From Clusters to Bulk Water: Coverage and Validation of LoRA-Adapted Foundation Potentials

Rohith Srinivaas Mohanakrishnan[1,2,†] **Joseph Park**[2,†] **Kristin A. Persson**[1,2,*]

[1]Department of Materials Science and Engineering,
University of California, Berkeley, Berkeley, California 94720, United States

[2]Materials Sciences Division, Lawrence Berkeley National Laboratory,
Berkeley, California 94720, United States

[†]These authors contributed equally to this work.

[*]Correspondence: kapersson@lbl.gov

Abstract

Parameter-efficient adaptation offers a route to specializing pretrained machine-learned interatomic potentials, but molecular accuracy alone does not establish reliable bulk behavior. We adapt MACE-OFF23(M) and MACELES-OFF using rank-32 LoRA with monomer–dimer (G1-G2) or monomer–hexamer (G1-G6) water-cluster coverage, and evaluate molecular validation errors, liquid density, radial distribution function, diffusion, and viscosity. Relative to monomer–dimer adaptation, broader coverage reduces the equally weighted G3–G6 force RMSE by 63–73%. Across four temperatures, mean absolute relative density errors decrease from the unadapted baselines of 18.98% and 14.93% to 8.96% and 4.63% for the broadly adapted OFF and LES models, respectively, although both liquids are underdense and overestimate the first O–O radial-distribution peak. Corrected bulk simulations use three training seeds, with three molecular-dynamics replicas per seed and temperature for each adapted model, and a common production of 2 ns NVT duration per replica. Improved molecular forces do not imply uniform transport agreement: at 275 K, the broadly adapted LES model has a mean viscosity approximately 60% above experiment. Stress-correlation diagnostics identify residual integration-window sensitivity. These results support separate validation of molecular accuracy, equilibrium properties, and transport for the tested water foundations.

1 Introduction

Machine-learned interatomic potentials (MLIPs) provide the energies and forces needed for molecular dynamics at scales impractical for on-the-fly first-principles calculations [1]. They are also increasingly used alongside reasoning models to generate structures, run simulations, and analyze trajectories [2, 3]. Pretrained foundation potentials extend these capabilities across broad chemical spaces [4, 5], but their accuracy may be insufficient for a particular system or reference method. Fine-tuning can reuse pretrained representations, although updating the full network may degrade pretrained behavior under a domain shift [5, 6]. Parameter-efficient adaptation instead freezes most parameters and optimizes a smaller task-specific subset. For example, ELoRA introduces low-rank modifications that preserve rotational equivariance in atomistic networks [7]. However, accurate molecular energies and forces alone do not establish reliable condensed-phase dynamics [8, 9].

40th Conference on Neural Information Processing Systems (NeurIPS 2026). Workshop: ML4Molecules: Agentic Systems for Molecular Sciences.

The coverage of adaptation data may influence this transfer. Monomers sample intramolecular deformation, dimers add pairwise environments, and larger clusters expose configurations containing multiple interacting neighbors. Such environments are relevant to water, whose properties depend on hydrogen bonding, polarization, dispersion, and nonadditive interactions [10, 11]. Water also provides complementary equilibrium and dynamical observables that can respond differently to changes in a potential [12, 13]. Evaluating these observables together tests whether molecular improvements extend to properties sampled along a bulk trajectory.

We examine rank-32 LoRA adaptation of MACE-OFF23(M), a short-range equivariant organic-molecule potential [4, 14], and MACELES-OFF, which includes Latent Ewald Summation (LES) contributions to the long-range energy [15, 16]. For each foundation, we compare the unadapted model with monomer–dimer (G1–G2) and monomer–hexamer (G1–G6) adaptation. We evaluate molecular validation errors, density, O–O radial distribution functions (RDFs), self-diffusion, and shear viscosity.

2 Methods

2.1 Dataset generation

Geometries were derived from 1,000 frames of the released 10,000-frame periodic liquid-water dataset of Chen et al. [17]. Parent frames were split into 800 training, 100 validation, and 100 test configurations, with all clusters from a parent retained in the same split. For each frame, one central water molecule was selected using random seed 0, and neighbors were ordered by minimum-image O–O distance. The central molecule and its nearest one through five neighbors define nested G2–G6 clusters. The six constituent molecules of each G6 cluster provide G1 geometries. G1–G2 and G1–G6 contain 5,600 and 8,800 training records, respectively. The raw G2–G6 total energies and forces are documented at the ωB97M-V/def2-TZVPD level [18–20], using Q-Chem [21]. The six constituent monomers of each G6 cluster additionally provide the corresponding G1 (monomer) geometries.

2.2 Parameter-efficient adaptation and molecular evaluation

Foundation weights were frozen, and low-rank updates with rank $r = 32$ and scaling $\alpha = 32$ were optimized following LoRA [22]. Updates preserve the irreducible-representation structure of equivariant linear maps and also target scalar radial-weight networks. Adapted components include the node embedding, interaction and product blocks, and energy readouts; LES additionally includes latent readouts. The audited target list and parameter counts are given in Appendix A. The force-component MSE loss coefficient was 100 times the per-atom energy MSE coefficient.

Training used seeds 17, 29, and 43. Molecular evaluation and bulk simulations used the exponential-moving-average checkpoint at 123,200 optimizer updates for each seed. Training and MD settings are summarized in Appendix Table A3. Adapters were fused before deployment, with energy, force, and stress parity checks between fused and unfused models. Molecular metrics are per-atom energy RMSE and Cartesian force-component RMSE. Cross-size summaries first average groupwise RMSEs equally within each seed and then report the mean and sample standard deviation across seeds.

2.3 Bulk-water simulations and observables

Each periodic system contained 64 water molecules. We used temperatures of 275, 295, 315, and 336 K, pressure 1.01325 bar, and a 1 fs NPT step, following the system and state-point choices of the MACELES-OFF density-isobar experiment [16]. NPT production lasted 1 ns [23, 24], and density and RDFs were computed over 500–1,000 ps. Each adapted model had three independently thermalized MD replicas per training seed and temperature; each baseline had three replicas per temperature. This gives 144 adapted and 24 baseline trajectories.

The NPT frame closest to each trajectory's mean analysis-window volume initialized fixed-volume Nosé–Hoover dynamics without changing its positions or momenta [25, 26]. Every NVT trajectory lasted 2 ns with a 0.5 fs step. Oxygen velocities and total stress, including kinetic and virial contributions, were recorded at every step.

Before evaluating the oxygen velocity autocorrelation function (VACF), we subtract the instantaneous mean oxygen velocity and then the time average of each oxygen Cartesian velocity component. The

velocity in the following expression denotes this processed velocity. For trajectory i, we evaluated

$$D_{L,i}(t) = \frac{1}{3}\int_0^t \langle \mathbf{v}_O(0)\cdot\mathbf{v}_O(t')\rangle_i \, dt', \tag{1}$$

$$\eta_i(t) = \frac{V_i}{k_\mathrm{B}T_i}\int_0^t C_{\sigma,i}(t')\, dt', \tag{2}$$

where T_i is the mean sampled temperature and $C_{\sigma,i}$ averages the autocorrelations of five demeaned shear/traceless stress channels [13, 27, 28]. Channel definitions are given in Appendix B. The primary scalar estimates, denoted $D_{L,i}$ and η_i, are arithmetic means of the sampled running integrals over 5–10 ps. Stress analysis extends to 50 ps to assess sensitivity to the integration window.

Figure 2 reports the trajectory-wise Yeh–Hummer diffusion correction for a cubic cell [29],

$$D_{\mathrm{YH},i} = D_{L,i} + \frac{2.837297\, k_\mathrm{B}T_i}{6\pi\eta_i L_i}, \qquad L_i = V_i^{1/3}. \tag{3}$$

The correction uses each trajectory's own viscosity and is applied before aggregation, retaining its covariance with raw diffusion. Density and viscosity references use the IAPWS liquid-water correlations at 0.1 MPa [30, 31], and diffusion uses the temperature-dependent fit of Holz et al. [32], evaluated at the nominal temperatures. The 295 K RDF is compared with the experiment-derived 298 K Soper curve [33].

2.4 Software and reproducibility

The MD campaign used the archived MACE/LES source snapshot declaring MACE 0.3.17, with PyTorch 2.7.1+cu128 and ASE 3.29.0, using float64 calculations [14, 24, 34]. Code and compact data are organized in the accompanying repository: `https://github.com/josephjoeypark/mlip-water-workshop-submission`. The materials include cluster geometries and labels, protocol and environment files, training and adapter-fusion code, MD and observable-analysis scripts, model manifests, and the per-seed and per-replica results needed to regenerate the figures. Full trajectories and checkpoint binaries are not bundled; figure reproduction uses the included compact data without them.

3 Results and Discussion

3.1 Adaptation capacity and molecular transfer

Within each foundation, the two coverage protocols use the same adapter rank and trainable parameter count (Appendix Table A1). Broad coverage substantially reduces validation force errors (Appendix Figure A1). The G2–G6 macro-average decreases from 12.066 ± 0.133 to 4.687 ± 0.009 meV Å^{-1} for OFF and from 15.202 ± 0.178 to 4.359 ± 0.045 meV Å^{-1} for LES. For G3–G6, which lie outside the explicit cluster-size coverage of G1–G2 training, the corresponding changes are 14.358 ± 0.169 to 5.267 ± 0.012 and 18.444 ± 0.221 to 4.979 ± 0.048 meV Å^{-1}: reductions of 63.3% and 73.0%. G2 force accuracy improves for both foundations, while G1 force errors remain similar with small increases.

Energy improvements are less uniform. Broad coverage lowers the G3–G6 energy macro-average from 0.927 ± 0.053 to 0.438 ± 0.026 meV atom^{-1} for OFF and from 1.323 ± 0.010 to 1.012 ± 0.040 meV atom^{-1} for LES, while increasing the G1–G2 energy macro-average for both foundations. LES G6 energy RMSE increases from 1.514 ± 0.019 to 1.621 ± 0.058 meV atom^{-1} despite a substantial force improvement. Broader coverage therefore redistributes accuracy across observables and cluster sizes. These changes characterize the complete adaptation protocols, whose dataset sizes and training implementations also differ.

3.2 Radial distribution function & Density

G1–G6 adaptation recovers a clearer first minimum and second-shell oscillation in the O–O RDF (Figure 1, upper row). Both broadly adapted models nevertheless overestimate the first peak. The improved shell organization therefore coexists with a residual local-structure bias. The mean absolute

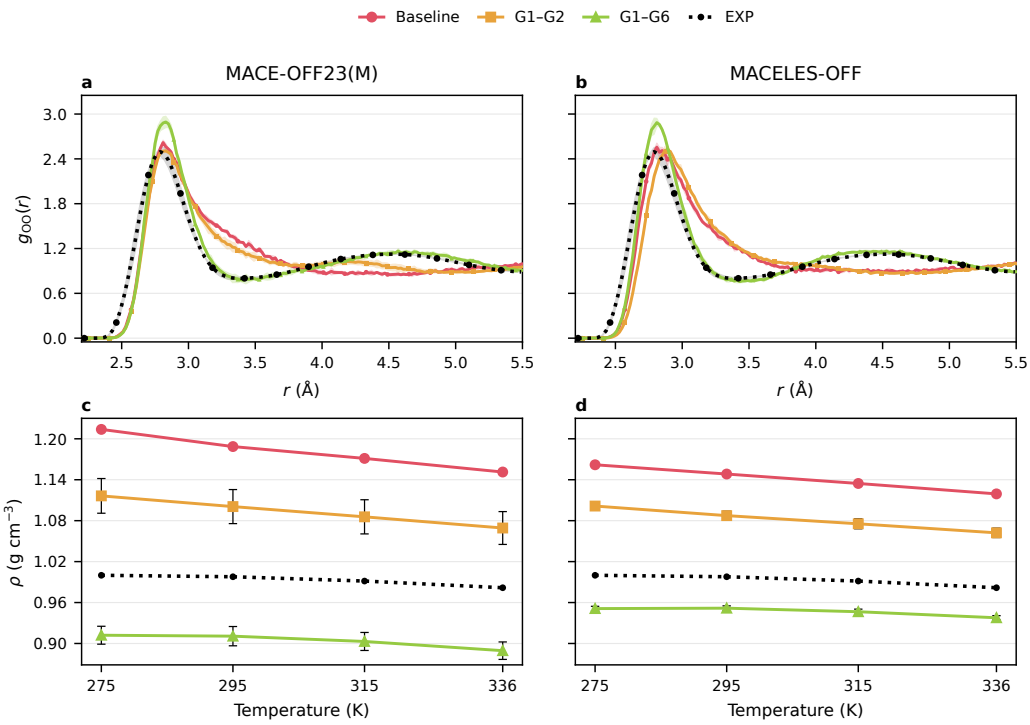


Figure 1: Equilibrium structure and density. Upper panels show 295 K O–O RDFs compared with the experiment-derived 298 K Soper reference [33]; lower panels show density isobars at 1.01325 bar compared with the IAPWS reference at 0.1 MPa [30]. OFF is left and LES is right. NPT analysis uses 500–1,000 ps. For adapted models, curves and markers average three training-seed means, each formed from three MD replicas; bands and error bars show the sample SD across those means. Baseline uncertainties are sample SDs across three MD replicas.

relative density errors over the four temperatures are 18.98%, 10.09%, and 8.96% for the OFF baseline, G1–G2, and G1–G6 models, respectively; the LES sequence is 14.93%, 8.95%, and 4.63%. Broad coverage reduces absolute density error while reversing the density bias: both G1–G6 liquids are underdense (Figure 1, lower row). At 295 K, OFF–G1–G6 and LES–G1–G6 have densities of 0.9107 ± 0.0141 and 0.9517 ± 0.0034 g cm^{-3}, respectively, compared with the reference value of 0.9978 g cm^{-3}.

3.3 Diffusion coefficient & Viscosity

Transport agreement depends on temperature and observable (Figure 2). At 295 K, OFF–G1–G6 gives $D_{\mathrm{YH}} = (2.262 \pm 0.259) \times 10^{-9}$ m^2 s^{-1} and $\eta = 0.977 \pm 0.145$ mPa s; LES–G1–G6 gives $D_{\mathrm{YH}} = (1.866 \pm 0.163) \times 10^{-9}$ m^2 s^{-1} and $\eta = 1.180 \pm 0.094$ mPa s. The corresponding reference values are 2.123×10^{-9} m^2 s^{-1} and 0.958 mPa s.

The broader-coverage protocol does not uniformly improve experimental agreement. At 275 K, LES–G1–G6 has $\eta = 2.694 \pm 0.473$ mPa s against a reference of 1.682 mPa s, a mean bias of $+60.2\%$; the narrow-coverage bias is -53.8%. At 336 K, OFF's corrected-diffusion bias increases from $+8.0\%$ under narrow coverage to $+23.8\%$ under broad coverage, while its viscosity bias changes from -5.1% to -18.5%. Moreover, the near-reference 295 K viscosity of OFF–G1–G6 coexists with substantial underdensity. The effect of adaptation coverage on bulk dynamics is therefore strongly dependent on the pretrained foundation model. Conversely, an isolated observable that falls near experiment should not be treated as general validation: for example, OFF–G1–G2 has a near-experimental 295 K viscosity while retaining a strongly overdense liquid. This is consistent with compensating errors rather than a uniformly accurate potential.

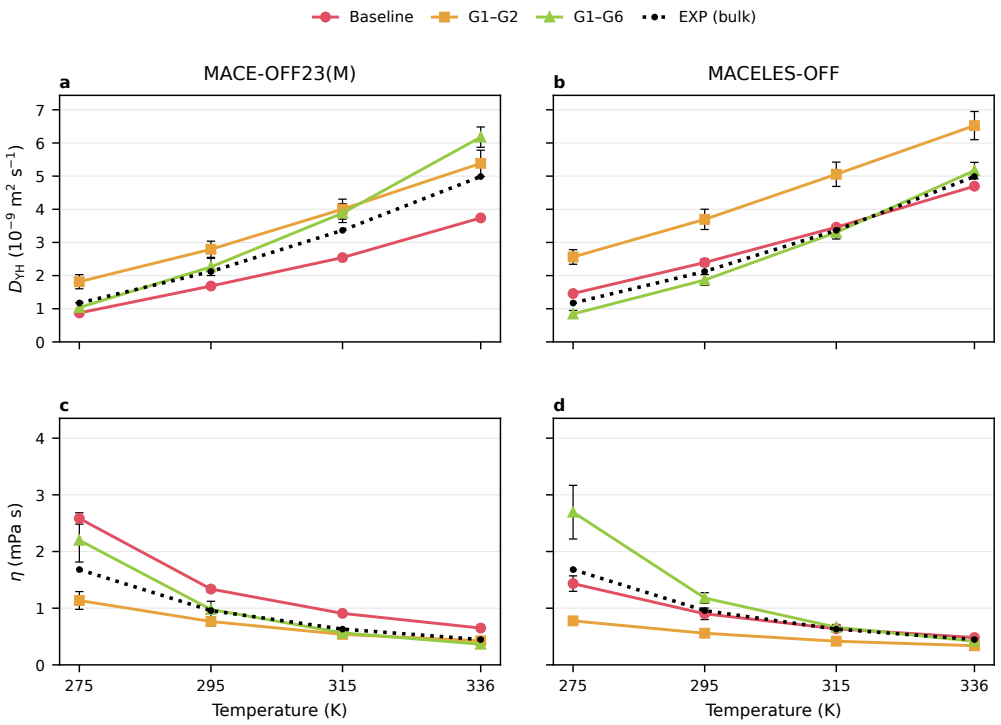


Figure 2: Transport from the corrected 2 ns NVT trajectories. Upper panels show oxygen VACF self-diffusion after the trajectory-wise Yeh–Hummer correction, D_{YH}; lower panels show Green–Kubo viscosity. OFF is left and LES is right. Both raw transport estimators use the arithmetic mean of their sampled running integrals over 5–10 ps. Black curves show the diffusion fit of Holz et al. [32] and the IAPWS viscosity reference at 0.1 MPa [30, 31], evaluated at nominal temperatures. Adapted error bars are sample SDs across three training-seed means after averaging three MD replicas per seed; baseline bars are MD-replica sample SDs. These are not confidence intervals. Raw and corrected 295 K diffusion values are listed in Appendix Table A4.

4 Conclusions

We tested whether parameter-efficient adaptation of equivariant foundation potentials preserves accuracy from molecular clusters to bulk water observables, across two pretrained foundations and two adaptation protocols differing in cluster coverage. Broadening coverage from dimers to hexamers substantially improved held-out multi-neighbor forces, corrected overdense zero-shot liquids, and recovered radial-distribution-function structure in both foundations, but these gains were uneven and did not guarantee dynamical accuracy: the two broadly adapted models, despite similar force accuracy and structural improvement, do not recover the experimental trend in dynamical properties like viscosity and diffusion coefficients. Hence, molecular and structural benchmarks alone should not be treated as sufficient validation for an adapted MLIP intended for condensed-phase dynamics.

References

[1] F. L. Thiemann, N. O'Neill, V. Kapil, A. Michaelides, and C. Schran. Introduction to machine learning potentials for atomistic simulations. *Journal of Physics: Condensed Matter*, 37:073002, 2025. doi: 10.1088/1361-648X/ad9657. URL `https://doi.org/10.1088/1361-648X/ad9657`.

[2] Bowen Deng, Bohan Li, Matthew Cox, Hoje Chun, Juno Nam, Artur Lyssenko, Sathya Edamadaka, Jurgis Ruza, Xiaochen Du, Nofit Segal, Jesus Diaz Sanchez, Mingrou Xie, Ty Perez, Yu Yao, Miguel Steiner, Sauradeep Majumdar, III Musgrave, Charles B., Anirban Chandra, Abhirup Patra, Detlef Hohl, Connor W. Coley, Ju Li, and Rafael Gómez-

Bombarelli. Harnessing AtomisticSkills for agentic atomistic research, 2026. URL https://arxiv.org/abs/2605.24002.

[3] Mingwei Ding, Chen Huang, Yibo Hu, Yifan Li, Zitian Lu, Xingtai Yu, Duo Zhang, Wenxi Zhai, Tong Zhu, Qiangqiang Gu, and Jinzhe Zeng. Automating computational chemistry workflows via OpenClaw and domain-specific skills, 2026. URL https://arxiv.org/abs/2603.25522.

[4] Dávid Péter Kovács et al. MACE-OFF: Short-range transferable machine learning force fields for organic molecules. *Journal of the American Chemical Society*, 147:17598–17611, 2025. doi: 10.1021/jacs.4c07099. URL https://doi.org/10.1021/jacs.4c07099.

[5] Yury Lysogorskiy, Anton Bochkarev, and Ralf Drautz. Graph atomic cluster expansion for foundational machine learning interatomic potentials. *npj Computational Materials*, 12:114, 2026. doi: 10.1038/s41524-026-01979-1. URL https://doi.org/10.1038/s41524-026-01979-1.

[6] Mariia Radova, Wojciech G. Stark, Connor S. Allen, Reinhard J. Maurer, and Albert P. Bartók. Fine-tuning foundation models of materials interatomic potentials with frozen transfer learning. *npj Computational Materials*, 11:237, 2025. doi: 10.1038/s41524-025-01727-x. URL https://doi.org/10.1038/s41524-025-01727-x.

[7] Chen Wang, Siyu Hu, Guangming Tan, and Weile Jia. ELoRA: Low-rank adaptation for equivariant GNNs. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 63113–63135. PMLR, 2025. URL https://proceedings.mlr.press/v267/wang25al.html.

[8] M. J. Gawkowski, M. Li, B. X. Shi, and V. Kapil. The good, the bad, and the ugly of atomistic learning for "clusters-to-bulk" generalization. *Machine Learning: Science and Technology*, 7:025004, 2026. doi: 10.1088/2632-2153/ae3c57. URL https://doi.org/10.1088/2632-2153/ae3c57.

[9] V. Svahn, I.-B. Magdău, S. P. Niblett, G. Csányi, K. Hermansson, and J. Kullgren. Limitations of cluster-trained MLIPs for liquid density and diffusivity. *Journal of Chemical Theory and Computation*, 22:3660–3671, 2026. doi: 10.1021/acs.jctc.5c02043. URL https://doi.org/10.1021/acs.jctc.5c02043.

[10] Marc Riera, Eleftherios Lambros, Thuong T. Nguyen, Andreas W. Götz, and Francesco Paesani. Low-order many-body interactions determine the local structure of liquid water. *Chemical Science*, 10:8211–8218, 2019. doi: 10.1039/C9SC03291F. URL https://doi.org/10.1039/C9SC03291F.

[11] Sandeep K. Reddy et al. On the accuracy of the MB-pol many-body potential for water: Interaction energies, vibrational frequencies, and classical thermodynamic and dynamical properties from clusters to liquid water and ice. *Journal of Chemical Physics*, 145:194504, 2016. doi: 10.1063/1.4967719. URL https://doi.org/10.1063/1.4967719.

[12] Tobias Hilpert and Georg Kresse. Accurate thermophysical properties of water using machine-learned potentials. *Journal of Chemical Physics*, 164:194504, 2026. doi: 10.1063/5.0325444. URL https://doi.org/10.1063/5.0325444.

[13] P. Montero de Hijes, L. Neubeck, G. Kresse, and C. Dellago. Dynamical properties of ab initio water from machine-learning potentials. *Journal of Chemical Physics*, 165:074502, 2026. doi: 10.1063/5.0348453. URL https://doi.org/10.1063/5.0348453.

[14] Ilyes Batatia, Dávid Péter Kovács, Gregor N. C. Simm, Christoph Ortner, and Gábor Csányi. MACE: Higher-order equivariant message passing neural networks for fast and accurate force fields. In *Advances in Neural Information Processing Systems*, volume 35, pages 11423–11436, 2022. URL https://papers.neurips.cc/paper_files/paper/2022/hash/4a36c3c51af11ed9f34615b81edb5bbc-Abstract-Conference.html.

[15] Bingqing Cheng. Latent ewald summation for machine learning of long-range interactions. *npj Computational Materials*, 11:80, 2025. doi: 10.1038/s41524-025-01577-7. URL https://doi.org/10.1038/s41524-025-01577-7.

[16] D. Kim, X. Wang, S. Vargas, P. Zhong, D. S. King, T. Jaffrelot Inizan, and B. Cheng. A universal augmentation framework for long-range electrostatics in machine learning interatomic potentials. *Journal of Chemical Theory and Computation*, 21:12709–12724, 2025. doi: 10.1021/acs.jctc.5c01400. URL https://doi.org/10.1021/acs.jctc.5c01400.

[17] Sijia Chen, Niamh O'Neill, Benjamin X. Shi, and Venkat Kapil. Distilling first-principles accuracy into compact machine learning potentials for condensed-phase chemistry, 2026. URL https://arxiv.org/abs/2606.06848.

[18] Narbe Mardirossian and Martin Head-Gordon. ωB97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *Journal of Chemical Physics*, 144:214110, 2016. doi: 10.1063/1.4952647. URL https://doi.org/10.1063/1.4952647.

[19] Florian Weigend and Reinhart Ahlrichs. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn. *Physical Chemistry Chemical Physics*, 7:3297–3305, 2005. doi: 10.1039/B508541A. URL https://doi.org/10.1039/B508541A.

[20] Dmitrij Rappoport and Filipp Furche. Property-optimized gaussian basis sets for molecular response calculations. *Journal of Chemical Physics*, 133:134105, 2010. doi: 10.1063/1.3484283. URL https://doi.org/10.1063/1.3484283.

[21] Evgeny Epifanovsky, Andrew T. B. Gilbert, Xintian Feng, Joonho Lee, Yuezhi Mao, Narbe Mardirossian, Pavel Pokhilko, Alec F. White, Marc P. Coons, Adrian L. Dempwolff, Zhengting Gan, Diptarka Hait, Paul R. Horn, Leif D. Jacobson, Ilya Kaliman, Jörg Kussmann, Adrian W. Lange, Ka Un Lao, Daniel S. Levine, Jie Liu, Simon C. McKenzie, Adrian F. Morrison, Kaushik D. Nanda, Felix Plasser, Dirk R. Rehn, Marta L. Vidal, Zhi-Qiang You, Ying Zhu, Bushra Alam, Benjamin J. Albrecht, Abdulrahman Aldossary, Ethan Alguire, Josefine H. Andersen, Vishikh Athavale, Dennis Barton, Khadiza Begam, Andrew Behn, Nicole Bellonzi, Yves A. Bernard, Eric J. Berquist, Hugh G. A. Burton, Abel Carreras, Kevin Carter-Fenk, Romit Chakraborty, Alan D. Chien, Kristina D. Closser, Vale Cofer-Shabica, Saswata Dasgupta, Marc de Wergifosse, Jia Deng, Michael Diedenhofen, Hainam Do, Sebastian Ehlert, Po-Tung Fang, Shervin Fatehi, Qingguo Feng, Triet Friedhoff, James Gayvert, Qinghui Ge, Gergely Gidofalvi, Matthew Goldey, Joe Gomes, Cristina E. González-Espinoza, Sahil Gulania, Anastasia O. Gunina, Magnus W. D. Hanson-Heine, Phillip H. P. Harbach, Andreas Hauser, Michael F. Herbst, Mario Hernández Vera, Manuel Hodecker, Zachary C. Holden, Shannon Houck, Xunkun Huang, Kerwin Hui, Bang C. Huynh, Maxim Ivanov, Ádám Jász, Hyunjun Ji, Hanjie Jiang, Benjamin Kaduk, Sven Kähler, Kirill Khistyaev, Jaehoon Kim, Gergely Kis, Phil Klunzinger, Zsuzsanna Koczor-Benda, Joong Hoon Koh, Dimitri Kosenkov, Laura Koulias, Tim Kowalczyk, Caroline M. Krauter, Karl Kue, Alexander Kunitsa, Thomas Kus, István Ladjánszki, Arie Landau, Keith V. Lawler, Daniel Lefrancois, Susi Lehtola, Run R. Li, Yi-Pei Li, Jiashu Liang, Marcus Liebenthal, Hung-Hsuan Lin, You-Sheng Lin, Fenglai Liu, Kuan-Yu Liu, Matthias Loipersberger, Arne Luenser, Aaditya Manjanath, Prashant Manohar, Erum Mansoor, Sam F. Manzer, Shan-Ping Mao, Aleksandr V. Marenich, Thomas Markovich, Stephen Mason, Simon A. Maurer, Peter F. McLaughlin, Maximilian F. S. J. Menger, Jan-Michael Mewes, Stefanie A. Mewes, Pierpaolo Morgante, J. Wayne Mullinax, Katherine J. Oosterbaan, Garrette Paran, Alexander C. Paul, Suranjan K. Paul, Fabijan Pavošević, Zheng Pei, Stefan Prager, Emil I. Proynov, Ádám Rák, Eloy Ramos-Cordoba, Bhaskar Rana, Alan E. Rask, Adam Rettig, Ryan M. Richard, Fazle Rob, Elliot Rossomme, Tarek Scheele, Maximilian Scheurer, Matthias Schneider, Nickolai Sergueev, Shaama M. Sharada, Wojciech Skomorowski, David W. Small, Christopher J. Stein, Yu-Chuan Su, Eric J. Sundstrom, Zhen Tao, Jonathan Thirman, Gábor J. Tornai, Takashi Tsuchimochi, Norm M. Tubman, Srimukh Prasad Veccham, Oleg Vydrov, Jan Wenzel, Jon Witte, Atsushi Yamada, Kun Yao, Sina Yeganeh, Shane R. Yost, Alexander Zech, Igor Ying Zhang, Xing Zhang, Yu Zhang, Dmitry Zuev, Alán Aspuru-Guzik, Alexis T. Bell, Nicholas A. Besley, Ksenia B. Bravaya, Bernard R. Brooks, David Casanova, Jeng-Da Chai, Sonia Coriani, Christopher J. Cramer, György Cserey, III DePrince, A. Eugene, Jr. DiStasio, Robert A., Andreas Dreuw, Barry D. Dunietz, Thomas R. Furlani, III Goddard, William A., Sharon Hammes-Schiffer, Teresa Head-Gordon, Warren J. Hehre, Chao-Ping Hsu, Thomas-C. Jagau, Yousung Jung, Andreas Klamt, Jing Kong, Daniel S. Lambrecht, WanZhen Liang, Nicholas J. Mayhall, C. William McCurdy, Jeffrey B. Neaton, Christian Ochsenfeld, John A. Parkhill, Roberto Peverati, Vitaly A.

Rassolov, Yihan Shao, Lyudmila V. Slipchenko, Tim Stauch, Ryan P. Steele, Joseph E. Subotnik, Alex J. W. Thom, Alexandre Tkatchenko, Donald G. Truhlar, Troy Van Voorhis, Tomasz A. Wesolowski, K. Birgitta Whaley, III Woodcock, H. Lee, Paul M. Zimmerman, Shirin Faraji, Peter M. W. Gill, Martin Head-Gordon, John M. Herbert, and Anna I. Krylov. Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. *J. Chem. Phys.*, 155(8):084801, 2021. doi: 10.1063/5.0055522.

[22] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=nZeVKeeFYf9.

[23] Glenn J. Martyna, Douglas J. Tobias, and Michael L. Klein. Constant pressure molecular dynamics algorithms. *Journal of Chemical Physics*, 101:4177–4189, 1994. doi: 10.1063/1.467468. URL https://doi.org/10.1063/1.467468.

[24] Ask Hjorth Larsen et al. The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter*, 29:273002, 2017. doi: 10.1088/1361-648X/aa680e. URL https://doi.org/10.1088/1361-648X/aa680e.

[25] Shuichi Nosé. A unified formulation of the constant temperature molecular dynamics methods. *Journal of Chemical Physics*, 81:511–519, 1984. doi: 10.1063/1.447334. URL https://doi.org/10.1063/1.447334.

[26] William G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31:1695–1697, 1985. doi: 10.1103/PhysRevA.31.1695. URL https://doi.org/10.1103/PhysRevA.31.1695.

[27] Melville S. Green. Markoff random processes and the statistical mechanics of time-dependent phenomena. II. irreversible processes in fluids. *Journal of Chemical Physics*, 22:398–413, 1954. doi: 10.1063/1.1740082. URL https://doi.org/10.1063/1.1740082.

[28] Ryogo Kubo. Statistical-mechanical theory of irreversible processes. I. general theory and simple applications to magnetic and conduction problems. *Journal of the Physical Society of Japan*, 12:570–586, 1957. doi: 10.1143/JPSJ.12.570. URL https://doi.org/10.1143/JPSJ.12.570.

[29] In-Chul Yeh and Gerhard Hummer. System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions. *Journal of Physical Chemistry B*, 108:15873–15879, 2004. doi: 10.1021/jp0477147. URL https://doi.org/10.1021/jp0477147.

[30] International Association for the Properties of Water and Steam. Revised supplementary release on properties of liquid water at 0.1 mpa. Technical Report SR6-08(2011), International Association for the Properties of Water and Steam, 2011. URL https://iapws.org/technical-guidance/release/LiquidWater.

[31] International Association for the Properties of Water and Steam. Release on the IAPWS formulation 2008 for the viscosity of ordinary water substance. Technical Report R12-08, International Association for the Properties of Water and Steam, Berlin, Germany, 2008. URL https://iapws.org/technical-guidance/release/viscosity.

[32] Manfred Holz, Stefan R. Heil, and Antonio Sacco. Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate ^{1}H NMR PFG measurements. *Physical Chemistry Chemical Physics*, 2(20):4740–4742, 2000. doi: 10.1039/B005319H. URL https://doi.org/10.1039/B005319H.

[33] Alan K. Soper. The radial distribution functions of water as derived from radiation total scattering experiments: Is there anything we can say for sure? *ISRN Physical Chemistry*, 2013: 279463, 2013. doi: 10.1155/2013/279463. URL https://doi.org/10.1155/2013/279463.

[34] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL `https://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library`.

A Adaptation specifications and molecular validation

Table A1 summarizes the foundation and adapter sizes. The MACELES-OFF adapter partition is obtained by summing the released per-module parameter counts: 339,840 parameters belong to the MACE path and 13,376 to the three LES readouts. Full module classes, parameter shapes, and source hashes accompany the released per-module audit.

Table A1: Foundation and adaptation specifications. Both coverage protocols use rank-32 LoRA with $\alpha = 32$ and the same adapter capacity within each foundation. Foundation parameters are frozen. Trainable fractions divide adapter counts by foundation plus adapter counts. Foundation specifications follow Kovács et al. [4], Kim et al. [16]; adapter counts follow the released module audit.

	MACE-OFF23(M)	MACELES-OFF
Local cutoff (Å)	5.0	4.5
Hidden channels	128	192
Foundation parameters	1,428,368	1,891,360
MACE-path adapter parameters	235,392	339,840
LES-readout adapter parameters	0	13,376
Total adapter parameters	235,392	353,216
Adapted modules	18	21
Trainable fraction	14.15%	15.74%

Table A2: LoRA target modules. Braces enumerate literal module indices. Each listed target contains a low-rank update. Equivariant maps use irrep-aware updates, while the radial-weight network layers operate on scalar features.

Module pattern	OFF	LES
`node_embedding.linear`	1	1
`interactions.{0,1}.linear_up`	2	2
`interactions.{0,1}.linear`	2	2
`interactions.{0,1}.conv_tp_weights.layer{0,1,2,3}`	8	8
`products.{0,1}.linear`	2	2
`readouts.0.linear`	1	1
`readouts.1.linear_1`	1	1
`readouts.1.linear_2`	1	1
`les_readouts.0.linear`	0	1
`les_readouts.1.linear_1`	0	1
`les_readouts.1.linear_2`	0	1

G1–G2 contains 5,600/700/700 training/validation/test records; G1–G6 contains 8,800/1,100/1,100. The 800/100/100 parent-frame split is shared. Both adaptation protocols are evaluated on the same G1–G6 validation configurations for the molecular comparison. For each training seed, a reported multi-group force metric is

$$\overline{\mathrm{RMSE}}_F(\mathcal{G}) = \frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} \mathrm{RMSE}_{F,g}, \tag{4}$$

with $\mathcal{G} = \{\mathrm{G3}, \ldots, \mathrm{G6}\}$. Energy summaries use the same averaging rule. All molecular uncertainties are sample SDs over training seeds 17, 29, and 43.

Table A3: Training and MD settings. G1–G2 uses Adam with AMSGrad; G1–G6 uses AdamW with AMSGrad; Both protocols retain a learning rate of 5×10^{-4} at the reported 123,200-update checkpoint. Equal update counts correspond to 44 and 28 training epochs for G1–G2 and G1–G6, respectively. NPT and NVT durations refer to each MD replica.

Setting	Value
Training precision	float64
G1–G2 optimizer	Adam with AMSGrad
G1–G6 optimizer	AdamW with AMSGrad
Training batch size	2 configurations
Validation batch size	2 configurations for G1–G2; 16 for G1–G6
Initial learning rate	5×10^{-4}
G1–G2 weight decay	10^{-8} on product-block adapters; zero on other adapters
G1–G6 weight decay	10^{-8} on all trainable adapters
Gradient-norm clipping	10
EMA decay	0.995
Reported checkpoint	123,200 optimizer updates
G1–G2 trainer	MACE training loop; validation every epoch
G1–G2 scheduler	`ReduceLROnPlateau`; factor 0.8, patience 50 evaluations
G1–G6 learning rate	Fixed at 5×10^{-4}
NPT implementation	Isotropic Martyna–Tobias–Klein
NPT production / analysis	1 ns / final 500 ps
NPT time step	1 fs
NPT thermostat / barostat damping	50 fs / 500 fs
NPT chains and substeps	Five elements and two substeps for each chain
NVT implementation	Nosé–Hoover at the selected fixed volume
NVT production / time step	2 ns / 0.5 fs
NVT thermostat time	10 fs
Velocity and stress cadence	Every NVT step (0.5 fs)

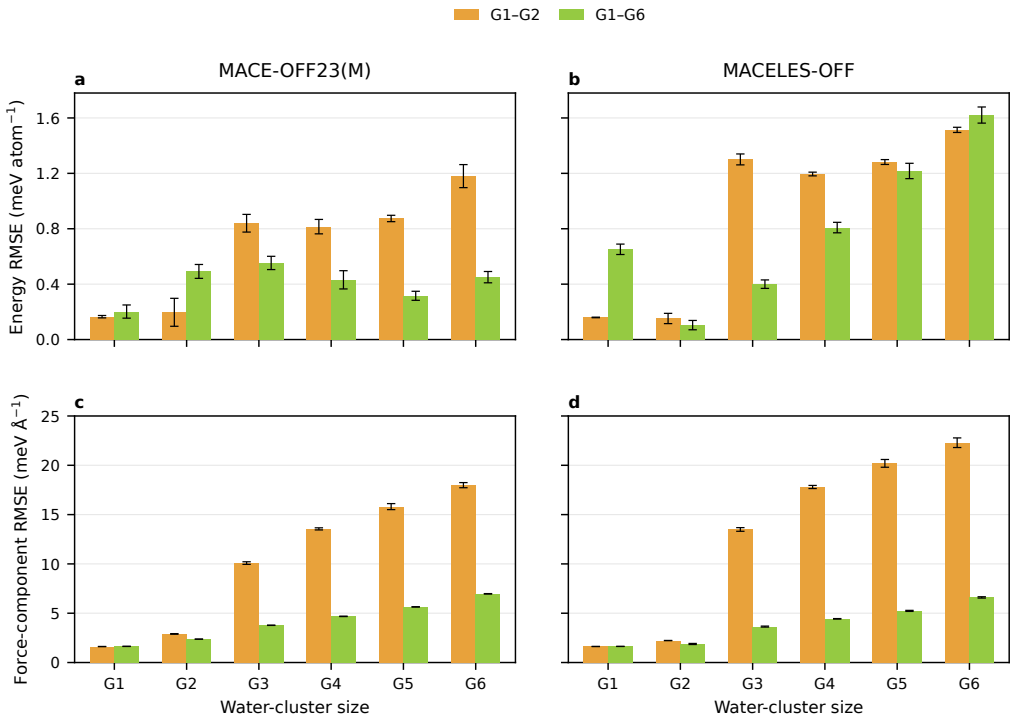


Figure A1: Molecular validation at 123,200 optimizer updates. Per-atom energy RMSE (upper panels) and Cartesian force-component RMSE (lower panels) are shown for G1–G6. OFF is left and LES is right. Bars and whiskers are means and sample SDs over training seeds 17, 29, and 43. These configurations belong to the parent-disjoint validation split, not the locked test split.

B Transport definitions & Sensitivity

The five stored stress channels are

$$\mathbf{s} = \left(\sigma_{xy},\, \sigma_{xz},\, \sigma_{yz},\, \frac{\sigma_{xx} - \sigma_{yy}}{2},\, \frac{\sigma_{yy} - \sigma_{zz}}{2} \right). \tag{5}$$

After subtracting each channel's trajectory mean,

$$C_\sigma(t) = \frac{1}{5} \sum_{a=1}^{5} \langle \delta s_a(0)\, \delta s_a(t) \rangle \,. \tag{6}$$

Time-origin averages use the available pairs at each lag. The diagonal stress combinations are not mutually orthogonal; channels and time origins are not treated as independent replicas. Correlations and trapezoidal integrations use the native 0.5 fs cadence. Duplicate observations at continuation-segment boundaries are removed, and the continuation implementation restores atomic momenta and thermostat state.

For a running integral $X(t)$ sampled at lag times t_j, the reported scalar is

$$\widehat{X}_{[a,b]} = \frac{1}{N_{[a,b]}} \sum_{j:\, a \le t_j \le b} X(t_j), \tag{7}$$

where $N_{[a,b]}$ is the number of sampled lag times in the interval. We use $[a, b] = [5, 10]$ ps for both raw diffusion and viscosity. The diffusion correction in Eq. (3) is then applied to each trajectory's scalar estimate. Table A4 gives the raw and corrected 295 K diffusion values alongside viscosity.

Table A4: Transport at 295 K. Diffusion units are 10^{-9} m^2 s^{-1} and viscosity units are mPa s. Adapted uncertainties are sample SDs across three training-seed means; baseline uncertainties are sample SDs across three MD replicas. The reference diffusion and viscosity are approximately 2.123 and 0.958 in these units, respectively.

Model	D_L	D_{YH}	η
OFF baseline	1.296 ± 0.046	1.685 ± 0.053	1.338 ± 0.053
OFF–G1–G2	2.125 ± 0.178	2.794 ± 0.244	0.765 ± 0.086
OFF–G1–G6	1.766 ± 0.191	2.262 ± 0.259	0.977 ± 0.145
LES baseline	1.818 ± 0.041	2.394 ± 0.109	0.901 ± 0.102
LES–G1–G2	2.781 ± 0.213	3.696 ± 0.305	0.558 ± 0.055
LES–G1–G6	1.453 ± 0.131	1.866 ± 0.163	1.180 ± 0.094

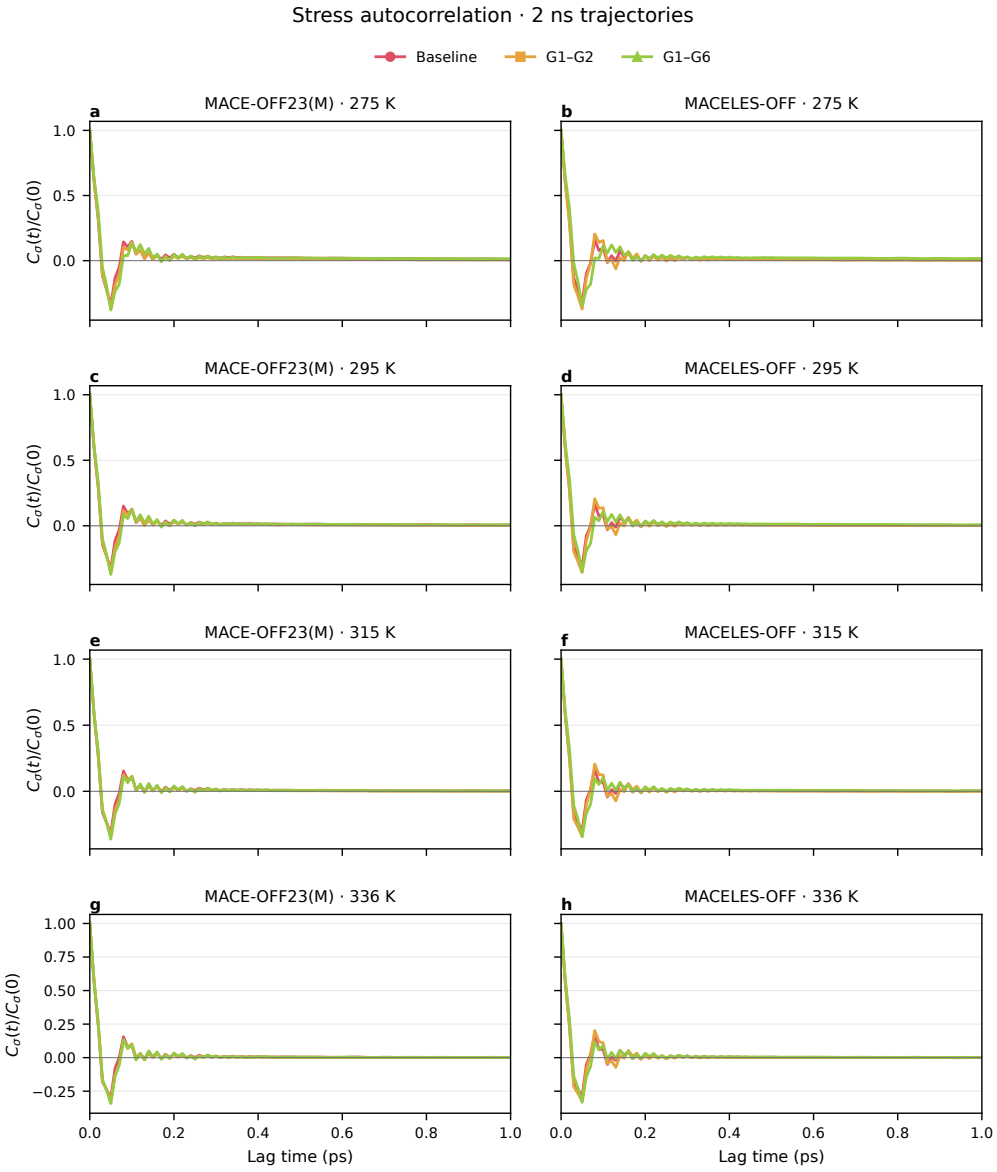


Figure A2: Signed normalized stress autocorrelation over 0–1 ps from the corrected 2 ns NVT trajectories. Rows correspond to 275, 295, 315, and 336 K; OFF is left and LES is right. Each trajectory is normalized by its own $C_\sigma(0)$ before averaging replicas within training seed and then across seeds. Adapted bands show sample SD across seed means; baseline bands show MD-replica sample SD.

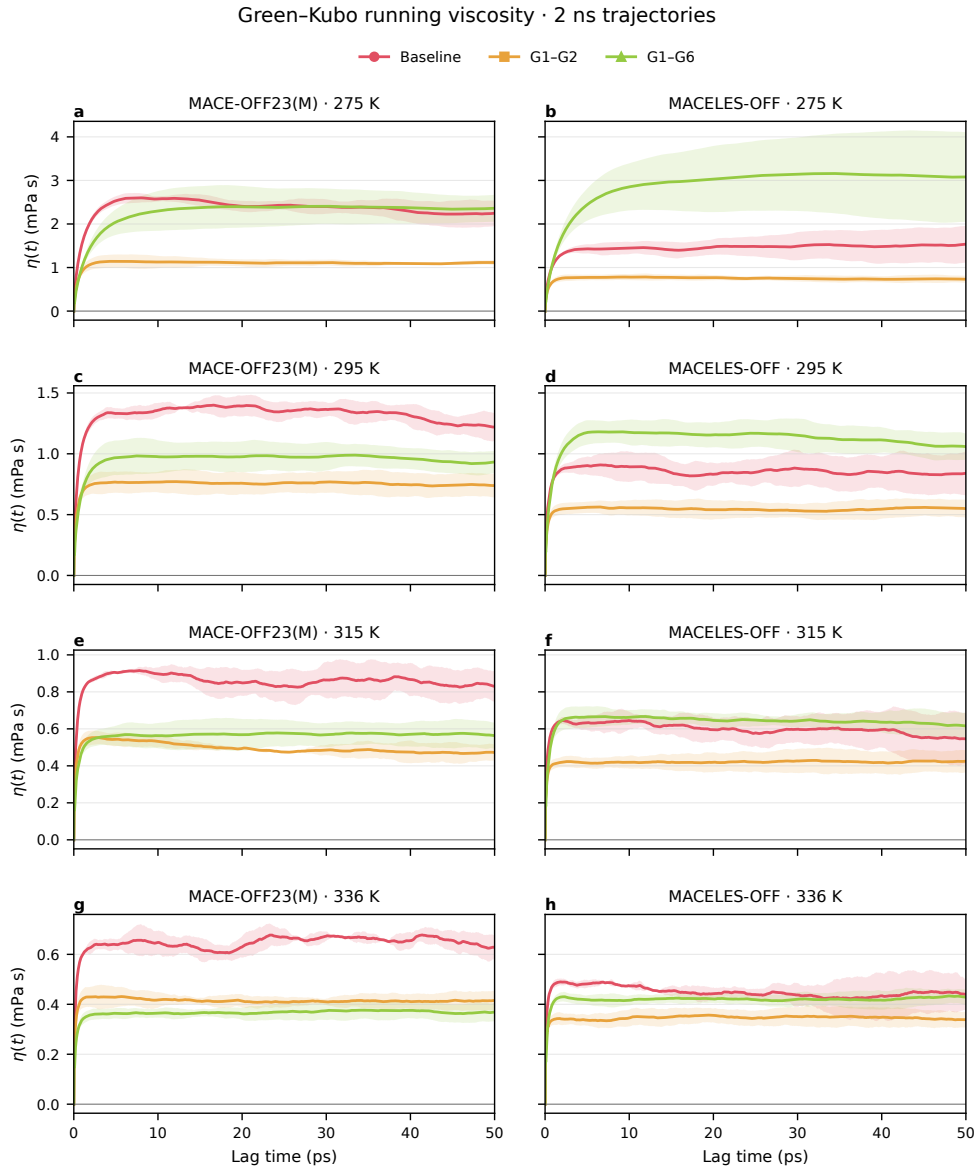


Figure A3: Green–Kubo running viscosity through 50 ps from the corrected 2 ns NVT trajectories. Rows correspond to 275, 295, 315, and 336 K; OFF is left and LES is right. Dark and light gray regions mark the 5–10 ps primary window and the 10–20 ps sensitivity window. Curves use native-cadence stress integration. Adapted curves and bands are means and sample SDs across training-seed means after replica averaging; baseline bands are MD-replica sample SDs.