Neural potentials of proteins extrapolate beyond training data

A Preprint

Geemi. P. Wellawatte
Department of Chemistry
University of Rochester
Rochester, NY, USA
gwellawa@ur.rochester.edu

Glen M. Hocky
Department of Chemistry
Simons Center for Computational Physical Chemistry
New York University
New York, NY, USA hockyg@nyu.edu

Andrew D. White[*]
Department of Chemical Engineering
University of Rochester
Rochester, NY, USA
andrew.white@rochester.edu

Abstract

We evaluate neural network coarse-grained force fields compared to traditional CG molecular mechanics force fields. We conclude neural network force fields are able to extrapolate and sample from unseen regions of the free energy surface free energy surfaces when trained with limited data. Our results come from 66 trained force fields trained on different combinations of clustered free energy surfaces across three proteins. We used total variation similarity as our metric, which assesses agreement between free energy surfaces of force fields. Additionally, force matching error was found to only be weakly correlated with a force field's ability to reconstruct the correct free energy surface. These conclusions support the common hypothesis that constructing force fields on one region of the protein free energy surface can extrapolate well.

1 Introduction

Coarse-grained (CG) molecular dynamics (MD) is a tool to complement experiments.[1,2] A CG model can be considered a "reduced model" as not all degrees of freedom are not considered explicitly. According to Noid[3], CG models provide a foundation to most scientific efforts by focusing over "essential" features of a system. CG MD enables sampling of thermodynamic systems at larger spacial and temporal scales, which are inaccessible at the all-atom resolution. As a result, CG MD is a useful tool to study phenomena such as protein folding pathways and multi-protein structure assemblies which require sampling at larger length or timescales.[4,5] CG models are based on the separation of time of complex systems, thereby providing a practical alternative to uncover the underlying Hamiltonian of these reduced models[6]

One such model consist of two main components – a CG representation (mapping) and a CG forcefield (FF). The first is a "coarser" representation of the all-atom system in the reduced CG phase space. CG atoms can be perceived as pseudo atoms which define the physicochemical character of given groups of atoms.[3,7] The CG FF represents the interactions between these pseudo CG atoms.[8] These interactions must be able to capture the eliminated atomistic level details.[3] Therefore, it can be defined as a potential of mean force (PMF) – a function of weighted averages of energies of the atomistic configurations.[9] This PMF must be able to compute any equilibrium property that is expressed as an ensemble average of the CG coordinates.[10] Finding a fitting approximation of this PMF is one of the key challenges associated with CG modeling.

[*] Corresponding Author

Traditionally, two common approaches are used in developing CG FF, 1) bottom-up approaches[2,11] and 2) top-down approaches.[12] A bottom-up approach relies on information from fine-grained models, while a top-down approach aim to reproduce macroscopic properties.[13] Both these approaches are based on the hypothesis that a CG model must reflect the "correct physics" of the all-atom system.[3] The study by Kidder et al.[14] provides a thorough perspective on CG FFs and the impact of entropic contributions.

The recent advances of deep learning have shown promising results in the field of CG molecular dynamics, where deep learning is used in CG mapping predictions[7] and in developing CG FFs.[15,16,16,17] Work by Behler and Parrinello[18] is one of the first in this research direction, where a generalized neural network (NN) was used to construct a DFT based potential energy surface (PES). Recent work by Majewski et al.[19] show that NN CG forcefields are transferrable among proteins - showed that one general NN FF is able to recover native conformations of multiple proteins. In this work, we focus on the applicability of NN as CG FFs and their limitations.

During training, we desire that the NN FF learn the underlying PMF as a function of the CG coordinates by generating forces based on CG bead positions which agree with the atomic forces mapped onto CG sites, using the following loss function[20,21] :

$$L_{FM} = \sum_{t} \|\nabla_{\mathbf{m}} \hat{F}(\mathbf{M}\mathbf{x_t}, \theta) + f^{\mathbf{m}}(\mathbf{x_t})\| \tag{1}$$

Here, $\mathbf{M}$ is the mapping matrix which takes the N atomic coordinates into n CG sites. $\nabla_{\mathbf{m}} \hat{F}(\mathbf{M}\mathbf{x_t}, \theta)$ represents the gradient of the learned free energy function (effective CG forces) where $\mathbf{m}$ are the CG variables. Instantaneous CG forces mapped from the all-atom trajectory are represented by the last term in the equation 1. Based on this, we may notice that, although NNs have shown to be promising as molecular FFs,[22–24] their performance is viewed as highly dependent on available training data.[21,25] Hence, it is a somewhat open question on how well they can extrapolate beyond training data, especially as newer NN FFs depart in functional form from molecular mechanics FFs. Zeni, *et al.*[25] explain that it is not trivial whether NN potentials are able to exploit the extrapolation regime, specifically when the potential energy surface is smoothened by CG representations.

In this work, we investigate the key question "are NN FFs able to extrapolate to unseen regions of the Free energy surface (FES)?". Furthermore, we investigate how to evaluate the impact of the amount of data used in training. We aim to discuss if NN are suitable to replace traditional, physics informed models to sample from low data regions of the FES. Additionally, we question if forces are an adequate benchmark to train CG FFs and if these FFs learn to sample only from "physically plausible" regions in the FES.

To study these research problems, we selected three proteins based on structural properties[26] – 1) a folded protein: P-Element somatic inhibitor miniprotein (PDB ID:2BN6)[27] 2) a half folded protein: Miniature Esterase (PDB ID: 1V1D)[28] and, 3) an unfolded protein: β-amyloid peptide residues 10-35 (PDB ID: 1HZ3)[29] α-Carbons of the backbone were used to represent each residue (CG representation). The data rich protein trajectories were processed using a Markov State Models (MSM) based approach which extract prominent clusters (meta stable states) from a given protein trajectory. Finally, various subsamples of the meta-states were used to train two NN FFs (CGSchNet[9,16] and, TorchMD-Net[24]) and to produce CG simulations. To evaluate the performance of the trained FFs we use a metric named total variation similarity[30] (TVS),

$$\text{TVS} = 1 - (\sum_{x=a}^{b} P_{mapped}(x) - P_{CG}(x) + \zeta) \tag{2}$$

Here, ζ is a penalty term which accounts for the number of frames in the CG trajectory that are beyond the regions of the mapped trajectory. This TVS metric is a system agnostic metric to evaluate the performances of trained FFs.

2 Methods

2.1 Simulation methods

For each protein, all-atom MD simulation inputs with the forcefield AMBER99SB*-ILDN[31,32] and TIP3P water model[33] were produced using GROMACS tools, with neutralizing potassium ions added. All simulations were performed in GROMACS 2020.4.[34]

All-atom simulations: Minimization and equilibration were performed according to a standard protocol[35] which involves up to 50000 steps of steepest descent minimization, followed by 100 ps of NVT equilibration with backbone atoms restrained. Production $15\mu s$ NPT simulations were performed for each protein, at T = 300K for 2BN6, T = 310K for 1HZ3 and T = 290K for 1V1D. These temperatures were selected empirically to ensure simulation temperature is below the melting point of each protein.[36–38] Production simulations used a 2 fs timestep, a 1 nm cutoff for electrostatics, the v-rescale thermostat[39] with a 0.1 ps time constant, and Parrinello-Rahman barostat,[40] using a 2 ps time constant. From these production runs, training data frames were generated by restarting from fixed points along these trajectories using the same MD parameters, but with velocities resampled for each run. Starting points for restart trajectories were checkpoint files separated every 50 ns starting after 2.5 microseconds. From these 250 checkpoint files, four 10 ns simulations were performed, where positions and forces were saved in double precision every 20 ps. Hence, for each protein we had 500,000 snapshots available for training (250 starting points x 4 simulations/point x 500 frames/simulation).

CG simulations: After training NN FFs (CGSchNet and TorchMD-Net models), each was used to conduct NVT CG simulations with Langevin dynamics at same temperatures as the all-atom simulations. (300K, 290K and 310K). A time step of 2fs were used for all FFs. Each CG trajectory was started from the centroid configuration of one of the testing clusters. We used this approach to avoid the impact of the starting configuration during the CG production. With CGSchNet FFs we were able to run 50 independent trajectories which were 0.02 ns - 0.2 ns long. Most simulation were not stable beyond 0.2ns. With TorchMD-Net we produced 2 ns long CG trajectories with 10 replicas. To perform NVT CG simulations with MARTINI FF, we employed the GROMACS simulation engine with leap-frog algorithm for integrating Newton's equations of motion. Each CG simulation was run for 2 ns with explicit water. For the CG simulations with OpenAWSEM FF, we used Langevin dynamics at constant temperatures 300K, 310K and 290K for 2BN6, 1V1D and 1HZ3 miniproteins. The CG mappings used in these analyses were the default mappings of MARTINI[41] and AWSEM[42] FFs. Each CG simulation was run for 1 ns. All simulations from MARTINI and OpenAWSEM FFs were stable and ran to completion. More details can be found in SI.

2.2 Training data

Firstly, the all-atom trajectories of 2BN6,[27] 1V1D[28] and 1HZ3[29] miniproteins were mapped into a CG representation, where each residue was represented with its α-carbon atom. Then each mapped trajectory was clustered into four meta stable states based on a Hidden-Markov State Model (HMSM).[43,44] using the PyEMMA python library as described next.[45,46] Finally, configurations (snapshots from the trajectory) from various subsets of the meta states were used for training separate FFs.

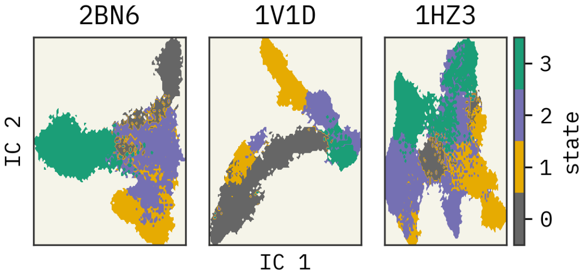

Figure 1: Meta stable clusters of the miniproteins in the low dimensional space projected using the TICA method.[47] P-Element somatic inhibitor miniprotein (PDB ID:2BN6), Miniature Esterase (PDB ID: 1V1D), and, β-amyloid peptide residues 10-35 (PDB ID: 1HZ3) were used in this study.

Clustering of protein trajectories to meta stable states in the FES is based on a MSMs. These are a powerful toolkit for analyzing dynamic data from MD simulations.[48,49] The main steps involved in building a MSM are, 1) featurization 2) dimensionality reduction 3) clustering and 4) estimation of the transition matrix.[50] They are extensively discussed in literature on approximating observables from MD simulations.[51–55]

We selected the α-Carbon pairwise distances to featurize mapped trajectories for the dimensionality reduction. Time-lagged independent component analysis (TICA)[47,56] was used for this step. Next, these projected spaces were discretized using K-means[57] clustering to estimate an initial Markov State Model (MSM). 50,

Table 1: Cluster combinations used for training

Cluster Percentage	100%	75%	50%	25%
FF label: **clusters used in training**	*FF0:* 1,2,3,4	*FF1:* 2,3,4 *FF2:* 1,3,4 *FF3:* 1,2,4 *FF4:* 1,2,3	*FF5:* 1,2 *FF6:* 3,4	*FF7:* 1 *FF8:* 2 *FF9:* 3 *FF10:* 4

75 and 200 cluster centers were used for 2BN6, 1V1D and 1HZ3 miniproteins respectively. These cluster numbers were selected based on the VAMP2 scores.[58] This is the sum of singular values of the symmetrized MSM transition matrix. Respective lags of 100, 100 and 10 were selected to build MSM. Lags were selected such that the implied timescales were constant with the statistical error (See SI). Furthermore, we validated the MSMs using Chapman–Kolmogorov tests.[59,60]

Finally, HMSMs were estimated based on the reference MSMs where each trajectory was clustered into four meta stable states – each frame of the trajectories were assigned to a cluster. Christoforou et al.[61] describe a HMSM as a "kinetic" coarse-graining model which group the microstates identified by the k-means clustering algorithm. We followed a similar approach as Christoforou et al.[61] to assign meta stable clusters. Figure 1 shows the four meta clusters of the reduced dimensional spaces along the first two independent components (IC1 and IC2) identified with TICA.[47] Further details for this procedure are given in the SI.

2.3 Training forcefields and running CG simulations

In this study, we selected two NN-based CG FFs; CGSchNet[9] and TorchMD-Net.[24] CGSchNet[9] is a modified version of the CGNet model[16] which learns the CG PES based on the force matching approach. In CGNet model, the inputs are hand-selected features such as bond distances, angles and dihedrals. However, in the CGSchNet model, the features are "learned" during training by leveraging the SchNet model.[62,63] TorchMD-Net[24] package provide a state-of-the-art graph neural network (GNN) and equivariant transformer (ET) based NN potentials for molecular simulations. Additionally, TorchMD[64] is a Python API for performing molecular dynamics. In this work, we used TorchMD-Net's GNN model and TorchMD for training CG FFs and for conducting CG simulations respectively.

The key objective of this work is to investigate if NN CG FFs are able to extrapolate to sample from unseen regions of the FES. Therefore, during training we subsampled different sets of meta-stable states and trained multiple independent FFs per miniprotein (listed in Table 1). For example, to train "FF1" we used 75% of clusters, those labeled as 2,3 and 4 – data from cluster 1 were withheld. We followed the same nomenclature for all three miniproteins and both CGSchNet[9] and TorchMD-Net FFs. The number of frames from each cluster were kept constant through downsampling. Other hyperparameters used in training and train-validation error plots can be found in SI. Finally, the trained FFs were used to produce CG simulation. See Simulation methods section for more details. Note that, due to the smoothness of the underlying CG FES, similar amount of sampling is obtained in these "ns" long simulations as compared to the original microseconds of training data, as well be shown shortly.

3 Results and discussion

First, we compared the performances of the *FF0* from CGSchNet and TorchMD-Net (trained with data from all four meta states) with state-of-the-art physics informed FFs MARTINI[41,65] and OpenAWSEM.[66] MARTINI is possibly the most popularly used FF in CG simulations[67] of lipids,[68,69] proteins,[70,71] sugars[72] and other biomolecules.[73,74] OpenAWSEM is the implementation of AWSEM[42] CG FF for proteins within the GPU compatible OpenMM framework. AWSEM contains physics informed many-body effects and employ an implicit solvent environment.[42] This FF has been successfully applied to study protein structure prediction.[75–77]

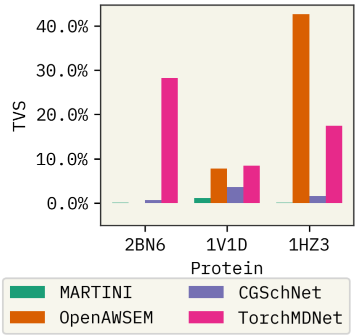

Figure 2: Comparison with state-of-the-art methods. Higher TVS refers to high similarity between mapped and CG simulation distributions in the projected TICA spaces.

Based on the comparison illustrated in Figure 2 we observe the following; a) Performance of the OpenAWSEM FF increases significantly with the increasing structural disorder of the miniproteins, b) CGSchNet has the lowest overall performance among all four FFs c) TorchMD-Net has the highest average performance. Note that MARTINI and OpenAWSEM mapped trajectories in the Figure 2 are visually different to CGSchNet and TorchMD-Net trajectories due to the differences in CG representations.

CG trajectories from forcefields *FF0* from CGSchNet, TorchMD-Net, MARTINI and OpenAWSEM were featurized and projected into a FES using TICA.[30] Figure 3 show the cartoon representations of the 3 miniproteins along with their projected trajectories. We observe in 3 that the TVS between the mapped and CG simulations from CGschNet FF0 is low because the trajectories explore a broader region in their 2D projected spaces. This observation indicates that the CGSchNet-FF0 tend to sample from physically non-meaningful regions. This hypothesis is further confirmed as the CG simulations from CGSchNet FFs did not run to completion for 2 out of 3 miniproteins. Total trajectory times were between 0.02 ns - 0.2 ns and none of the CG simulations were stable beyond 0.2 ns. However, with TorchMD-Net we were able to produce longer simulations for 2 ns each. We selected this cut off due to time and resource constraints. Additionally, we notice that TorchMD-Net FFs strictly explores the region around the mapped trajectories, thus avoiding physically non-plausible regions. With these observations, we conclude that TorchMD-Net outperforms CGSchNet, MARTINI and OpenAWSEM FFs when trained with all available data – TorchMD-Net has the highest average TVS between mapped and CG FES. Therefore, we proceeded to focus on the impact of training data only on the performance of TorchMD-Net FFs.

Figure 4 illustrates the performances of TorchMD-Net FFs trained with various combinations of meta-states. Surprisingly, we observe that the percentage of meta stable clusters used in training does not strongly impact the performance of the FFs. For example, we see that the TVS of *FF0* trained with data from all 4 meta-states is comparable to *FF7-10* trained with data from only one meta-cluster. Note the number of frames is kept constant for each example by downsampling. This shows that exploration of configurational space has little impact on the FF quality – the FFs can extrapolate.

Finally, to study if training on force matching error is a predictor for NN FF correctness, we compared the force error (validation error) of all 11 CGschNet FFs and 11 TorchMD-Net FFs per miniprotein. Figure 5 shows that force errors from both CGschNet TorchMD-Net models only differ by ± 0.2 kcal/(mol.). However, their TVS differ by $\sim 20\%$. Based on this observation, we can conclude there is weak or no correlation between force matching error and configuration free energy surface – even within the same model architecture. This observation aligns with the findings by Fu et al.[20], which showed that the models having best force matching error are not necessarily the best at predicting properties like diffusivity or radial distribution function agreement.

4 Conclusions

Based on our results, we observe that TorchMD-Net significantly outperforms the two physics informed FFs (MARTINI[41] and OpenAWSEM[66]) and CGSchNet FFs. Unlike the other FFs, TorchMD-Net FFs strictly explore around the same FES as the mapped trajectories indicating TorchMD-Net FFs tend to avoid

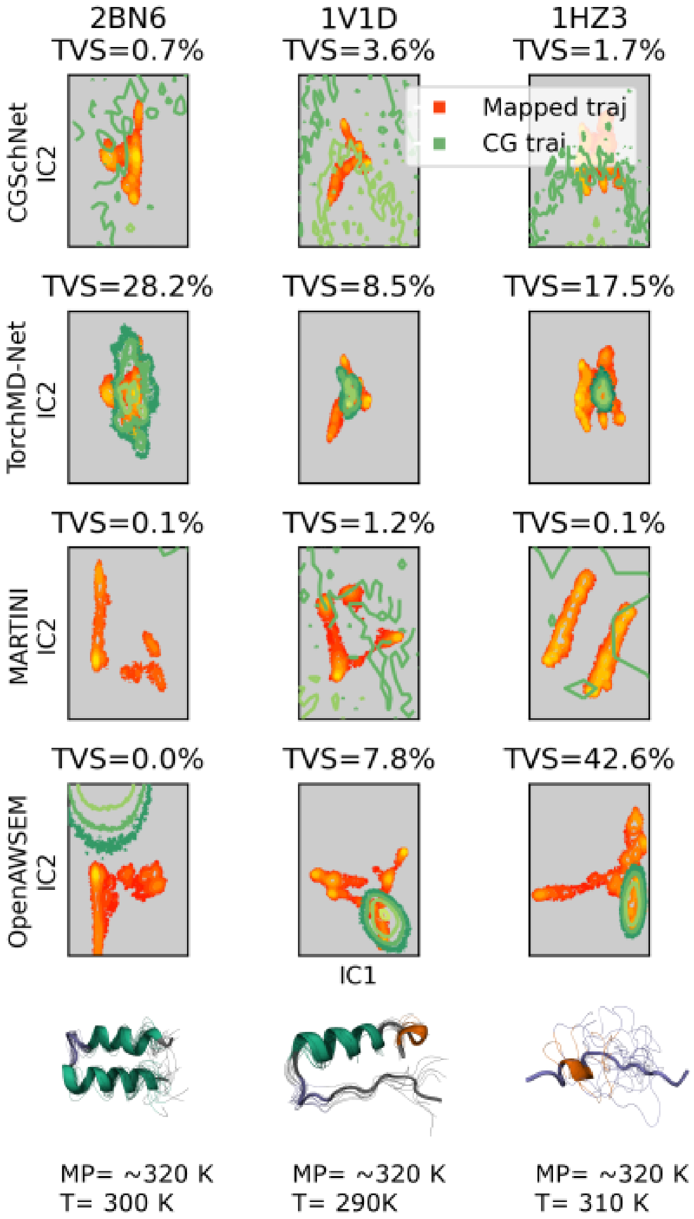

Figure 3: Mapped and CG FES from *FF0* – FFs trained with all 4 meta-states. Top: projected miniprotein trajectories from CGSchNet, TorchMD-Net, MARTINI and OpenAWSEM FFs. Bottom: cartoon representations of miniprotein trajectories annotated with approximate melting and simulation temperatures. Conformational ensembles are 20 random frames after a weighted iterative alignment following the procedure of Ref. 78.

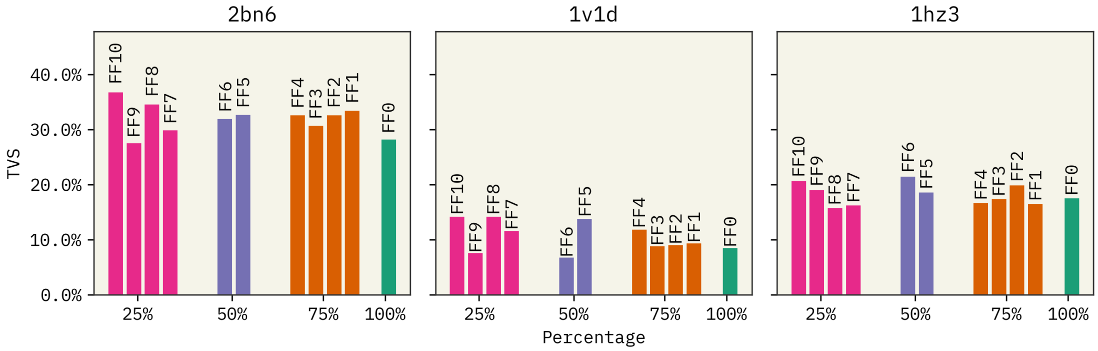

Figure 4: Impact of data in training of forcefields. Labels of the FFs indicate the Meta-states used in training. See Table 1.

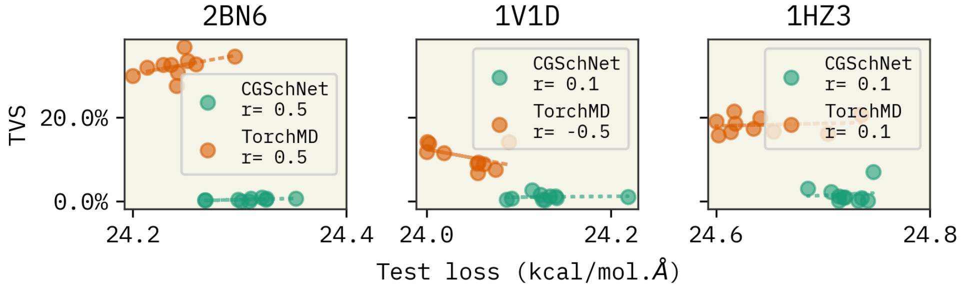

Figure 5: Variation of TVS with force matching error of all CGSchNet and TorchMD-Net FFs. The x-axis denotes average validation error of the last 3 epochs. TVS indicates the similarity between mapped and CG trajectories from the trained FFs.

physically improbable configurations. Mainly, we observe that the number of meta stable clusters used in training does not impact the overall performance of FFs trained with TorchMD-Net, for proteins ranging from fully ordered to fully disordered. Therefore, we conclude that NN FFs are able to extrapolate to unseen regions of the FES. Furthermore, we note that these NNs are comparable to the physics informed FFs or even able to outperform them. Additionally, we find support that maximizing agreement between forces is not a strong predictor of model accuracy in other metrics, as supported by other recent findings[20,79]

5 acknowledgement

Research was supported by the National Institutes of Health through the award R35GM137966 (for GPW and ADW) and R35GM138312 (for GMH). This work was supported in part through the NYU IT High Performance Computing resources, services, and staff expertise, and simulations were partially executed on resources supported by the Simons Center for Computational Physical Chemistry at NYU (SF Grant No 839534).

References

[1] Gary S. Ayton and Gregory A. Voth. Simulation of biomolecular systems at multiple length and time scales. *International Journal for Multiscale Computational Engineering*, 2(2), 2004. ISSN 1543-1649.

[2] Sergei Izvekov and Gregory A. Voth. A multiscale coarse-graining method for biomolecular systems. *The Journal of Physical Chemistry B*, 109(7):2469–2473, 2005. doi:10.1021/jp044629q. URL `https://doi.org/10.1021/jp044629q`. PMID: 16851243.

[3] W. G. Noid. Perspective: Coarse-grained models for biomolecular systems. *The Journal of Chemical Physics*, 139(9):090901, 2013. doi:10.1063/1.4818908. URL `https://doi.org/10.1063/1.4818908`.

[4] Sebastian Kmiecik, Dominik Gront, Michal Kolinski, Lukasz Wieteska, Aleksandra Elzbieta Dawid, and Andrzej Kolinski. Coarse-grained protein models and their applications. *Chemical Reviews*, 116(14): 7898–7936, 2016. doi:10.1021/acs.chemrev.6b00163. URL `https://doi.org/10.1021/acs.chemrev.6b00163`. PMID: 27333362.

[5] Troy Cellmer and Nicolas L. Fawzi. *Coarse-Grained Simulations of Protein Aggregation*, pages 453–470. Humana Press, Totowa, NJ, 2012. ISBN 978-1-61779-921-1. doi:10.1007/978-1-61779-921-1_27. URL `https://doi.org/10.1007/978-1-61779-921-1_27`.

[6] Cecilia Clementi. Coarse-grained models of protein folding: toy models or predictive tools? *Current Opinion in Structural Biology*, 18(1):10–15, 2008. ISSN 0959-440X. doi:https://doi.org/10.1016/j.sbi.2007.10.005. URL `https://www.sciencedirect.com/science/article/pii/S0959440X07001753`. Folding and Binding / Protein-nucleic acid interactions.

[7] Zhiheng Li, Geemi P. Wellawatte, Maghesree Chakraborty, Heta A. Gandhi, Chenliang Xu, and Andrew D. White. Graph neural network based coarse-grained mapping prediction. *Chem. Sci.*, 11:9524–9531, 2020. doi:10.1039/D0SC02458A. URL `http://dx.doi.org/10.1039/D0SC02458A`.

[8] Jay W. Ponder and David A. Case. Force fields for protein simulations. In *Protein Simulations*, volume 66 of *Advances in Protein Chemistry*, pages 27–85. Academic Press, 2003. doi:https://doi.org/10.1016/S0065-3233(03)66002-X. URL `https://www.sciencedirect.com/science/article/pii/S006532330366002X`.

[9] Brooke E. Husic, Nicholas E. Charron, Dominik Lemm, Jiang Wang, Adrià Pérez, Maciej Majewski, Andreas Krämer, Yaoyi Chen, Simon Olsson, Gianni de Fabritiis, Frank Noé, and Cecilia Clementi. Coarse graining molecular dynamics with graph neural networks. *The Journal of Chemical Physics*, 153 (19):194101, 2020. doi:10.1063/5.0026133. URL `https://doi.org/10.1063/5.0026133`.

[10] Jonas Köhler, Yaoyi Chen, Andreas Krämer, Cecilia Clementi, and Frank Noé. Force-matching coarse-graining without forces. *arXiv preprint arXiv:2203.11167*, 2022.

[11] Dirk Reith, Mathias Pütz, and Florian Müller-Plathe. Deriving effective mesoscale potentials from atomistic simulations. *Journal of Computational Chemistry*, 24(13):1624–1636, 2003. doi:https://doi.org/10.1002/jcc.10307. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.10307`.

[12] Marissa G. Saunders and Gregory A. Voth. Coarse-graining methods for computational biology. *Annual Review of Biophysics*, 42(1):73–93, 2013. doi:10.1146/annurev-biophys-083012-130348. URL `https://doi.org/10.1146/annurev-biophys-083012-130348`. PMID: 23451897.

[13] Zack Jarin, James Newhouse, and Gregory A. Voth. Coarse-grained force fields from the perspective of statistical mechanics: Better understanding of the origins of a martini hangover. *Journal of Chemical Theory and Computation*, 17(2):1170–1180, 2021. doi:10.1021/acs.jctc.0c00638. URL `https://doi.org/10.1021/acs.jctc.0c00638`. PMID: 33475352.

[14] Katherine M. Kidder, Ryan J. Szukalo, and W. G. Noid. Energetic and entropic considerations for coarse-graining. *The European Physical Journal B*, 94(7):153, Jul 2021. ISSN 1434-6036. doi:10.1140/epjb/s10051-021-00153-4. URL `https://doi.org/10.1140/epjb/s10051-021-00153-4`.

[15] Linfeng Zhang, Jiequn Han, Han Wang, Roberto Car, and Weinan E. Deepcg: Constructing coarse-grained models via deep neural networks. *The Journal of Chemical Physics*, 149(3):034101, 2018. doi:10.1063/1.5027645. URL `https://doi.org/10.1063/1.5027645`.

[16] Jiang Wang, Simon Olsson, Christoph Wehmeyer, Adrià Pérez, Nicholas E. Charron, Gianni de Fabritiis, Frank Noé, and Cecilia Clementi. Machine learning of coarse-grained molecular dynamics force fields. *ACS Central Science*, 5(5):755–767, 2019. doi:10.1021/acscentsci.8b00913. URL `https://doi.org/10.1021/acscentsci.8b00913`. PMID: 31139712.

[17] Stefan Doerr, Maciej Majewski, Adrià Pérez, Andreas Krämer, Cecilia Clementi, Frank Noe, Toni Giorgino, and Gianni De Fabritiis. Torchmd: A deep learning framework for molecular simulations. *Journal of Chemical Theory and Computation*, 17(4):2355–2363, 2021. doi:10.1021/acs.jctc.0c01343. URL `https://doi.org/10.1021/acs.jctc.0c01343`. PMID: 33729795.

[18] Jörg Behler and Michele Parrinello. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.*, 98:146401, Apr 2007. doi:10.1103/PhysRevLett.98.146401. URL `https://link.aps.org/doi/10.1103/PhysRevLett.98.146401`.

[19] Maciej Majewski, Adrià Pérez, Philipp Thölke, Stefan Doerr, Nicholas E. Charron, Toni Giorgino, Brooke E. Husic, Cecilia Clementi, Frank Noé, and Gianni De Fabritiis. Machine learning coarse-grained potentials of protein thermodynamics. 12 2022. doi:10.48550/arxiv.2212.07492. URL `https://arxiv.org/abs/2212.07492v1`.

[20] Xiang Fu, Zhenghao Wu, Wujie Wang, Tian Xie, Sinan Keten, Rafael Gomez-Bombarelli, and Tommi Jaakkola. Forces are not enough: Benchmark and critical evaluation for machine learning force fields with molecular simulations. *arXiv preprint arXiv:2210.07237*, 2022.

[21] Frank Noé, Alexandre Tkatchenko, Klaus-Robert Müller, and Cecilia Clementi. Machine learning for molecular simulation. *Annual Review of Physical Chemistry*, 71(1):361–390, 2020. doi:10.1146/annurev-physchem-042018-052331. URL `https://doi.org/10.1146/annurev-physchem-042018-052331`. PMID: 32092281.

[22] Alireza Khorshidi and Andrew A. Peterson. Amp: A modular approach to machine learning in atomistic simulations. *Computer Physics Communications*, 207:310–324, 2016. ISSN 0010-4655. doi:https://doi.org/10.1016/j.cpc.2016.05.010. URL `https://www.sciencedirect.com/science/article/pii/S0010465516301266`.

[23] Stefan Chmiela, Alexandre Tkatchenko, Huziel E. Sauceda, Igor Poltavsky, Kristof T. Schütt, and Klaus-Robert Müller. Machine learning of accurate energy-conserving molecular force fields. *Science Advances*, 3(5):e1603015, 2017. doi:10.1126/sciadv.1603015. URL `https://www.science.org/doi/abs/10.1126/sciadv.1603015`.

[24] Philipp Thölke and Gianni De Fabritiis. Torchmd-net: Equivariant transformers for neural network based molecular potentials. *arXiv preprint arXiv:2202.02541*, 2022.

[25] Claudio Zeni, Andrea Anelli, Aldo Glielmo, and Kevin Rossi. Exploring the robust extrapolation of high-dimensional machine learning potentials. *Phys. Rev. B*, 105:165141, Apr 2022. doi:10.1103/PhysRevB.105.165141. URL `https://link.aps.org/doi/10.1103/PhysRevB.105.165141`.

[26] Paul Robustelli, Stefano Piana, and David E. Shaw. Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences*, 115(21): E4758–E4766, 2018. doi:10.1073/pnas.1800690115. URL `https://www.pnas.org/doi/abs/10.1073/pnas.1800690115`.

[27] Tijana Ignjatovic, Ji-Chun Yang, Jonathan Butler, David Neuhaus, and Kiyoshi Nagai. Structural basis of the interaction between p-element somatic inhibitor and u1-70k essential for the alternative splicing of p-element transposase. *Journal of Molecular Biology*, 351(1):52–65, 2005. ISSN 0022-2836. doi:https://doi.org/10.1016/j.jmb.2005.04.077. URL `https://www.sciencedirect.com/science/article/pii/S0022283605005437`.

[28] Andrew J. Nicoll and Rudolf K. Allemann. Nucleophilic and general acid catalysis at physiological ph by a designed miniature esterase. *Org. Biomol. Chem.*, 2:2175–2180, 2004. doi:10.1039/B404730C. URL `http://dx.doi.org/10.1039/B404730C`.

[29] S. Zhang, K. Iwata, M.J. Lachenmann, J.W. Peng, S. Li, E.R. Stimson, Y. a. Lu, A.M. Felix, J.E. Maggio, and J.P. Lee. The alzheimer's peptide aβ adopts a collapsed coil structure in water. *Journal of Structural Biology*, 130(2):130–141, 2000. ISSN 1047-8477. doi:https://doi.org/10.1006/jsbi.2000.4288. URL `https://www.sciencedirect.com/science/article/pii/S1047847700942886`.

[30] Sourav Chatterjee. Sourav9990, Jul 2008. URL `https://archive.org/details/@sourav9990?tab=web-archive&amp;sort=-date`.

[31] Robert B Best and Gerhard Hummer. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J. Phys. Chem. B*, 113(26):9004–9015, 2009.

[32] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L Klepeis, Ron O Dror, and David E Shaw. Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins*, 78(8):1950–1958, 2010.

[33] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2): 926–935, 1983.

[34] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C Smith, Berk Hess, and Erik Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25, 2015.

[35] `http://www.mdtutorials.com/gmx/lysozyme/index.html`.

[36] Nikolas H Chmiel, Donald C Rio, and Jennifer A Doudna. Distinct contributions of kh domains to substrate binding affinity of drosophila p-element somatic inhibitor protein. *Rna*, 12(2):283–291, 2006.

[37] S. Zhang, K. Iwata, M.J. Lachenmann, J.W. Peng, S. Li, E.R. Stimson, Y. a. Lu, A.M. Felix, J.E. Maggio, and J.P. Lee. The alzheimer's peptide a adopts a collapsed coil structure in water. *Journal of Structural Biology*, 130(2):130–141, 2000. ISSN 1047-8477. doi:https://doi.org/10.1006/jsbi.2000.4288. URL `https://www.sciencedirect.com/science/article/pii/S1047847700942886`.

[38] Andrew J. Nicoll and Rudolf K. Allemann. Nucleophilic and general acid catalysis at physiological ph by a designed miniature esterase. *Org. Biomol. Chem.*, 2:2175–2180, 2004. doi:10.1039/B404730C. URL `http://dx.doi.org/10.1039/B404730C`.

[39] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126(1):014101, 2007.

[40] Michele Parrinello and Aneesur Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. App. Phys.*, 52(12):7182–7190, 1981.

[41] Siewert J. Marrink, Alex H. de Vries, and Alan E. Mark. Coarse grained model for semiquantitative lipid simulations. *The Journal of Physical Chemistry B*, 108(2):750–760, 2004. doi:10.1021/jp036508g. URL `https://doi.org/10.1021/jp036508g`.

[42] Aram Davtyan, Nicholas P. Schafer, Weihua Zheng, Cecilia Clementi, Peter G. Wolynes, and Garegin A. Papoian. Awsem-md: Protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. *The Journal of Physical Chemistry B*, 116(29):8494–8503, 2012. doi:10.1021/jp212541y. URL `https://doi.org/10.1021/jp212541y`. PMID: 22545654.

[43] Leonard E Baum and Ted Petrie. Statistical inference for probabilistic functions of finite state markov chains. *The annals of mathematical statistics*, 37(6):1554–1563, 1966.

[44] L.R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989. doi:10.1109/5.18626.

[45] Martin K. Scherer, Benjamin Trendelkamp-Schroer, Fabian Paul, Guillermo Pérez-Hernández, Moritz Hoffmann, Nuria Plattner, Christoph Wehmeyer, Jan-Hendrik Prinz, and Frank Noé. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *Journal of Chemical Theory and Computation*, 11:5525–5542, October 2015. ISSN 1549-9618. doi:10.1021/acs.jctc.5b00743. URL `http://dx.doi.org/10.1021/acs.jctc.5b00743`.

[46] Robert Zwanzig. From classical dynamics to continuous time random walks. *Journal of Statistical Physics*, 30(2):255–262, 1983.

[47] L. Molgedey and H. G. Schuster. Separation of a mixture of independent signals using time delayed correlations. *Phys. Rev. Lett.*, 72:3634–3637, Jun 1994. doi:10.1103/PhysRevLett.72.3634. URL `https://link.aps.org/doi/10.1103/PhysRevLett.72.3634`.

[48] Vijay S. Pande, Kyle Beauchamp, and Gregory R. Bowman. Everything you wanted to know about markov state models but were afraid to ask. *Methods*, 52(1):99–105, 2010. ISSN 1046-2023. doi:https://doi.org/10.1016/j.ymeth.2010.06.002. URL `https://www.sciencedirect.com/science/article/pii/S1046202310001568`. Protein Folding.

[49] Brooke E. Husic and Vijay S. Pande. Markov state models: From an art to a science. *Journal of the American Chemical Society*, 140(7):2386–2396, 2018. doi:10.1021/jacs.7b12191. URL `https://doi.org/10.1021/jacs.7b12191`. PMID: 29323881.

[50] Christopher Kolloff and Simon Olsson. Machine learning in molecular dynamics simulations of biomolecular systems. *arXiv preprint arXiv:2205.03135*, 2022.

[51] Frank Noé, Illia Horenko, Christof Schütte, and Jeremy C Smith. Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states. *The Journal of chemical physics*, 126(15):04B617, 2007.

[52] Frank Noé. Probability distributions of molecular observables computed from markov models. *The Journal of chemical physics*, 128(24):244103, 2008.

[53] John D Chodera and Frank Noé. Probability distributions of molecular observables computed from markov models. ii. uncertainties in observables and their time-evolution. *The Journal of chemical physics*, 133(10):09B606, 2010.

[54] Nicolae-Viorel Buchete and Gerhard Hummer. Coarse master equations for peptide folding dynamics. *The Journal of Physical Chemistry B*, 112(19):6057–6069, 2008.

[55] Christof Schütte, Frank Noé, Jianfeng Lu, Marco Sarich, and Eric Vanden-Eijnden. Markov state models based on milestoning. *The Journal of chemical physics*, 134(20):05B609, 2011.

[56] Guillermo Pérez-Hernández, Fabian Paul, Toni Giorgino, Gianni De Fabritiis, and Frank Noé. Identification of slow molecular order parameters for markov model construction. *The Journal of Chemical Physics*, 139(1):015102, 2013. doi:10.1063/1.4811489. URL `https://doi.org/10.1063/1.4811489`.

[57] S. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982. doi:10.1109/TIT.1982.1056489.

[58] Hao Wu and Frank Noé. Variational approach for learning markov processes from time series data. *Journal of Nonlinear Science*, 30(1):23–66, 2020.

[59] Jan-Hendrik Prinz, Hao Wu, Marco Sarich, Bettina Keller, Martin Senne, Martin Held, John D. Chodera, Christof Schütte, and Frank Noé. Markov models of molecular kinetics: Generation and validation. *The Journal of Chemical Physics*, 134(17):174105, 2011. doi:10.1063/1.3565032. URL `https://doi.org/10.1063/1.3565032`.

[60] A Papoulis. Bayes' theorem in statistics and bayes' theorem in statistics (reexamined). *Probability, random variables, and stochastic processes. 2nd ed. New York, NY: McGraw-Hill*, pages 38–114, 1984.

[61] Emmanouil Christoforou, Hari Leontiadou, Frank Noé, Jannis Samios, Ioannis Z. Emiris, and Zoe Cournia. Investigating the bioactive conformation of angiotensin ii using markov state modeling revisited with web-scale clustering. *Journal of Chemical Theory and Computation*, 18(9):5636–5648, 2022. doi:10.1021/acs.jctc.1c00881. URL `https://doi.org/10.1021/acs.jctc.1c00881`. PMID: 35944098.

[62] Kristof Schütt, Pieter-Jan Kindermans, Huziel Enoc Sauceda Felix, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *Advances in neural information processing systems*, 30, 2017.

[63] KT Schutt, Pan Kessel, Michael Gastegger, KA Nicoli, Alexandre Tkatchenko, and K-R Muller. Schnetpack: A deep learning toolbox for atomistic systems. *Journal of Chemical Theory and Computation*, 15 (1):448–455, 2018.

[64] Stefan Doerr, Maciej Majewsk, Adrià Pérez, Andreas Krämer, Cecilia Clementi, Frank Noe, Toni Giorgino, and Gianni De Fabritiis. Torchmd: A deep learning framework for molecular simulations, 2020.

[65] Siewert J. Marrink, H. Jelger Risselada, Serge Yefimov, D. Peter Tieleman, and Alex H. de Vries. The martini force field: Coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B*, 111(27):7812–7824, 2007. doi:10.1021/jp071097f. URL `https://doi.org/10.1021/jp071097f`. PMID: 17569554.

[66] Wei Lu, Carlos Bueno, Nicholas P Schafer, Joshua Moller, Shikai Jin, Xun Chen, Mingchen Chen, Xinyu Gu, Aram Davtyan, Juan J de Pablo, et al. Openawsem with open3spn2: A fast, flexible, and accessible framework for large-scale coarse-grained biomolecular simulations. *PLoS computational biology*, 17(2): e1008308, 2021.

[67] Bart M. H. Bruininks, Paulo C. T. Souza, and Siewert J. Marrink. *A Practical View of the Martini Force Field*, pages 105–127. Springer New York, New York, NY, 2019. ISBN 978-1-4939-9608-7. doi:10.1007/978-1-4939-9608-7_5. URL `https://doi.org/10.1007/978-1-4939-9608-7_5`.

[68] Siewert J Marrink and Alan E Mark. The mechanism of vesicle fusion as revealed by molecular dynamics simulations. *Journal of the American Chemical Society*, 125(37):11144–11145, 2003.

[69] Siewert J Marrink and Alan E Mark. Molecular dynamics simulation of the formation, structure, and dynamics of small phospholipid vesicles. *Journal of the American Chemical Society*, 125(49):15233–15242, 2003.

[70] Luca Monticelli, Senthil K Kandasamy, Xavier Periole, Ronald G Larson, D Peter Tieleman, and Siewert-Jan Marrink. The martini coarse-grained force field: extension to proteins. *Journal of chemical theory and computation*, 4(5):819–834, 2008.

[71] Djurre H de Jong, Gurpreet Singh, WF Drew Bennett, Clement Arnarez, Tsjerk A Wassenaar, Lars V Schafer, Xavier Periole, D Peter Tieleman, and Siewert J Marrink. Improved parameters for the martini coarse-grained protein force field. *Journal of chemical theory and computation*, 9(1):687–697, 2013.

[72] César A López, Giovanni Bellesia, Antonio Redondo, Paul Langan, Shishir PS Chundawat, Bruce E Dale, Siewert J Marrink, and S Gnanakaran. Martini coarse-grained model for crystalline cellulose microfibers. *The Journal of Physical Chemistry B*, 119(2):465–473, 2015.

[73] Jaakko J Uusitalo, Helgi I Ingólfsson, Parisa Akhshi, D Peter Tieleman, and Siewert J Marrink. Martini coarse-grained force field: extension to dna. *Journal of chemical theory and computation*, 11(8):3932–3945, 2015.

[74] Djurre H De Jong, Nicoletta Liguori, Tom Van Den Berg, Clement Arnarez, Xavier Periole, and Siewert J Marrink. Atomistic and coarse grain topologies for the cofactors associated with the photosystem ii core complex. *The Journal of Physical Chemistry B*, 119(25):7791–7803, 2015.

[75] Xun Chen, Mingchen Chen, and Peter G. Wolynes. Exploring the interplay between disordered and ordered oligomer channels on the aggregation energy landscapes of -synuclein. *The Journal of Physical Chemistry B*, 126(28):5250–5261, 2022. doi:10.1021/acs.jpcb.2c03676. URL `https://doi.org/10.1021/acs.jpcb.2c03676`. PMID: 35815598.

[76] Xun Chen, Mingchen Chen, Nicholas P Schafer, and Peter G Wolynes. Exploring the interplay between fibrillization and amorphous aggregation channels on the energy landscapes of tau repeat isoforms. *Proceedings of the National Academy of Sciences*, 117(8):4125–4130, 2020.

[77] Weihua Zheng, Min-Yeh Tsai, Mingchen Chen, and Peter G Wolynes. Exploring the aggregation free energy landscape of the amyloid-β protein (1–40). *Proceedings of the National Academy of Sciences*, 113 (42):11835–11840, 2016.

[78] Heidi Klem, Glen M Hocky, and Martin McCullagh. Size-and-shape space gaussian mixture models for structural clustering of molecular dynamics trajectories. *Journal of chemical theory and computation*, 2022.

[79] Michael Schaarschmidt, Morgane Riviere, Alex M Ganose, James S Spencer, Alexander L Gaunt, James Kirkpatrick, Simon Axelrod, Peter W Battaglia, and Jonathan Godwin. Learned force fields are ready for ground state catalyst discovery. *arXiv preprint arXiv:2209.12466*, 2022.