Supporting Information

Accelerated Simulations of Molecular Systems through Learning of Effective Dynamics

Pantelis R. Vlachas,[†] Julija Zavadlav,[‡,¶] Matej Praprotnik,[§,‖] and Petros Koumoutsakos[*,⊥]

†*Computational Science and Engineering Laboratory, ETH, Zurich CH-8092, Switzerland*
‡*Professorship of Multiscale Modeling of Fluid Materials, TUM School of Engineering and Design, Technical University of Munich, 85748 Garching bei München, Germany.*
¶*Munich Data Science Institute, Technical University of Munich, 85748 Garching bei München, Germany.*
§*Laboratory for Molecular Modeling, National Institute of Chemistry, SI-1001 Ljubljana, Slovenia*
‖*Department of Physics, Faculty of Mathematics and Physics, University of Ljubljana, SI-1000 Ljubljana, Slovenia*
⊥*John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA*

E-mail: petros@seas.harvard.edu

S1 Müller-Brown Potential

The MBP has the form

$$V(\boldsymbol{x}) = \sum_{k=1}^{4} A_k \exp\big(\alpha_k(x_1 - \hat{X}_{1,k}) + b_k(x_1 - \hat{X}_{1,k})(x_2 - \hat{X}_{2,k}) + c_k(x_2 - \hat{X}_{2,k})\big), \tag{S1}$$

where $\boldsymbol{x} = [x_1, x_2]^T$ is the position. The parametrization

$$\begin{aligned} \alpha &= [-1, -1, -6.5, 0.7]^T, \\ b &= [0, 0, 11, 0.6]^T, \\ c &= [-10, -10, -6.5, 0.7]^T, \\ A &= [-200, -100, -170, 15]^T, \\ \hat{X} &= \begin{bmatrix} 1 & 0 & -0.5 & -1 \\ 0 & 0.5 & 1.5 & 1 \end{bmatrix}, \end{aligned} \tag{S2}$$

is followed according to Ref.[1]

S1.1 Definition of Metastable States

The metastable states of the MB potential are defined as ellipses in the $\boldsymbol{x} \in \mathbb{R}^2$ space. The centers and axes given in Table S1.

Table S1: Metastable states in the MBP modeled as ellipses $x_1^2/\alpha^2 + x_2^2/\beta^2 \leq 1$. The ellipses are rotated by θ.

State	Center (x_1, x_2)	Axes $\alpha, \beta)$	θ
0	$(-0.57, 1.45)$	$(0.15, 0.3)$	$\pi/4$
1	$(0.45, 0.05)$	$(0.35, 0.15)$	0

S1.2 LED Hyperparameters

In order to prepare the dataset for training, validation, and testing of the LED in the MBP, 96 initial conditions are sampled from $\boldsymbol{x} \in [-1.5, 1.2] \times [-0.2, 2]$. The dynamics are solved with the Velocity Verlet algorithm, with timestep $\delta t = 10^{-2}$ up to $T = 5000$, after an initial

transient period of $\tilde{T} = 10^3$ discarded from the data. The data are sub-sampled to $\Delta t = 0.5$, keeping every 50^{th} data point. In this way, 96 trajectories of $N = 10^4$ samples, each corresponding to $T = 5000$ time units are created. LED is trained on 32 of these trajectories. 32 trajectories are used for validation, while all 96 trajectories are used for testing.

The number and size of hidden layers are the same for the encoder $\mathcal{E}$, the decoder $\mathcal{D}$, and the latent MDN $\mathcal{Z}$. In the first phase, the MDN-AE is trained, tuning its hyperparameters based on a grid search reported in Table S2. The autoencoder with the smallest error on the state statistics on the validation dataset is picked. Next, the MDN-LSTM is trained, tuning its hyperparameters based on a grid search reported in Table S3. The LED model with the smallest error on the state statistics on the validation dataset is picked. Both networks are trained with validation based early stopping. The LED is tested on the total 96 initial conditions. For more information of the training technicalities the interested reader is referred to Ref. [2]

Table S2: Hyperparameter tuning of AE for MBP

Hyperparameter	Values
Batch size	32
Initial learning rate	0.001
Weight decay rate	$\{0, 10^{-5}\}$
Number of AE layers	$\{2, 3\}$
Size of AE layers	$\{10, 20, 40\}$
Activation of AE layers	selu, tanh
Latent dimension	$\{1\}$
Input/Output data scaling	$[0, 1]$
MDN-AE kernels	$\{2, 3\}$
MDN-AE hidden units	50
MDN-AE multivariate	1
MDN-AE covariance scaling factor	$\{0.4, 0.6, 0.8\}$

S1.3 Marginal State Distributions

The marginal distributions of the MBP states from trajectories sampled from the LED is compared with the groundtruth (test data) in Figure S1.

Table S3: Hyperparameter tuning of LSTM for MBP

Hyperparameter	Values
Batch size	32
Initial learning rate	10^{-3}
BPTT sequence length	$\{200, 400\}$
Number of LSTM layers	1
Size of LSTM layers	$\{10, 20, 40\}$
Activation of LSTM Cell	tanh
MDN-LSTM kernels	$\{4, 5, 6\}$
MDN-LSTM hidden units	$\{10, 20\}$
MDN-LSTM multivariate	0
MDN-LSTM covariance scaling factor	$\{0.1, 0.2, 0.3, 0.4\}$

Table S4: Hyperparameters of LED model with lowest validation error on MBP

Hyperparameter	Values
Number of AE layers	3
Size of AE layers	40
Activation of AE layers	tanh
Latent dimension	1
MDN-AE kernels	3
MDN-AE hidden units	50
MDN-AE multivariate	1
MDN-AE covariance scaling factor	0.6
Weight decay rate	0.0
BPTT sequence length	400
Number of LSTM layers	1
Size of LSTM layers	20
Activation of LSTM Cell	tanh
MDN-LSTM kernels	4
MDN-LSTM hidden units	20
MDN-LSTM multivariate	0
MDN-LSTM covariance scaling factor	0.4

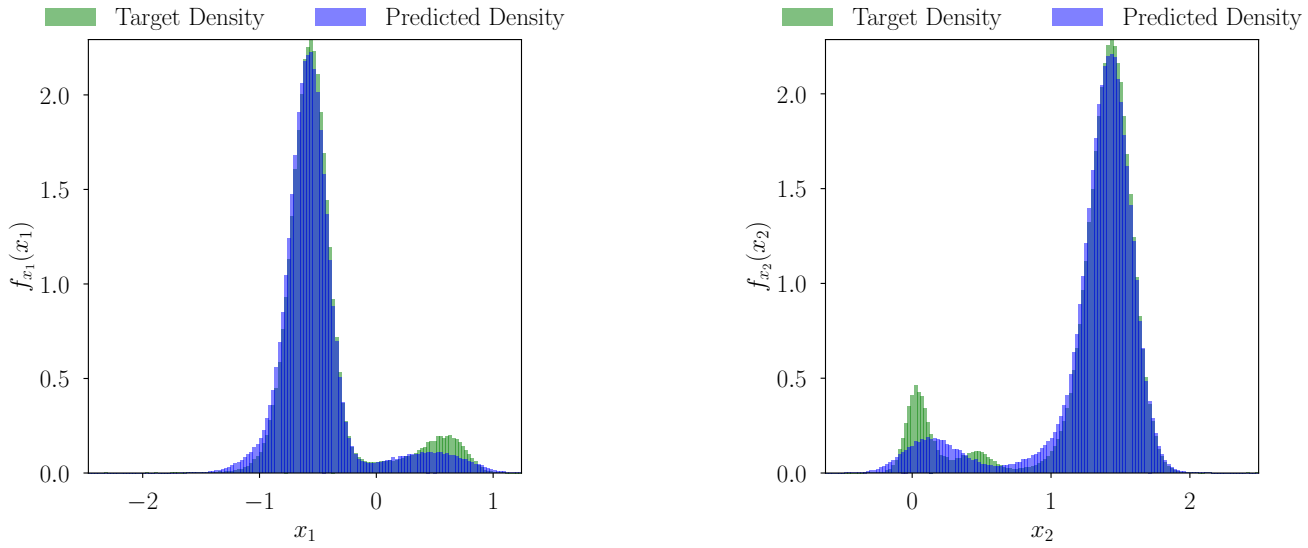

Figure S1: Comparison of the marginal distributions of the MBP states x_1 and x_2 between the test data and trajectories of LED. The LED is propagating the dynamics on a one dimensional reduced order latent state, i.e., $d_{\boldsymbol{z}} = 1$.

S1.4 Time-scales in the LED Latent Space

The latent space learned by LED can be utilized to identify low-energy metastable states without the need for prior knowledge. The definition of the metastable states in the rotationally and translationally invariant space constitutes such prior knowledge. Minima in the free energy projection on the LED latent space, constitute probable metastable states.

The trajectories sampled with LED are clustered based on these latent metastable clusters depicted in Figure 4 (main text). An MSM is fitted on the clustered trajectories. The time-lag of the MSM is set to 100 time units to ensure Markovianity. The timescales computed by MSM are $\overline{T}_{0\to1} = 49$ and $\overline{T}_{1\to0} = 321$. LED is overestimating $\overline{T}_{1\to0}$ and underestimating $\overline{T}_{0\to1}$. The order of the timescales, however, is captured. In contrast, an MSM with a time-lag of $\Delta_t = 0.5$, which is the timestep of the LED, fails to capture the order of the timescales due to the violated Markovianity assumption ($\overline{T}_{0\to1} = 3$ and $\overline{T}_{0\to1} = 21$).

S2 Trp Cage

S2.1 LED Hyperparameters

In the LED architecture, the number and size of hidden layers are the same for the encoder $\mathcal{E}$, the decoder $\mathcal{D}$, and the latent MDN $\mathcal{Z}$. The MDN-AE is trained, tuning its hyperparameters based on the grid search reported in Table S5. The latent space of the MDN-AE is $\boldsymbol{z} \in \mathbb{R}^2$, i.e., $d_{\boldsymbol{z}} = 2$. The MDN-AE model with the lowest error on the state statistics in the validation dataset is picked. Then, the MDN-AE is coupled with the MDN-LSTM as LED. The MDN-LSTM is trained to minimize the latent data likelihood. The hyperparameters of the MDN-LSTM are tuned according to the grid search reported in Table S6. The LED model with the lowest error on the state statistics in the validation dataset is selected. Its hyperparameters are reported in Table S7. The LED is tested in 248 initial conditions randomly sampled from the testing data. Starting from these initial conditions, we utilize the iterative propagation in the latent space of the LED to forecast $T = 400$ps.

Table S5: Hyperparameter tuning of AE for Trp Cage

Hyperparameter	Values
Batch size	32
Initial learning rate	10^{-3}
Weight decay rate	$\{0, 10^{-4}, 10^{-5}, 10^{-6}\}$
Number of AE layers	$\{4, 6\}$
Size of AE layers	$\{100, 200, 500\}$
Activation of AE layers	selu, tanh
Latent dimension	2
Input/Output data scaling	$[0, 1]$
MDN-AE kernels	$\{3, 4, 5\}$
MDN-AE hidden units	$\{20, 50\}$
MDN-AE covariance scaling factor	0.8

S2.2 Marginal State Distributions

The marginal distributions of the trajectories generated by LED match the ground-truth ones (MD data) closely, as depicted in Figure S2. In Figure S3, a sample from MD data of the TRP cage is compared with a close sample (in

Table S6: Hyperparameter tuning of LSTM for Trp Cage

Hyperparameter	Values
Batch size	32
Initial learning rate	10^{-3}
BPTT sequence length	$\{200, 400\}$
Number of LSTM layers	1
Size of LSTM layers	$\{10, 20, 40\}$
Activation of LSTM Cell	tanh
MDN-LSTM kernels	$\{4, 8, 12, 24\}$
MDN-LSTM hidden units	$\{10, 20, 40, 80\}$
MDN-LSTM multivariate	$\{0, 1\}$
MDN-LSTM covariance scaling factor	$\{0.1, 0.2, 0.3, 0.4\}$

Table S7: Hyperparameters of LED model with lowest validation error on Trp Cage

Hyperparameter	Values
Number of AE layers	6
Size of AE layers	500
Activation of AE layers	tanh
Latent dimension	2
MDN-AE kernels	5
MDN-AE hidden units	50
MDN-AE multivariate	0
MDN-AE covariance scaling factor	0.8
Weight decay rate	0
BPTT sequence length	400
Number of LSTM layers	1
Size of LSTM layers	40
Activation of LSTM Cell	tanh
MDN-LSTM kernels	4
MDN-LSTM hidden units	20
MDN-LSTM multivariate	0
MDN-LSTM covariance scaling factor	0.2

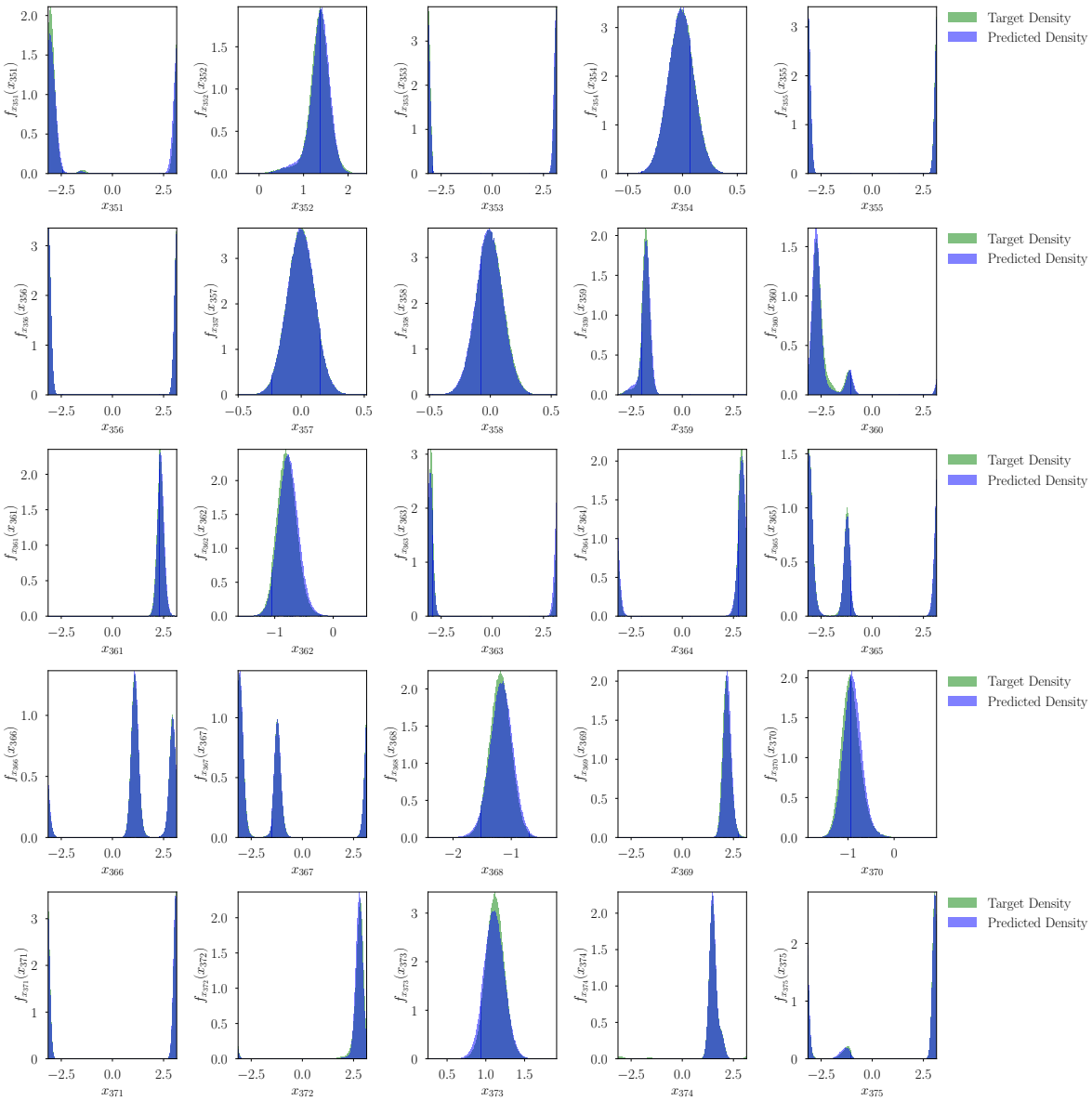

Figure S2: Plot of the marginal state distributions $\boldsymbol{s}_{351} - \boldsymbol{s}_{375}$ in the Trp Cage miniprotein. Comparison of the state distributions estimated from the MD data (test dataset) and from trajectories sampled from LED.

terms of the latent space) of LED. The RMSD is 2.784$\mathring{A}$.

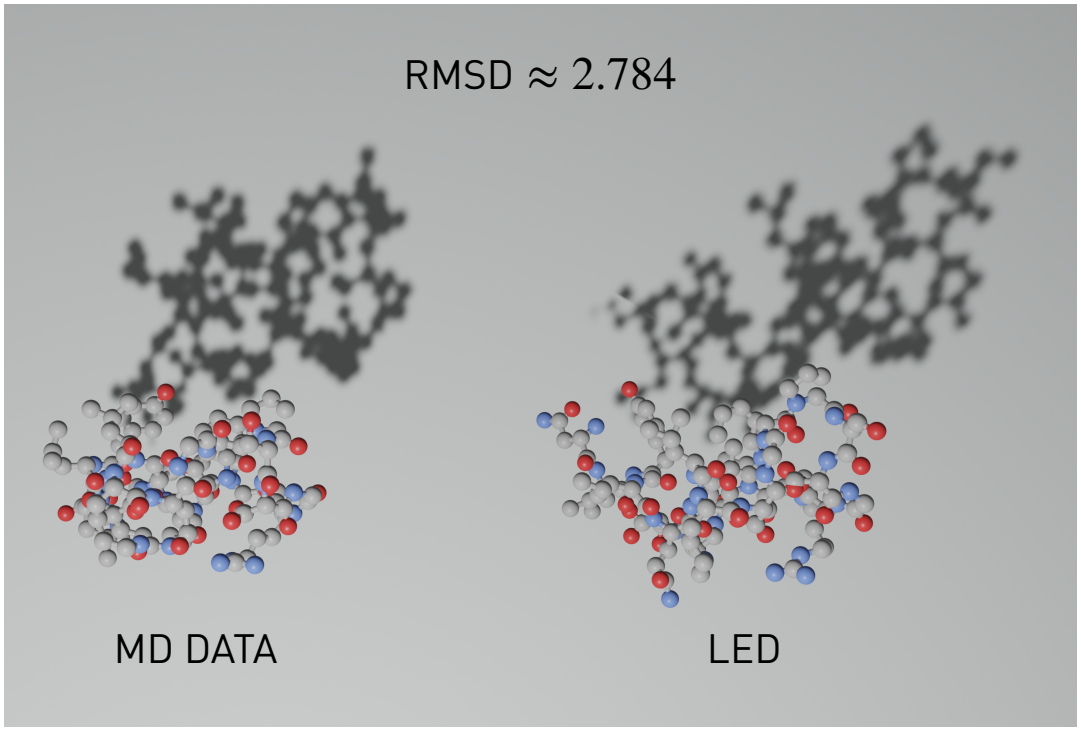

Figure S3: Trp Cage protein configurations found in the MD data compared to a sample of LED that is is in close proximity in the latent space. The RMSD error between the two configurations is 2.784$\mathring{A}$.

S3 Alanine Dipeptide

A molecule of alanine dipeptide in water is simulated with MD.[3] The peptide is modeled with the AMBER03 force field,[4] while the water is modeled with TIP3P/Fs.[5] The Velocity Verlet algorithm is employed for the integration. The simulation domain is a cubic box (edge length 2.7 nm) with periodic boundary conditions and minimum image convention. The temperature is maintained at 298 K with a local Langevin thermostat,[6] with the value of the friction constant equal to 1.0/ps. The cutoff distance for the nonbonded interactions is $r_c = 0.9$ nm. The reaction field method[7] is used for the electrostatic interaction beyond the cutoff, with the dielectric permittivity of inner and outer regions equal to 1 and 80, respectively.

A timestep of $\delta t = 1$fs is considered, and the dynamics are integrated up to a total time of $T = 100$ns, creating a dataset with a total of 10^8 data samples. The data are subsampled, keeping every 100^{th} datapoint, creating a trajectory with $N = 10^6$ samples. The coarse time-step of LED is thus $\Delta t = 0.1$ps. The protein positions are transformed into rototranslational invariant features (internal coordinates), composed of bonds, angles, and dihedral angles. The data are split to 248 trajectories of 4000 samples (each trajectory corresponds to $T = 400$ps of MD data), discarding the remaining data. The first 96 trajectories (corresponding to a total of 38.4ns of MD data) are used for training and the next 96 trajectories for validation. All 248 initial conditions are used for testing.

S3.1 Metastable State Definition

The protein is considered to lie in each of the five metastable states $\{C5, P_{II}, \alpha_R, \alpha_L, C_7^{ax}\}$ if the distance in the Ramachandran plot between the protein state and the metastable state center is smaller than 10 degrees. The metastable state centers are defined in Table S8.

Table S8: Centers of the metastable states in the Ramachandran plot.

Metastable state	Center (ϕ, ψ)
P_{II}	$(-75, 150)$
$C5$	$(-155, 155)$
α_R	$(-75, -20)$
α_L	$(67, 5)$
C_7^{ax}	$(70, 160)$

S3.2 LED Hyperparameters

Regarding the LED architecture, the number and size of hidden layers are the same for the encoder $\mathcal{E}$, the decoder $\mathcal{D}$, and the latent MDN $\mathcal{Z}$. The MDN-AE is trained, tuning its hyperparameters based on the grid search reported in Table S9. The latent space of the MDN-AE is $\boldsymbol{z} \in \mathbb{R}^2$, i.e., $d_{\boldsymbol{z}} = 1$. The MDN-AE model with the lowest validation error on the state statistics is picked. Then, the MDN-AE is coupled with the MDN-LSTM in LED. The MDN-LSTM is trained to minimize the latent data likelihood. The hyperparameters of the MDN-LSTM are tuned according to the grid search reported in Table S10. The LED model with the lowest error on the state statistics in the validation dataset is selected. Its hyperparameters are reported in Table S11. The LED is tested in 248 initial conditions randomly sampled from the testing data. Starting from these initial conditions, we utilize the iterative propagation in the latent space of the LED to forecast $T = 400$ps.

Table S9: Hyperparameter tuning of AE for alanine dipeptide

Hyperparameter	Values
Batch size	32
Initial learning rate	10^{-3}
Weight decay rate	$\{0, 10^{-5}\}$
Number of AE layers	$\{4, 6\}$
Size of AE layers	$\{50, 100\}$
Activation of AE layers	selu, tanh
Latent dimension	2
Input/Output data scaling	$[0, 1]$
MDN-AE kernels	5
MDN-AE hidden units	$\{20, 50\}$
MDN-AE multivariate	0
MDN-AE covariance scaling factor	0.8

Table S10: Hyperparameter tuning of LSTM for alanine dipeptide

Hyperparameter	Values
Batch size	32
Initial learning rate	10^{-3}
BPTT sequence length	$\{200, 400\}$
Number of LSTM layers	1
Size of LSTM layers	$\{10, 20, 40\}$
Activation of LSTM Cell	tanh
MDN-LSTM kernels	$\{4, 5, 6\}$
MDN-LSTM hidden units	$\{10, 20\}$
MDN-LSTM multivariate	$\{0, 1\}$
MDN-LSTM covariance scaling factor	$\{0.1, 0.2, 0.3, 0.4\}$

Table S11: Hyperparameters of LED model with lowest validation error on alanine dipeptide

Hyperparameter	Values
Number of AE layers	4
Size of AE layers	50
Activation of AE layers	tanh
Latent dimension	2
MDN-AE kernels	5
MDN-AE hidden units	50
MDN-AE multivariate	0
MDN-AE covariance scaling factor	0.8
Weight decay rate	0
BPTT sequence length	400
Number of LSTM layers	1
Size of LSTM layers	20
Activation of LSTM Cell	tanh
MDN-LSTM kernels	5
MDN-LSTM hidden units	20
MDN-LSTM multivariate	0
MDN-LSTM covariance scaling factor	0.4

S3.3 Marginal State Distributions

The marginal distributions of the trajectories generated by LED match the ground-truth ones (MD data) closely, as depicted in Figure S4.

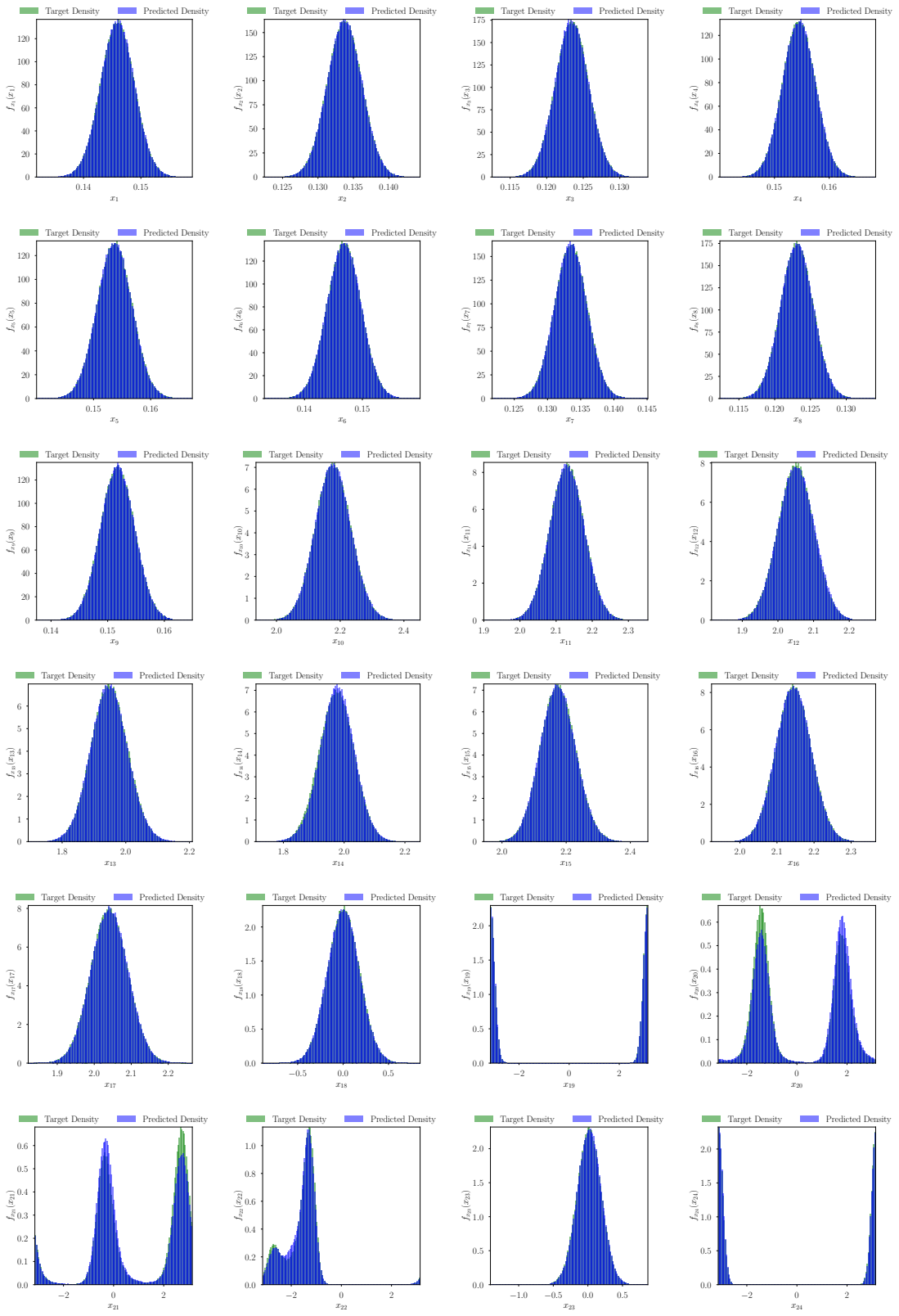

Figure S4: Plot of the marginal state distributions. Comparison of the state distributions estimated from the MD data (test dataset) and from trajectories sampled from LED.

In Figure S5, a configuration randomly sampled from MD data is given for each metastable state. The closest configuration sampled from LED is compared with the MD data sample in terms of the Root Mean Square Deviation (RMSD) score. The LED samples realistic configurations with low RMSD errors for all metastable states. The mean and standard deviation of the RMSD scores of the 10 closest neighbors sampled from LED are $\mu \pm \sigma = 0.148 \pm 0.021\mathring{A}$ for the C_5 MD sample configuration (Figure S5 top left). This score for the rest of the metastable states is $0.340 \pm 0.463\mathring{A}$ for P_{II}, $0.101 \pm 0.019\mathring{A}$ for α_R, $0.885 \pm 0.162\mathring{A}$ for α_L, and $0.383 \pm 0.125\mathring{A}$ for C_7^{ax}. The LED samples similar configurations with low RMSD scores for the most frequently observed metastable states $\{C_5, P_{II}, \alpha_R\}$. The average RMSD error is slightly higher and fluctuates more for the less frequently observed $\{\alpha_R, C_7^{ax}\}$.

S3.4 Metastable States on the Latent Space and Mean First Passage Times

The metastable states can be defined on the latent space of LED, by projecting the free energy on the latent space, and identifying the local minima. This alleviates the need for expert knowledge (definition of the metastable states). The MFPTs between the metastable states on the latent space of the LED are compared with the MFPTs between the corresponding metastable states on the Ramachadran space in Table S12. Note that the results depend on how the latent metastable states are defined. However, in order to capture the order of the timescales without the need of prior expert knowledge, a rough approximation (small region around the minima in the latent space) is adequate. The LED is able to capture the order of the timescales, alleviating the need for expert knowledge on the definition of the metastable states.

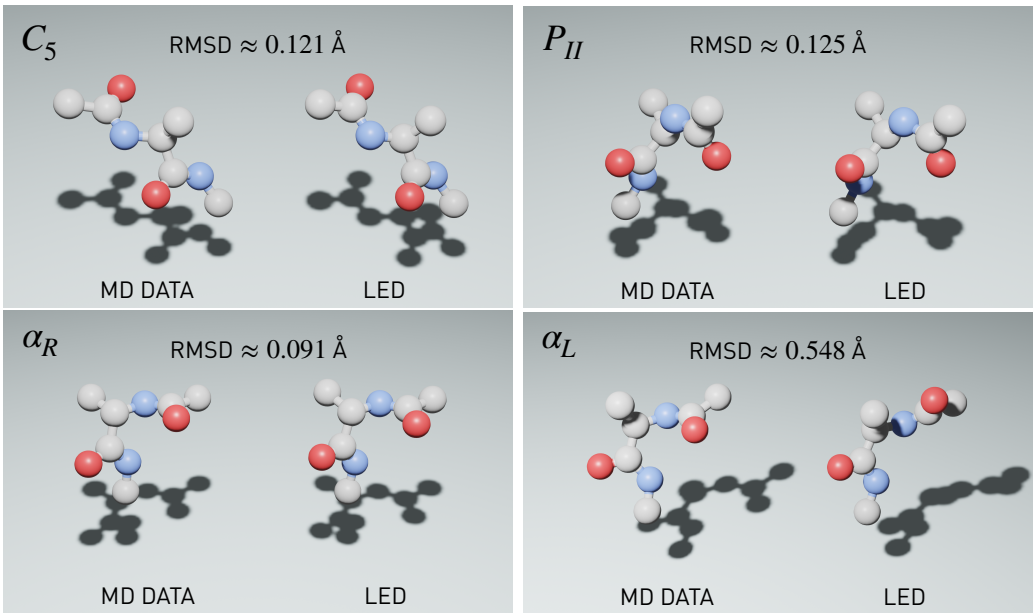

Figure S5: For each metastable state, a random alanine dipeptide configuration sampled from MD data is compared against the closest configuration sampled from the LED with $d_{\mathbf{z}} = 1$. The Root Mean Square Deviation (RMSD) in $\mathring{A}$ between the two is plotted for reference.

Table S12: Mean first-passage times (MFPT) between the metastable states of alanine dipeptide in water in [ns]. MFPTs are estimated by fitting MSMs with a time-lag of 10ps on MD trajectories. In LED, the metastable states are considered as regions around the local minima of the free energy projection on the latent space. The average relative error is given for reference.

		Metastable states on **Ramachandran Space**		Metastable states on **LED Latent Space**	
MFPT [ns]	MSM – 10ps on MD data Reference	MSM – 10ps on LED – 0.1ps data MFPT	Error (%)	MSM – 10ps on LED – 0.1ps data MFPT	Error (%)
$T_{C_5 \to P_{II}}$	0.105	0.103	**2**	0.143	36
$T_{C_5 \to \alpha_R}$	0.104	0.082	**21**	0.124	19
$T_{P_{II} \to C_5}$	0.226	0.242	**7**	0.356	57
$T_{P_{II} \to \alpha_R}$	0.105	0.083	**21**	0.123	18
$T_{\alpha_R \to C_5}$	0.236	0.258	**9**	0.361	53
$T_{\alpha_R \to P_{II}}$	0.116	0.119	**2**	0.148	27
Average Relative Error			10.51%		35.21%

References

(1) Müller, K.; Brown, L. D. Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theor. Chim. Acta* **1979**, *53*, 75–93.

(2) Vlachas, P. R.; Arampatzis, G.; Uhler, C.; Koumoutsakos, P. Learning the Effective Dynamics of Complex Multiscale Systems. *arXiv preprint arXiv:2006.13431* **2020**,

(3) Guzman, H. V.; Tretyakov, N.; Kobayashi, H.; Fogarty, A. C.; Kreis, K.; Krajniak, J.; Junghans, C.; Kremer, K.; Stuehn, T. ESPResSo++ 2.0: Advanced methods for multiscale molecular simulation. *Comput. Phys. Commun.* **2019**, *238*, 66–76.

(4) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T., et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(5) Schmitt, U. W.; Voth, G. A. The computer simulation of proton transport in water. *J. Chem. Phys.* **1999**, *111*, 9361.

(6) Grest, G. S.; Kremer, K. Molecular dynamics simulation for polymers in the presence of a heat bath. *Phys. Rev. A* **1986**, *33*, 3628.

(7) Neumann, M. The dielectric constant of water. Computer simulations with the MCY potential. *J. Chem. Phys.* **1985**, *82*, 5663.