

Figure 4. Experiments on 2AA test dipeptides QW (top row) and HT (bottom row). (a) TICA plots for a long MD chain and samples generated with the Timewarp MCMC algorithm (Algorithm 1). (b) Free energy comparison for the MD trajectory, Timewarp MCMC (Algorithm 1), and Timewarp exploration (Algorithm 2).

experiments and results can be found in Appendix B. We train three flow models on three datasets: (i) AD, consisting of simulations of alanine dipeptide, (ii) 2AA, with peptides with 2 amino acids, and (iii) 4AA, with peptides with 4 amino acids. All datasets are created with MD simulations performed with the same parameters (see Table 1 and Appendix D). For AD, we do not test transferability as there is only a single molecule. For 2AA and 4AA, we train on a randomly selected trainset of peptides, and evaluate on unseen test peptides. The relative frequencies of the amino acids in 2AA and 4AA are similar across the splits. For 4AA, the training set consists of about 1% of the total number of possible tetrapeptides (20^4), making the generalisation task significantly more difficult than for 2AA.

Alanine dipeptide (AD) We first investigate alanine dipeptide, a small (22 atoms) single peptide molecule commonly used to benchmark MD software. We train Timewarp on AD as described in Section 5 and sample new states using Algorithm 1 for a chain length of 10 million, accepting roughly 2% of the proposals. In Figure 3a we visualise the samples using a *Ramachandran plot* (Ramachandran et al., 1963), which shows the distribution of the backbone dihedral angles φ and ψ . Each mode in the plot represents a metastable state. We see that the Timewarp samples closely match MD, visiting all the metastable states with the correct relative weights. In Figure 3b we plot the free energy (*i.e.*, the relative log probability) of the φ and ψ angles, again showing close agreement. The roughness in the plot is due to some regions of state space having very few samples. In Figure 3c we show, for an initial state $x(t)$, the conditional distribution of MD obtained by integrating Equation (2), $\mu(x(t+\tau)|x(t))$, compared with the model $p_\theta(x(t+\tau)|x(t))$, demonstrating close agreement. Finally, Figure 3d shows the time-evolution of the φ angle for MD and Timewarp. Timewarp exhibits significantly more tran-

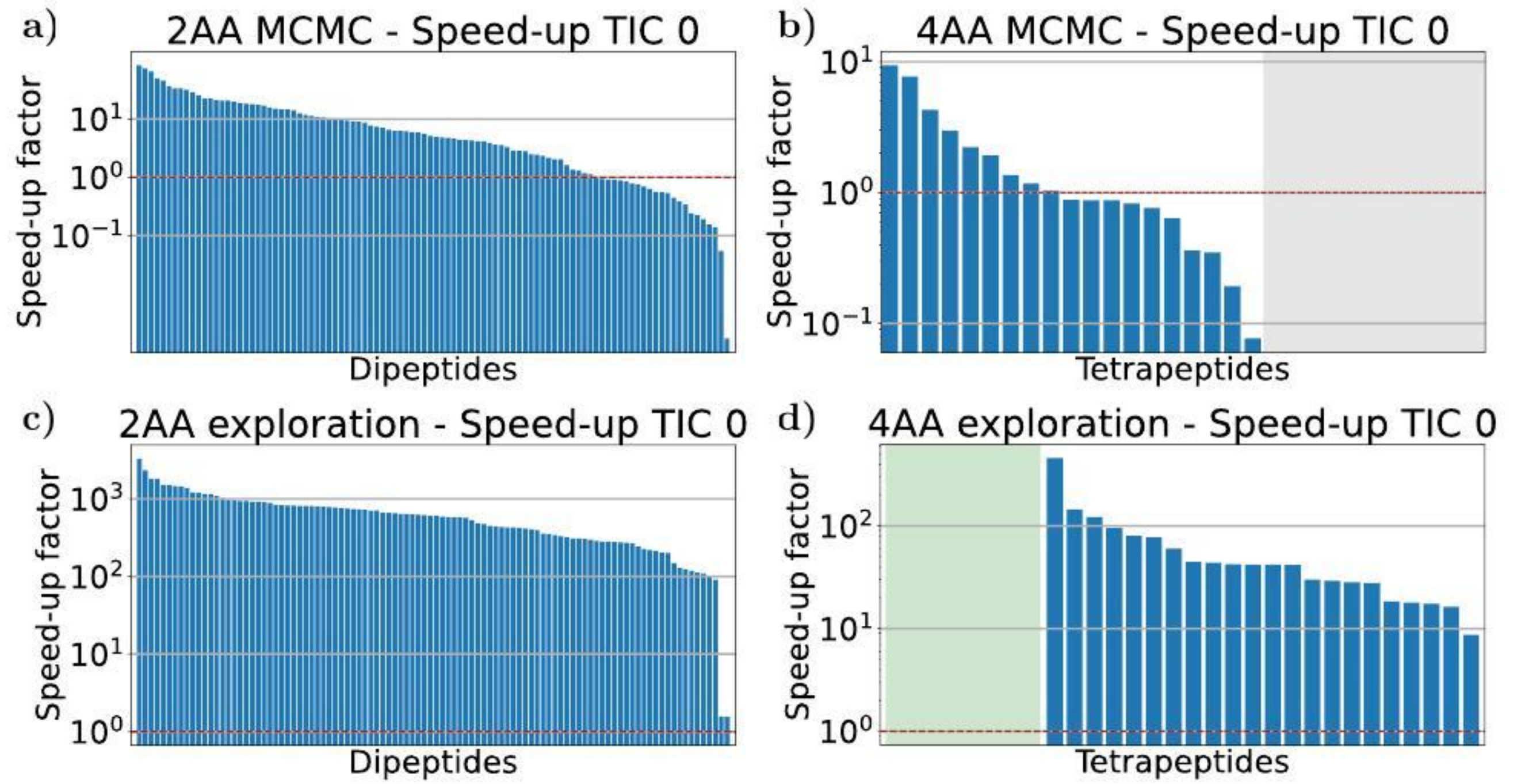


Figure 5. Speed-up factors in terms of ESS/s ratios for the slowest TICA component for the Timewarp MCMC and exploration algorithms, compared to MD. The dashed red line shows a speed-up factor of one. Gray areas depict peptides where Timewarp fails to explore all meta-stable states within 20 million steps, but MD does. Green areas depict peptides where MD fails to find all metastable states, but Timewarp does. (a), (b) Speed-up for the Timewarp MCMC algorithm (Algorithm 1) on test dipeptides (2AA) and tetrapeptides (4AA), respectively. (c), (d) Speed-up for the Timewarp exploration algorithm (Algorithm 2) on test dipeptides (2AA) and tetrapeptides (4AA), respectively.

sitions between the metastable states than MD. As a result, the autocorrelation along the φ angle decays much faster in terms of wall-clock time, resulting in a $\approx 7\times$ speed-up in terms of ESS/s compared to MD (see Appendix B.4).

Dipeptides (2AA) Next, we demonstrate transferability on dipeptides in 2AA. After training on the train dipeptides, we deploy Timewarp with Algorithm 1 on the test dipeptides for a chain length of 20 million. Timewarp achieves acceptance probabilities between 0.03% and 2% and explores all metastable states (Appendix B.1). The results are shown for the dipeptides QW and HT in Figure 4, showing close agreement between Timewarp and long MD chains ($1 \mu s = 2 \times 10^9$ timesteps). For these dipeptides, Timewarp achieves ESS/s speed-up factors over MD of 5 and 33 respectively (Appendix B.4). In Figure 5, Left, we show the speed-up factors for Timewarp versus MD for each of the 100 test dipeptides. Timewarp provides a median speed-up factor of about five across these peptides. In addition, we generate samples with the Timewarp model *without* the MH correction as detailed in Section 3.5. We sample 100 parallel chains for only 10000 steps starting from the same initial state for each test peptide. For each peptide we select *only one* of these chains that finds all meta-stable states for evaluations. As before, we compute the ESS/s to compare with MD, showing a median speedup factor of ≈ 600 (Figure 5c). Note that the actual speedup when using all the chains sampled in parallel will be much larger. Timewarp exploration leads to free energy estimates that are qualitatively similar to MD, but less accurate than Timewarp MCMC (Figure 4b).