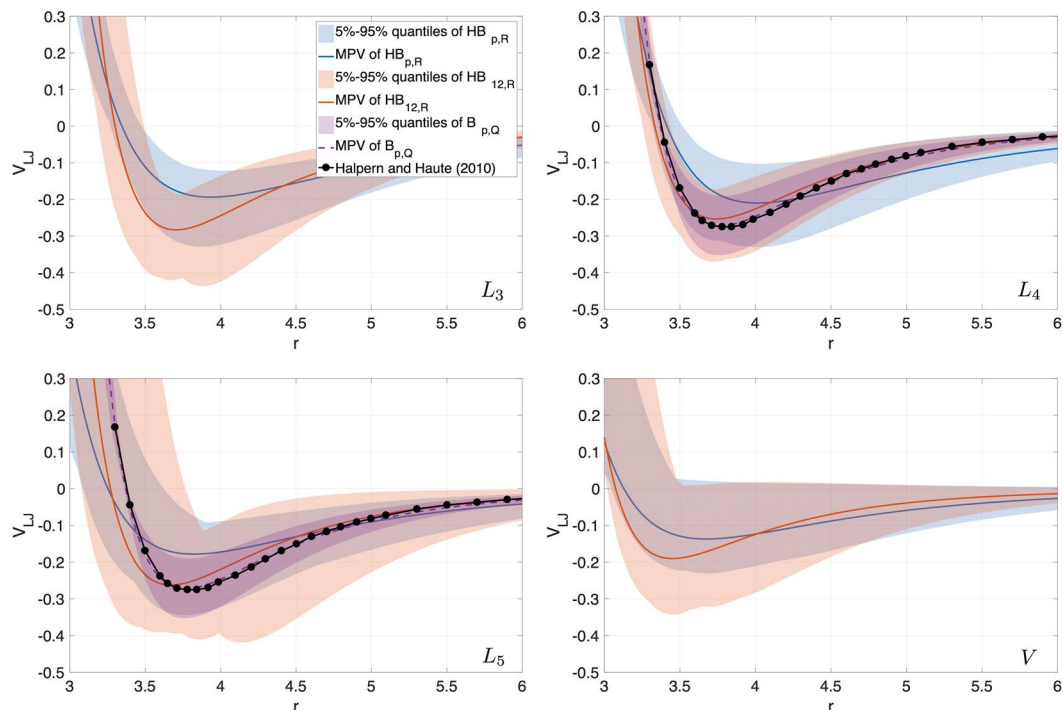


Figure 4. Posterior LJ potentials comparison: MPVs along with 5–95% quantiles obtained in $HB_{p,R}$ (blue), $HB_{12,R}$ (red), $B_{p,Q}$ (purple). Black line with dots: quantum dimer calculations.

	Δ RDF			$\Delta\rho$			ΔD		
	S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
$B_{p,Q}$	0.087	—	—	0.118	—	—	0.136	—	—
$HB_{12,R}$	0.071	0.050	—	0.029	0.108	—	0.009	1.573	—
$HB_{p,R}$	0.016	0.024	0.027	0.011	0.068	0.049	0.043	1.230	0.898

Table 6. Errors of robust posterior predictions of RDF, density and diffusion coefficient using LJ 6–12 and LJ 6- p . We denote $S_1 = \{L_4, L_5\}$ (all inferences produce the correct argon phase), $S_2 = \{L_3 - L_5, V\}$ ($B_{p,Q}$ produces wrong phase), $S_3 = \{L_1 - L_5, V\}$ ($B_{p,Q}$ and $HB_{12,R}$ produce wrong phase).

be applied to explain the inability of LJ 6-12 to drive L_1 and L_2 to the liquid phase: the potential is too repulsive, frustrating the liquid packing, and the system behaves either like a gas or like a solid (note that L_1 is close to the argon triple point). The full set of the inferred LJ 6-12 and LJ 6- p potentials is given in Fig. 4.

Robust posterior prediction. The quality of the predictions, made for a quantity of interest (QoI) different than the one used for the inference, quantifies the predictive power of the model (see Supplementary Material S1.2). We obtain robust predictions of the RDF, density ρ and diffusion coefficient D of argon by propagating the posterior LJ parameters uncertainty into these quantities.

We measure the error Δg of the prediction of the scalar quantity g as

$$\Delta g = \frac{1}{N} \sum_{k=1}^N \left(\frac{g_k - r_k}{r_k} \right)^2,$$

where $N \leq 6$ is the number of thermodynamic conditions for which the prediction can be made, g_k is the prediction made using the MPV and r_k is the reference value. The reference values for ρ are experimental measurements taken from ref.¹². The reference values for D are computed analytically using the equations from ref.¹⁴. The accuracy of the fit for these computations is 0.7%. The error of the RDF is computed as an average over all the thermodynamic conditions of the mean squared error between the computed and the experimental RDF. The predictions are compared on three different sets of conditions: (1) the conditions which can be simulated using MPVs obtained in all the three inferences $HB_{12,R}$, $HB_{p,R}$, and $B_{p,Q}$ (L_4, L_5), (2) the conditions which can be simulated using MPVs obtained in the inferences $HB_{12,R}$ and $HB_{p,R}$ ($L_3 - L_5, V$), (3) the conditions which can be simulated using MPVs obtained in the inference $HB_{p,R}$ ($L_1 - L_5, V$).

The predictions made using the results of $HB_{p,R}$ are the most accurate for all the QoIs considered and for all but one case where $HB_{12,R}$ gives a better result (see Table 6). On the other hand, the predictions made using the results of $B_{p,Q}$ are the least accurate for all the QoIs. Additionally, the inferences $HB_{12,R}$ and $B_{p,Q}$ result in LJ potentials which cannot be used to simulate all the thermodynamic conditions. We conclude that the $HB_{p,R}$ produces a better LJ model than $HB_{12,R}$, and that $B_{p,Q}$ does not result in a good model for liquid argon or saturated argon vapor.

We note that the values of D differ by an order of magnitude for liquid and vapor which explains the huge deterioration of the predictions on the sets of conditions that include V .

The MPVs of D and ρ along with the corresponding quantiles are presented in Fig. 2. The same values for RDF are given in Fig. 2.

Discussion

We have performed a systematic study of the modified 6- p Lennard-Jones potential for liquid and gaseous argon using Hierarchical Bayesian inference with data from experiments and quantum mechanics simulations. Our results show that the value $p = 12$ of the repulsive exponent needs to be calibrated together with the other LJ parameters. In the case of liquid argon we obtain much better predictions with a smaller value $p \approx 6.5$, while for the gaseous argon the classical $p = 12$ or a slightly bigger $p \approx 12.7$ results in a better agreement with the data.

Our results contradict the conclusion of ref.¹¹, where LJ potentials with p ranging from 10 to 20 with a step 2 were fit to viscosity and pressure data, and the potential with $p = 12$ showed better predictions for different thermodynamic conditions. Even though the conclusion of ref.¹¹ formally contradicts our findings, we still consider it a good match, for the following three reasons. Firstly, in ref.¹¹, the average errors of pressure and viscosity predictions for argon strictly increase with the increase of the exponent from 12 to 20, which is consistent with our results (in the sense that exponents smaller than 14 are preferred by all our inferences). Secondly, for the exponents $p = 10$ and $p = 12$ the authors of ref.¹¹ report quite close errors of approximately 2% and 4%, respectively. Finally, different data was used in the calibration process, which may, at least partially, account for the observed difference.

Taking into account the differences in ε and σ (Fig. 1), as well as in p , we conclude that one cannot use the same values of the LJ parameters for liquid, saturated vapor and gas. This issue is attributed to the simplicity of the LJ description: the potential, and thus the underlying physical model of the intermolecular interactions is not flexible or accurate enough to allow for the transferability of the same parameters between different physical environments. The model can be potentially improved by including more complex physics in it, such as many-body interactions with geometrical considerations or quantum corrections. For argon however, the quantum corrections are not expected to be significant¹⁵. At the same time, including the triple-dipole interactions was found to provide better results than the LJ 6-12 for all thermodynamic states of argon². We emphasize however that ab-initio derived forcefields including multi-body terms have also suffered from non perfect transferability of their parameters between gaseous and condensed phase environments. AMOEBA^{16,17} has been an example of a hybrid force-field that includes additional physics, and it was also calibrated using both quantum and experimental data. In the limit of a perfect model, perfect transferability should be anticipated between all physical environments. However, since all force field models are approximations to reality and QM interaction energies even at the Coupled Cluster Stochastic Dynamics (CCSD(T)) level are uncertain, we expect the same lack of transferability for their optimal parameters between different physical environments.

Smaller values of the repulsion exponent ($p \in (6, 9)$) in the Lennard-Jones potential provide better predictions for RDF, density and diffusion data (Fig. 2) than the conventional $p = 12$. Additionally, these new LJ exponents allow to simulate a larger variety of thermodynamic conditions, as compared to the classical 12.

We have also examined whether the smaller exponent allows for bigger time steps in MD simulations. However, it appears that the exponent is not a critical factor for the stability of the system. That is, the energy conservation and the RDF were unaffected by the increase of the time step from 1 fs to approximately 16 fs for both LJ 6-12 and LJ 6-6.5. Further increase of the time step lead to the crash of both MD simulations. We observed similar execution times for the simulations with MPVs of LJ 6-12 and LJ 6- p .

From the computational point of view, usage of low computational cost interpolation-based models replacing the real simulations (called surrogates) resulted in a speed-up of 28%.

Methods

Molecular Dynamics. We perform MD simulations of argon using LAMMPS package¹⁸. The argon atoms are modeled as spheres which interact with LJ 6- p potential:

$$V_{LJ}(r; \varepsilon, \sigma, p) = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^p - \left(\frac{\sigma}{r} \right)^6 \right],$$

where r is the distance between the interacting atoms and p is the repulsion exponent usually taken to be 12. The parameters ε , σ and p are to be chosen according to the available measurements. As the Lennard-Jones interactions quickly decay with the distance, an additional computational cut-off parameter r_c is usually introduced, after which the values of the potential are defined in a different fashion. As was shown in ref.⁷, the values of the cut-off are connected with the values of ε and σ . However, in our work we would like to focus on the effect of the repulsion exponent, and thus leave the value of the cut-off unchanged. We assign $r_c = 3\sigma$ and set the LJ values for $r > r_c$ to 0. The thermodynamic state of the system is defined by the temperature and the pressure of the argon atoms. We ensure that argon is in the liquid/vapor state by checking the self-diffusion coefficient and the density.

The simulation starts with energy minimization followed by 5×10^6 steps, of 2 fs, in an NPT ensemble. Then the RDF is computed in the production run consisting of 10^5 NVE integration steps of 2 fs each. The boundary conditions are periodic in each direction, the domain contains 666 argon atoms. The self-diffusion coefficient is calculated via the mean-squared displacement of the atoms, the RDF is discretized using 100 bins. The units used in the current work are given in Table 2.

Bayesian Uncertainty Quantification. This section presents a brief description of the Bayesian inference theory. The details are given in Supplementary Material S1. Here and further in the text, small bold letters represent vectors while big bold letters represent matrices. Each random variable ξ is assumed to be continuous with a probability density function (PDF) denoted as ξ .

Let $f(\mathbf{x}; \boldsymbol{\vartheta}) \in \mathbb{R}^M$ denote the output, or a QoI, of a computational model with input $\mathbf{x} \in \mathbb{R}^{N_x}$ and parameters $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_{N_\vartheta}) \in \mathbb{R}^{N_\vartheta}$. Let also $\mathbf{d} \in \mathbb{R}^{N_d}$ be a vector of experimental data corresponding to the QoI f and input parameters \mathbf{x} . The experimental data are linked with the computational model through the likelihood function, $\mathbf{d}[\boldsymbol{\vartheta}, \mathbf{x}]$. A usual model assumption for the likelihood function involves a Gaussian,

$$p(\mathbf{d}|\boldsymbol{\vartheta}, \mathbf{x}) = \mathcal{N}(\mathbf{d}|f(\mathbf{x}; \boldsymbol{\vartheta}), \boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma}$ is a covariance matrix that may be a function of $\boldsymbol{\vartheta}$. To simplify the notations, the conditioning on \mathbf{x} is omitted below. Prior information on the parameters $\boldsymbol{\vartheta}$ is encoded into the probability distribution with PDF $\boldsymbol{\vartheta}|\mathcal{M}$. We assume $\boldsymbol{\Sigma} = \sigma_n^2 \mathbf{I}$, where \mathbf{I} is the identity matrix in $\mathbb{R}^{N_\vartheta \times N_\vartheta}$ and $\sigma_n \in \mathbb{R}$ is *a priori* unknown. In this work, we infer the parameters of the LJ potential together with the parameter of the covariance matrix: $\boldsymbol{\vartheta} = (\varepsilon, \sigma, \sigma_n)$ or $\boldsymbol{\vartheta} = (\varepsilon, \sigma, p, \sigma_n)$ depending on whether the exponent p is being inferred or not.

Bayes' theorem provides a tool for the inference of the parameters $\boldsymbol{\vartheta}$ conditioned on the observations \mathbf{d} ,

$$p(\boldsymbol{\vartheta}|\mathbf{d}, \mathcal{M}) = \frac{p(\mathbf{d}|\boldsymbol{\vartheta}, \mathcal{M})p(\boldsymbol{\vartheta}|\mathcal{M})}{p(\mathbf{d}|\mathcal{M})},$$

where $p(\mathbf{d}|\mathcal{M}) = \int p(\mathbf{d}|\boldsymbol{\vartheta}, \mathcal{M})p(\boldsymbol{\vartheta}|\mathcal{M})d\boldsymbol{\vartheta}$ is a normalization constant and \mathcal{M} stands for “model”, which is a set of the assumptions regarding the likelihood and the prior. We remark that the denominator $p(\mathbf{d}|\mathcal{M})$, called model evidence, is used for model selection (see Supplementary Material S1.3).

In certain cases the data may correspond to different input variables \mathbf{x} of the model, one of the examples is pressure and temperature used in this work. Let $\mathbf{d} = \{\mathbf{d}_1, \dots, \mathbf{d}_N\}$ be the set of all provided data with $\mathbf{d}_i \in \mathbb{R}^{N_{\vartheta_i}}$, where each \mathbf{d}_i corresponds to different input \mathbf{x}_i . In this case one wishes to infer different parameters, $\boldsymbol{\vartheta}_i \in \mathbb{R}^{N_{\vartheta}}$, for each dataset \mathbf{d}_i . Here, we assume that the parameters $\boldsymbol{\vartheta}_i$ depend on hyper-parameters $\boldsymbol{\psi} \in \mathbb{R}^{N_\psi}$, which encode the variability of $\boldsymbol{\vartheta}_i$ between the datasets and should also be inferred.

For the sampling of the distributions we use the Transitional Markov Chain Monte Carlo (TMCMC) algorithm¹⁹ (see Supplementary Material S1.1). We perform all the inferences using the open-source library Π 4U⁸ on Brutus cluster of the ETH Zurich and Piz Daint cluster of the Swiss National Supercomputing Center (CSCS). We use 2000 samples per TMCMC stage for LJ 6-12 and 4000 samples per stage for LJ 6- p . The parallelisation is made with MPI and internal worker threads of the Π 4U library. The task-based parallelism and the load balancing mechanisms of Π 4U provide the necessary flexibility for running MD simulations with very different execution time within TMCMC.

In order to reduce the computational cost of the simulations, we apply kriging surrogates following the methodology proposed in ref.²⁰. Namely, for each Markov chain leader we build a kriging interpolating surface using the samples from the leader's bounding box. We select the size of the box to be equal to a quarter of the current domain. The surrogate value is rejected if the kriging error is greater than 5% of the predicted value. In addition, we do not allow the kriging predictions which are outside the 5–95% quantile range of all the values obtained from MD simulations.

Availability of materials and data. We use an open-source framework Π 4U available at <http://www.cse-lab.ethz.ch/software/Pi4U>. The data we used come from refs^{12,13}.

References

- Jones, J. E. On the determination of molecular fields. ii. from the equation of state of a gas. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **106**, 463–477, <https://doi.org/10.1098/rspa.1924.0082> (1924).
- Barker, J. A., Fisher, R. A. & Watts, R. O. Liquid argon: Monte Carlo and molecular dynamics calculations. *Mol. Phys.* **21**, 657–673 (1971).
- Rahman, A. Correlations in the motion of atoms in liquid argon. *Phys. Rev.* **136**, A405–A411 (1964).
- Rowley, L. A., Nicholson, D. & Parsonage, N. G. Monte Carlo grand canonical ensemble calculation in a gas-liquid transition region for 12-6 argon. *J. Comp. Phys.* **17**, 401–414 (1975).
- White, J. A. Lennard-Jones as a model for argon and test of extended renormalization group calculations. *J. Chem. Phys.* **111**, 9352–9356 (1999).
- Cailliez, F. & Pernot, P. Statistical approaches to forcefield calibration and prediction uncertainty in molecular simulation. *J. Chem. Phys.* **134**, 054124 (2011).
- Angelikopoulos, P., Papadimitriou, C. & Koumoutsakos, P. Bayesian uncertainty quantification and propagation in molecular dynamics simulations: A high performance computing framework. *J. Chem. Phys.* **137**, 144103 (2012).
- Hadjidoukas, P. E., Angelikopoulos, P., Papadimitriou, C. & Koumoutsakos, P. Pi4U: A high performance computing framework for Bayesian uncertainty quantification of complex models. *J. Comp. Phys.* **284**, 1–21 (2015).
- Wu, S., Angelikopoulos, P., Papadimitriou, C., Moser, R. & Koumoutsakos, P. A hierarchical Bayesian framework for force field selection in molecular dynamics simulations. *Phil. Trans. R. Soc. A* **374**, 20150032 (2015).
- Wu, S., Angelikopoulos, P., Tauriello, G., Papadimitriou, C. & Koumoutsakos, P. Fusing heterogeneous data for the calibration of molecular dynamics force fields using hierarchical Bayesian models. *J. Chem. Phys.* **145**, 244112 (2016).

11. Galliéro, G., Boned, C., Baylaucq, A. & Montel, F. Molecular dynamics comparative study of Lennard-Jones a-6 and exponential α -6 potentials: Application to real simple fluids (viscosity and pressure). *Phys. Rev. E* **73**, 061201–1 (2006).
12. Eisenstein, A. & Gingrich, N. S. The diffraction of X-rays by argon in the liquid, vapor, and critical regions. *Phys. Rev.* **62**, 261–270 (1942).
13. Halpern, A. M. & Haute, T. Structural and thermodynamic properties of the argon dimer a computational chemistry exercise in quantum and statistical mechanics. *J. Chem. Educ.* **87** (2010).
14. Kestin, J. *et al.* Equilibrium and transport properties of the noble gases and their mixtures at low density. *J. Phys. Chem. Ref. Data* **13**, 229–303 (1984).
15. Ashcroft, N. & Mermin, N. *Solid State Physics* (Saunders College Publishing, Fort Worth, 1976).
16. Ren, P. & Ponder, J. W. Polarizable atomic multipole water model for molecular mechanics simulation. *The Journal of Physical Chemistry B* **107**, 5933–5947 (2003).
17. Shi, Y. *et al.* Polarizable atomic multipole-based amoeba force field for proteins. *Journal of chemical theory and computation* **9**, 4046–4063 (2013).
18. <http://lammps.sandia.gov>.
19. Ching, J. & Chen, Y. Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging. *J. Eng. Mech.* **133**, 816–832 (2007).
20. Angelikopoulos, P., Papadimitriou, C. & Koumoutsakos, P. X-TMCMC: Adaptive kriging for Bayesian inverse modeling. *Comp. Meth. Appl. Mech. Eng.* **289**, 409–428 (2015).

Acknowledgements

We would like to acknowledge helpful discussions with Dr. S. Litvinov, Dr. J. Zavadlav and Dr. E. Cruz-Chu. We would like to acknowledge the computational time at Swiss National Supercomputing Center (CSCS) under the project s659. We gratefully acknowledge support from the European Research Council (ERC) Advanced Investigator Award (No. 34117).

Author Contributions

L.K. ran the simulations, prepared the figures and tables, wrote the Results and Molecular Dynamics sections of the manuscript. G.A. prepared the single process HB code, the Supporting Information and the Bayesian Uncertainty Quantification text. P.A. prepared the LAMMPS script and guided the MD part of the research. P.K. wrote the Abstract, the Introduction and the Discussion sections. P.C. wrote the high performance computing implementation of the HB code and assisted in running the simulations. C.P. guided the Bayesian part of the research. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-017-16314-4>.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017