

Machine Learned Coarse-Grained Protein Force-Fields: Are We There Yet?

Aleksander E. P. Durumeric^a, Nicholas E. Charron^{b,c,e}, Clark Templeton^c, Félix Musil^c, Klara Bonneau^c, Aldo S. Pasos-Trejo^c, Yaoyi Chen^a, Atharva Kelkar^a, Frank Noé^{d,a,c,f}, Cecilia Clementi^{c,e,f,b}

^a*Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 12, Berlin, 14195, Berlin, Germany*

^b*Department of Physics and Astronomy, Rice University, 6100 Main Street, Houston, 77005, Texas, USA*

^c*Department of Physics, Freie Universität Berlin, Arnimallee 12, Berlin, 14195, Berlin, Germany*

^dMicrosoft Research AI4Science, Karl-Liebknecht Str. 32, Berlin, 10178, Berlin, Germany

^eCenter for Theoretical Biological Physics, Rice University, 6100 Main Street, Houston, 77005, Texas, USA

*f*Department of Chemistry, Rice University, 6100 Main Street, Houston, 77005, Texas, USA

26 Abstract

The successful recent application of machine learning methods to scientific problems includes the learning of flexible and accurate atomic-level force-fields for materials and biomolecules from quantum chemical data. In parallel, the machine learning of force-fields at coarser resolutions is rapidly gaining relevance, as an efficient way to represent the higher-body interactions needed in coarse-grained force-fields to compensate for the omitted degrees of freedom. Coarse-grained models are important for the study of systems at time and length scales exceeding those of atomistic simulations. However, the development of transferable coarse-grained models via machine learning still presents significant challenges. Here we discuss recent developments in this field and current efforts to address the remaining challenges.

²⁷ *Keywords:* Protein Dynamics, Machine Learning, Coarse-Graining

28 **1. Introduction**

29 The definition of simplified models is central to physical sciences; proteins
30 are no exception [1, 2]. Statistical mechanical approaches to describe pro-
31 tein folding and dynamics [3, 4, 5], as well as the analysis of long molecular
32 dynamics (MD) trajectories [6, 7, 8], have demonstrated that slow processes
33 in large biomolecular systems can be described by a reduced number of vari-
34 ables despite hundreds of thousands of atoms comprising the full system. In
35 this spirit, many coarse-grained (CG) models have been proposed to study
36 proteins through MD and energy minimization. These CG models have been
37 used to investigate the principles underlying protein folding [9, 10, 11, 12], in-
38 termolecular binding/interactions [13, 14], protein-mediated membrane phe-
39 nomena [15, 16], and to make predictions about novel biological systems of
40 immediate medical interest [17, 18].

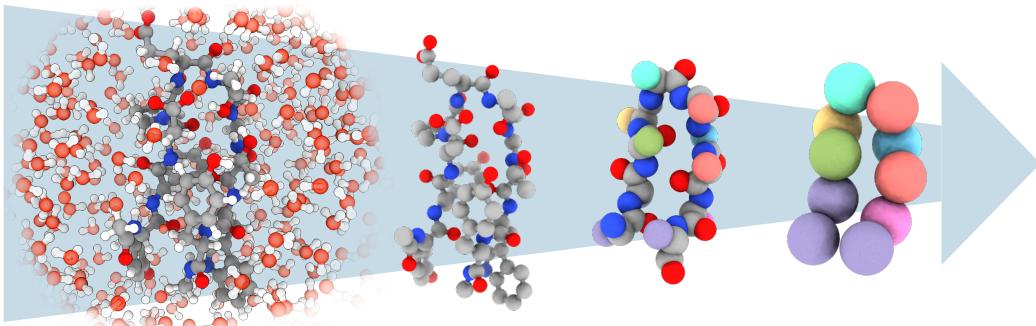


Figure 1: Sequential reduction in resolution of a variant of the miniprotein Chignolin (CLN025) from a solvated all-atom representation containing many thousands of atoms, to an implicit solvent representation, to a heavy-backbone representation with C_β beads, and finally to a C_α CG representation containing 10 beads.

41 Despite their successes, CG models of proteins have not yet achieved
42 the predictive performance of their atomistic counterparts. CG models are
43 primarily designed by specifying their resolution, which defines the coarse
44 degrees of freedom (referred to as “sites” or “beads”, see Fig. 1), and by
45 their effective energy function, which dictates how these beads interact. Tra-
46 ditionally, the resolution is first chosen using either chemical intuition or
47 through optimization designed to reproduce chosen properties (*e.g.*, [19]).
48 The model’s effective energy function is then parameterized to reproduce
49 experimental or simulation data. The fundamental goal of the *transferable*

CG models discussed in this article is to predict the conformational landscape of proteins not used for their parameterization, ideally using only the primary structure of the proteins of interest. Atomistic models have been able to explore the relevant landscape of small globular proteins [20, 21, 22]; however, it is still an open question as to whether there exists a resolution at which a chemically transferable CG model can quantitatively describe the configurational landscape of arbitrary proteins.

A transferable CG protein model would have significant consequences. By employing special-purpose supercomputers [23] or distributed simulation combined with Markov State Models (MSMs)[24, 21, 25], the dynamics of small solvated proteins can be simulated over millisecond timescales [20]. However, biological phenomena routinely involve larger complexes and span longer timescales (seconds or more). CG models promise to reach such scales by reducing the computational cost via decreasing the number of degrees of freedom and increasing the effective simulation timestep. This increased efficiency would vastly improve the use of MD for both fundamental research and applications, *e.g.*, in protein design.

There has been a surge of interest in using machine learning (ML) methods for molecular simulation [26], including learning of CG models from large amounts of data. In a sense, the development of ML CG models can be seen as an extension of ongoing research on the design of accurate atomistic force-fields from quantum mechanical calculations. In this area, ML has already produced highly accurate force-fields which have facilitated groundbreaking computational studies [27, 28, 29]. When combined with the field of bottom-up CG [30, 31], these approaches provide a seemingly clear strategy to leverage ML to learn a CG force-field from existing atomistic MD trajectories. Indeed, thanks to the flexibility of ML algorithms, some frameworks developed for the atomistic resolution [32, 33] have been transferred to the CG resolution [34, 35, 36, 37].

Despite these advances, a completely bottom-up transferable CG model still does not exist for proteins or other biopolymers. This limited progress is due to multiple outstanding challenges, which together firmly differentiate the creation of ML bottom-up CG force-fields from their atomistic counterparts. We here discuss these difficulties and current efforts to overcome them.

85 **2. Thermodynamic Consistency: Why is it difficult?**

86 Bottom-up coarse-graining typically models the following free-energy sur-
87 face (U) [30, 26, 31] referred to as the effective CG (free) energy:

$$U(\mathbf{R}) = -\beta^{-1} \ln \int \delta [\mathbf{R} - \mathcal{M}(\mathbf{r})] \exp [-\beta u(\mathbf{r})] d\mathbf{r}, \quad (1)$$

88 where \mathcal{M} maps all-atom configurations $\mathbf{r} \in \mathbb{R}^{3n}$ to their CG counterparts
89 $\mathbf{R} = \mathcal{M}(\mathbf{r}) \in \mathbb{R}^{3N}$, u is the reference all-atom energy, and β is the inverse
90 temperature. Intuitively, \mathcal{M} defines the CG resolution and U defines how
91 particles at this resolution interact; the design of a bottom-up CG model
92 then entails defining \mathcal{M} and approximating U , and the two tasks are in-
93 terdependent. A CG energy that, up to a constant, equals U is said to
94 be *thermodynamically consistent* with the atomistic counterpart. Such a U
95 produces free energy landscapes identical to the reference in any reaction co-
96 ordinates that are a function of the CG coordinates. We note that the phrase
97 *thermodynamically consistent* here does not refer to thermodynamic observ-
98 ables (*e.g.*, pressure), but instead considers the configurational distributions
99 of the CG and atomistic force-fields. For information on thermodynamic
100 properties in CG models we defer to recent articles [38, 39, 40, 31].

101 Although the thermodynamically consistent CG energy is uniquely de-
102 fined by Eq. (1), the integral cannot be solved for non-trivial systems [30].
103 As a result, multiple strategies [41, 30, 31] to approximate U have been pro-
104 posed. Traditionally, the functional forms for CG (free) energy functions
105 have been low body-order with physically motivated terms [30, 31]. How-
106 ever, recent studies have employed higher body-order terms parameterized
107 by neural networks with success [42, 43, 44, 36, 35]. Kernel methods have
108 also been proposed (*e.g.*, [34]) but have not been applied to proteins. While
109 exceptions exist [42, 45, 46, 47], for reasons of computational efficiency exist-
110 ing ML CG models [44, 36, 26] have been primarily based on the Multiscale
111 Coarse-Graining [48] (“force-matching”) variational framework, primarily due
112 to the fact that it does not require the CG model to be simulated during its
113 parameterization.

114 In principle, Eq. (1) suggests that once the CG resolution has been chosen
115 the creation of the CG model should be straightforward. However, designing
116 an accurate ML CG model is not trivial (see Fig. 2). The choices of reference
117 atomistic system (u), of the resolution (\mathcal{M}), and of the different terms of U
118 all entail challenges unique to models designed at a CG resolution. These

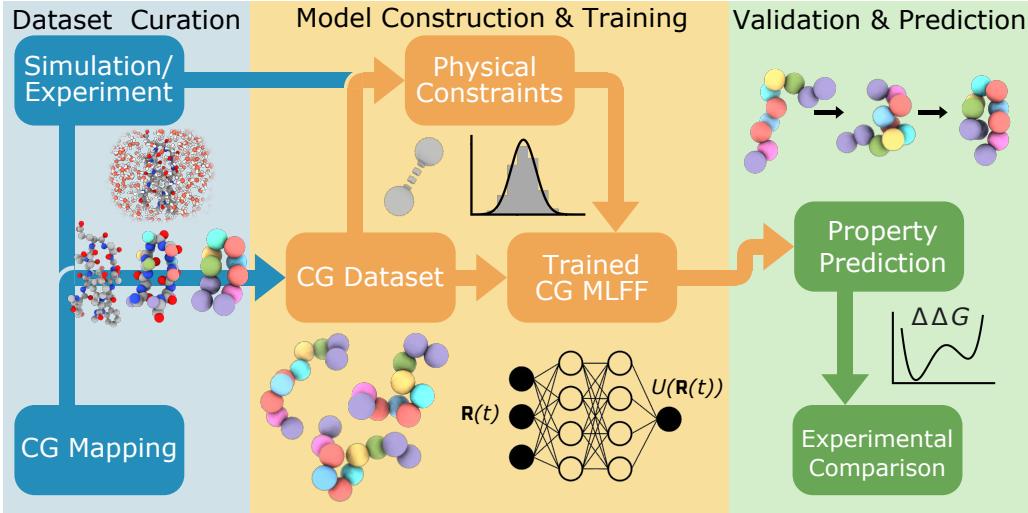


Figure 2: A pipeline for creating and using ML CG models from atomistic simulation data and experimental measurements. A chosen CG mapping can reduce reference information into a CG dataset that can be used to train ML CG models. This training can rely on both simulation and experimental observables in order to reduce the complexity of the learning task and respect physical constraints. A trained ML CG model can then be validated through CG MD and used for general property predictions.

119 challenges are compounded by difficulties with validation, which are also
 120 present in the development of ML atomistic force-fields from quantum chem-
 121 ical data. We address these challenges in detail: beginning with the data
 122 used for training, continuing by discussing the design of U and subsequently
 123 the resolution of the CG model, and finishing by discussing validation and
 124 robustness. For brevity we only discuss algorithms which are applicable to
 125 ML CG force-fields; for a more comprehensive introduction to bottom-up
 126 coarse-graining we refer readers to recent perspective articles [30, 31].

127 *2.1. The difficulty in training CG force-fields*

128 The principal challenge in bottom-up coarse-graining with machine-learned
 129 force-fields lies in finding a suitable ML formulation that directly or indirectly
 130 estimates the intractable integral described in Eq. (1). The situation is more
 131 difficult compared to learning atomistic potential energy surfaces from quan-
 132 tum mechanical data, where reference energies and forces are known: When

133 learning a CG free energy, neither U nor its gradient for a given CG structure
134 \mathbf{R} are known because the integral Eq. (1) is intractable.

135 The most straightforward approach to Eq. (1) is to directly estimate the
136 behavior of the probability density proportional to $\exp(-\beta U(\mathbf{R}))$ from simu-
137 lation data. This requires an equilibrium sample of atomistic conformations
138 \mathbf{r} , *e.g.*, obtained by MD simulations. After mapping them to the CG res-
139 olution, a ML model is then trained to approximate $U(\mathbf{R})$, by minimizing
140 the Kullback–Leibler divergence between the coarse-grained and atomistic
141 probability densities. This is called relative entropy minimization in the
142 coarse-graining literature [49, 50, 47] and maximum likelihood estimation in
143 the ML energy-based model community [51]. Similar approaches [42, 45, 46]
144 which estimate and reduce the difference between a CG force-field and U or
145 optimize selected observables in turn expand on other approaches from the
146 ML community (*e.g.*, [52]).

147 The difficulty with most of these approaches is that the CG model must
148 be periodically re-simulated during training in order to evaluate the equilib-
149 rium density generated by the current model of $U(\mathbf{R})$. While this is feasible
150 for quickly equilibrating CG models, such as those of liquids [47], it is ex-
151 tremely challenging for models that exhibit rare events, such as realistic CG
152 models of protein folding. This limitation is even more problematic for com-
153 plex parameterizations of U (*e.g.*, neural networks) and significantly impedes
154 simultaneously training over multiple molecules when creating transferable
155 models. Approaches have attempted to reduce this burden by, for example,
156 reweighting the density of previous iterations [50, 45, 53] or by modifying the
157 sampling of the atomistic system [46]. However, these approaches have not
158 yet reached the simplicity and applicability of non-iterative approaches.

159 The most common bottom-up approach for approximating U is force-
160 matching [48], which fits a CG free energy such that its negative gradient
161 matches projected instantaneous atomistic forces on average. Critically, this
162 does not require simulations of the CG model during training, and was pro-
163 posed for ML CG protein force-fields in [44]. As many atomistic coordinates
164 \mathbf{r} map to the same CG coordinate \mathbf{R} , the instantaneous force is noisy, and the
165 signal-to-noise ratio becomes smaller the more degrees of freedom are “CGed
166 away”; thus CG force-matching requires more data as compared to atomistic
167 force-fields. A second difficulty comes from the fact that U is obtained by
168 implicitly integrating the mean force, and as a result, obtaining the free en-
169 ergy difference between minima depends on estimates of the forces along the
170 transition path, where the uncertainties are the largest.

171 The recently proposed flow-matching method [54] combines relative en-
172 tropy estimation and force-matching by employing generative deep learn-
173 ing: the CG density is estimated by a normalizing flow, a neural network
174 that can generate one-shot samples of equilibrium CG conformations. This
175 flow can then generate samples to train a downstream ML CG force-field
176 by force-matching. The limitation of this approach still lies in finding flow
177 architectures that can scale to large macromolecules.

178 The distribution of atomistic configurations is fundamental to the dis-
179 cussed algorithms. Rare events are important but infrequently sampled in
180 the canonical distribution; directing atomistic sampling towards barriers and
181 areas of “high uncertainty” may be beneficial. While ML models are expres-
182 sive than, *e.g.*, pair potentials, they require more data. For example, ML
183 CG force-matching may use upwards of one million canonically distributed
184 samples covering the configurational space for small proteins [36], in con-
185 trast to harmonic models parameterized using short trajectories in the folded
186 state [55]. Modifying the distribution of samples may reduce data require-
187 ments [56], but it is unclear how such approaches scale with system complex-
188 ity.

189 Concurrently, iterative methods may overcome their computational bar-
190 riers if non-canonical sampling is used; expanding discriminative training
191 may remove the need for repeated training simulations [46], and biasing po-
192 tentials may promote diversity and produce more accurate parameters [57].
193 However, these approaches often require data to be drawn from a modi-
194 fied distribution, which impedes the use of preexisting atomistic trajectories.
195 Nevertheless, these approaches will be critical to expanding current ML CG
196 success to multi-domain proteins.

197 For a transferable ML CG model, more requirements for the training
198 dataset arise. It is straightforward, and important, to simultaneously force-
199 match a model using reference data from multiple proteins as evidence has
200 shown that extended ensembles can act as regularization [58]. Previous pi-
201 oneering work developing bottom-up transferable CG models used this ap-
202 proach, but fell short of unassisted folding and relied on artificially lower-
203 ing simulation temperatures to stabilize states of interest [59]; we associate
204 these inaccuracies to limitations of the force-field basis and training set. We
205 anticipate that the proportion of structural motifs in the dataset plays an
206 important role. In the ideal case, a general CG model of proteins would likely
207 include globular, fibrous, and intrinsically disordered proteins in its training
208 procedure. Such a transferable training setup naturally expands the amount

209 of atomistic data available to train a given model; whether this will improve
210 predictions on individual proteins remains to be seen.

211 *2.2. Choice of the coarse-grained representation*

212 In the design of an atomistic force-field, the Born-Oppenheimer approx-
213 imation justifies the separation between electronic and nuclear degrees of
214 freedom and provides the framework for effective nuclear potential energy
215 surfaces. However, the separation of scales is less clear for CG models. Con-
216 sequently, the selection of the CG resolution (\mathcal{M} from Eq. (1)) is non-trivial
217 and influences the free energy surface that must be learned. The fundamen-
218 tal questions in this area are which resolutions are “easy to learn” and which
219 are conducive to creating transferable models. These points highlight the
220 challenges of validating ML CG models that are capable of extrapolating to
221 unseen systems. For certain resolutions, it may be easier to learn an effec-
222 tive CG energy and extrapolate into unknown regions of phase space. On
223 the other hand, certain resolutions may be conducive to accurate ML CG
224 models but may be difficult to interpret.

225 Current successful CG ML protein applications [42, 44, 36, 46, 54] typi-
226 cally focus on a single site per resolution resolution (typically C_α); however,
227 this appears to be mostly due to simplicity and not systematic validation.
228 “Optimal” resolutions have been studied [60, 61], but it is unclear how they
229 impact ML CG models. Back-mapping, *i.e.*, reconstructing details from CG
230 models, is a current area of investigation [62] and may alleviate interpretabil-
231 ity constraints on the CG resolution. Views which link back-mapping with
232 potential optimization can facilitate a joint optimization of the representa-
233 tion alongside the CG energy model [63]; however, these approaches do not
234 yet in themselves search for transferable resolutions.

235 *2.3. Functional Form of the Many-body Effective CG (Free) Energy*

236 In practice, training ML CG models via force-matching from equilibrium
237 data requires a baseline (or “prior”) potential to reduce catastrophically in-
238 correct extrapolation in unphysical regions of phase space [44, 36, 64]. Ulti-
239 mately, a good prior potential incorporates physical principles, reduces learn-
240 ing complexity, and allows for stable simulation. Similar to Δ -learning [65]
241 for atomistic force-fields [66, 67], the CG energy is usually decomposed into:

$$U(\mathbf{R}; \boldsymbol{\theta}) = U_{prior}(\mathbf{R}) + U_{net}(\mathbf{R}; \boldsymbol{\theta}) \quad (2)$$

242 where $U_{net}(\mathbf{R}; \boldsymbol{\theta})$ is a trainable multibody potential expressed by an ML
243 model with parameters $\boldsymbol{\theta}$ and $U_{prior}(\mathbf{R})$ is the prior energy.

244 Designing the priors is non-trivial as it depends on an interplay between
245 the CG mapping, the ML architecture, and the training data. Poor choices
246 of priors can significantly reduce the performance of an ML force-field [47,
247 68]. Currently, the prevailing strategy involves proposing a prior inspired
248 by the low body-order terms from classical force-fields, and then iteratively
249 developing an ML CG model over both the prior terms and hyperparameters
250 [44, 36, 35, 37, 47]. Systematic strategies have yet to be developed to design
251 prior energies that are transferable to different molecules.

252 While priors help enforce important physical asymptotic interactions, the
253 ML model architecture itself should respect basic physical constraints. These
254 include invariance with respect to permutations of particles of the same type,
255 invariance to translations and rotations of the reference frame, and curl-free
256 force predictions [69, 44]. A way to allow the learnable energy U_{net} from
257 Eq. (2) to be transferable [36] is to decompose it bead-wise such that:

$$U_{net}(\mathbf{R}, \mathbf{a}) = \sum_i^N u_{net}(\mathbf{R} - \mathbf{R}_i, a_i), \quad (3)$$

258 where \mathbf{R}_i is the i^{th} bead (with type a_i) in the configuration \mathbf{R} so that $\mathbf{R} - \mathbf{R}_i$
259 are the relative displacements of all beads around bead \mathbf{R}_i , and u_{net} is the
260 bead-wise contribution to the potential.

261 On top of these constraints, bottom-up coarse-graining involves additional
262 architectural challenges. Coarse-graining a variety of different groups
263 of atoms leads to a large number of CG bead types, *e.g.*, at least 20 types
264 (one for each amino acid) for proteins at the C_α resolution. As mentioned
265 above, training with noisy forces requires a large number of training config-
266 urations. As a result, the ML approach must accommodate large training
267 sets (indicating that neural networks may be preferable over kernel methods)
268 and should not scale with the number of bead types so that evaluation times
269 do not increase when considering transferable models. This constraint favors
270 the use of deep learning architectures like SchNet [33] over models based on
271 fixed representation, *e.g.*, symmetry functions [70, 71].

272 2.4. Validation and Robustness

273 While atomistic ML force-field development has matured, there exists no
274 appropriate set of best practices for probing stability and robustness. It is

275 common to assess atomistic model accuracy with pointwise metrics, such as
276 the mean force error, over fixed test datasets[27]. However, without an un-
277 derstanding of how models extrapolate into data-poor regions, these metrics
278 can not be used as indicators of simulation stability or accuracy [72, 64], as
279 simulations may explore uncovered configurations. For ML CG models, even
280 with the use of prior energy terms, force error does not guarantee a stable
281 model [68].

282 Due to the difficulty in constructing comprehensive test sets, the robust-
283 ness and accuracy of a trained ML model can only be ascertained through
284 extensive sampling, *e.g.*, by using the model to run long MD simulations. Re-
285 cent investigations into ML architectures have revealed the need for such met-
286 rics for both atomistic and CG ML models [64, 68]. Unfortunately, obtaining
287 a converged CG MD simulation can require several million force evaluations;
288 for large systems and complex architectures this may present a computational
289 bottleneck [73]. Validation difficulties impede hyperparameter optimization
290 (*e.g.*, regularization strength, cutoff, or prior potential), as searches may be-
291 come prohibitively expensive. We note, however, that existing applications
292 provide suitable initial choices of hyperparameters for select architectures
293 and resolutions (see [36]), but that the introduction of novel architectures
294 naturally requires substantial effort for the initial hyperparameter search.

295 Even once MD has been used to characterize a ML CG model, validation
296 still poses difficulties. When characterizing atomistic force-fields on selected
297 configurations it is typically possible to compare the model’s energy and
298 force predictions to reference values; unfortunately, these are not available
299 at the CG resolution (Eq. (1)). Instead, analysis typically projects CG con-
300 figurations onto low-dimensional collective variables (*e.g.*, Fig. 3). However,
301 as ML CG models are now able to reproduce such collective variable sur-
302 faces, the need for more rigorous validation is emerging. Recent work [75]
303 has proposed classification as an approach to generate energy-like errors for
304 CG models and may provide an avenue for connecting atomistic and CG
305 force-field validation.

306 A related challenge is presented by model uncertainty: How robust is
307 a ML model to different training seeds or data partitioning strategies? For
308 neural networks, these can be expensive questions to answer. However, recent
309 advances have started to enable estimates of uncertainty [76, 77]. A promising
310 strategy involves estimating the uncertainty of predictions and minimizing it
311 either before or during model deployment, either through iterative training
312 or through “on-the-fly” frameworks [78] where data is added to the training

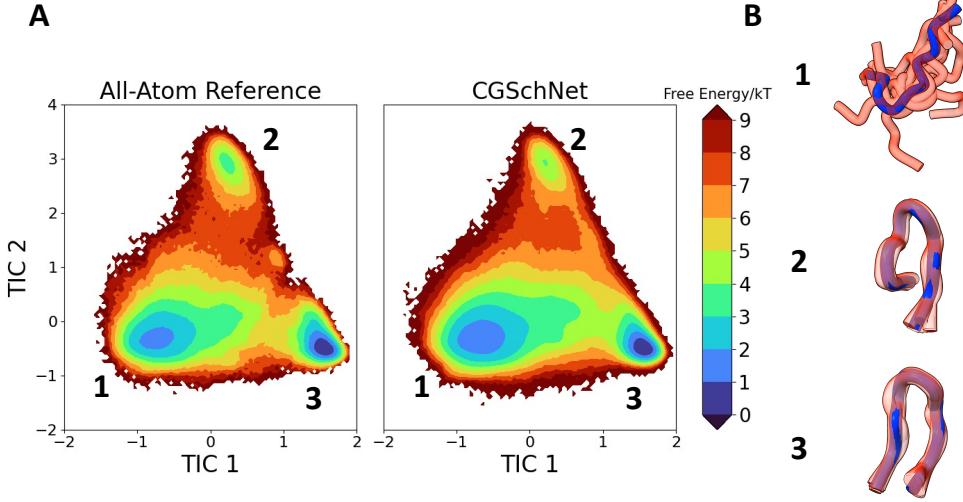


Figure 3: State-of-the-art performance for a C_α CG ML model on the benchmark protein CLN025. A) Comparison of the CG free energy landscape of CLN025 (produced using MD) for a learned CG ML model with the corresponding free energy for the reference all-atom dataset projected onto slow degrees of freedom (TICA) [74]. B) Ensembles of structures sampled from the CG ML model MD simulation (in red) are superimposed onto all-atom reference structure counterparts (in blue). Basin 1 represents the unfolded state, basin 2 the misfolded state, and basin 3 the folded state.

313 set based on such estimates.

314 3. Conclusion

315 At the moment of writing, state-of-the-art ML CG models can quantita-
316 tively reproduce the behavior of small proteins, as shown in Fig. 3 for Chig-
317 nolin and in Ref. [54] for Trpcage, BBA, and Villin. Currently, the largest
318 barrier to describe larger proteins is gathering sufficient training data. To
319 what extent such an approach can be extended to define *transferable* CG
320 models remains an open question. It may be possible only for a class of pro-
321 teins, or at particular resolutions. Before the advent of ML methods, these
322 questions remained challenging to answer, as thermodynamic consistency be-
323 tween an atomistic and a CG model (Eq. (1)) could only be approximately
324 enforced; it was not clear whether the limitations of transferable models
325 [59, 79, 80, 81, 39] were due to the limited expressivity of the CG energy and
326 limited reference data or to more fundamental problems with transferability.

327 Now, as ML CG models can quantitatively enforce thermodynamic consistency
328 for single proteins (as shown in Fig. 3), we have the tools to address
329 these questions and explore the trade-off between accuracy and transferabil-
330 ity. Here, we have discussed the practical challenges towards this goal, but
331 we remain optimistic that such a line of research can be pursued.

332 Even if a transferable bottom-up ML CG model can be defined, even-
333 tually, the success of a computational model relies on its comparison to ex-
334 periments. Bottom-up CG models rely on the reference atomistic models
335 and necessarily inherit their inaccuracies and flaws. With the improvements
336 in atomistic force-fields, we expect CG models to also become more accu-
337 rate. However, even small inconsistencies between the CG and atomistic
338 models may compound into a significant deviation from experimental data.
339 We believe that ultimately bottom-up ML CG will need to be merged with
340 top-down models for their useful and predictive applications.

341 **4. Conflict of interest statement**

342 Nothing declared.

343 **5. Acknowledgments**

344 Molecular visualizations for the graphical abstract and all in-text figures
345 were produced using UCSF ChimeraX, a software developed by the Resource
346 for Biocomputing, Visualization, and Informatics at the University of Califor-
347 nia, San Francisco [82]. The free energy surfaces in Fig. 3 were produced using
348 the Python package Matplotlib [83]. We acknowledge the financial support of
349 Deutsche Forschungsgemeinschaft DFG (SFB/TRR 186, Project A12; SFB
350 1114, Projects B03 and A04; SFB 1078, Project C7; and RTG 2433, Project
351 Q04 and Q05), the National Science Foundation NSF (CHE-1900374 and
352 PHY-2019745), the NLM Training Program in Biomedical Informatics and
353 Data Science (Grant No. 5T15LM007093-27), and the Einstein Foundation
354 Berlin (Project 0420815101). We also thank the members of the Clementi
355 and Noé groups for their helpful discussions.

356 F.M. acknowledges support from the SNSF under the Postdoc.Mobility
357 fellowship P500PT_203124.

358 **References**

- 359 [1] M. Levitt, A. Warshel, Computer simulation of protein folding, *Nature*
360 253 (5494) (1975) 694–698.
- 361 [2] C. Clementi, Coarse-grained models of protein folding: toy models or
362 predictive tools?, *Curr. Opin. Struct. Biol.* 18 (1) (2008) 10–15.
- 363 [3] J. D. Bryngelson, P. G. Wolynes, Spin glasses and the statistical me-
364 chanics of protein folding., *Proc. Natl. Acad. Sci. USA* 84 (21) (1987)
365 7524–7528.
- 366 [4] J. N. Onuchic, Z. Luthey-Schulten, P. G. Wolynes, Theory of Protein
367 Folding: The energy landscape perspective, *Annu. Rev. Phys. Chem.*
368 48 (1) (1997) 545–600.
- 369 [5] K. A. Dill, S. Bromberg, K. Yue, H. S. Chan, K. M. Ftebig, D. P.
370 Yee, P. D. Thomas, Principles of protein folding — a perspective
371 from simple exact models, *Protein Science* 4 (4) (1995) 561–602.
372 doi:10.1002/pro.5560040401.
- 373 [6] R. B. Best, G. Hummer, Reaction coordinates and rates from transition
374 paths, *Proc. Natl. Acad. Sci. USA* 102 (19) (2005) 6732–6737.
- 375 [7] J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill, W. C. Swope,
376 Automatic discovery of metastable states for the construction of markov
377 models of macromolecular conformational dynamics, *J. Chem. Phys.*
378 126 (15) (2007) 155101.
- 379 [8] F. Noé, C. Clementi, Collective variables for the study of long-time
380 kinetics from molecular trajectories: theory and methods, *Curr. Opin.*
381 *Struct. Biol.* 43 (2017) 141–147.
- 382 [9] C. Clementi, H. Nymeyer, J. N. Onuchic, Topological and energetic
383 factors: what determines the structural details of the transition state
384 ensemble and “en-route” intermediates for protein folding? an investi-
385 gation for small globular proteins, *J. Mol. Biol.* 298 (5) (2000) 937–953.
- 386 [10] A. Liwo, M. Khalili, H. A. Scheraga, Ab initio simulations of protein-
387 folding pathways by molecular dynamics with the united-residue model
388 of polypeptide chains, *Proc. Natl. Acad. Sci. USA* 102 (7) (2005)
389 2362–2367.

- 390 [11] A. Davtyan, N. P. Schafer, W. Zheng, C. Clementi, P. G. Wolynes, G. A.
391 Papoian, Awsem-md: Protein structure prediction using coarse-grained
392 physical potentials and bioinformatically based local structure biasing,
393 The Journal of Physical Chemistry B 116 (29) (2012) 8494–8503, pMID:
394 22545654. doi:10.1021/jp212541y.
- 395 [12] T. Bereau, M. Deserno, Generic coarse-grained model for protein fold-
396 ing and aggregation, The Journal of chemical physics 130 (23) (2009)
397 06B621.
- 398 [13] P. C. T. Souza, S. Thallmair, P. Conflitti, C. Ramírez-Palacios,
399 R. Alessandri, S. Raniolo, V. Limongelli, S. J. Marrink, Protein–ligand
400 binding with the coarse-grained martini model, Nature Communications
401 11 (11) (2020) 3714. doi:10.1038/s41467-020-17437-5.
- 402 [14] J. Roel-Touris, C. G. Don, R. V. Honorato, J. P. G. L. M. Rodrigues,
403 A. M. J. J. Bonvin, Less is more: Coarse-grained integrative modeling
404 of large biomolecular assemblies with HADDOCK, J. Chem. Theory
405 Comput. 15 (11) (2019) 6358–6367. doi:10.1021/acs.jctc.9b00310.
- 406 [15] M. Louhivuori, H. J. Risselada, E. van der Giessen, S. J. Marrink, Re-
407 lease of content through mechano-sensitive gates in pressurized lipo-
408 somes, Proc. Natl. Acad. Sci. USA 107 (46) (2010) 19856–19860.
- 409 [16] K. M. Davies, C. Anselmi, I. Wittig, J. D. Faraldo-Gómez,
410 W. Kühlbrandt, Structure of the yeast F1Fo-ATP synthase dimer and
411 its role in shaping the mitochondrial cristae, Proc. Natl. Acad. Sci. USA
412 109 (34) (2012) 13602–13607.
- 413 [17] W. Zheng, M.-Y. Tsai, M. Chen, P. G. Wolynes, Exploring the aggrega-
414 tion free energy landscape of the amyloid- β protein (1–40), Proc. Natl.
415 Acad. Sci. USA 113 (42) (2016) 11835–11840.
- 416 [18] A. J. Pak, A. Yu, Z. Ke, J. A. G. Briggs, G. A. Voth, Cooperative mul-
417 tivalent receptor binding promotes exposure of the SARS-CoV-2 fusion
418 machinery core, Nat. Commun. 13 (1) (2022).
- 419 [19] M. Giulini, R. Menichetti, M. S. Shell, R. Potestio, An information-
420 theory-based approach for optimal model reduction of biomolecules, J.
421 Chem. Theory. Comput. 16 (11) (2020) 6795–6813.

- 422 [20] K. Lindorff-Larsen, S. Piana, R. O. Dror, D. E. Shaw, How fast-folding
423 proteins fold, *Science* 334 (6055) (2011) 517–520.
- 424 [21] N. Plattner, S. Doerr, G. De Fabritiis, F. Noé, Complete protein–protein
425 association kinetics in atomic detail revealed by molecular dynamics
426 simulations and markov modelling, *Nat. Chem.* 9 (10) (2017) 1005–1011.
- 427 [22] S. Bottaro, K. Lindorff-Larsen, Biophysical experiments and biomolec-
428 ular simulations: A perfect match?, *Science* 361 (6400) (2018) 355–360.
- 429 [23] D. E. Shaw, J. Grossman, J. A. Bank, B. Batson, J. A. Butts, J. C. Chao,
430 M. M. Deneroff, R. O. Dror, A. Even, C. H. Fenton, et al., Anton 2: rais-
431 ing the bar for performance and programmability in a special-purpose
432 molecular dynamics supercomputer, in: SC’14: Proceedings of the Inter-
433 national Conference for High Performance Computing, Networking,
434 Storage and Analysis, IEEE, 2014, pp. 41–53.
- 435 [24] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D.
436 Chodera, C. Schütte, F. Noé, Markov models of molecular kinetics: Gen-
437 eration and validation, *J. Chem. Phys.* 134 (17) (2011) 174105.
- 438 [25] B. E. Husic, V. S. Pande, Markov state models: From an art to a science,
439 *J. Am. Chem. Soc.* 140 (7) (2018) 2386–2396.
- 440 [26] F. Noé, A. Tkatchenko, K.-R. Müller, C. Clementi, Machine learning for
441 molecular simulation, *Annu. Rev. Phys. Chem.* 71 (2020) 361–390.
- 442 [27] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky,
443 K. T. Schütt, A. Tkatchenko, K.-R. Müller, Machine Learning Force
444 Fields, *Chem. Rev.* 121 (16) (2021) 10142–10186.
- 445 [28] V. Kapil, C. Schran, A. Zen, J. Chen, C. J. Pickard, A. Michaelides, The
446 first-principles phase diagram of monolayer nanoconfined water, *Nature*
447 609 (7927) (2022) 512–516.
- 448 [29] L. Gigli, M. Veit, M. Kotiuga, G. Pizzi, N. Marzari, M. Ceriotti, Ther-
449 modynamics and dielectric response of BaTiO₃ by data-driven modeling,
450 *npj Comput. Mater.* 8 (1) (2022) 1–17.
- 451 [30] W. G. Noid, Perspective: Coarse-grained models for biomolecular sys-
452 tems, *J. Chem. Phys.* 139 (9) (2013) 090901.

- 453 [31] J. Jin, A. J. Pak, A. E. Durumeric, T. D. Loose, G. A. Voth, Bottom-up
454 Coarse-Graining: Principles and Perspectives, *J. Chem. Theory Comput.* (2022).
- 455
- 456 [32] A. P. Bartók, M. C. Payne, R. Kondor, G. Csányi, Gaussian approxi-
457 mation potentials: The accuracy of quantum mechanics, without the
458 electrons, *Phys. Rev. Lett.* 104 (13) (2010) 136403.
- 459
- 460 [33] K. T. Schütt, H. E. Sauceda, P. J. Kindermans, A. Tkatchenko, K. R.
461 Müller, SchNet - A deep learning architecture for molecules and mate-
462 rials, *J. Chem. Phys.* 148 (24) (2018) 241722.
- 463
- 464 [34] S. T. John, G. Csányi, Many-Body Coarse-Grained Interactions Using
465 Gaussian Approximation Potentials, *J. Phys. Chem. B* 121 (48) (2017)
466 10934–10949.
- 467
- 468 [35] J. Wang, N. Charron, B. Husic, S. Olsson, F. Noé, C. Clementi, Multi-
469 body effects in a coarse-grained protein force field, *J. Chem. Phys.*
470 154 (16) (2021) 164113.
- 471
- 472 [36] B. E. Husic, N. E. Charron, D. Lemm, J. Wang, A. Pérez, M. Majewski,
473 A. Krämer, Y. Chen, S. Olsson, G. de Fabritiis, F. Noé, C. Clementi,
474 Coarse graining molecular dynamics with graph neural networks, *J.
475 Chem. Phys.* 153 (194101) (2020).
- 476
- 477 [37] Y. Chen, A. Krämer, N. E. Charron, B. E. Husic, C. Clementi, F. Noé,
478 Machine learning implicit solvation for molecular dynamics, *J. Chem.
479 Phys.* 155 (8) (2021) 084101.
- 480
- 481 [38] J. W. Wagner, J. F. Dama, A. E. Durumeric, G. A. Voth, On the repre-
482 sentability problem and the physical meaning of coarse-grained models,
483 *J. Chem. Phys.* 145 (4) (2016) 044108.
- 484
- 485 [39] N. J. Dunn, T. T. Foley, W. G. Noid, Van der Waals perspective on
486 coarse-graining: Progress toward Solving Representability and Trans-
487 ferability Problems, *Acc. Chem. Res.* 49 (12) (2016) 2832–2840.
- 488
- 489 [40] J. Jin, A. J. Pak, G. A. Voth, Understanding missing entropy in coarse-
490 grained systems: Addressing issues of representability and transferabil-
491 ity, *J. Phys. Chem. Lett.* 10 (16) (2019) 4549–4557.
- 492

- 484 [41] G. Tóth, Interactions from diffraction data: historical and comprehensive
485 overview of simulation assisted methods, *J. Phys.: Condens. Matter*
486 19 (33) (2007) 335220.
- 487 [42] T. Lemke, C. Peter, Neural network based prediction of conformational
488 free energies - a new route toward coarse-grained simulation models, *J.*
489 *Chem. Theory Comput.* 13 (12) (2017) 6213–6221.
- 490 [43] L. Zhang, J. Han, H. Wang, R. Car, W. E, DeePCG: Constructing
491 coarse-grained models via deep neural networks, *J. Chem. Phys.* 149 (3)
492 (2018) 034101.
- 493 [44] J. Wang, S. Olsson, C. Wehmeyer, A. Pérez, N. E. Charron, G. de Fabritiis,
494 F. Noé, C. Clementi, Machine learning of coarse-grained molecular
495 dynamics force fields, *ACS Cent. Sci.* (2019).
- 496 [45] S. Thaler, J. Zavadlav, Learning neural network potentials from exper-
497 imental data via Differentiable Trajectory Reweighting, *Nat. Commun.*
498 12 (1) (2021) 6884.
- 499 [46] X. Ding, B. Zhang, Contrastive learning of coarse-grained force fields,
500 *J. Chem. Theory Comput.* 18 (10) (2022) 6334–6344.
- 501 [47] S. Thaler, M. Stupp, J. Zavadlav, Deep Coarse-grained Potentials via
502 Relative Entropy Minimization, *arXiv e-prints* (2022) arXiv:2208.10330.
- 503 [48] W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A.
504 Voth, A. Das, H. C. Andersen, The multiscale coarse-graining method.
505 I. A rigorous bridge between atomistic and coarse-grained models, *J.*
506 *Chem. Phys.* 128 (24) (2008) 244114.
- 507 [49] M. S. Shell, The relative entropy is fundamental to multiscale and inverse
508 thermodynamic problems, *J. Chem. Phys.* 129 (14) (2008) 144108.
- 509 [50] S. P. Carmichael, M. S. Shell, A new multiscale algorithm and its ap-
510 plication to coarse-grained peptide models for self-assembly, *J. Phys.*
511 *Chem. B.* 116 (29) (2012) 8383–8393.
- 512 [51] G. E. Hinton, Training products of experts by minimizing contrastive
513 divergence, *Neural Comput.* 14 (8) (2002) 1771–1800.

- 514 [52] M. Gutmann, A. Hyvärinen, Noise-contrastive estimation: A new es-
515 timation principle for unnormalized statistical models, in: Proceedings
516 of the thirteenth international conference on artificial intelligence and
517 statistics, JMLR Workshop and Conference Proceedings, 2010, pp. 297–
518 304.
- 519 [53] M. Wieder, J. Fass, J. D. Chodera, Teaching free energy calculations to
520 learn from experimental data, bioRxiv 2021.08.24.457513 (2021).
- 521 [54] J. Köhler, Y. Chen, A. Krämer, C. Clementi, F. Noé, Flow-
522 matching – efficient coarse-graining molecular dynamics without forces,
523 arXiv:2203.11167 (2022).
- 524 [55] E. Lyman, J. Pfaendtner, G. A. Voth, Systematic multiscale parameter-
525 ization of heterogeneous elastic network models of proteins, Biophys. J.
526 95 (9) (2008) 4183–4192.
- 527 [56] E. V. Podryabinkin, E. V. Tikhonov, A. V. Shapeev, A. R. Oganov, Ac-
528 celerating crystal structure prediction by machine-learning interatomic
529 potentials with active learning, Phys. Rev. B 99 (6) (2019) 64114.
- 530 [57] K. Shen, N. Sherck, M. Nguyen, B. Yoo, S. Köhler, J. Speros, K. T. De-
531 laney, G. H. Fredrickson, M. S. Shell, Learning composition-transferable
532 coarse-grained models: Designing external potential ensembles to max-
533 imize thermodynamic information, The Journal of Chemical Physics
534 153 (15) (2020) 154116.
- 535 [58] K. H. Kanekal, J. F. Rudzinski, T. Bereau, Broad chemical transfer-
536 ability in structure-based coarse-graining, J. Chem. Phys. 157 (2022)
537 104102.
- 538 [59] R. D. Hills, L. Lu, G. A. Voth, Multiscale coarse-graining of the protein
539 energy landscape, PLoS Comput. Biol. 6 (6) (2010) e1000827.
- 540 [60] W. Wang, R. Gómez-Bombarelli, Coarse-graining auto-encoders for
541 molecular dynamics, npj Comput. Mater. 5 (1) (2019) 125.
- 542 [61] T. T. Foley, K. M. Kidder, M. S. Shell, W. Noid, Exploring the land-
543 scape of model representations, Proc. Natl. Acad. Sci. Unit. States Am.
544 117 (39) (2020) 24061–24068.

- 545 [62] W. Wang, M. Xu, C. Cai, B. K. Miller, T. Smidt, Y. Wang, J. Tang,
546 R. Gómez-Bombarelli, Generative coarse-graining of molecular conformations, arXiv:2201.12176 (2022).
- 548 [63] S. Chennakesavalu, D. J. Toomer, G. M. Rotskoff, Ensuring thermodynamic consistency with invertible coarse-graining, arXiv preprint
549 arXiv:2210.07882 (2022).
- 551 [64] X. Fu, Z. Wu, W. Wang, T. Xie, S. Keten, R. Gomez-Bombarelli,
552 T. Jaakkola, Forces are not enough: Benchmark and critical evaluation
553 for machine learning force fields with molecular simulations, arXiv
554 e-prints (2022) arXiv:2210.07237.
- 555 [65] R. Ramakrishnan, P. O. Dral, M. Rupp, O. A. von Lilienfeld, Big data
556 meets quantum chemistry approximations: The δ -machine learning approach,
557 J. Chem. Theory Comput. 11 (5) (2015) 2087–2096.
- 558 [66] P. E. Dolgirev, I. A. Kruglov, A. R. Oganov, Machine learning scheme for
559 fast extraction of chemically interpretable interatomic potentials, AIP
560 Advances 6 (8) (2016) 085318.
- 561 [67] V. L. Deringer, G. Csányi, Machine learning based interatomic potential
562 for amorphous carbon, Phys. Rev. B 95 (2017) 094203.
- 563 [68] E. Ricci, G. Giannakopoulos, V. Karkaletsis, D. N. Theodorou, N. Vergadou,
564 Developing Machine-Learned Potentials for Coarse-Grained
565 Molecular Simulations: Challenges and Pitfalls, in: Proceedings of the
566 12th Hellenic Conference on Artificial Intelligence, 2022, p. 1–6.
- 567 [69] F. Musil, A. Grisafi, A. P. Bartók, C. Ortner, G. Csányi, M. Ceriotti,
568 Physics-Inspired Structural Representations for Molecules and Materials,
569 Chem. Rev. 121 (16) (2021) 9759–9815.
- 570 [70] J. Behler, M. Parrinello, Generalized neural-network representation of
571 high-dimensional potential-energy surfaces, Phys. Rev. Lett. 98 (14)
572 (2007).
- 573 [71] J. S. Smith, O. Isayev, A. E. Roitberg, ANI-1: an extensible neural
574 network potential with DFT accuracy at force field computational cost,
575 Chem. Sci. 8 (4) (2017) 3192–3203.

- 576 [72] S. Stocker, J. Gasteiger, F. Becker, S. Günnemann, J. T. Margraf, How
577 robust are modern graph neural network potentials in long and hot
578 molecular dynamics simulations?, ChemRxiv (Apr 2022).
- 579 [73] O. T. Unke, M. Stöhr, S. Ganscha, T. Unterthiner, H. Maennel,
580 S. Kashubin, D. Ahlin, M. Gastegger, L. M. Sandonas, A. Tkatchenko,
581 K.-R. Müller, Accurate machine learned quantum-mechanical force fields
582 for biomolecular simulations, arXiv e-prints (2022) arXiv:2205.08306.
- 583 [74] G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, F. Noé, Iden-
584 tification of slow molecular order parameters for markov model construc-
585 tion, The Journal of Chemical Physics 139 (1) (2013) 015102.
- 586 [75] A. E. P. Durumeric, G. A. Voth, Explaining classifiers to understand
587 coarse-grained models, arXiv preprint arXiv:2109.07337 (2021).
- 588 [76] Y. Gal, Z. Ghahramani, Dropout as a Bayesian Approximation: Rep-
589 resenting Model Uncertainty in Deep Learning, in: Proceedings of the
590 33rd International Conference on International Conference on Machine
591 Learning - Volume 48, 2016, p. 1050–1059.
- 592 [77] B. Lakshminarayanan, A. Pritzel, C. Blundell, Simple and Scalable Pre-
593 dictive Uncertainty Estimation Using Deep Ensembles, in: Proceedings
594 of the 31st International Conference on Neural Information Processing
595 Systems, 2017, p. 6405–6416.
- 596 [78] J. Vandermause, S. B. Torrisi, S. Batzner, Y. Xie, L. Sun, A. M. Kolpak,
597 B. Kozinsky, On-the-fly active learning of interpretable Bayesian force
598 fields for atomistic rare events, npj Comput. Mater. 6 (1) (2020) 1–11.
- 599 [79] T. Sanyal, J. Mittal, M. S. Shell, A hybrid, bottom-up, structurally
600 accurate, Go-like coarse-grained protein model, J. Chem. Phys. 151 (4)
601 (2019) 044111.
- 602 [80] T. D. Potter, J. Tasche, M. R. Wilson, Assessing the transferability of
603 common top-down and bottom-up coarse-grained molecular models for
604 molecular mixtures, Phys. Chem. Chem. Phys. 21 (2019) 1912–1927.
- 605 [81] D. Rosenberger, N. F. Van Der Vegt, Addressing the temperature trans-
606 ferability of structure based coarse graining models, Phys. Chem. Chem.
607 Phys. 20 (9) (2018) 6617–6628.

- 608 [82] T. D. Goddard, C. C. Huang, E. C. Meng, E. F. Pettersen, G. S. Couch,
609 J. H. Morris, T. E. Ferrin, UCSF ChimeraX: Meeting modern challenges
610 in visualization and analysis, *Protein Science* 27 (1) (2018) 14–25.
- 611 [83] J. D. Hunter, Matplotlib: A 2D graphics environment, *Computing in
612 Science & Engineering* 9 (3) (2007) 90–95.

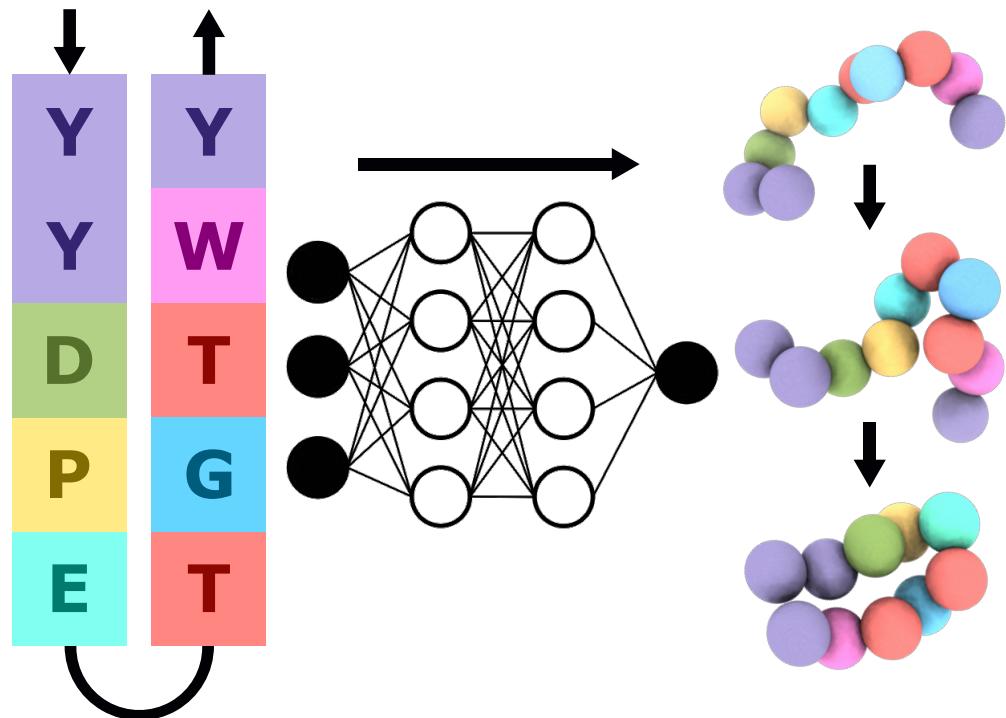
613 *Reference annotations*

- 614 ★ [61] The effect of different choices of resolution and mapping on coarse-
615 graining model systems
- 616 ★ [69] An extensive review of different representations and their properties
617 for the modeling of molecules and materials with machine learning methods.
- 618 ★ [26] A review of the application of machine learning on different aspects
619 of molecular simulation
- 620 ★ [31] A review of the theory, implementation, and open challenges for
621 bottom-up coarse-graining
- 622 ★ [46] An alternative approach for bottom-up coarse-graining of molec-
623 ular systems, that does not require atomistic forces nor expensive sampling
624 iterations.
- 625 ★ [64] The authors show that the force error is not a good metric to
626 evaluate the performance of machine-learned force-fields and introduce a new
627 benchmark suite for ML MD simulation.
- 628 ★ [45] Coarse-graining with machine learning methods; in particular, the
629 parameters of a neural network are tuned to reproduce selected observables
630 using iterative simulations.
- 631 ★ [36] A graph neural network is used to model the effective energy of
632 an ML CG model. The network architecture makes the model in principle
633 transferable across chemical space.
- 634 ★★ [40] A comprehensive discussion on the challenges of representability
635 and transferability in coarse-grained systems.
- 636 ★★ [54] Data efficient method for learning coarse-grained force-fields by
637 employing generative deep neural networks. It is shown to work well on a
638 set of fast-folding proteins.
- 639 ★★ [44] Introduction of machine learning into protein coarse-graining, by
640 training a neural network with a physical prior potential with variational
641 force-matching.

¹ Graphical Abstract

² Machine Learned Coarse-Grained Protein Force-Fields: Are We
³ There Yet?

⁴ Aleksander E. P. Durumeric, Nicholas E. Charron, Clark Templeton, Félix
⁵ Musil, Klara Bonneau, Aldo S. Pasos-Trejo, Yaoyi Chen, Atharva Kelkar,
⁶ Frank Noé, Cecilia Clementi



7 Highlights

8 Machine Learned Coarse-Grained Protein Force-Fields: Are We 9 There Yet?

**10 Aleksander E. P. Durumeric, Nicholas E. Charron, Clark Templeton, Félix
11 Musil, Klara Bonneau, Aldo S. Pasos-Trejo, Yaoyi Chen, Atharva Kelkar,
12 Frank Noé, Cecilia Clementi**

- 13 • Coarse-graining is a powerful tool for modeling complex macromolecu-
14 lar systems**
- 15 • Machine learning is now enabling the definition of accurate bottom-up
16 coarse-grained protein models**
- 17 • Bottom-up protein coarse-grained models that are fully transferable in
18 sequence space still do not exist**
- 19 • We discuss the outstanding challenges towards the design of transfer-
20 able coarse-grained protein models and possible ways forward**