1

Alchemical Free-Energy Calculations by Multiple-replica λ-dynamics: The Conveyor Belt Thermodynamic Integration

> *"Let us learn to dream, gentlemen, and then perhaps we shall learn the truth."*
>
> August Kekulé, 1865

A new method is proposed to calculate alchemical free-energy differences based on molecular dynamics (MD) simulations, called the conveyor belt thermodynamic integration (CBTI) scheme. As in thermodynamic integration (TI), K replicas of the system are simulated at different values of the alchemical coupling parameter λ. The number K is taken to be even and the replicas are equally spaced on a forward-turn-backward-turn path, akin to a conveyor belt (CB) between the two physical end-states. And as in λ-dynamics (λD), the λ-values associated with the individual systems evolve in time along the simulation. However, they do so in a concerted fashion, determined by the evolution of a single dynamical variable Λ of period 2π controlling the advance of the entire CB. Thus, a change of Λ is always associated with $K/2$ equispaced replicas moving forward and $K/2$ equispaced replicas moving backward along λ. As a result, the effective free-energy profile of the replica system along Λ is periodic of period $2\pi K^{-1}$ and the magnitude of its variations decreases rapidly upon increasing K, at least as K^{-1} in the limit of large K. When a sufficient number of replicas is used, these variations become small, which enables a complete

and quasi-homogeneous coverage of the λ-range by the replica system, without application of any biasing potential. If desired, a memory-based biasing potential can still be added to further homogenize the sampling, the preoptimization of which is computationally inexpensive. The final free-energy profile along λ is calculated similarly to TI, by binning of the Hamiltonian λ-derivative as a function of λ considering all replicas jointly, followed by quadrature integration. The associated quadrature error can be kept very low owing to the continuous and quasi-homogeneous λ-sampling. The CBTI scheme can be viewed as a continuous/deterministic/dynamical analog of the Hamiltonian replica-exchange/permutation (HRE/HRP) schemes, or as a correlated multiple-replica analog of the λD or λ-local elevation umbrella sampling (λ-LEUS) schemes. Compared to TI, it shares the advantage of the latter schemes in terms of enhanced orthogonal sampling, *i.e.* the availability of variable-λ paths to circumvent conformational barriers present at specific λ-values. Compared to HRE/HRP, it permits a deterministic and continuous sampling of the λ-range, and bypasses the need to carefully preselect a λ-ladder and a swapping-attempt frequency. Compared to λ-LEUS, it eliminates (or drastically reduces) the dead time associated with the preoptimization of a biasing potential. The goal of this chapter is to provide the mathematical/physical formulation of the proposed CBTI scheme, along with an initial application of the method to the calculation of the hydration free energy of methanol.

1.1   INTRODUCTION

Classical molecular dynamics (MD) simulations provide insight into (bio-)molecular systems at atomic resolution, thereby explaining and complementing experimental observations. This often involves the calculation of free-energy differences,[?][?][?] which characterize the relative stabilities of two or more macroscopic states of the system. These states may differ thermodynamically (different pressures, temperatures or numbers of molecules), conformationally (different regions in a space spanned by a set of specific generalized coordinates) or alchemically (different Hamiltonian functions).

Alchemical free-energy calculations involve atom mutations or interaction alterations that have no experimental counterpart. However, by comparing the results of two such calculations in different environments (*e.g.* mutation of a molecule into another one in vacuum or in a solvent) *via* a thermodynamic cycle,[?][?] the calculated values can be converted to experimentally accessible differences (*e.g.* relative solvation free energies of the two molecules in the given solvent). This indirect pathway *via* a cycle typically offers a strong sampling advantage relative to the calculation over a direct conformational path (*e.g.* reversibly displacing the two molecules from vacuum into the solvent across the liquid surface), while giving the same result at full convergence.[?][?][?] Over the last decades, numerous methods have been proposed to calculate alchemical free-energy differences between states A, B, C, ... of a molecular system involving the same numbers of atoms but different Hamiltonian functions. They can be roughly classified as reference-state methods and pathway-dependent methods.[?]

In reference-state methods, a single simulation is performed at a reference state R and the relative free energy of a target state A is calculated using one-step free-energy perturbation[?][?][?][?][?][?][?] (OSP) as a free-energy estimator. The state R may be unphysical, in which case the free-energy difference between two physical states A and B is obtained by comparing the results of two such calculations (R to A and R to B). The accuracy of the method, *i.e.* its convergence at finite sampling times,

depends crucially on the extent of Boltzmann-weighted phase-space overlap between the reference and target states,[?][?][?][?] *i.e.* whether configurations relevant for *A* are well sampled in *R*. Methods to enhance this overlap include in particular: (*i*) the use of a reference state with soft-core[?] (SC) sites[?][?][?] or other softened force-field terms;[?][?][?][?] (*ii*) the construction of a reference state encompassing all the target states, as in enveloping distribution sampling[?][?][?] (EDS); (*iii*) the extension to the use of multiple reference states,[?] along with the application of the (multi-state[?]) Bennet acceptance ratio[?][?] (BAR or MBAR) estimator instead of OSP. In principle, reference-state methods bear the promise of enabling the extrapolative calculation of the relative free energies of numerous arbitrary states (*A*, *B*, *C*, ...) based on a single reference-state simulation. In practice, however, they seldom hold this promise (or not in a sufficiently robust fashion based on finite simulations), because the design of a suitable reference state requires considerable (*e.g.* SC approach) or even complete (*e.g.* EDS approach) *a priori* knowledge of the target states it has to be appropriate for.

In pathway-dependent methods, a hybrid Hamiltonian is constructed by employing a coupling parameter λ that defines a continuous transformation between the Hamiltonians of the physical end-states *A* and *B*. Given such a path, the most established, robust and still frequently used method is multi-configuration[?] thermodynamic integration[?][?] (MCTI or, simply, TI). In TI, a set of independent simulations are performed at different constant λ-values, and the ensemble average of the derivative of the hybrid Hamiltonian with respect to λ is subsequently integrated by numerical quadrature[?][?] or curve fitting.[?][?][?] A common alternative to TI is multi-configuration[?] free-energy perturbation[?][?][?][?][?][?][?] (MC-FEP or, simply, FEP), where the OSP estimator is used to evaluate the free-energy difference from one λ-point to the previous or/and next one. Refinements and improvements of the original TI protocol include in particular: (*i*) the design of Hamiltonian coupling schemes leading to high sampling efficiencies;[?][?][?][?][?][?] (*ii*) the partial automation of the TI protocol,[?][?][?][?][?] *i.e.* of the selection of λ-points along with associated initial configurations, equilibration times and sampling times; (*iii*) the use of free-

energy estimators with improved statistical efficiencies over plain quadrature, *e.g.* extended TI (EXTI) estimator[?] or MBAR estimator;[?] (*iv*) the design of alternative single- or multiple-replica schemes where the λ-values are no longer fixed during the simulation.[? ? ? ? ?]

Although the possibility of changing the λ-values over the course of a simulation may appear at first sight to represent an unnecessary complication of the TI protocol, it often leads to a significant enhancement of the sampling efficiency. This is because it increases the likelihood of crossing barriers in the orthogonal space,[? ?] that is, the space spanned by all the degrees of freedom of the system excluding λ (*i.e.* all the conformational ones). These orthogonal barriers are typically higher at certain λ-values than at others. The highest ones may be seldom crossed in corresponding simulations at fixed λ, and allowing for variations of λ may open pathways to circumvent them. Note that, in many cases, this enhancement is not sufficient *per se* and other techniques must be applied to further improve the orthogonal sampling along the λ-path.[? ? ? ? ? ? ? ? ? ? ? ? ?] There exist two main routes for performing variable-λ simulations: (*i*) Hamiltonian replica exchange (HRE) or permutation (HRP), which involve multiple system replicas; (*ii*) λ-dynamics (λD) or λ-Monte Carlo (λMC), which consider a single system.

In the HRE scheme,[? ? ?] a series of system copies (walkers) are distributed over a set of fixed λ-values (replicas), as is the case in TI. However, at regular time intervals, swaps are attempted between pairs of systems corresponding to different λ-values (typically adjacent ones). These attempts are accepted or rejected according to a Metropolis-Hastings criterion[? ?] depending on the Boltzmann factor ratio of the replica system before and after the swap. Although the trajectories at each λ-point become discontinuous, the TI-like statistics is preserved and the data can be analyzed in the same way as for TI. Recent extensions of the method involve in particular the consideration of more advanced exchange schemes,[? ? ?] of replica reservoirs,[? ? ? ? ? ? ?] of frozen replicas,[? ?] of heating-quenching steps between the sampling periods,[? ? ? ? ?] of the infinite-swapping limit,[? ? ? ? ? ?] and of generalized-ensemble distributions.[? ? ? ? ? ?] They also include the implementation

of λ-moves that go beyond pairwise swaps with a selection based on a Suwa-Todo criterion,[?][?][?] as implemented in the HRP method.[?][?][?] In the latter case, enabling arbitrary permutations and abandoning the detailed-balance constraint leads to a significant increase in the probability of exchange acceptance.

In the λD scheme[?][?][?][?][?][?][?][?] (see also its ancestor[?][?] λMC), a single system is considered for which the λ-value evolves dynamically in the course of the simulation, *i.e.* λ is treated as an extra pseudo-conformational degree of freedom with an assigned mass parameter m_λ and momentum p_λ. This momentum enters an extended Hamiltonian that includes an additional term for the corresponding kinetic energy. This results in a continuous sampling of the λ-range instead of the discrete sampling underlying TI or HRE/HRP. The free-energy difference is then typically estimated from the probability distribution along λ, with the drawback of requiring a threshold to define the end-states.[?] This issue can be alleviated by introducing a coordinate transformation with plateaus,[?][?] or by using a TI-like formula[?][?] or a Rao-Blackwell estimator,[?] the latter conceptually similar to MBAR.

The major advantages of λD compared to HRE/HRP are that it involves a simpler single-system setup, is deterministic, samples the λ-range continuously, and does not require the specification (and optimization[?][?][?][?][?][?][?][?][?][?][?][?][?]) of a λ-ladder and of a swapping-attempt frequency. The main drawback[?] is that the sampling probability along λ is no longer imposed in the form of a fixed set of λ-points with equal sampling times, but entirely controlled by the free-energy profile along λ. As a result, λ-values with high relative free energies may be poorly represented (or not at all) and λ-barriers with high relative free energies may be seldom crossed (or not at all) in the course of a finite simulation. In addition, care must be taken to avoid the sampling of λ-values beyond the physical end-states of the alchemical coupling[?] (*i.e.* below 0 or above 1). Both the sampling inhomogeneity and the end-point issues can in principle be remedied[?] by employing an appropriate coordinate transformation[?][?][?][?][?][?][?] or/and by applying a biasing potential along λ.[?][?][?][?][?][?] A combination of these two approaches underlies the λ-local elevation umbrella sampling method[?][?][?] (λ-LEUS), which relies in particular on an adaptive memory-based biasing potential.

A similar principle is also at the heart of numerous other methods such as the flat-histogram,[?][?] λ-metadynamics,[?][?][?] adaptive integration,[?] adaptive biasing force,[?] adaptively biased[?] and expanded-ensemble[?][?][?][?][?][?][?] methods.

In λ-LEUS, a local elevation[?] (LE) build-up phase is used to construct a suitable biasing potential, and followed by an umbrella sampling[?][?] (US) phase where this potential is frozen[?] and biased equilibrium statistics gathered with quasi-homogeneous sampling of the λ-range. Clearly, the LE phase represents an efficiency loss in the method, *i.e.* a dead time. Note, however, that a similar (but generally shorter) dead time also exists in TI and HRE/HRP in the form of discarded initial equilibration times for all replicas. Another drawback of λ-LEUS is that it is not parallelizable in the same fashion as TI and HRE/HRP, where the simulations at different λ-points can be carried out in parallel (see, however, various swarm,[?][?][?][?] multiple-walker[?][?] and flying Gaussian[?][?] variants of memory-based biasing methods). Finally, some care must be taken to select an appropriate mass and thermostat-coupling scheme for the λ-variable, so as to ensure that this variable is adequately coupled to the configurational degrees of freedom.[?]

In the present chapter, we propose a new alchemical free-energy calculation scheme with the goal of combining the advantages and alleviating the shortcomings of both HRE/HRP and λD. This scheme is termed conveyor belt thermodynamic integration (CBTI) and relies on the coupled λD of a set of system replicas, in which the λ-distance between successive replicas is kept fixed along a forward-turn-backward-turn path, akin to a conveyor belt between the two physical end-states. The basic principle of CBTI is illustrated schematically in Fig. 1.1.

Considering a free-energy profile $G(\lambda)$ presenting a constant uphill slope between A and B, a single system subjected to λD in the absence of a biasing potential will "roll down" to A and keep sampling the neighborhood of this state (Fig. 1.1a). To circumvent this problem, one may decide to couple the λD of two replicas 0 and 1 of the system in such a way that any downhill displacement of replica 0 implies an equivalent uphill displacement of replica 1 (Fig. 1.1c). This working principle is exploited in practice by vehicles like cable cars or funiculars, where the motion

of the two vehicles are controlled by a stretched cable connected to two pulleys. It has also been exploited previously for MD in the context of twin-system[?][?] EDS, which couples forward and reverse alchemical changes performed in two different environments. In this situation, one may describe the λ-values of the two replicas by means of a single periodic angular variable Λ representing the advance of the cable from 0 (replica 0 in state A, replica 1 in state B) to π (0 in B, 1 in A) and then 2π (back to the starting situation). Note that the π to 2π return range of Λ deviates from the cable car or funicular analogy, where the vehicles never go "over the pulley". If the free-energy profile $G(\lambda)$ has a constant slope, it is easily seen that there will be no net driving force on the cable. In the λD context, this means that the Λ-variable will undergo random diffusion with a homogeneous sampling of Λ, *i.e.* the corresponding free-energy profile $G_\Lambda(\Lambda)$ will be flat (Fig. 1.1d). Accordingly, each of the two replicas will sample the entire λ-range homogeneously.

Consider now a free-energy profile $G(\lambda)$ presenting a constant uphill slope from 0 to 1/2 and a constant downhill slope from 1/2 to 1 (Fig. 1.1e). In a setup with two replicas, the $G_\Lambda(\Lambda)$ profile will no longer be flat (Fig. 1.1f, blue curve). From 0 to $\pi/2$ and from π to $3\pi/2$, both replicas move uphill, whereas from $\pi/2$ to π and from $3\pi/2$ to 2π, they both move downhill. This can be remedied by using four instead of only two replicas, placed at equal distances along the cable. This working principle is now more reminiscent of the real-life situation of a conveyor belt. In this four-replica setup, the $G_\Lambda(\Lambda)$ profile will again be flat (Fig. 1.1f, red curve), and each of the four replicas will sample the entire λ-range homogeneously.

Finally, consider a more realistic free-energy profile $G(\lambda)$ (Fig. 1.1g). In the general case, the $G_\Lambda(\Lambda)$ profile will never be exactly flat irrespective of the number K of replicas. However, by adding more and more replicas to the conveyor belt, the features of the $G_\Lambda(\Lambda)$ profile can be progressively reduced in magnitude (as illustrated for $K = 2$, 4, 8 or 16 replicas by the blue, red, green and orange curves of Fig. 1.1h). Note that we restrict the choice of K to even values, in order to have always the same number of replicas moving forward and backward. Although the choice of an odd number of replicas would be acceptable, it is likely to be less
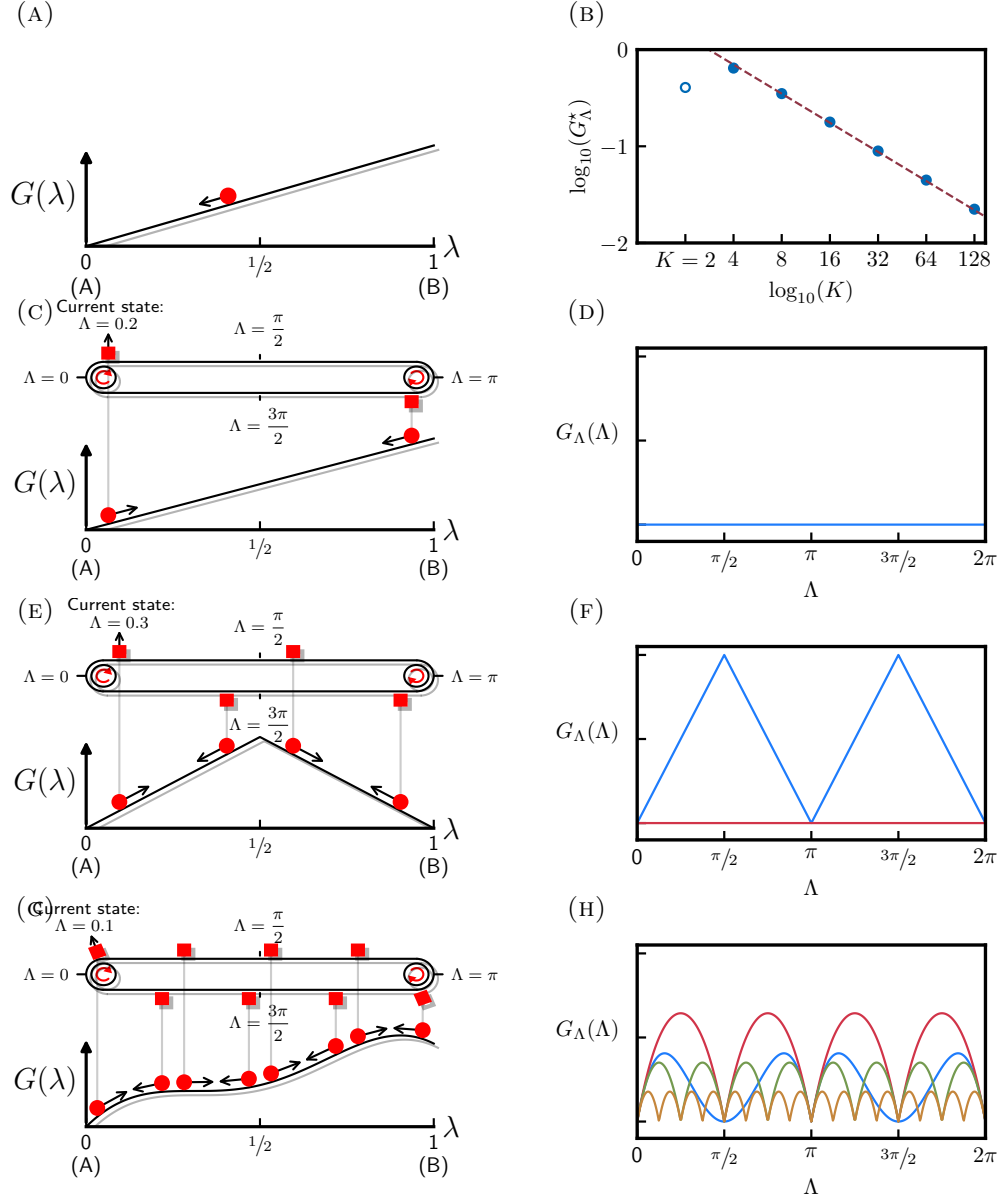
FIGURE 1.1: *Schematic illustration of the conveyor belt thermodynamic integration (CBTI) approach for different free-energy profiles $G(\lambda)$ and numbers of replicas K. The free-energy profile $G(\lambda)$ is shown on the left along with the conveyor belt for an inclined plane (c), a piecewise-linear curve (e) and a more realistic curve (g), the latter curve corresponding to the function $G(\lambda) = \sin(\pi\lambda)(\lambda - 0.5)^2 + 0.25\lambda$. The corresponding free-energy profiles $G_\Lambda(\Lambda)$ along the conveyor belt advance variable Λ (Eq. 1.16) are shown on the right for $K = 2$ (blue), $K = 4$ (red), $K = 8$ (green) or $K = 16$ (orange). The top-left panel (a) illustrates the situation of plain unbiased λD, where the system would "roll down" the slope and keep sampling the neighborhood of the state with the lowest free energy. The top-right panel (b) shows the height $G_\Lambda^\star$ of the residual barriers in the free-energy profile $G_\Lambda(\Lambda)$ considering this illustrative function and increasing values of K. The data is represented in logarithmic form and a linear line of slope -1 is fitted to the filled circles.*

favorable (especially at small K), because one extra replica would always move in either of the two directions.

Considering the illustrative $G(\lambda)$ of Fig. 1.1g, the magnitude $G_\Lambda^\star$ of the variations in $G_\Lambda(\Lambda)$ is shown in Fig. 1.1b as a function of K. As discussed in Sect. 1.A, these variations can be interpreted as the residual of a K-point trapezoid quadrature approximation to the vanishing integral of the derivative of an even periodic function over one period, namely that of the $G(\lambda)$ profile after mirroring. Quantitatively, this interpretation shows that the magnitude $G_\Lambda^\star$ of these variations decreases at least as K^{-1} in the limit of large K. Note that the convergence to a flat profile (*i.e.* the decrease of the barrier heights towards zero upon increasing K) does not need to be regular for small K (see *e.g.* the comparatively low variations for the blue curve in Fig. 1.1h) and that it can be stronger than K^{-1} for large K if $G(\lambda)$ presents particular continuity/symmetry properties. By using a sufficient number of replicas, one may thus ensure a quasi-homogeneous sampling of the λ-range by each replica, even in the absence of coordinate transformation or biasing potential. If desired, a memory-based biasing potential may still be applied to further homogenize the sampling. However, the LE build-up time can be considerably reduced relative to that needed in a corresponding single-system λ-LEUS simulation, because this biasing potential only needs to be applied to the Λ-variable and no longer to the λ-variable. Thus, it has to compensate for comparatively small $G_\Lambda(\Lambda)$ variations. Furthermore, it only needs to be constructed over a limited Λ-range, considering that $G_\Lambda(\Lambda)$ is periodic with period $2\pi K^{-1}$ as well as even over this $2\pi K^{-1}$ interval.

Irrespective whether a biasing potential is employed or not, the definition of a free-energy estimator for the CBTI scheme, *i.e.* a procedure to construct $G(\lambda)$ based on the simulation results for the multiple-replica system, is not as trivial[?] as in the single-system λD or λ-LEUS cases. Here, CBTI is analyzed using a TI-like estimator[?] relying on the quadrature integration of the average Hamiltonian λ-derivative binned along λ considering all replicas simultaneously. The associated quadrature error can be kept very low by using a large number of bins, which is rendered possible by the continuous and quasi-homogeneous λ-sampling. Although this estimator is very robust, we note that it may not be optimal in terms of statistical efficiency.[? ? ? ? ? ? ? ?]

In the present chapter, we provide the mathematical/physical formulation of the proposed CBTI scheme, and report an initial application of the method to an illustrative alchemical transformation: the mutation of a methanol molecule to a dummy (non-interacting) skeleton in water, giving access to the hydration free energy of the molecule.

1.2   THEORY

1.2.1 ALCHEMICAL FREE-ENERGY CALCULATIONS

The goal of path-dependent alchemical free-energy calculations is to evaluate the free-energy difference ΔG between two states A and B of a molecular system, by introducing a coupling scheme relying on a parameter λ, and sampling along the so-defined λ-path. The two states have the same number $3N$ of degrees of freedom, but distinct Hamiltonian functions $\mathcal{H}_A(\mathbf{x})$ and $\mathcal{H}_B(\mathbf{x})$, respectively, where $\mathbf{x} = (\mathbf{r}, \mathbf{p})$ is the $6N$-dimensional phase-space vector representative of a microscopic system configuration, $\mathbf{r}$ and $\mathbf{p}$ being the corresponding coordinate and momentum vectors. The coupling parameter is introduced into a hybrid Hamiltonian $\mathcal{H}(\mathbf{x}; \lambda)$ satisfying the boundary conditions $\mathcal{H}(\mathbf{x}; 0) = \mathcal{H}_A(\mathbf{x})$ and $\mathcal{H}(\mathbf{x}; 1) = \mathcal{H}_B(\mathbf{x})$, where the semi-colon indicates a parametric dependence.

Since the proposed CBTI scheme encompasses features of both λD and TI, these two approaches are summarized briefly in the next two subsections. The following three subsections then describe in turn the basis of the CBTI scheme, its free-energy estimator and the application of a biasing potential.

1.2.2 λ-DYNAMICS (λD)

In the λD scheme,[? ? ? ? ? ? ? ?] the coupling parameter λ is assigned a mass m_λ and a momentum p_λ, and considered to be an additional pseudo-conformational degree of

freedom of the system. The Hamiltonian of the extended system is defined as

$$\mathcal{H}^{\star}(\mathbf{x}, \lambda) = \mathcal{H}(\mathbf{x}; \lambda) + \frac{p_{\lambda}^{2}}{2m_{\lambda}} \ , \tag{1.1}$$

where the star refers to an extended system in which λ is now a variable and no longer a parameter (thus the replacement of the semi-colon by a comma). This leads to the additional equation of motion

$$\ddot{\lambda} = \frac{\dot{p}_{\lambda}}{m_{\lambda}} = -\frac{1}{m_{\lambda}} \frac{\partial \mathcal{H}(\mathbf{x}; \lambda)}{\partial \lambda} \tag{1.2}$$

for propagating the λ-variable, where a dot over a variable indicates its time derivative.

The free-energy difference ΔG between the two physical end-states can then in principle be calculated based on a single thermostated MD simulation of the extended system, as

$$\Delta G = -\frac{1}{\beta} \ln \frac{\langle \delta(\lambda - 1) \rangle^{\star}}{\langle \delta(\lambda) \rangle^{\star}} \ , \tag{1.3}$$

where $\beta = (k_B T)^{-1}$, k_B being the Boltzmann constant and T the absolute temperature, δ is the Dirac delta function, and $\langle \cdots \rangle^{\star}$ denotes ensemble averaging for the extended system (*i.e.* over the joint trajectories of $\mathbf{x}$ and λ). In practice, the δ-functions in Eq. 1.3 must be replaced by two finite end-state bins (λ-cutoff), sufficiently large for proper statistics but also sufficiently small for avoiding distortions due to averaging at the end-states.[?]

1.2.3 THERMODYNAMIC INTEGRATION (TI)

In the original TI scheme,[? ? ?] a set of K replicas of the system are simulated in parallel at fixed predefined λ-values in the range $[0, 1]$. Since the replicas are entirely decoupled from each other, the simulations can be performed serially as well. However, TI extensions including the HRE scheme[? ? ?] and the HRP scheme[? ? ?] introduce a coupling in the form of λ-value exchanges, in which case the simulations must really be carried out in parallel. The same will apply to the proposed CBTI scheme, where

the coupling involves a synchronization of the dynamical λ-variations.

Considering all replicas $k = 0 \ldots K-1$ as the members of a replica system, one may note the corresponding $6K \times N$-dimensional phase-space vector as $\mathbf{X} = \{\mathbf{x}_k\}$ and the corresponding K-dimensional vector containing the fixed λ-values as $\boldsymbol{\lambda} = \{\lambda_k\}$. In plain TI, the Hamiltonian of the replica system is defined as

$$\mathcal{H}^\dagger(\mathbf{X}; \boldsymbol{\lambda}) = \sum_{k=0}^{K-1} \mathcal{H}(\mathbf{x}_k; \lambda_k) \ , \tag{1.4}$$

where the dagger refers to a replica system, and $\boldsymbol{\lambda}$ is here a parameter vector (thus the semi-colon). Because the Hamiltonian of Eq. 1.4 involves no coupling term between the replicas, the dynamics of a replica k is independent from that of the other replicas and solely depends on λ_k.

The free energy difference ΔG between the two states can then be calculated based on a single thermostated MD simulation of the replica system, as

$$\Delta G = \int_0^1 \mathrm{d}\lambda' \left\langle \frac{\partial \mathcal{H}(\mathbf{x}; \lambda)}{\partial \lambda} \right\rangle_{\lambda'} \approx \sum_{k=0}^{K-1} w_k \left\langle \frac{\partial \mathcal{H}^\dagger(\mathbf{X}; \boldsymbol{\lambda})}{\partial \lambda_k} \right\rangle^\dagger = \sum_{k=0}^{K-1} w_k \left\langle \frac{\partial \mathcal{H}(\mathbf{x}_k; \lambda_k)}{\partial \lambda_k} \right\rangle^\dagger ,$$

$$\tag{1.5}$$

where the w_k are quadrature weights for the numerical integration,[?][?][?] $\langle \cdots \rangle_\lambda$ denotes ensemble averaging for a single system (*i.e.* over $\mathbf{x}$) at the given λ value, and $\langle \cdots \rangle^\dagger$ denotes ensemble averaging for the replica system (*i.e.* over $\mathbf{X}$).

In the above form, TI has long been the workhorse of alchemical free-energy calculations. The method is extremely robust in the sense that the accuracy of the calculated ΔG can always be systematically improved (more λ-points, longer equilibration or/and sampling times). However, it is not necessarily the most efficient method to determine ΔG up to a certain accuracy, due to possible sub-optimalities in the coupling scheme,[?][?][?][?][?][?] the protocol design[?][?][?][?] the free-energy estimator,[?][?][?][?][?][?][?] and the orthogonal sampling.[?][?][?][?][?]

1.2.4 CONVEYOR BELT THERMODYNAMIC INTEGRATION (CBTI)

The proposed CBTI scheme encompasses features of both λD and TI. Similarly to TI, it is based on the simulation of a replica system involving K copies of the molecular system of interest, where K is taken to be even. And similarly to λD, the individual replicas are extended systems, for which the associated λ_k-variable is allowed to evolve along the simulation. However, the evolutions of these λ_k-variables are not independent. They are coupled to each other by means of a sequence of hard constraints, so that they follow the course of a conveyor belt (CB). Thus, they are entirely determined by a single dynamical variable Λ, following the scenario depicted in Fig. 1.1 and discussed in the Introduction section.

The variable Λ is a continuous real variable representing the overall advance of the CB, successive multiples of 2π corresponding to as many full rotations. Given Λ and K, the λ-value λ_k associated with a system k on the CB is obtained as

$$\lambda_k(\Lambda) = \zeta\left(\Lambda + k\Delta\Lambda\right) \ , \tag{1.6}$$

with

$$\Delta\Lambda = 2\pi K^{-1}. \tag{1.7}$$

Here, the function ζ is a continuous and periodic zig-zag function of period 2π and image range $[0, 1]$, defined over the reference period $[0, 2\pi)$ as

$$\zeta(\theta) = \begin{cases} \pi^{-1}\theta & \text{if } \theta < \pi \\ 2 - \pi^{-1}\theta & \text{if } \theta \geq \pi \end{cases} \quad \text{for } \theta \in [0, 2\pi) \ , \tag{1.8}$$

where the $[\cdot, \cdot)$ indicates an interval that is open to the right side, *e.g.* $[0, 2\pi)$ includes 0 but excludes 2π. An advance of the CB by $\Delta\Lambda$ corresponds to a cyclic permutation of the K replicas, each system moving by one position forward along the CB, *i.e.* $\lambda_k(\Lambda + \Delta\Lambda) = \lambda_{k+1}(\Lambda)$ for $k < K - 1$, along with $\lambda_{K-1}(\Lambda + \Delta\Lambda) = \lambda_0(\Lambda)$. For this reason, the increment $\Delta\Lambda$ will be further referred to as one shift of the CB.

The system $k = 0$ can be viewed as a reference system, as $\lambda_0 = \pi^{-1}\Lambda$ for $0 \le \Lambda < \pi$ and $\lambda_0 = 2 - \pi^{-1}\Lambda$ for $\pi \le \Lambda < 2\pi$. Since K is chosen to be even, an increase of Λ always corresponds to an increase of λ_k for half of the systems (forward-moving side of the CB) and a decrease of λ_k for the other half of the systems (backward-moving side of the CB). This choice also implies that Λ-values which are integer multiples of the CB shift $\Delta\Lambda$ correspond to situations where there is one system in state A and one system in state B. Since an advance of the CB variable by 2π leaves the replica system invariant, the variable Λ will commonly be refolded into the reference period $[0, 2\pi)$ when illustrating the results of the CBTI method.

The Hamiltonian of the extended replica system is defined as

$$\mathcal{H}^{\dagger\star}(\mathbf{X}, \Lambda) = \mathcal{H}^{\dagger}(\mathbf{X}; \boldsymbol{\lambda}) + \frac{p_\Lambda^2}{2m_\Lambda} \quad \text{with} \quad \boldsymbol{\lambda} = \boldsymbol{\lambda}(\Lambda), \tag{1.9}$$

where $\mathcal{H}^{\dagger}$ is defined as in TI by Eq. 1.4 and $\boldsymbol{\lambda}(\Lambda)$ by Eq. 1.6. In analogy with Eq. 1.2, the resulting equation of motion for Λ reads

$$\ddot{\Lambda} = \frac{\dot{p}_\Lambda}{m_\Lambda} = -\frac{1}{m_\Lambda}\frac{\partial \mathcal{H}^{\dagger}(\mathbf{X}; \boldsymbol{\lambda})}{\partial \Lambda} \tag{1.10}$$

where

$$\frac{\partial \mathcal{H}^{\dagger}(\mathbf{X}; \boldsymbol{\lambda})}{\partial \Lambda} = \sum_{k=0}^{K-1} \frac{\partial \mathcal{H}(\mathbf{x}_k; \lambda_k)}{\partial \lambda_k}\frac{\mathrm{d}\lambda_k}{\mathrm{d}\Lambda} = \sum_{k=0}^{K-1} \frac{\partial \mathcal{H}(\mathbf{x}_k; \lambda_k)}{\partial \lambda_k}\zeta'\left(\Lambda + 2\pi K^{-1}k\right) \ . \tag{1.11}$$

Here, the function ζ' is the derivative of the zig-zag function of Eq. 1.8, given over the reference period $[0, 2\pi)$ by

$$\zeta'(\theta) = \begin{cases} \pi^{-1} & \text{if } \theta < \pi \\ -\pi^{-1} & \text{if } \theta \ge \pi \end{cases} \quad \text{for } \theta \in [0, 2\pi) \ . \tag{1.12}$$

Formally, the derivative is not defined when θ is an integer multiple of π, *i.e.* for a system that is exactly in one of the physical end-states A or B. In this case, the value of ζ' has been arbitrarily set to π^{-1} for even multiples and $-\pi^{-1}$ for odd multiples. This has a negligible impact in practice, as it only concerns a series of

infinitesimal points over the entire Λ-range, *i.e.* infinitesimally few configurations along a CBTI simulation. For example, when using double-precision floating-point arithmetics (including denormalized numbers), their probability of occurrence is on the order of $10^{-324}K$ (*i.e.* a single expected occurrence over a simulation lasting about $10^{292}K^{-1}$ times the age of the universe with a 2 fs timestep). Neither does going over the discontinuity within a timestep represent a source of non-conservativeness. The concerned replica will merely bounce back the corresponding physical end-state with a reversion of its velocity, akin to a particle reflected elastically by a hard wall (delta-function force). If desired, these exceptional points could be handled more formally by altering the definition of ζ, *e.g.* by smoothing its tips in a narrow range around 0 and π.

The λ_k-dynamics of the individual systems is entirely specified by Eq. 1.10 to propagate Λ along with Eq. 1.6 to calculate the λ_k-values from the current Λ. Alternatively, one may write an equation of motion for the λ_k-variables of the individual replicas by combining the two equations (along with Eq. 1.11) as

$$\ddot{\lambda}_k = \ddot{\Lambda}\frac{\mathrm{d}\lambda_k}{\mathrm{d}\Lambda} + \dot{\Lambda}^2\frac{\mathrm{d}^2\lambda_k}{\mathrm{d}\Lambda^2}$$
$$= -\frac{1}{m_\Lambda}\left(\sum_{l=0}^{K-1}\frac{\partial\mathcal{H}(\mathbf{x}_l;\lambda_l)}{\partial\lambda_l}\zeta'\left(\Lambda + 2\pi K^{-1}l\right)\right)\zeta'\left(\Lambda + 2\pi K^{-1}k\right) \ . \qquad (1.13)$$

Note that the term in $\dot{\Lambda}^2$ vanishes since the second derivative ζ'' of ζ is zero (except at the exceptional singular points). However, it should not be overlooked if one decides to use a different function ζ. Introducing the vector $\mathbf{D}$ and the symmetric Λ-dependent matrix $\underline{\mathbf{C}}$ defined by their components as

$$D_k = \frac{\partial\mathcal{H}(\mathbf{x}_k;\lambda_k)}{\partial\lambda_k} \quad \text{and} \quad C_{kl}(\Lambda) = \pi^2\zeta'\left(\Lambda + 2\pi K^{-1}k\right)\zeta'\left(\Lambda + 2\pi K^{-1}l\right) \ , \ (1.14)$$

Eq. 1.13 can be rewritten in an elegant matrix form as

$$\ddot{\boldsymbol{\lambda}} = -\frac{\underline{\mathbf{C}}(\Lambda)}{\pi^2 m_\Lambda}\mathbf{D} \ . \tag{1.15}$$

The elements of the symmetric matrix $\underline{\mathbf{C}}$ are either -1 (pair of systems currently on opposite sides of the CB, and thus moving in opposite directions) or +1 (pair of systems currently on the same side of the CB, and thus moving in the same direction). The diagonal elements are all +1, and the other +1 values surround the diagonal (line- and column-wise), the rest being -1 values. Because K is even, the two types of values are always equally represented in the matrix, specific locations depending on Λ. Note that the variable Λ itself still needs to be explicitly propagated using Eq. 1.10.

For a given configuration $\mathbf{X}$ of the replica system, the Hamiltonian $\mathcal{H}^\dagger$ of Eq. 1.4 (together with Eq. 1.6) is periodic in Λ with a period 2π corresponding to a full rotation of the CB. However, because the Hamiltonians of the individual replicas are identical, upon ensemble averaging over $\mathbf{X}$, one expects the calculated properties to be periodic over Λ with a smaller period $\Delta\Lambda$, corresponding to one shift of the CB. This is in particular the case for the probability distribution $P(\Lambda)$ along Λ and the associated free-energy profile $G_\Lambda(\Lambda)$, given by

$$
G_\Lambda(\Lambda) = G_\Lambda(0) + \int_0^\Lambda \mathrm{d}\Lambda' \left\langle \frac{\partial \mathcal{H}^\dagger(\mathbf{X}; \boldsymbol{\lambda})}{\partial \Lambda} \right\rangle_{\Lambda'}^\dagger = \tilde{G}_\Lambda(0) + \sum_{k=0}^{K-1} G(\lambda_k) \quad \text{with } \boldsymbol{\lambda} = \boldsymbol{\lambda}(\Lambda) ,
$$

(1.16)

where $\boldsymbol{\lambda}(\Lambda)$ is defined by Eq. 1.6, $\langle \cdots \rangle_\Lambda^\dagger$ denotes ensemble averaging for the replica system (*i.e.* over $\mathbf{X}$) at the given Λ value, and the second equality follows from Eqs. 1.4 and 1.5 (the unknown constant $\tilde{G}_\Lambda(0)$ is equal to $G_\Lambda(0)$ increased by a sum of $-G(\lambda_k(0))$ offsets).

Owing to this periodicity over a smaller interval, it is convenient to introduce a fractional advance variable $\tilde{\Lambda}$ defined as

$$
\tilde{\Lambda} = \gamma(\Lambda, \Delta\Lambda),
$$

(1.17)

where

$$
\gamma(\theta, \theta_o) = \theta_o \left(\theta_o^{-1}\theta - \lfloor \theta_o^{-1}\theta \rfloor \right)
$$

(1.18)

returns the part of θ in excess of the closest lower integer multiple of θ_o. In contrast to Λ, which is an unbounded variable, $\tilde{\Lambda}$ only spans a finite definition interval $[0, \Delta\Lambda)$. At full convergence, any average property binned as a function of Λ over the interval $[0, 2\pi)$ will consist of K successive repeats of the same property binned as a function of $\tilde{\Lambda}$ over its definition interval $[0, \Delta\Lambda)$, as observed in Fig. 1.1h for the free energy $G_\Lambda(\Lambda)$. Accordingly, in the absence of full convergence along Λ, binning as a function of $\tilde{\Lambda}$ over the interval $[0, \Delta\Lambda)$ followed by K-fold replication provides an efficient way to construct a more accurate representation of any Λ-resolved average quantity. In fact, the definition interval of $\tilde{\Lambda}$ could be further halved by noting that, upon ensemble averaging over $\mathbf{X}$ and for any Λ value that is an integer multiple of $\Delta\Lambda$, a forward move of the CB produces the same result as a backward move of the same magnitude. Consequently, Λ-resolved average properties are even over successive $2\pi K^{-1}$ intervals, as also observed in Fig. 1.1h for the free energy $G_\Lambda(\Lambda)$. The corresponding information is thus entirely encompassed in an interval of size $\Delta\Lambda/2$.

The normalized probability distribution $p(\lambda)$ along the coupling variable λ considering all the replicas is defined by

$$p(\lambda) = K^{-1} \sum_{k=0}^{K-1} \langle \delta(\lambda_k - \lambda) \rangle^{\dagger\star} \ , \tag{1.19}$$

where $\langle \cdots \rangle^{\dagger\star}$ denotes ensemble averaging for the extended replica system (*i.e.* over the joint trajectories of $\mathbf{X}$ and $\boldsymbol{\lambda}$). At full convergence, this probability over the interval $[0, 1]$ will consist of $K/2$ successive repeats of the corresponding distribution over the interval $[0, 2K^{-1})$. More precisely, the distribution $p(\lambda)$ is related to the distribution $\tilde{P}(\tilde{\Lambda})$ of $\tilde{\Lambda}$ over interval $[0, \Delta\Lambda)$ as

$$p(\lambda) = \Delta\Lambda \, \tilde{P}(\pi\gamma(\lambda, \pi^{-1}\Delta\Lambda)) \ . \tag{1.20}$$

In plain words, this means that $\tilde{P}(\tilde{\Lambda})$ is the relevant quantity in terms of sampling along the coupling variable λ. If it is close to uniform over the range $[0, \Delta\Lambda)$, then $p(\lambda)$ will also be close to uniform over the range $[0, 1]$. Here again, it is noted that $p(\lambda)$ is also even over the interval $[0, 2K^{-1})$, and could be mapped to a $\tilde{\Lambda}$ value

defined over an interval of size $\Delta\Lambda/2$ instead of $\Delta\Lambda$ if desired, as

$$
p(\lambda) = \begin{cases} \Delta\Lambda\,\tilde{P}(\pi\gamma(\lambda,\pi^{-1}\Delta\Lambda)) & \text{if } \gamma(\lambda,\pi^{-1}\Delta\Lambda) < (2\pi)^{-1}\Delta\Lambda \\ \Delta\Lambda\,\tilde{P}(\pi\gamma(1-\lambda,\pi^{-1}\Delta\Lambda)) & \text{otherwise} \end{cases} . \qquad (1.21)
$$

1.2.5 CBTI Free-energy Estimator

Due to the constraints coupling the λ_k-values of the K replicas, the function $p(\lambda)$ of Eq. 1.19 is by no means a Boltzmann distribution in terms of the single-system Hamiltonian. In fact, as seen above, it consists at full convergence of $K/2$ successive repeats of the same even curve. In addition, compared to the Boltzmann distribution, it will be significantly flatter. On the one hand, the smaller amplitude of variations are desired, as they will lead to more homogeneous sampling and are expected to ease transitions along Λ (up to the limit imposed by the speed of random diffusion). On the other hand, it is no longer possible to evaluate the free-energy difference ΔG directly from $p(\lambda)$ in analogy with the λD expression of Eq. 1.3. However, since the dynamics remains Hamiltonian and the coupling between replicas does not involve the configurational degrees of freedom, the change from TI to CBTI does not affect the conditional probabilities $\mathcal{P}(\mathbf{x}|\lambda)$. Thus, configurational ensemble averages sorted by λ-values will remain identical to those one would obtain from TI (or from HRE/HRP or λD). As a result, ΔG can still be obtained by integrating over the average Hamiltonian derivative binned as a function of λ considering all replicas simultaneously, in analogy with the TI expression of Eq. 1.5. Note that the exceptional points of the function ζ' (discussed previously in the context of Eq. 1.12) have no influence on the integration, as they represent finite discontinuities over infinitesimal ranges.

In practice, ΔG is calculated here based on a single thermostated MD simulation

of the extended replica system, as

$$\Delta G = \int_0^1 d\lambda' K^{-1} \sum_{k=0}^{K-1} \left\langle \frac{\mathcal{H}(x_k; \lambda_k)}{\partial \lambda_k} \delta(\lambda_k - \lambda') \right\rangle^{\dagger\star} \tag{1.22}$$

$$\approx \sum_{j=0}^{J-1} \left\langle \frac{\sum_{k=0}^{K-1} \frac{\mathcal{H}(x_k; \lambda_k)}{\partial \lambda_k} \alpha(\lambda_k, j; J)}{\sum_{k=0}^{K-1} \alpha(\lambda_k, j; J)} \right\rangle^{\dagger\star},$$

where

$$\alpha(\theta, j; J) = \begin{cases} 1 & \text{if } j \leq J\theta < j+1 \\ 0 & \text{otherwise} \end{cases} \tag{1.23}$$

is a binning function corresponding to a discretization of the λ-interval $[0, 1]$ using J bins. The approximation in Eq. 1.22 corresponds to a simple forward rectangular quadrature, where the Hamiltonian derivative is averaged over the J successive bins considering all replicas. Since $\tilde{P}(\tilde{\Lambda})$, and thus $p(\lambda)$, will typically be close to homogeneous, J can be taken very large, resulting in a negligible quadrature error. For example, if K replicas sample L configurations each, the number of data points per bin will be close to KL/J, with limited variations across bins. Defining the maximal allowed value $J_{\max}$ as the highest value of J for which empty bins (vanishing denominator in the ensemble average of Eq. 1.22) never occur, a graph of ΔG evaluated upon increasing J from 1 to $J_{\max}$ will rapidly level off to a plateau when quadrature errors become negligible. Two variants which do not require the specification of a number of bins are also proposed in Sect. **??** (Eqs. 1.C.1 and 1.C.2).

1.2.6 CBTI with Memory-based Biasing Potential

When using a large number of replicas, the sampling along λ afforded by the CBTI scheme will be close to homogeneous. However, for practical reasons (*e.g.* number of processors available on a computer node), one may wish to use a small number of replicas. In this case, the sampling homogeneity can be enhanced by addition of a biasing potential. It is sufficient to apply this potential to the fractional advance

variable $\tilde{\Lambda}$ over the range $[0, \Delta\Lambda)$. With inclusion of a biasing potential $\mathcal{B}$, Eq. 1.10 becomes

$$\ddot{\Lambda} = -\frac{1}{m_\Lambda}\frac{\partial}{\partial\Lambda}\left(\mathcal{H}^\dagger(\mathbf{X};\boldsymbol{\lambda}) + \mathcal{B}(\tilde{\Lambda})\right) \quad \text{with } \boldsymbol{\lambda} = \boldsymbol{\lambda}(\Lambda) \text{ and } \tilde{\Lambda} = \tilde{\Lambda}(\Lambda) \ , \qquad (1.24)$$

where $\mathcal{H}^\dagger$ is defined by Eq. 1.4, $\boldsymbol{\lambda}(\Lambda)$ by Eq. 1.6 and $\tilde{\Lambda}(\Lambda)$ by Eq. 1.17.

In analogy with the λ-LEUS scheme,[?][?][?][?] this biasing potential can be expressed as a sum of local grid-based spline functions, built in a LE preoptimization phase and frozen in a subsequent US sampling phase. However, the duration of the LE phase can be considerably reduced compared to a single-system λ-LEUS simulation, considering that $\tilde{P}(\tilde{\Lambda})$ is already close to homogeneity in the absence of biasing and that the support interval is reduced to the $\tilde{\Lambda}$-range $[0, \Delta\Lambda)$. The latter interval can actually be further restricted to $[0, \Delta\Lambda/2]$ considering the even symmetry of $\tilde{P}(\tilde{\Lambda})$, *i.e.* by enforcing an even symmetry of $\mathcal{B}$ as well.

Since the application of a biasing potential that only involves the λ_k-variables does not alter the conditional probabilities $\mathcal{P}(\mathbf{x}|\lambda)$, Eq. 1.22 (or the variants of Eqs. 1.C.1 and 1.C.2) can still be employed without any modification to evaluate the free-energy change. In other words, in contrast to the λ-LEUS scheme, the CBTI scheme with the presented TI-like free-energy estimator does not require any reweighting.

1.3   COMPUTATIONAL DETAILS

1.3.1 TEST SYSTEM

As an initial application of the proposed CBTI scheme, we considered here a relatively simple perturbation, namely the conversion of methanol from a fully interacting molecule to a dummy skeleton (no intermolecular interactions) in an aqueous environment at $P = 1\,\text{bar}$ and $T = 298.15\,\text{K}$. The calculations were performed using a modified version of the GROMOS11 program[?][?] along with the parameters of the GROMOS-compatible 2016H66 force field[?] for methanol[?] (united-atom, rigid

bonds, flexible bond-angle) and the simple point charge (SPC) model[?] for water (fully rigid). Since the dummy skeleton retains the intramolecular interactions (here, only the bond-angle), the calculated free-energy change ΔG corresponds directly to minus the hydration free energy of methanol.

Possible issues related to the existence of a singularity[?,?,?] and the insufficient solute-solvent kinetic-energy exchange[?,?,?,?] close to $\lambda = 1$ were alleviated in the usual way, by means of a soft-core scheme[?] for the alchemical coupling and of stochastic dynamics[?,?] (SD) for thermostating the solute and solvent conformational degrees of freedom. In most CBTI simulations, the instantaneous temperature T_Λ of the CB advance variable Λ was also controlled separately by means of a Nosé-Hoover chain thermostat[?] at 298.15 K (eight successive thermostat variables), with a coupling time τ_Λ.

1.3.2 SIMULATIONS SETS

The exploration of the CBTI scheme and the comparison of its performance with that of existing methods was carried out in five successive steps: (1) establishing reference TI results; (2) analyzing the influence of the CBTI parameters (number K of replicas along with the mass-parameter m_Λ and thermostat coupling time τ_Λ of the CB advance variable); (3) investigating the use of a biasing potential; (4) examining the features of the TI-like free-energy estimator (effect of the number J of integration bins and use of equations without specification of J); (5) comparing the results of CBTI with those of existing methods.

The reference TI calculations (Step 1) were performed using $K_{\text{TI}} = 2^n + 1$ equidistant λ-points covering the range $[0, 1]$ with $n = 1,2,...,7$. They involved initial configurations equilibrated for $0.2\,\text{ns}$ starting from the equilibrated configuration at the previous λ-point, and a simulation time of $100 K_{\text{TI}}^{-1}$ ns per λ-point. Each of these calculations, involving a total single-system sampling time of 100 ns, was repeated ten times using different random initial velocities. The integration over the average Hamiltonian derivative was performed based on Eq. 1.5 using the Simpson quadrature rule.[?,?,?]

To explore the influence of the CBTI parameters (Step 2), various combinations of K, m_Λ and τ_Λ were considered in three series of calculations, namely: (i) the choices $m_\Lambda = 16, 160, 800, 1600$ or $3200\,\mathrm{u\,nm}^2$ (where u stands for atomic mass unit, *i.e.* g mol^{-1}), along with $K = 16$ replicas in the absence of thermostat coupling for Λ, *i.e.* with $\tau_\Lambda \to \infty$; (ii) the choices $\tau_\Lambda = 0.05, 0.1, 0.5, 1$ or $2\,\mathrm{ps}$ along with $K = 16$ replicas and $m_\Lambda = 160\,\mathrm{u\,nm}^2$; ($iii$) the choices $K = 8, 16, 32, 64$ or 128 along with $m_\Lambda = 40K^{1/2}\,\mathrm{u\,nm}^2$ and $\tau_\Lambda = 0.5\,\mathrm{ps}$; The parameters (and results) of these three series of simulations, including their durations t_sim, are summarized in Tab. 1.1 (entries 1-15). All these simulations were preceded by $0.2\,\mathrm{ns}$ equilibration. For the the third series (entries 11-15), m_Λ was made proportional to $K^{1/2}$, an arbitrary parameter choice justified by arguments provided in Sect. 1.B, and the five simulations relied on the the same total single-system sampling time of 256 ns. This exploration showed that the CBTI method is rather robust with respect to the choice of its parameters. The values $K = 16$, $m_\Lambda = 160\,\mathrm{u\,nm}^2$ and $\tau_\Lambda = 0.5\,\mathrm{ps}$ were retained as a good combination for the alchemical perturbation considered. For comparison with the TI results of Step 1, ten repeats of the calculation involving this specific choice were performed using different random initial velocities and a total single-system sampling time of 100 ns after $0.2\,\mathrm{ns}$ equilibration.

TABLE 1.1: *Influence of the CBTI parameters in simulations of the aqueous methanol-to-dummy mutation at 298.15 K and 1 bar with the 2016H66 force field.* This table investigates the influence of the parameters selected for the CBTI scheme on the temperature and dynamics of the CB advance variable Λ and on the calculated free-energy change ΔG (for the latter, considering a constant total single-system sampling time of 100 ns). For each simulation, the successive entries are: the index of the simulation (sim), the number K of replicas, the simulation time t_{sim} for the replica system, the mass-parameter m_Λ, the thermostat coupling time τ_Λ (∞ indicates that no coupling is applied), the average temperature T_Λ, the root-mean-square fluctuation $\sigma_{\dot\Lambda}$ of $\dot\Lambda$, the autocorrelation time $\tau_{\dot\Lambda}$ of $\dot\Lambda$, the diffusion coefficient D_Λ (Eq. 1.26), the free-energy difference ΔG calculated using Eq. 1.22 with $J = 500$ (except for entry 11, $J = 200$), the alternative free-energy difference ΔG_{alt} calculated using Eq. 1.C.1, and the approximate free-energy difference ΔG_{app} calculated using Eq. 1.C.2. Error estimates obtained by bootstrapping (no Student t-factor included) are also reported between parentheses for ΔG, ΔG_{alt} and ΔG_{app}. Note that the simulations differ in terms of total single-system sampling time Kt_{sim}. To enable a fair comparison, the free energy-changes and associated errors have been calculated after truncating the all simulations to 100 ns single-system sampling time evenly distributed over all replicas. Associated graphs for the distributions $P(\Lambda)$ of Λ, $P_{\dot\Lambda}$ of $\dot\Lambda$ and $P_{\ddot\Lambda}$ of $\ddot\Lambda$, as well as the mean-square displacements d_Λ of Λ and autocorrelation functions $c_{\dot\Lambda}$ of $\dot\Lambda$ can be found in Figs. 1.3, 1.5 and 1.7, or in Figs. Fig. 1.F.1 - Fig. 1.F.15 and Fig. 1.I.2-Fig. 1.I.4. Simulations 11, 13, 15, 16 and 17 are discussed in the main text. The other simulations are discussed in Sect. 1.B.

sim	K	t_{sim}[ns]	m_Λ[u nm²]	τ_Λ[ps]	T_Λ[K]	$\sigma_{\dot\Lambda}$[ps⁻¹]	$\tau_{\dot\Lambda}$[ps]	D_Λ[ns⁻¹]	ΔG[kJ mol⁻¹]	ΔG_{alt}[kJ mol⁻¹]	ΔG_{app}[kJ mol⁻¹]
1	16	10	16	∞	308.0	0.40	0.02	16.7	21.14 (0.16)	21.23 (0.17)	20.21 (0.32)
2	16	10	160	∞	297.9	0.12	0.61	13.2	21.27 (0.17)	21.32 (0.15)	20.29 (0.31)
3	16	10	800	∞	296.3	0.05	1.90	10.2	21.25 (0.15)	21.34 (0.16)	20.27 (0.32)
4	16	10	1600	∞	304.0	0.04	3.47	10.3	21.28 (0.15)	21.25 (0.17)	20.32 (0.32)
5	16	10	3200	∞	297.7	0.03	5.63	7.4	21.23 (0.16)	21.24 (0.18)	20.11 (0.31)
6	16	10	160	0.05	284.0	0.12	0.37	6.9	21.23 (0.16)	21.04 (0.15)	20.22 (0.31)
7	16	10	160	0.10	291.7	0.12	0.36	7.4	21.48 (0.13)	21.41 (0.16)	20.47 (0.29)
8	16	10	160	0.50	296.4	0.12	0.41	9.1	21.42 (0.14)	21.48 (0.16)	20.45 (0.30)
9	16	10	160	1.00	299.4	0.12	0.49	12.9	21.57 (0.16)	21.48 (0.17)	20.59 (0.30)
10	16	10	160	2.00	296.8	0.12	0.56	13.1	21.48 (0.13)	21.73 (0.15)	20.45 (0.29)
11	8	32	113	0.50	298.4	0.15	0.23	4.4	21.69 (0.40)	21.69 (0.44)	15.65 (0.31)
12	16	16	160	0.50	298.3	0.12	0.41	10.2	21.42 (0.16)	21.48 (0.17)	20.45 (0.33)
13	32	8	226	0.50	300.6	0.10	0.38	6.2	21.44 (0.14)	21.33 (0.15)	21.34 (0.32)
14	64	3	320	0.50	297.2	0.09	0.31	2.5	21.32 (0.14)	21.27 (0.14)	21.29 (0.30)
15	128	2	452	0.50	304.1	0.08	0.24	2.3	21.43 (0.12)	21.61 (0.14)	21.41 (0.32)
16	8	22	113	0.50	296.4	0.15	0.43	15.3	21.48 (0.16)	21.61 (0.17)	19.65 (0.34)
17	16	22	160	0.50	299.3	0.12	0.43	9.4	21.30 (0.13)	21.43 (0.16)	20.61 (0.34)

The application of CBTI with a biasing potential (Step 3) was investigated in the context of simulations with $K = 8$ or 16, both with $m_\Lambda = 40K^{1/2}\,\mathrm{u\,nm}^2$ and $\tau_\Lambda = 0.5\,\mathrm{ps}$. For $K = 8$, the biasing potential $\mathcal{B}$ (Eq. 1.24) was constructed using $N_{\mathrm{gp}} = 34$ basis functions centered at equidistant grid-points $i = 0, ..., N_{\mathrm{gp}} - 1$ over the $\tilde\Lambda$-range $[0, \pi/4]$. The coefficients of the basis-functions i and $N_{\mathrm{gp}} - 1 - i$ with $i = 0, ..., N_{\mathrm{gp}}/2 - 1$ were constrained to be identical, considering the expected even symmetry of $\tilde P(\tilde\Lambda)$. In terms of the CB advance variable Λ, this means that the biasing potential relied in effect on $K(N_{\mathrm{gp}} - 1) = 264$ local functions covering the Λ-range $[0, 2\pi)$, these functions being defined by only 17 independent coefficients. For $K = 16$, $\mathcal{B}$ relied on $N_{\mathrm{gp}} = 18$ basis functions over the $\tilde\Lambda$-range $[0, \pi/8]$, leading to 272 functions

over the Λ-range $[0, 2\pi)$ defined by 9 independent coefficients. Second-order splines[?][?] (of range $\pm 2\delta$ with $\delta = \pi/132$ or $\pi/136$ for $K = 8$ and 16, respectively) were employed[?] as basis functions. An initial build-up force constant $c_{\mathrm{LE}} = 10^{-3}\,\mathrm{kJ\,mol^{-1}}$ was used, which was multiplied by a reduction factor $f_{\mathrm{red}} = 0.1$ after each double-sweep of half the $\tilde{\Lambda}$-range $[0, \pi K^{-1}]$. The duration t_{LE} of the LE build-up phase for the replica system was $= 0.15\,\mathrm{ns}$ for $K = 8$ and $0.07\,\mathrm{ns}$ for $K = 16$, corresponding to only $1.1 - 1.2\,\mathrm{ns}$ total single-system simulation time. The parameters (and results) of these two simulations are summarized in Tab. 1.1 (entries 16 an 17). The duration t_{sim} of the US sampling phases for the replica system were $22\,\mathrm{ns}$, corresponding to total single-system sampling times of $176\,\mathrm{ns}$ and $352\,\mathrm{ns}$ for $K = 8$ and 16, respectively.

To examine the features of the TI-like free-energy estimator (Step 4), the number J of integration bins used in the rectangular quadrature to calculate ΔG (Eq. 1.22) was varied considering the simulations of Steps 2 and 3 above (17 simulations of Tab. 1.1). The resulting ΔG values were also compared to those of the variants ΔG_{alt} (Eq. 1.C.1) and ΔG_{app} (Eq. 1.C.2) proposed in Sect. **??**.

Finally, the results of the CBTI simulations were compared to those of other methods (Step 5), namely TI or HRE using Simpson quadrature as estimator as well as TI using EXTI and MBAR as estimator. For Simpson quadrature, the TI simulations relied on $K_{\mathrm{TI}} = 3, 5, 9, 17, 65$ or 129 equispaced λ-points, and the HRE simulations relied on $K_{\mathrm{HRE}} = 17, 33$ or 65 equispaced replicas with exchange attempts every $\tau_{\mathrm{HRE}} = 0.2\,\mathrm{ps}$. The use of EXTI and MBAR was explored based on TI-like simulations relying on $K_{\mathrm{TI}} = 9$ and 17 equispaced λ-points. For EXTI, the average Hamiltonian derivative was extrapolated during the K_{TI} simulations to 129 equispaced virtual λ-points, and the latter 129 values used in the Simpson quadrature. For MBAR, the Hamiltonian was calculated at 129 equispaced virtual λ-points considering all the configurations sampled in the K_{TI}, and the data combined using the MBAR equation[?] as implemented in pymbar.[?] All the above comparisons were performed at a total single-system sampling time of $100\,\mathrm{ns}$ distributed evenly over all replicas.

Error bars on the calculated ΔG values were estimated in two different ways. For

the calculations involving ten repeats (all TI calculations plus one CBTI simulation), the standard deviation σ of the mean was calculated by scaling that of the ten estimates by the square-root of nine, and the error ϵ on the mean was calculated as $\epsilon = 2.262\sigma$, where 2.262 is the Student t-factor[?] for nine degrees of freedom and a two-sigma confidence interval of 95%. For the individual calculations that were not repeated, the statistical error was estimated by bootstrapping[?][?] (no Student t-factor included) using 100 bootstrap samples. If K replicas (CBTI) or λ-points (all other methods) have generated as many sets of L data points, a sample consists here of K sets of L data points selected randomly (possibly multiple times) from the K original data sets. Note that for TI/EXTI and TI/MBAR, it is essential to perform the bootstrapping based on the data from the K_{TI} real λ-points, and not from the $129 K_{\text{TI}}$ predicted values (the latter procedure would result in underestimated errors due to correlation in the derived data). The bootstrapping error will only be accurate provided that the data from the simulations, written to file every $2\,\text{ps}$, is uncorrelated in time. Normalized autocorrelation functions and characteristic times for the average Hamiltonian derivative in the different TI-simulations are provided in Fig. 1.D.1 and Tab. 1.D.1 to support this assumption.

1.3.3 Simulation Parameters

The simulations involved a cubic computational box containing one methanol and 1000 water molecules under periodic boundary conditions in the isothermal-isobaric ensemble at $P = 1\,\text{bar}$ and $T = 298.15\,\text{K}$. They were performed using SD by integrating the Langevin equation of motion[?] using the leap-frog scheme[?] (SD variant[?]) with a timestep $\Delta t = 2\,\text{fs}$ and a friction coefficient $\gamma = 10\,\text{ps}^{-1}$. Since the kinetics of the system is irrelevant in this work, SD instead of thermostated MD was used to avoid problems related to insufficient solute-solvent kinetic-energy exchange[?][?][?][?] close to $\lambda = 1$ (dummy-skeleton state). The value of γ corresponds to the coupling time of $0.1\,\text{ps}$ commonly employed in GROMOS simulations[?][?] relying on a weak-coupling[?] thermostat. The average pressure was maintained close to its reference value by isotropic weak coupling[?] using a molecular virial, a coupling time

$\tau_P = 0.5\,\mathrm{ps}$ and a compressibility $\kappa = 4.575 \cdot 10^4\,\mathrm{kJ\,mol^{-1}\,nm^{-3}}$ as commonly used in GROMOS for aqueous biomolecular systems.[?][?] The bond rigidity of methanol and the full rigidity of water were enforced by application of the SHAKE algorithm[?] with a relative geometric tolerance of 10^{-4}. The energies and Hamiltonian derivatives were written to file every $2\,\mathrm{ps}$ for analysis.

The non-bonded interactions were handled by means of a molecule-based twin-range cutoff scheme[?] with short- and long-range cutoff distances set to 0.8 and $1.4\,\mathrm{nm}$, respectively, and an update frequency of 5 timesteps for the short-range pairlist and intermediate-range interactions. The molecule center was the center of geometry for methanol and the oxygen atom for water. A reaction-field correction[?][?] was applied to account for the mean effect of the electrostatic interactions beyond the long-range cutoff distance, using a relative dielectric permittivity of 61 as appropriate for the SPC model.[?] To alleviate issues related to the existence of a singularity[?][?][?] close to $\lambda = 1$ (dummy-skeleton state), the alchemical transformation relied on a soft-core scheme,[?] applied with the parameters $\alpha_{\mathrm{LJ}} = 0.5$ and $\alpha_{\mathrm{CRF}} = 0.5\,\mathrm{nm^2}$.

For the CBTI calculations, the propagation of the Λ variable (Eq. 1.10) preceded that of the conformational degrees of freedom, and was performed with the same timestep Δt. More precisely, the following leap-frog steps were carried out in sequence: $\dot{\Lambda}(t - \Delta t/2) \rightarrow \dot{\Lambda}(t + \Delta t/2)$, $\Lambda(t) \rightarrow \Lambda(t + \Delta t)$, calculate $\boldsymbol{\lambda}$ from Λ using Eq. 1.6, $\dot{\mathbf{r}}(t - \Delta t/2) \rightarrow \dot{\mathbf{r}}(t + \Delta t/2)$ and $\mathbf{r}(t) \rightarrow \mathbf{r}(t + \Delta t)$. Unless otherwise specified (explorative simulations), the CBTI simulations relied on a mass-parameter m_Λ set to $m_\Lambda = 40 K^{1/2}\,\mathrm{u\,nm^2}$, on a coupling time τ_Λ set to $0.5\,\mathrm{ps}$, and on the use of $J = 500$ bins for evaluating ΔG based on Eq. 1.22 ($J = 200$ for the unbiased simulation with $K = 8$).

1.3.4 Trajectory Analysis

The dynamics of the replica system in the CBTI simulations was characterized by monitoring the distribution P of the Λ variable, the average temperature T_Λ, the distribution $P_{\dot{\Lambda}}$ of $\dot{\Lambda}$, the associated root-mean-square fluctuation $\sigma_{\dot{\Lambda}}$, the normalized autocorrelation function $c_{\dot{\Lambda}}$ of $\dot{\Lambda}$, the associated autocorrelation time $\tau_{\dot{\Lambda}}$, the mean-square displacement d_Λ of Λ as a function of time, the associated diffusion coefficient

D_Λ, and the distribution $P_{\ddot{\Lambda}}$ of $\ddot{\Lambda}$.

The distribution $P_{\dot{\Lambda}}$ of $\dot{\Lambda}$ can be compared to the analytical one-dimensional Maxwell-Boltzmann velocity distribution[?]

$$P_{\dot{\Lambda}}^{\mathrm{MB}}(\dot{\Lambda}) = \left(\frac{\beta m_\Lambda}{2\pi}\right)^{\frac{1}{2}} \mathrm{e}^{-\frac{\beta m_\Lambda}{2}\dot{\Lambda}^2}. \tag{1.25}$$

The diffusion coefficient D_Λ was calculated from the mean-square displacement d_Λ of Λ as a function of time t according to the one-dimensional Einstein equation[?]

$$D_\Lambda = \lim_{t\to\infty} \frac{d_\Lambda(t)}{2t} \qquad \text{with} \qquad d_\Lambda(t) = \left\langle [\Lambda(\tau + t) - \Lambda(\tau)]^2 \right\rangle_t, \tag{1.26}$$

where $\langle\cdots\rangle_t$ denotes averaging over τ (all possible time origins) at constant t. Note that this equation must be applied to the unbounded variable Λ, *i.e.* without refolding to the reference interval $[0, 2\pi)$. The infinite-time limit was replaced in practice by a linear least-squares fit over the time range 0 to $0.15\,\mathrm{ns}$.

All the graphs presented in this chapter were generated with Python (www.python.org) and the Matplotlib library.[?]

1.4   RESULTS AND DISCUSSION

1.4.1 REFERENCE TI CALCULATIONS

The curve for the average Hamiltonian derivative as a function of λ corresponding to the aqueous methanol-to-dummy mutation is shown in Fig. 1.2a (blue curve, scale on the left), as averaged over the ten repeats of the $100\,\mathrm{ns}$ TI calculation with 129 equispaced λ-points. The corresponding running integral, *i.e.* the free-energy profile $G(\lambda)$, is also shown (orange curve, scale on the right). In view of the large amount of statistics (total $1\,\mu\mathrm{s}$ sampling, *i.e.* clearly an overkill for such a calculation), the curve is perfectly smooth and extremely well converged. The average Hamiltonian derivative is large and positive below about 0.5 and becomes smaller and negative

thereafter, leading to a maximum in the free-energy profile. The shape of these curves is determined both by the hydration physics of methanol and by the choice of the employed (soft-core) coupling scheme.
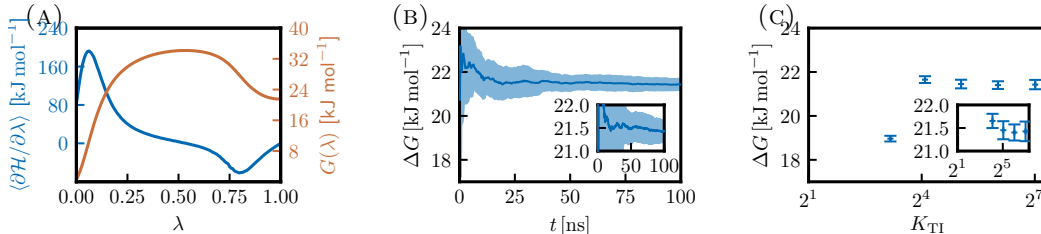

FIGURE 1.2: *Convergence properties of the TI calculations for the aqueous methanol-to-dummy mutation at 298.15 K and 1 bar with the 2016H66 force field.* This figure provides the reference free-energy profile for the perturbation considered, and investigates the convergence properties of TI with different numbers K_{TI} of equidistant λ-points (based on ten repeats of 100 ns TI calculations). Panel (a) shows the average Hamiltonian derivative as a function of λ (blue curve, scale on the left) and the corresponding free-energy profile (orange, scale on the right). The curves are averaged over ten repeats of 100 ns TI calculations with 129 equidistant λ-points. Panel (b) shows the running ΔG estimates for each of the ten repeats (dashed curves) as a function of the total sampling time t per repeat, along with the corresponding mean and error on the mean (95% confidence interval) over the repeats (thick blue curve and shaded area). Panel (c) shows ΔG estimates considering ten repeats of 100 ns TI calculations using $K_{\text{TI}} = 2^n + 1$ equidistant λ-points with $n = 1,2,...,7$ (individual colored circles), along with the corresponding mean and error on the mean (95% confidence interval) over the repeats (blue crosses and error bars). Figs. equivalent to Panel (b) for the TI calculations with fewer λ-points are provided in Fig. 1.E.1. Numerical values at full sampling time are reported in Tab. 1.2 and Tab. 1.E.1.

The convergence properties of these TI calculations as a function of the total sampling time per repeat are illustrated in Fig. 1.2b, in the form of running ΔG estimates for each of the ten repeats (dashed curves), along with the associated mean and error (95% confidence interval) on the mean over the repeats (thick blue curve and shaded area). The latter mean and error at full sampling time, reported in Tab. 1.2, evaluate to $21.43 \pm 0.21\,\text{kJ}\,\text{mol}^{-1}$. This result (with a minus sign) is in quantitative agreement with the experimental value for the hydration free energy of methanol at 298.15 K and 1 bar, namely[?] $-21.4\,\text{kJ}\,\text{mol}^{-1}$.

Corresponding graphs for TI calculations involving different numbers of equidistant λ-points ($K_{\text{TI}} = 2^n + 1$ with $n = 1, 2, \ldots, 7$) at identical total sampling time (100 ns for each of the ten repeats) are shown in Fig. 1.E.1. The associated ΔG means and errors at full sampling time are reported numerically in Tab. 1.2 (see Tab. 1.E.1 for the values of the individual repeats). The results of Tab. 1.2 also include a comparison of different error estimates (all excluding the Student t-factor), namely:

TABLE 1.2: *Repeat statistics over TI and CBTI calculations for the aqueous methanol-to-dummy mutation at* $298.15\,\mathrm{K}$ *and* $1\,\mathrm{bar}$ *with the 2016H66 force field.* This table investigates the convergence properties of the calculated free-energy change ΔG considering TI calculations with different numbers of λ-points along with a CBTI simulation with 16 replicas (average and various error estimates, based on ten calculation repeats involving each at a constant total single-system sampling time of $100\,\mathrm{ns}$) The TI calculations involved $K_{\mathrm{TI}} = 2^n + 1$ equispaced λ-points with $n = 1, 2, ..., 7$. The CBTI simulations involved $K = 16$ replicas, along with $m_\Lambda = 160\,\mathrm{u\,nm^2}$ and $\tau_\Lambda = 0.5\,\mathrm{ps}$ (no biasing potential). In both cases, the total single-system sampling time was $100\,\mathrm{ns}$, and the calculation was repeated ten times. Considering each set of repeats, the successive entries are: the free-energy change ΔG_{avg} averaged over the repeats; the corresponding standard deviation σ of the mean; the associated error ϵ on the mean; alternative error estimates σ_{B} and σ_{G} calculated by bootstrapping (no Student t-factor included). The quantity σ is calculated by scaling the standard deviation of the ten ΔG values by the square-root of nine. The error estimate ϵ is then calculated as $\epsilon = 2.262\sigma$, corresponding to a two-sigma confidence interval of 95%. The estimate σ_{B} corresponds to a bootstrap error calculated from the concatenated trajectories of the ten repeats (for TI, concatenated separately at each λ-point). The estimate σ_{G} (TI only) is obtained by calculating the bootstrap errors on the average Hamiltonian derivatives at all λ-point, and propagating them assuming a Gaussian distribution of the estimates, as proposed in Refs.[?][?] The ΔG values for the individual repeats can be found in Tabs. 1.E.1 and 1.G.1.

	TI							CBTI
n	1	2	3	4	5	6	7	-
K_{TI} or K	3	5	9	17	33	65	129	16
$\Delta G_{\mathrm{avg}}\,[\mathrm{kJ\,mol^{-1}}]$	15.61	4.81	18.97	21.65	21.45	21.40	21.43	21.39
$\epsilon\,[\mathrm{kJ\,mol^{-1}}]$	0.05	0.18	0.13	0.15	0.20	0.17	0.21	0.10
$\sigma\,[\mathrm{kJ\,mol^{-1}}]$	0.02	0.08	0.06	0.07	0.09	0.08	0.09	0.04
$\sigma_{\mathrm{B}}\,[\mathrm{kJ\,mol^{-1}}]$	0.03	0.06	0.05	0.05	0.07	0.08	0.09	0.05
$\sigma_{\mathrm{G}}\,[\mathrm{kJ\,mol^{-1}}]$	0.03	0.06	0.05	0.05	0.05	0.05	0.05	-

(*i*) the standard deviation σ of the mean over the repeats; (*ii*) a bootstrap error σ_{B} calculated from the concatenated trajectories of the repeats at each λ-point; (*iii*) an alternative error σ_{G} obtained similarly by calculating the bootstrap errors on the average Hamiltonian derivatives at all λ-points, and propagating them assuming a Gaussian distribution of the estimates, as proposed in Refs.[?][?] The three values turn out to be very similar, so that further discussions will refer to the error ϵ defined as σ amplified by the Student t-factor.

The effect of K_{TI} on the convergence properties is illustrated graphically in Fig. 1.2c, which shows the ΔG estimates at full sampling time as a function of K_{TI} for each of the ten repeats (individual colored circles), along with the associated mean and error on the mean over the repeats (blue crosses and error bars). Using too few λ-points leads to quadrature errors, which are very significant for $n \leq 4$ ($K_{TI} \leq 17$), *i.e.* the shape of the average Hamiltonian derivative curve is not well captured by the Simpson quadrature. For $n \geq 5$ ($K_{\mathrm{TI}} \geq 33$), the quadrature error becomes essentially

negligible compared to the sampling error.

Interestingly, at constant total sampling time, the latter error appears to be essentially insensitive to the number of λ-points employed, with very similar ΔG estimates and error bars of $21.45 \pm 0.20, 21.40 \pm 0.17$ and $21.43 \pm 0.21 \, \text{kJ} \, \text{mol}^{-1}$ for $K_{\text{TI}} = 33, 65$ and 129, respectively (Tab. 1.2). Thus, for the system considered here and at constant total sampling time, the use of a larger number of λ-points with a reduced simulation time is advantageous as it reduces the quadrature error without incurring any penalty in terms of sampling error. However, it also induces an extra cost in the form of the preequilibration dead time, which is proportional to the number of λ-points employed. In addition, the insensitivity of the sampling error to the number of λ-points may still break down when the simulation times at the different λ-points become shorter than the corresponding orthogonal relaxation times.

The use of numerous λ-points with insufficient preequilibration is the plague affecting the slow-growth (SG) method,[?][?] which is rather inaccurate in pratice[?][?][?] (large error and hysteresis), unless the results are exponentially averaged over repeats with different Boltzmann-distributed initial conditions, as in fast growth[?][?][?][?] (FG). In essence, the CBTI scheme aims at achieving this limit of a very large number of λ-points, but without suffering from issues related to the preequilibration dead time (TI), to an insufficient preequilibration (SG), or to the requirement of exponential averaging over initial conditions (FG).

1.4.2 RESULTS USING THE CBTI SCHEME

The exploration of the influence of the CBTI mass parameter m_Λ and thermostat coupling time τ_Λ considering different numbers K of replicas is presented in details in Sect. 1.B. The results are summarized in Tab. 1.1 (entries 1-10). They show that when selected within reasonable ranges, the parameters m_Λ and τ_Λ have only a limited influence on the kinetic-energy exchange between the CB advance variable Λ and the conformational degrees of freedom, on the average temperature T_Λ, on the diffusion constant D_Λ, and on the calculated free-energy change ΔG. Based on this

exploration, a working choice $m_\Lambda = 40K^{1/2}\,\mathrm{u}\,\mathrm{nm}^2$ and $\tau_\Lambda = 0.5\,\mathrm{ps}$ was selected for the following CBTI calculations.

Using this parameter setting, CBTI calculations were performed using $K = 8, 16, 32, 64$ or 128 replicas along with $256K^{-1}\,\mathrm{ns}$ simulation time for the replica system, *i.e.* $256\,\mathrm{ns}$ total single-system sampling time in all cases. The detailed results are shown graphically in Fig. 1.F.11 - Fig. 1.F.15 and summarized numerically in Tab. 1.1 (entries 11-15). Illustrative results for $K = 8, 32$ or 128 (entries 11, 13 and 15) are also shown in Figs. 1.3 and 1.5, and discussed in details below. The time series and probability distributions of Λ for these three simulations are displayed in Fig. 1.3a. For $K = 8$ replicas (left panel), Λ covers the entire range $[0, 2\pi)$, but presents 8 regularly spaced probability peaks at integer multiples of $2\pi/8$. This corresponds to CB configurations where two of the replicas are at the physical end-states $\lambda = 0$ and $\lambda = 1$, the six others being by pairs at $\lambda = 0.25, 0.5$ and 0.75. The occurrence of such probability peaks is not surprising considering the shape of the free-energy profile (Fig. 1.2a). This profile presents minima at $\lambda = 0$ and $\lambda = 1$, the former one being the fully interacting methanol state, which is particularly favored. This significant inhomogeneity in the sampling of Λ correlates with a stepwise (hopping) dynamics in the Λ time series.

In contrast, for $K = 128$ replicas (right panel), Λ only spans a limited subrange of the interval $[0, 2\pi)$ and no longer presents regularly spaced probability peaks. For longer simulations, one would expect the probability curve to also cover the entire $[0, 2\pi)$ range and present regularly spaced peaks at integer multiples of $2\pi/128$, but these would be of much smaller magnitudes compared to the $K = 8$ case. For $K = 128$, the effect of these residual variations becomes almost negligible, *i.e.* there is essentially no driving force on the Λ variable. As a result, instead of a hopping dynamics, the Λ variable now follows a purely diffusive (random-walk) trajectory.

The situation for $K = 32$ replicas (middle panel) is intermediate. The $[0, 2\pi)$ range is fully covered, but the distribution is irregular, presenting some small indentations at $2\pi/32$ spacing.

The reduction of the coverage and homogeneity in terms of Λ over the interval
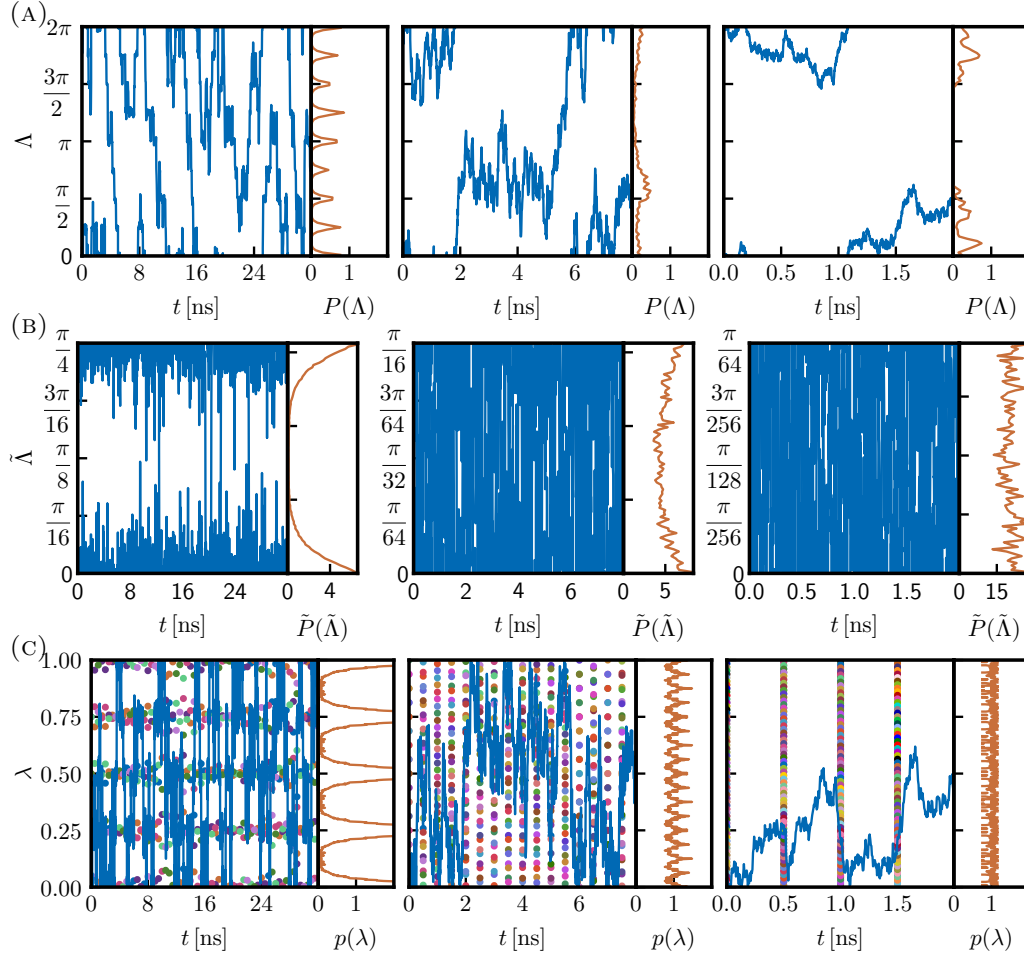
FIGURE 1.3: *Time series and probability distributions of the relevant variables in unbiased CBTI simulations of the aqueous methanol-to-dummy mutation at 298.15 K and 1 bar with the 2016H66 force field.* This figure compares the dynamics and distributions of the CB variables Λ (a), $\tilde{\Lambda}$ (b) and λ (c) in CBTI simulations with $K = 8$ (left), 32 (middle) or 128 (right) replicas (based on simulations with a total single-system sampling time of 256 ns). Panel (a) shows the time series $\Lambda(t)$ and probability distribution $P(\Lambda)$ of the CB advance variable Λ. Panel (b) shows the time series $\tilde{\Lambda}(t)$ and probability distribution $\tilde{P}(\tilde{\Lambda})$ of the CB fractional advance variable $\tilde{\Lambda}$ (Eq. 1.17). Panel (c) shows the time series $\lambda(t)$ and probability distribution $p(\lambda)$ of the coupling variable λ for all replicas (Eq. 1.19). In Panel (c), the time series is shown as a blue curve for one reference replica $k = 0$, and as individual colored points at 0.5 ns interval for the $K - 1$ other replicas, and the probability distribution $p(\lambda)$ is calculated considering all replicas according to Eq. 1.19. All probability distributions are normalized to one. The CBTI simulations relied on $m_\Lambda = 40K^{1/2}\,\mathrm{u\,nm}^2$ and $\tau_\Lambda = 0.5\,\mathrm{ps}$ (no biasing potential). See also Tab. 1.1 entries 11, 13 and 15 for relevant numerical results. Corresponding graphs also including $K = 16$ and 64 can be found in Figs. 1.F.11 - 1.F.15.

$[0, 2\pi)$ when going from $K = 8$ to 32, and then to 128, is not due to a progressively slower (hopping *vs.* diffusive) dynamics, but simply to the different time spans of the simulations (32, 8 and 2 ns, respectively), imposed by the constant total single-system sampling time of 256 ns in the present simulations. This can be seen in Fig. 1.5.

Fig. 1.5a shows the time series and probability distribution of $\dot{\Lambda}$, the velocity

associated with the Λ-variable. The Maxwell-Boltzmann distribution of Eq. 1.25 is also displayed for comparison. The observed width $\sigma_{\dot{\Lambda}}$ of the distribution is reported numerically in Tab. 1.1. As expected from Eq. 1.25, it is proportional to $m_{\Lambda}^{-1/2}$ which, in the present simulations, was itself made proportional to $K^{1/2}$ (an arbitrary parameter choice justified in Sect. 1.B). Thus, the average magnitude of $\dot{\Lambda}$ decreases somewhat when going from $K = 8$ to 32, and then to 128 (scaling factor $2^{-1/2} \approx 0.7$ each time). Fig. 1.4 shows the normalized autocorrelation functions $c_{\dot{\Lambda}}$ of $\dot{\Lambda}$ for different m_{Λ} in the absence of thermostat coupling (entries 1-5 of Tab. 1.1). The corresponding correlation times $\tau_{\dot{\Lambda}}$ are reported numerically in Tab. 1.1. Here, one observes that the correlation time increases with m_{Λ} as a result of the higher inertia of the CB. In addition, for the two lowest values of m_{Λ}, the initial decay of $c_{\dot{\Lambda}}$ to negative values suggests that the velocities are frequently reversed at short time, when barriers cannot be crossed. Note, however, that this connection between m_{Λ} and $\tau_{\dot{\Lambda}}$ is somewhat altered by the thermostat coupling in the other simulations (including those of entries 11-15 in Tab. 1.1).
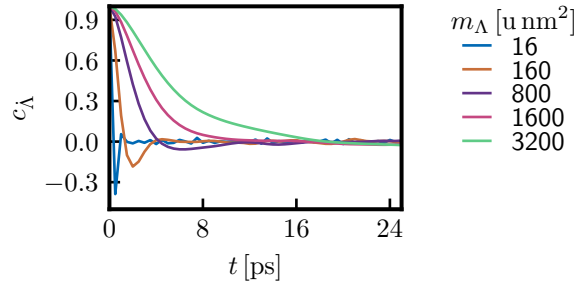
FIGURE 1.4: *Velocity autocorrelation functions of the CB advance variable in unbiased CBTI simulations of the aqueous methanol-to-dummy mutation at* 298.15 K *and* 1 bar *with the 2016H66 force field.* This figure shows the increase in the autocorrelation time of the CB velocity upon increasing the mass parameter in CBTI simulations with $K = 16$ replicas and no thermostat coupling. The normalized autocorrelation function $c_{\dot{\Lambda}}$ of the velocity $\dot{\Lambda}$ is shown for different values of the mass-parameter m_{λ}, in the absence of thermostating of the Λ-variable ($\tau_{\Lambda} \to \infty$). See also Tab. 1.1 entries 1-5 for relevant numerical results, including the corresponding autocorrelation times $\tau_{\dot{\Lambda}}$. Corresponding graphs for all simulations listed in Tab. 1.1 are shown in Figs. 1.F.1 - 1.F.15

Because the two above trends act in opposite directions, the actual rate of diffusion along Λ is not very different in the three simulations.

This is seen in Fig. 1.5c, which shows the mean-square displacement d_{Λ} of Λ as a function of time. The curves are essentially linear, as expected from the Einstein

model (Eq. 1.26). The corresponding diffusion constants D_Λ are reported in Tab. 1.1, namely 4.4, 6.2 and $2.3\,\mathrm{ns}^{-1}$ for $K = 8$, 32 and 128 replicas, respectively, *i.e.* not dramatically different. A similar observation holds for the probability distribution of $\ddot{\Lambda}$, the acceleration associated with the Λ-variable, shown in Fig. 1.5b. The choice of taking m_Λ proportional to $K^{1/2}$ was made here precisely to achieve a CB acceleration that is essentially independent of K (see discussion in Sect. 1.A).

The evolution of D_Λ with K is clearly non-monotonic (see values for entries 11-15 in Tab. 1.1), which results in large part from the two opposite effects mentioned above concerning the effect of m_Λ on $\sigma_{\dot{\Lambda}}$ and $\tau_{\dot{\Lambda}}$. On the one hand, $\sigma_{\dot{\Lambda}}$ (*i.e.* the average magnitude of the CB velocity) decreases upon increasing m_Λ, and the dynamics along the Λ-variable becomes intrinsically slower. On the other hand, $\tau_{\dot{\Lambda}}$ (*i.e.* the dynamical inertia of the CB) increases upon increasing m_Λ, and the ability to cross residual barriers in the $G_\Lambda(\Lambda)$ free-energy profile is enhanced. The magnitude of these barriers itself decreases upon increasing K. In addition, inertia along Λ is also affected by the thermostat coupling. The combination of these effects is complex and the resulting trends in D_Λ upon varying K would be difficult to rationalize in details. Most importantly, however, they appear not to be extremely pronounced.

The time series and probability distributions of the CB fractional advance variable $\tilde{\Lambda}$ are shown in Fig. 1.3b. In contrast to the distributions of Λ over the range $[0, 2\pi)$ (Fig. 1.3a), the distributions of $\tilde{\Lambda}$ over the range $[0, \Delta\Lambda)$ are converged and cover the entire range, also for $K = 128$. For $K = 8$, the distribution of $\tilde{\Lambda}$ is significantly biased towards 0 and $2\pi/8$, *i.e.* CB configurations with two of the replicas at the physical end-states. For $K = 32$, it is still somewhat biased towards 0 and $2\pi/32$, which corresponds to the same situation. And for $K = 128$, the probability distribution becomes very close to homogeneous.

Finally, the time series and probability distributions of the coupling variable λ are shown in Fig. 1.3c for all replicas. The time series is shown for one reference replica $k = 0$ (blue curve) as well as the other replicas $k = 1 \dots K - 1$ (individual colored points at $0.5\,\mathrm{ns}$ interval). Compared to the reference replica, these representative points are merely shifted by $2K^{-1}k$, refolded by periodicity into the range $[0, 2)$, and
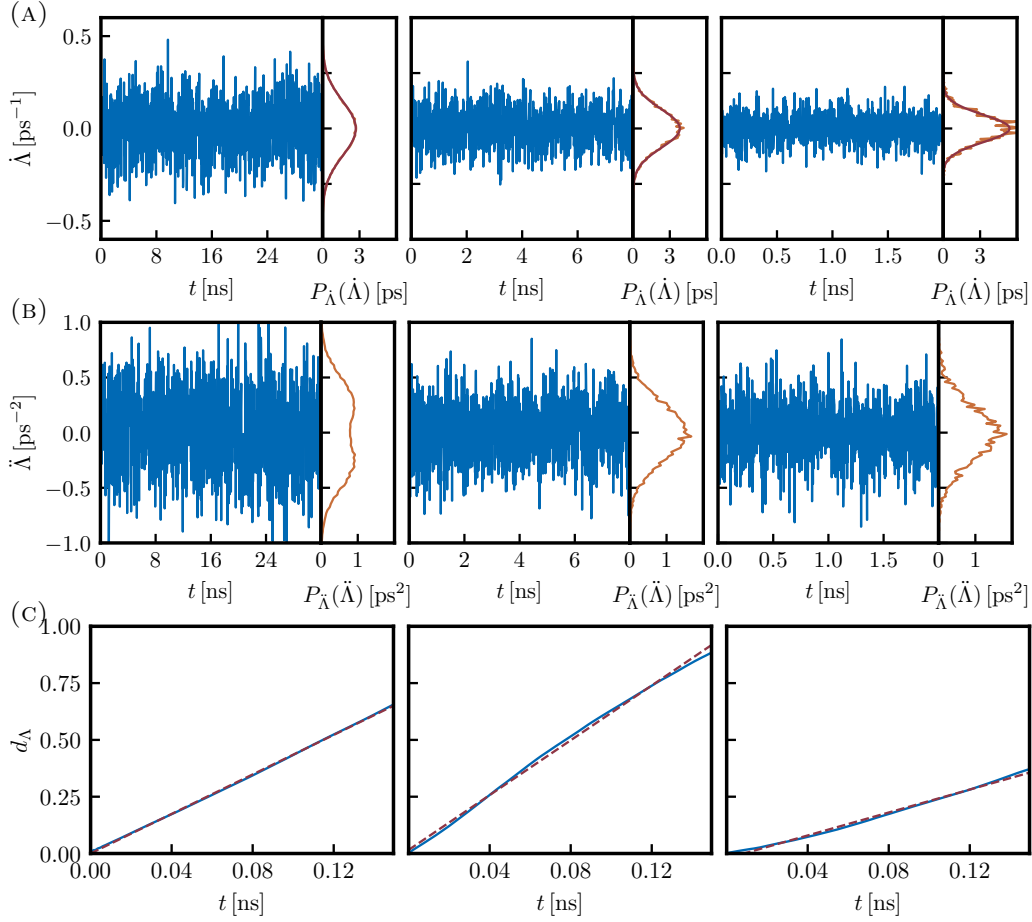
FIGURE 1.5: *Dynamical properties of the CB advance variable in unbiased CBTI calculations of the aqueous methanol-to-dummy mutation at 298.15 K and 1 bar with the 2016H66 force field.* This figure compares the dynamics and distributions of the CB variables $\dot{\Lambda}$ (a) and $\ddot{\Lambda}$ (b), as well as the mean-square displacement d_Λ along Λ (c), in CBTI simulations with $K = 8$ (left), 32 (middle) or 128 (right) replicas (based on simulations with a total single-system sampling time of 256 ns). Panel (a) shows the time series $\dot{\Lambda}(t)$ and probability distribution $P_{\dot{\Lambda}}(\dot{\Lambda})$ of the velocity associated with the CB advance variable Λ. Panel (b) shows the time series $\ddot{\Lambda}(t)$ and probability distribution $P_{\ddot{\Lambda}}(\ddot{\Lambda})$ of the corresponding acceleration. Panel (c) shows the mean-square displacement $d_\Lambda(t)$ of Λ (Eq. 1.26). In Panel (a), the probability distribution from the simulation (blue curve) is shown together with the analytical Maxwell-Boltzmann distribution $P_{\dot{\Lambda}}^{\mathrm{MB}}(\dot{\Lambda})$ of Eq. 1.25 (orange curve). In Panel (c), the mean-square displacement d_Λ from the simulation (blue curve) is shown along with a linear least-squares fit over the interval 0 - 0.15 ns (brown dashed line). All probability distributions are normalized to one. The CBTI simulations relied on $m_\Lambda = 40K^{1/2}\,\mathrm{u\,nm^2}$ and $\tau_\Lambda = 0.5\,\mathrm{ps}$ (no biasing potential). See also Tab. 1.1, entries 11, 13 and 15 for relevant numerical results. Corresponding graphs also including $K = 16$ and 64 can be found in Figs. 1.F.11 - 1.F.15.

the points in the range $[1, 2)$ mirrored into the range $[0, 1]$ according to Eqs. 1.6- 1.8.

The probability distribution $p(\lambda)$ is calculated according to Eq. 1.19, *i.e.* considering

all replicas. Although the Λ-space in the range $[0, 2\pi)$ is sampled only partially given

the simulation lengths (Fig. 1.3a), the λ-space in the range $[0, 1]$ (Fig. 1.3c), just

like the $\tilde{\Lambda}$-space in the range $[0, \Delta\Lambda)$ (Fig. 1.3b), is completely sampled in all cases.

However, for $K = 8$, the barriers are still too high and the sampling is considerably biased towards λ-values that are integer multiples of $1/4$. The corresponding bias towards integer multiples of $1/16$ is far less pronounced for $K = 32$, and the simulation with $K = 128$ is very close to a homogeneous sampling of λ.

In summary, the sampling of Λ only needs to cover a $\Delta\Lambda$ range to ensure the sampling of all possible $\tilde{\Lambda}$-values by the CB and thus, of all possible λ-values by at least one replica. However, the homogeneity of the sampling in λ depends crucially on K, quasi-homogeneous sampling requiring a large number of replicas.

This can be seen in Fig. 1.6, which shows in a logarithmic form the height $G_\Lambda^\star$ of the peaks in $G_\Lambda(\Lambda)$ as a function of the number K of replicas (see also Fig. 1.H.1 for the individual curves). The dependency is approximately linear with a slope of minus one, suggesting that $G_\Lambda^\star$ scales as K^{-1}. Such a scaling is expected based on the considerations made in Sect. 1.A. Note that the proportionality constant depends on the derivatives of $G(\lambda)$ at the two physical end-states. In a situation where these derivatives vanished, a scaling as K^{-2} would be expected instead.
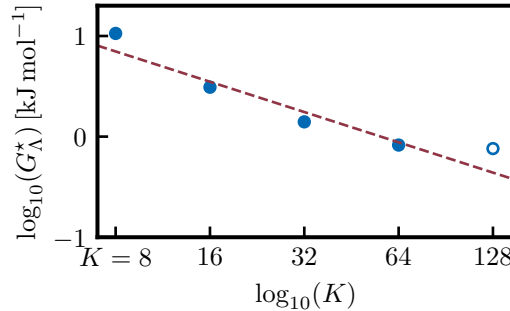
FIGURE 1.6: *Height $G_\Lambda^\star$ of the residual barriers in the free-energy profile $G_\Lambda(\Lambda)$ for unbiased CBTI simulations of the aqueous methanol-to-dummy mutation at* $298.15\,\mathrm{K}$ *and* $1\,\mathrm{bar}$ *with the 2016H66 force field.* This figure shows the progressive decrease in the height of the barriers along Λ upon increasing the number K of replicas in a CBTI simulation. The data is represented in logarithmic form and a linear line of slope -1 is fitted to the filled circles (intercept 1.75 at $\log_{10}(K) = 0$). The open circle was omitted from the fit because $G_\Lambda^\star$ cannot be determined sufficiently accurately (within the noise). The simulations considered rely on $K = 8, 16, 32, 64$ or 128 replicas, along with $m_\Lambda = 40K^{1/2}\,\mathrm{u\,nm}^2$ and $\tau_\Lambda = 0.5\,\mathrm{ps}$ (no biasing potential). The corresponding free-energy profiles $G_{\tilde{\Lambda}}(\tilde{\Lambda})$ are shown graphically in Fig. 1.H.1.

1.4.3 CBTI with Biasing Potential

As seen above, simulations employing only few replicas may still be affected by a significant sampling inhomogeneity along λ due to high residual barriers along Λ. To prevent the undersampling of specific λ-ranges in this case, a memory-based biasing potential can be employed. This renders the CBTI method more robust since it becomes insensitive to the chosen number of replicas, and more compatible with possible constraints related to *e.g.* the number of cores in a specific computing node during the calculation. As discussed in the Theory section, it is sufficient to define such a biasing potential over a $[0, \Delta\Lambda/2]$ range of $\tilde{\Lambda}$, and its magnitude is limited by the already largely homogeneous sampling of $\tilde{\Lambda}$. These features speed up the build-up procedure drastically, leading to a comparatively small simulation time for the LE build-up phase relative to single-system λ-LEUS.

Fig. 1.7 illustrates the sampling in a biased compared to that in the unbiased CBTI simulation with $K = 8$ replicas. The biasing relied on 34 basis functions over half the $\tilde{\Lambda}$-range $[0, \pi/4]$ (17 free coefficients) and a build-up phase of 0.15 ns (*i.e.* only $1.2\,\mathrm{ns}$ total single-system build-up time). During this time, the replica system performed three double sweeps over all grid-points, leading to a biasing potential with an energetic resolution of $10^{-6}\,\mathrm{kJ\,mol^{-1}}$ (Figs. 1.7a and 1.7b).

The Λ-sampling in the biased CBTI simulation is much closer to uniform compared to that in the unbiased one (Fig. 1.7c *vs.* 1.7d), *i.e.* the relative probabilities between most and least sampled Λ-values has been reduced from 70 to 4.6. The time series of λ and the sampling distribution $p(\lambda)$ show that the motion of the conveyor belt was hindered in the unbiased simulation, which it is no longer the case in the biased one (Fig. 1.7e *vs.* 1.7f).

Alternative build-up protocols were also explored for $K = 8$ and $K = 16$, differing in the number N_{gp} of grid points, in the duration t_{LE} of the build-up phase, and in the build-up parameters c_{LE} and f_{red}. The corresponding results are shown graphically in Figs. 1.I.2 and 1.I.4 and summarized numerically in Tab. 1.I.1. All the tested protocols reduce the sampling inhomogeneity to an acceptable level, although residual
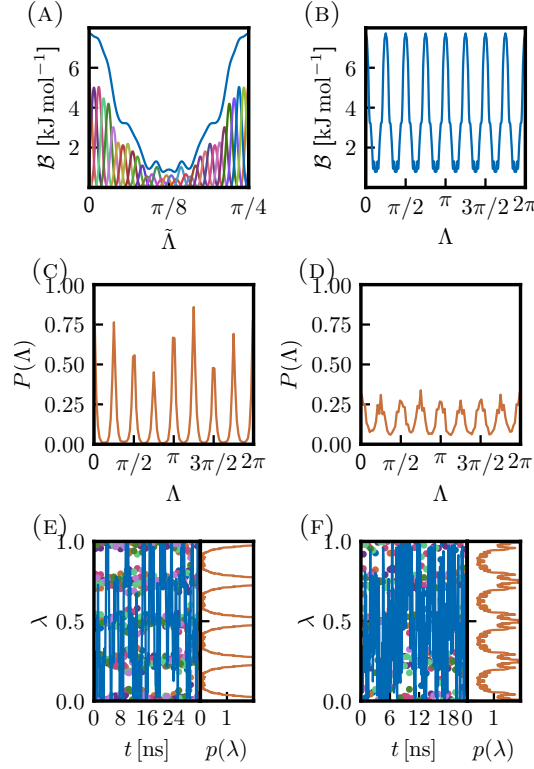
FIGURE 1.7: *Biasing potential and simulation results from unbiased vs. biased CBTI calculations of the aqueous methanol-to-dummy mutation at 298.15 K and 1 bar with the 2016H66 force field.* This figure illustrates how the use of a biasing potential permits to drastically reduce the sampling inhomogeneity of CBTI simulations performed with a small number of replicas (here, $K = 8$ replicas). Panels (a) and (b) show the biasing potential along the $\tilde{\Lambda}$- and Λ-ranges, respectively. Panels (c) and (d) show the probability distribution $P(\Lambda)$ of the CB advance variable Λ in the unbiased and biased simulation, respectively. Panels (e) and (f) show the time series $\lambda(t)$ and probability distribution $p(\lambda)$ of the coupling variable λ for all replicas in the unbiased and biased simulation, respectively. In Panel (a), the symmetric biasing potential (blue curve) is represented using 34 basis functions (curves of different colors) involving 17 free coefficients. In Panels (e) and (f), the time series is shown as a blue curve for one reference replica $k = 0$, and as single colored points at 0.5 ns spacing for the $K - 1$ other replicas, and the probability distribution $p(\lambda)$ is calculated considering all replicas according to Eq. 1.19. All probability distributions are normalized to one. The CBTI simulations were performed with or without biasing potential, using $m_\Lambda = 113\,\mathrm{u\,nm^2}$ and $\tau_\Lambda = 0.5\,\mathrm{ps}$. For the biased simulation, the biasing potential was constructed in a build-up phase of 0.15 ns duration. Results for biased simulations with $K = 16$ and alternative build-up protocols are shown in Fig. 1.I.2 and Fig. 1.I.4, the numerical results being reported in Tab. 1.I.1.

inhomogeneities are difficult to remove entirely. Compared to the TI results with a sufficient number of λ-points (Tab. 1.2, entries with $K_{\mathrm{TI}} \geq 33$) and the unbiased CBTI results with a sufficient number of replicas (Tab. 1.1, entries 12-15 with $K \geq 16$), the biased CBTI results with $K = 8$ or 16 (entries 16 and 17 in Tab. 1.1) deliver similar ΔG values and error bars, provided that an appropriate build-up protocol is employed.

1.4.4 CBTI Free-Energy Estimator

The calculation of free-energy differences from CBTI simulations according to Eq. 1.22 relies on the choice of a number J of bins. This number can be taken very large, thereby reducing the (rectangular) quadrature error to an essentially negligible amount, but should not exceed a threshold $J_{\max}$ where empty bins occur. The influence of J is illustrated in Fig. 1.8, which shows the average Hamiltonian derivative curve obtained from the unbiased CBTI simulation with $K = 32$ replicas (entry 13 in Tab. 1.1, restricted to total 100 ns single-system sampling time) for $J = 10$ (Fig. 1.8a), $J = 500$ (Fig. 1.8b) and $J = 5000$ (Fig. 1.8c). A too low number of bins (*e.g.* $J = 10$) will lead to an insufficient number of integration points, thereby increasing the quadrature error. A too high number of bins will lead to insufficient sampling of each bin, thereby increasing the statistical error. The optimal value of J will depend on the total sampling time and on the degree of homogeneity of the sampling along λ.

To investigate the influence of J on the calculated ΔG, this number was systematically increased from 1 to $J_{\max}$ in increments ranging from 10 to 1000. The results are shown in Fig. 1.8d, considering the unbiased CBTI simulations with $K = 8, 16, 32, 64$ or 128 (entries 11-15 in Tab. 1.1, restricted to 100 ns total single-system sampling time) as well as the biased CBTI simulations with $K = 8$ or 16 (entries 16 and 17 in Tab. 1.1, same restriction). For the unbiased simulation with $K = 8$ (blue curve), the convergence of ΔG upon increasing J is irregular and the value of $J_{\max}$ is comparatively low (about 600), due to the inhomogeneous sampling along λ. All other curves present qualitatively similar convergence features. For a low J ($J < 100$), the quadrature error leads to strong variations. For intermediate J ($100 \leq J \leq 3000$), the curve presents a plateau, *i.e.* the ΔG estimate becomes insensitive to changes in J. For larger J ($3000 < J \leq J_{\max}$, where $J_{\max}$ ranges from about 5000 to 8000) the curves become somewhat noisy again due to undersampled bins. These results suggest that, as a rule of thumb, the value of J for the application of Eq. 1.22 should be selected between at least about 100 and at most about $J_{\max}/2$. For most of the ΔG results presented in this chapter, a default value $J = 500$ was selected as leading
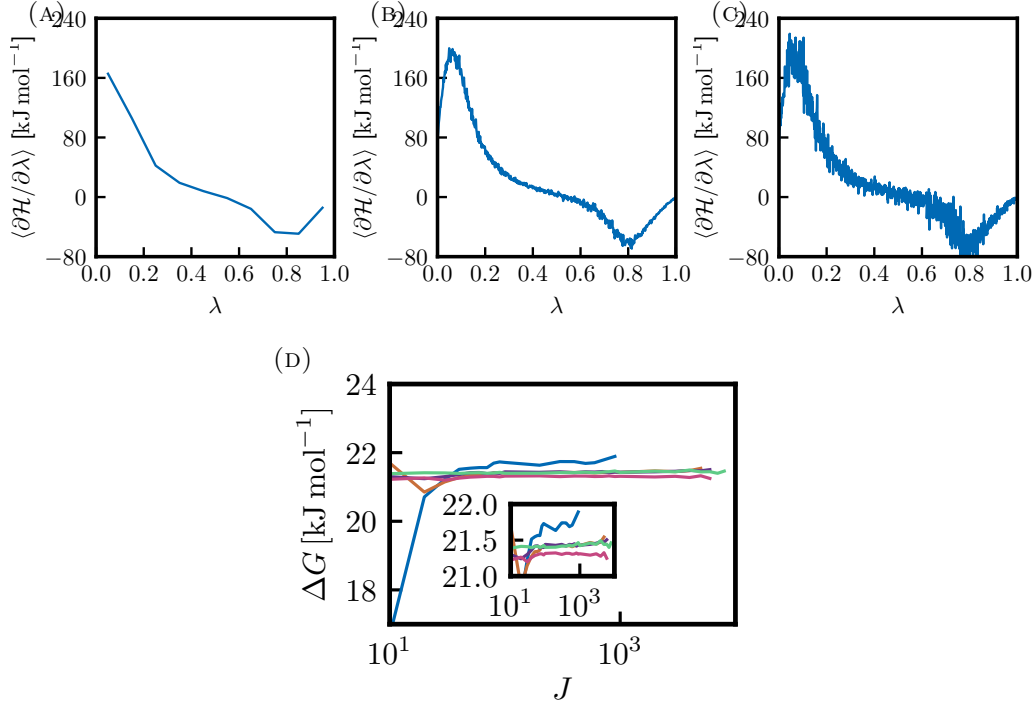
FIGURE 1.8: *Influence of the number of bins used in the free-energy estimator for CBTI simulations of the aqueous methanol-to-dummy mutation at* $298.15\,\mathrm{K}$ *and* $1\,\mathrm{bar}$ *with the 2016H66 force field.* This figure illustrates how the number J of bins employed in the estimator affects the quadrature and sampling quality (top) and the resulting estimated ΔG (bottom). Panels (a), (b) and (c) show the average Hamiltonian derivative curve constructed with different numbers J of bins, namely $J = 10$ (a), 100 (b) or 5000 (c). The curves are based on the CBTI simulation using $K = 32$ replicas (entry 13 in Tab. 1.1). Panel (d) shows the ΔG estimate from Eq. 1.22 for CBTI simulations with $K = 8$, 16, 32, 64 or 128 replicas (the two former either unbiased or biased; entries 11-17 in Tab. 1.1) as a function of J from 1 to J_{max}, where J_{max} corresponds to the first occurence of empty bins. All CBTI simulations were performed using $m_\Lambda = 40K^{1/2}\,\mathrm{u\,nm^2}$ and $\tau_\Lambda = 0.5\,\mathrm{ps}$, and their analysis is restricted here to 100 ns total single-system sampling time.

to a reasonably smooth Hamiltonian derivative curve (Fig. 1.8b) while keeping the quadrature error negligible (except for the CBTI simulation with $K = 8$, where $J = 200$ was used). The results for the variants ΔG_{alt} and ΔG_{app} to ΔG from Eq. 1.22 which do not require the specification of a number of bins (Eqs. 1.C.1 and 1.C.2) are also included in Tab. 1.1 for comparison. They are discussed in Sect. **??**.

1.4.5 CBTI CONVERGENCE PROPERTIES

The convergence properties of TI and CBTI are compared in Figs. 1.9a and 1.9b, which shows the ΔG values calculated from ten simulation repeats (individual dashed curves) along with the corresponding mean and error (95% confidence interval) on the mean (thick blue curve and shaded area) as a function of the total single-system

sampling time per repeat. The simulation protocols compared are TI with $K_{TI} = 129$ λ-points (Fig. 1.9a, which is identical to Fig. 1.2b) and unbiased CBTI with $K = 16$ (Fig. 1.9b). Note that in the latter case, the curves do not start at time zero because the application of Eq. 1.22 with $J = 500$ to estimate ΔG only becomes possible when $J_{\max}$ exceeds 500 (*i.e.* each of the 500 bins encompasses at least one sampled λ-value). The mean ΔG values over the repeats and the associated error bars at full single-system sampling time (100 ns per repeat) are reported in Tab. 1.2 (last two columns; see also Tab. 1.E.1 and Tab. 1.G.1 for the results of the individual repeats).
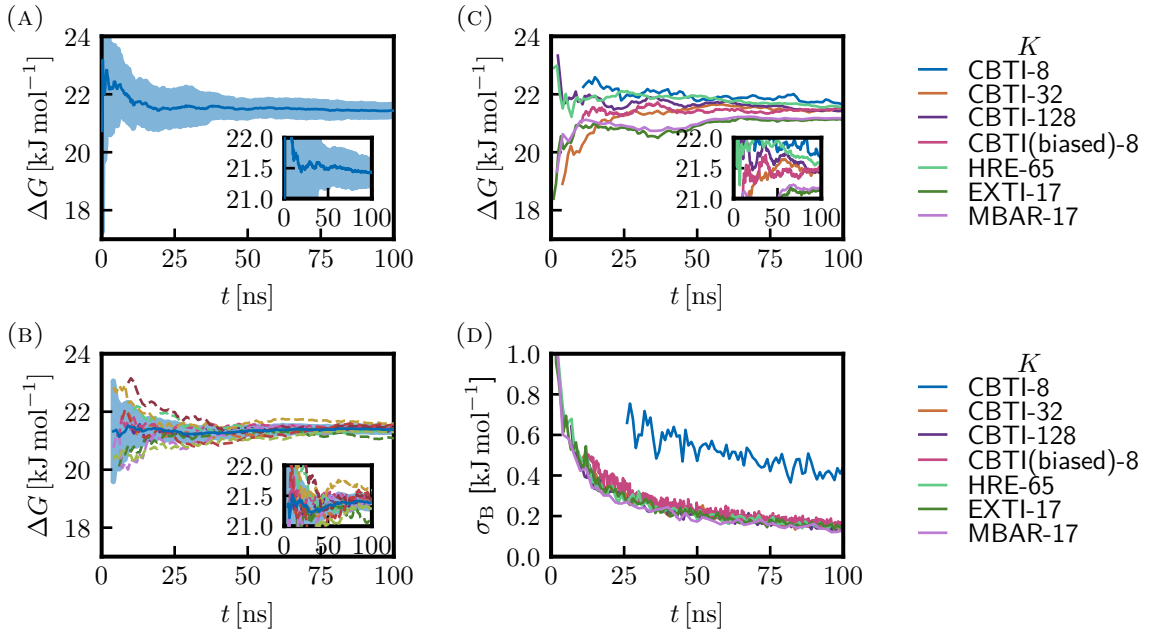
FIGURE 1.9: *Convergence properties of the CBTI calculations compared to other methods for the aqueous methanol-to-dummy mutation at* 298.15 K *and* 1 bar *with the 2016H66 force field.* This figure compares the convergence properties of the calculated free-energy change ΔG considering TI (a) *vs.* CBTI (b) based on ten repeats of calculations involving 100 ns total single-system sampling time, or considering CBTI (biased or unbiased) *vs.* TI, HRE, TI/EXTI and TI/MBAR in terms of estimated ΔG (c) and error (d) based on one calculation involving 100 ns total single-system sampling time. Panel (a) is identical to Fig. 1.2b and shows the convergence of ΔG for TI calculations with $K_{TI} = 129$ λ-points. Panel (b) displays the corresponding convergence of ΔG for unbiased CBTI calculations with $K = 16$ replicas. Panels (a) and (b) show the running ΔG estimates for each of the ten repeats (dashed curves) as a function of the total sampling time t per repeat, along with the corresponding mean and error on the mean (95% confidence interval) over the repeats (thick blue curve and shaded area). The final ΔG values for the individual repeats are reported in the Tab. 1.G.1, and the statistics over the repeats can be found in Tab. 1.2. Panel (c) compares convergence properties in terms of ΔG based on single simulations. Panel (d) displays the associated errors σ_B calculated by bootstrapping (no Student t-factor included). Panels (c) and (d) consider unbiased CBTI ($K = 8, 32$ or 128), biased CBTI ($K = 8$) HRE ($K_{HRE} = 65$), along with TI with EXTI or MBAR estimators ($K_{TI} = 17$ and 129 virtual λ-points in both cases). The corresponding final ΔG estimates are reported in Tab. 1.3.

The comparison clearly evidences an enhanced convergence for CBTI at identical

sampling time, both in terms of the running averages and of the final values. These final values are $21.43 \pm 0.21 \, \text{kJ} \, \text{mol}^{-1}$ for TI, compared to $21.39 \pm 0.10 \, \text{kJ} \, \text{mol}^{-1}$ for CBTI, *i.e.* identical within error bars, but with an error reduced by about a factor two for CBTI. This improved convergence is likely due to the improved orthogonal sampling, *i.e.* the diffusion of the systems along λ permits to circumvent orthogonal barriers present at specific λ-values.

The same observation holds for the CBTI simulations involving different numbers of replicas or/and a biasing potential. This is shown in Figs. 1.9c and 1.9d for $K = 8$, 32 or 128 (unbiased) and for $K = 8$ (biased). Here, a single simulation was performed (no repeats), and the running ΔG value (Fig. 1.9c) is displayed along with the corresponding bootstrap error estimate (Fig. 1.9d; no Student t-factor included) as a function of the total single-system sampling time. The ΔG estimates and bootstrap error bars at full single-system sampling time (100 ns) are reported in Tab. 1.3 (second and third columns). Except for the unbiased CBTI simulation with $K = 8$, where the sampling along λ is still heterogeneous, all the CBTI protocols follow very similar error curves upon increasing the sampling time (Fig. 1.9d), with final errors of $0.12 - 0.16 \, \text{kJ} \, \text{mol}^{-1}$. In contrast, the convergence of TI is significantly slower, with corresponding final errors of $0.17 - 0.30 \, \text{kJ} \, \text{mol}^{-1}$ (fourth column in Tab. 1.3).

1.4.6 Comparison of CBTI with other Methods

The ΔG values and associated bootstrap errors (no Student t-factor included) calculated using various CBTI protocols are compared to the results of other approaches in Tab. 1.3. The comparison involves calculations relying on 100 ns total single-system sampling time, *i.e.* essentially identical computational costs. The protocols compared are unbiased CBTI (with $K = 8, 16, 32, 64$ or 128), biased CBTI (with $K = 8$ or 16), TI (with $K_{\text{TI}} = 9, 17, 33, 65$ or 129), HRE (with $K_{\text{HRE}} = 17, 33$ or 65), along with TI/EXTI and TI/MBAR (both with $K_{\text{TI}} = 9$ or 17, and using 129 virtual λ-points). More details on the HRE, TI/EXTI and TI/MBAR calculations can be found in Figs. 1.J.1 - 1.L.1.

TABLE 1.3: *Free-energy estimates from the CBTI calculations compared to other methods for the aqueous methanol-to-dummy mutation at* $298.15\,\mathrm{K}$ *and* $1\,\mathrm{bar}$ *with the 2016H66 force field.* This table compares the convergence properties of the calculated free-energy change ΔG considering CBTI (biased or unbiased), TI, HRE, EXTI and MBAR calculations with different numbers of replicas (all at a constant total single-system sampling time of 100 ns). The CBTI calculations involved K replicas, along with $m_\Lambda = 40 K^{1/2}\,\mathrm{u\,nm^2}$ and $\tau_\Lambda = 0.5\,\mathrm{ps}$ (with or without biasing potential). The TI calculations involved K_{TI} equispaced λ-points. The HRE calculations involve K_{HRE} equispaced λ-points, along with $\tau_{\mathrm{HRE}} = 0.2\,\mathrm{ps}$. For EXTI and MBAR, the calculations relied on K_{TI} equispaced real λ-points, and consideration of 129 equispaced virtual λ-points. In all cases, the total single-system sampling time was 100 ns. The corresponding convergence properties are illustrated graphically in Fig. 1.9. Additional numerical results regarding the CBTI simulations are reported in Tab. 1.1, entries 11-15 (unbiased CBTI) and entries 16-17 (biased CBTI).

| | $\Delta G\,[\mathrm{kJ\,mol^{-1}}]$ | | | | | |
$K, K_{\mathrm{TI}} - 1$ or $K_{\mathrm{HRE}} - 1$	CBTI (unbiased)	CBTI (biased)	TI	HRE	EXTI	MBAR
8	21.69 (0.40)	21.48 (0.16)	19.22 (0.18)	-	21.24 (0.37)	21.37 (0.15)
16	21.42 (0.16)	21.30 (0.13)	21.87 (0.17)	21.52 (0.16)	21.12 (0.14)	21.17 (0.13)
32	21.44 (0.14)	-	21.48 (0.20)	21.39 (0.15)	-	-
64	21.32 (0.14)	-	21.18 (0.24)	21.47 (0.13)	-	-
128	21.43 (0.12)	-	21.45 (0.30)	-	-	-

As previously discussed, the results of the TI calculations with $K_{\mathrm{TI}} = 9$ or 17 are affected by large quadrature errors (see Fig. 1.2c and Tab. 1.2), and those of the unbiased CBTI simulation with $K = 8$, are affected by a significant heterogeneity of the sampling along λ (see Fig. 1.3). Excepting these three calculations, all free-energy estimates of Tab. 1.3 are essentially compatible, taking into account that the error bars exclude a Student t-factor.

The convergence of the HRE ($K_{\mathrm{HRE}} = 65$), TI/EXTI ($K_{\mathrm{TI}} = 17$) and TI/MBAR ($K_{\mathrm{TI}} = 17$) protocols are also compared to those of TI ($K_{\mathrm{TI}} = 129$), unbiased CBTI ($K = 8, 32$ or 128) and biased CBTI ($K = 8$) in Figs. 1.9c and 1.9d. Considering 1.9d, the uncertainty of the unbiased CBTI simulation with $K = 8$ is found to be particularly large, owing to the sampling heterogeneity. Excepting this simulation, the uncertainty is the largest for the TI estimate. The different CBTI protocols reduce the error by about a factor two compared to TI, presumably due to enhanced orthogonal sampling. This sampling advantage is shared by HRE, for which the error curve is very similar. Advanced free-energy estimators such as EXTI and MBAR also improve the convergence over plain TI in a different way, namely by increasing the statistical efficiency of the ΔG estimation. The error reduction is also about

a factor two compared to TI, which is probably not entirely coincidental. If two fixed-λ simulations are trapped in different orthogonal configurational wells in TI, the variable-λ approaches (CBTI, HRE) will promote well-transitions, whereas the advanced-estimator approaches (EXTI, MBAR) will import statistical information on the other well from the neighboring λ-point. One might refer to these two types of effects orthogonal-sampling *vs.* orthogonal-statistics advantages, respectively. For the simple system considered here, both results in comparable convergence improvements. This suggests that the two effects may not be cumulative if one generalizes the current CBTI estimator of Eq. 1.22 to an EXTI- or MBAR-type estimator.

1.5 CONCLUSION

In the present chapter, we proposed a new method called conveyor belt thermodynamic integration (CBTI) to calculate alchemical free-energy differences based on MD simulations. This approach borrows and combines ideas from thermodynamic integration[?][?][?] (TI), replica exchange[?][?][?] (HRE) or permutation[?][?][?] (HRP), and λ-dynamics[?][?][?][?][?][?][?][?] (λD), along with the real-life working principle of the funicular.

In CBTI, one simulates in parallel a set of K equally spaced replicas (with K even) on a forward-turn-backward-turn path along the alchemical coupling variable λ, akin to a conveyor belt (CB) between the two physical end states. Because the λ-forces (Hamiltonian λ-derivative) exerted by the individual replicas on the CB largely compensate each other, the overall Λ-force on the CB advance variable Λ becomes increasingly small when K is made increasingly large (residual free-energy barriers decreasing at least as K^{-1} in the limit of large K, as shown in Sect. 1.A). As a result, for a sufficient number K, quasi-homogeneous sampling of the λ-range can be achieved without application of any biasing potential. If a smaller K is employed, a memory-based biasing potential can still be added to further homogenize the sampling, the preoptimization of which is computationally inexpensive. The results of a CBTI simulation (whether biased or not) can be analyzed similarly to TI, by binning of the average Hamiltonian λ-derivative as a function of λ considering all replicas jointly, followed by quadrature integration. In this case, the continuous and quasi-homogeneous sampling of the λ-range permits to use a large number of bins, thereby essentially eliminating quadrature errors.

As a first application, the CBTI scheme was employed here to calculate the hydration free energy of methanol. It was shown that the method is rather robust with respect to the choice of its parameters (K as well as the mass-parameter m_Λ and thermostat coupling time τ_Λ of the CB), the most important sensitivity being relative to K. Upon increasing K, the distribution/dynamics of Λ evolves from regularly spaced preferential values with a hopping dynamics to quasi-homogeneous coverage

with a diffusive dynamics. For the smallest number of replicas considered ($K = 8$), application of a biasing potential is recommended. For larger numbers of replicas ($K \geq 16$), it becomes unnecessary. The calculated ΔG values compare well with those obtained using other methods.

The convergence is accelerated relative to TI with Simpson quadrature (smaller error bar at identical total single-system sampling time), owing to improved orthogonal sampling and reduced quadrature errors. It is comparable to HRE, which shares the same orthogonal-sampling advantage. It is also similar to TI with EXTI or MBAR as free-energy estimator, which achieve a similar improvement *via* an orthogonal-statistics advantage, *i.e.* by effectively mixing information concerning distinct configurational wells across λ-points. It should be stressed, however, that the present mutation is rather non-challenging in terms of orthogonal sampling. Work is in progress to investigate other types of systems with more complicated orthogonal spaces: (*i*) the side-chain mutation in the central residue of a tripeptide considered in Refs.;[?][?] (*ii*) the hydrogen-to-bromine mutation in the base of a nucleotide considered in Refs.[?][?][?] Here, it is expected that CBTI alone (just like HRE) will help overcoming barriers in the orthogonal space when these barriers are low at some λ-value (as in the first system mentioned), but may be insufficient on its own when these barriers are high at all λ-values (as in the second system mentioned), in which case additional modifications must be applied to create artificially an orthogonal tunnel at least over a limited λ-range.

Compared to existing MD-based alchemical free-energy calculation methods, the CBTI scheme can be viewed in at least three different ways: (*i*) as a continuous/deterministic/dynamical (instead of discrete/stochastic) analog of the HRE scheme[?][?][?] or the HRP scheme;[?][?][?] (*ii*) as a correlated multiple-replica analog (reminiscent of other swarm,[?][?][?][?] multiple-walker[?][?] or flying-Gaussian[?][?] approaches) of the λ-local elevation umbrella sampling (λ-LEUS) scheme[?][?][?][?] (or the conceptually similar flat-histogram[?][?] λ-metadynamics,[?][?][?] adaptive integration,[?] adaptive biasing force,[?] adaptively biased[?] and expanded-ensemble[?][?][?][?][?][?][?] methods); (*iii*) as an equilibrium multiple-replica variant of the slow-growth[?][?] (SG) method

(bypassing the associated hysteresis issues[?][?][?] or the requirement for exponential averaging over multiple repeats[?][?][?][?]).

Compared to plain TI, it shares the advantage of HRE/HRP and λ-LEUS in terms of enhanced orthogonal sampling.[?][?][?][?] Compared to HRE/HRP, it permits a deterministic and continuous sampling of the λ-range, and bypasses the need for a careful preselection[?][?][?][?][?][?][?][?][?] of the λ-ladder and exchange-attempt interval. Compared to both TI and HRE/HRP, the quasi-homogeneous λ-sampling also essentially removes quadrature errors. Finally, compared to λ-LEUS, it eliminates (or drastically reduces) the dead time associated with the preoptimization of a biasing potential[?] or, alternatively, the use of this formally non-equilibrium statistics[?] in the production calculation.[?] For the above reasons, the CBTI scheme certainly represents a useful addition to the alchemical free-energy calculation toolkit.

Like TI and HRE/HRP, the CBTI method is also intrinsically parallel. However, assuming that the replicas are assigned to separate processors (including possible GPU implementations or/and cloud-computing applications), the requirement of an all-to-all information exchange between processors at every timestep might represent a drawback of the method relative to the no-exchange and infrequent exchange situations of TI and HRE/HRP, respectively. Although the communication is lightweight (Hamiltonian λ-derivative, *i.e.* a single real number), the synchronization requirement may cause a performance loss (reduced scalability and fault tolerance). Unless asynchronous variants[?][?] or multiple-timestep schemes[?][?] can be developed, this performance loss may represent a problem for parallel applications in situations involving many replicas of a small system, as these will involve more data exchange at a more frequent rate.

This scheme opens the way to at least two types of generalizations and extensions. First, a number of components of the scheme can be modified/generalized and in particular the following: (*i*) different functions ζ may be used in Eq. 1.6 to modulate the replica density along λ (*e.g.* smoothing the tips or adding plateaus[?] for a denser sampling close to or at the physical end-states); (*ii*) different matrices $\underline{\mathbf{C}}$ may be used in Eq. 1.14 corresponding to a to non-uniform weighting of the λ-forces

into the Λ-force, (*e.g.* canceling the effect of higher-order derivatives by alternating non-integer weights in analogy to standard quadrature methods); (*iii*) the coupling between replicas may be generalized from sequential pairwise constraints to possibly non-pairwise potentials (*e.g.* collective or harmonic); (*iv*) the TI-like free-energy estimator of Eq. 1.22 may be replaced by a statistically more powerful one of the MBAR type;[?,?,?,?,?,?,?] (*v*) CBTI would benefit from the use of an alchemical coupling path presenting a vanishing free-energy derivative at the physical end-states (residual free-energy barriers along Λ decaying at least in K^{-2} instead of K^{-1} for large K, as shown in Sect. 1.A).

Second, the application range of CBTI, restricted here to alchemical processes, can be extended to encompass either thermodynamic or conformational free-energy changes. In the former case, extension to a CB variant of parallel tempering[?] (*i.e.* a CBPT scheme) appears relatively straightforward considering that scaling the temperature is equivalent to scaling the system potential energy. Such a CBPT scheme would represent a form of multicanonical sampling.[?,?,?] In the latter case, extension to a CB variant or umbrella sampling[?,?] (*i.e.* a CBUS scheme) could be designed *e.g.* by anchoring harmonic potentials to the CB and propagating the corresponding restraining forces onto the CB. Finally the extension of the CB approach to multistate problems[?,?,?,?] (*e.g.* path or network of CBs connecting the different states of a system) as well as multidimensional problems (sub-CBs anchored to a main CB) can also be envisioned. Work is currently in progress along these lines.

1.A Appendix RELATIONSHIP BETWEEN CBTI AND QUADRATURE INTEGRATION

This appendix investigates the convergence properties of the CBTI scheme upon increasing the number K of replicas based on an analogy with quadrature integration. More precisely, it is shown that the variation amplitude $G_\Lambda^\star$ of $G_\Lambda(\Lambda)$ decreases at least as K^{-1} in the limit of large K (with K even), as suggested by Fig. 1.1b and 1.6.

To this purpose, we first observe that the situation of a conveyor belt (CB) spanning the λ-range $[0, 1]$ of the free-energy profile $G(\lambda)$ with K equidistant replicas at spacings $2K^{-1}$ along the cable (Fig. 1.A.1a) is mathematically equivalent to the situation of a train spanning a periodic function $g(x)$ of period 2π with K equidistant carriages at spacings $2\pi K^{-1}$ spanning one period (Fig. 1.A.1b). The function $g(x)$ is obtained by mirroring $G(\lambda)$ at $\lambda = 0$ or 1, periodically translating the mirrored function by integer multiples of two, and stretching λ by a factor of π to define the new variable x.
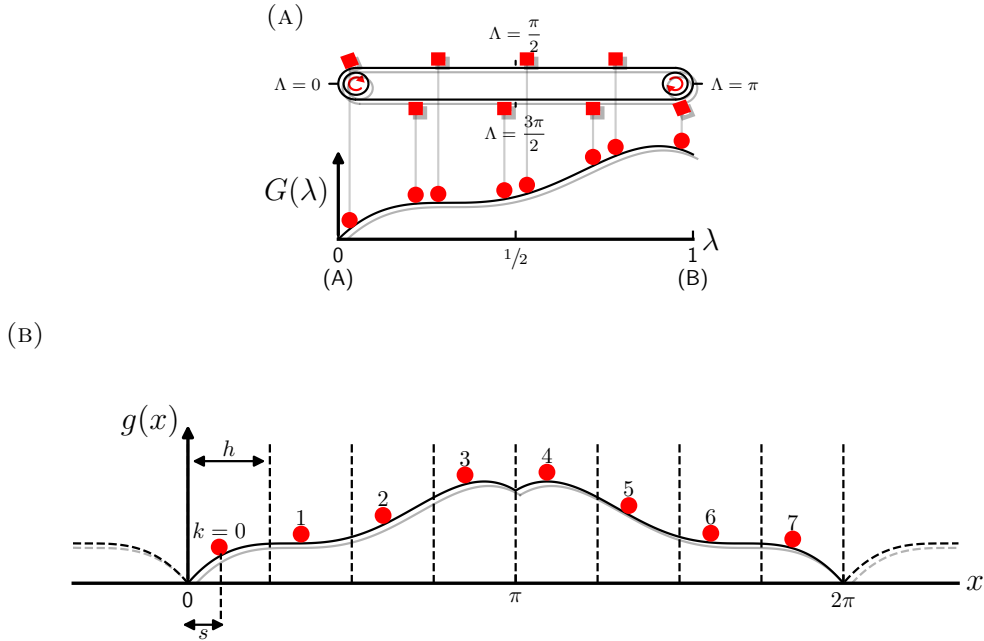
FIGURE 1.A.1: *Analogy between the CBTI scheme and the sampling of an even periodic function by a train of equispaced points.* The illustration considers a conveyor belt (CB) with $K = 8$ replicas, analog to a train with 8 carriages.

In this analogy, we now have a periodic function $g(x)$ of period 2π that is even over the reference interval $[0, 2\pi)$. This function is sampled by K equidistant points at locations

$$x_k = s + kh \quad \text{with} \quad k = 0 \ldots K - 1 \quad \text{and} \quad h = 2\pi K^{-1}, \tag{1.A.1}$$

where s represents the offset position of the first point $k = 0$. Note that s can always be selected in the interval $[0, h)$ by consideration of periodicity, along with an appropriate renumbering of the points carriages. We also introduce the negative first derivative of the function

$$f(x) = -g'(x). \tag{1.A.2}$$

In practical applications of CBTI, the Hamiltonian-coupling scheme will be defined by a continuously differentiable function of λ, in which case $G(\lambda)$ is also continuously differentiable. The same applies to $g(x)$ in the analogy of Fig. 1.A.1b, except for the possible occurrence of kinks at integer multiples of π. These occur because the derivative $G'(\lambda)$ of $G(\lambda)$ generally differs from zero at the physical end-states $\lambda = 0$ or $\lambda = 1$. As a result, $f(x)$ is continuous except at these points, where it will present jumps. An illustrative example for functions $g(x)$ and $f(x)$ with the above properties is shown in Figs. 1.A.2a and 1.A.2b, respectively.

In the analogy of Fig. 1.A.1b to the CB situation, the total force on the train of points (analog to the mean force on the Λ-variable) is given for a specific value of s (analog to the $\tilde{\Lambda}$-variable) by

$$F(s) = \sum_{k=0}^{K-1} f(kh + s). \tag{1.A.3}$$

As illustrated in Fig. 1.A.2c, this sum has a simple interpretation. It represents a trapezoidal quadrature estimate to the integral of the periodic function $f(x)$ over the

period $[s, 2\pi + s)$. Because $f(x)$ is the negative derivative of the periodic function $g(x)$, its exact integral over one period must be zero. Thus, for any s, we expect $F(s)$ to converge to zero in the limit $K \to \infty$. The following paragraphs investigate the corresponding convergence rate.

The function $F(s)$ is periodic of period h, odd over the reference interval $[0, h)$ and generally discontinuous at s values which are integer multiples of h. Introducing

$$\tilde{s} = h^{-1}s \quad \text{and} \quad \tilde{F}(\tilde{s}) = F(h\tilde{s}) = \sum_{k=0}^{K-1} f((k + \tilde{s})h), \tag{1.A.4}$$

the function $\tilde{F}(\tilde{s})$ is periodic of period 1, odd over the reference interval $[0, 1)$ and generally discontinuous at integer values of $\tilde{s}$.

The corresponding potential energy of the train of points (analog to the free-energy profile $G_\Lambda(\Lambda)$ of the CB over the interval $[0, 2\pi K^{-1}]$) is given by

$$G(s) = -\int_0^s \mathrm{d}s' F(s'). \tag{1.A.5}$$

The function $G(s)$ is periodic of period h, even over the reference interval $[0, h)$ with an extremum at $h/2$, and possibly presents kinks at s values which are integer multiples of h. Using the variable $\tilde{s}$ and introducing

$$\tilde{G}(\tilde{s}) = -\int_0^{\tilde{s}} \mathrm{d}\tilde{s}' \tilde{F}(\tilde{s}') \,, \tag{1.A.6}$$

one may write

$$G(s) = h\tilde{G}(h^{-1}s). \tag{1.A.7}$$

The function $\tilde{G}(\tilde{s})$ is periodic of period 1, even over the reference interval $[0, 1)$ with an extremum at $1/2$, and possibly presents kinks at integer values of $\tilde{s}$.

The functions $\tilde{F}(\tilde{s})$ and $\tilde{G}(\tilde{s})$ corresponding to the illustrative functions $g(x)$ and $f(x)$ are shown in Figs. 1.A.2d and 1.A.2e, respectively, considering the choices

$K = 2, 4, 8, 16, 32$ or 64. Convergence is observed upon increasing K, in which $\tilde{F}(\tilde{s})$ becomes linear and $\tilde{G}(\tilde{s})$ parabolic over the reference interval $[0, 1)$. Thus, in the limit of large K, one expects the maximal variation $\tilde{G}^\star$ in $\tilde{G}(\tilde{s})$ to converge to a constant and, *via* Eq. 1.A.7, the maximal variation $G^\star$ in $G(s)$ (analog to the corresponding maximal variation $G_\Lambda^\star$ of $G_\Lambda(\Lambda)$ in the CB scheme) to scale as h, *i.e.* as K^{-1}.

This observation, made here in a special case, can be generalized as follows. Given

$$R(\tilde{s}) = h\tilde{F}(\tilde{s}) = h \sum_{k=0}^{K-1} f(h(k + \tilde{s})), \tag{1.A.8}$$

$G^\star$ is the absolute extremum (largest absolute value) of the function

$$S(\tilde{s}) = h\tilde{G}(\tilde{s}) = \int_0^{\tilde{s}} d\tilde{s}' R(\tilde{s}') \tag{1.A.9}$$

over the $\tilde{s}$-interval $[0, 1/2]$. According to Fig. 1.A.2c, the quantity $R(\tilde{s})$ is the residual of the trapezoidal quadrature approximation to the integral of $f(x)$ over one period, where the first grid point is offset from the origin by a fraction $\tilde{s}$ of the spacing h. If $f(x)$ was continuous everywhere, the convergence would be quadratic in h (*i.e.* scale as K^{-2}), as expected from a trapezoidal quadrature. However, the presence of discontinuities in $f(x)$ at 0 and π introduces an error that is linear in h. To recover quadratic convergence, one would need to introduce a correcting term for the points $0, K/2 - 1, K/2$ and $K - 1$ surrounding the discontinuities, namely

$$\begin{aligned} R'(\tilde{s}) = R(\tilde{s}) + h\Bigg\{ &\frac{\tilde{s} - 1}{2}\left[f(h\tilde{s}) + f\left(h\left(\tilde{s} + \frac{K}{2}\right)\right)\right] \\ &- \frac{\tilde{s}}{2}\left[f(h(\tilde{s} + K - 1)) + f\left(h\left(\tilde{s} + \frac{K}{2} - 1\right)\right)\right] \\ &+ \left(\tilde{s} - \frac{1}{2}\right)(\Delta_0 + \Delta_\pi)\Bigg\} \end{aligned} \tag{1.A.10}$$

where Δ_0 and Δ_π account for the magnitude of the discontinuities, *i.e.*

53

$$\Delta_0 = \lim_{x \xrightarrow{\geq} 0} f(x) \quad \text{and} \quad \Delta_\pi = \lim_{x \xrightarrow{\geq} \pi} f(x). \tag{1.A.11}$$

The convergence of $R(\tilde{s})$ and $R'(\tilde{s})$ for the illustrative functions $g(x)$ and $f(x)$ upon increasing K along with the choices $\tilde{s} = 0.01, 0.2, 0.3, 0.4, 0.49$ is shown in logarithmic form in Figs. 1.A.2f and 1.A.2g, respectively. As expected, for large K, R shows a linear convergence (slope -1) and R' a quadratic one (slope -2). The limiting cases $\tilde{s} = 0$ and $\tilde{s} = 1/2$ are special. For $\tilde{s} = 0$, Eq. 1.A.8 implies an evaluation of $f(x)$ at the discontinuity. If one uses an average value of 0 for these points, $R(0)$ evaluates to 0 for any K by symmetry. The same applies for $\tilde{s} = 1/2$, $R(1/2)$. For all other values for $\tilde{s}$, the convergence is as expected, *i.e.* linear for $R(\tilde{s})$ and quadratic for $R'(\tilde{s})$.

If $R(\tilde{s})$ converges to zero as K^{-1} for (nearly) all $\tilde{s}$ values, the function $S(\tilde{s})$ of Eq. 1.A.9 will converge to zero with the same scaling, and so will its absolute extremum $G^\star$. We conclude that the magnitude $G^\star_\Lambda$ of the residual variations in $G_\Lambda(\Lambda)$ for the CBTI scheme scales at least as K^{-1} in the limit of large K. And if $G(\lambda)$ has a vanishing derivative at the physical end-states, it will scale at least as K^{-2}. It should be stressed, however, that these are worst-case scalings. A higher level of continuity or specific symmetry properties of $G(\lambda)$ may tighten the scaling. To give an extreme case, for a free-energy profile $G(\lambda) = \cos(2\pi\lambda)$, one shows easily that the residual variations $G^\star_\Lambda$ entirely vanish irrespective of K. This follows from

$$\sum_{k=0}^{K-1} \cos\left(2\pi K^{-1} k + c\right) = 0 \quad \forall\, c \quad \text{(for } K \text{ even)}. \tag{1.A.12}$$

The present analysis connects the convergence of the CBTI scheme with K to the properties of a quadrature integration. This connection opens interesting tracks for future work. In particular, it suggests that these convergence properties could be improved by altering the coupling scheme (*e.g.* to enforce vanishing $G(\lambda)$ derivatives at the end-states), the function ζ in Eq. 1.6 (*e.g.* higher system density close to discontinuities) and the CBTI weighting in Eq. 1.4 (*e.g.* from trapezoidal to

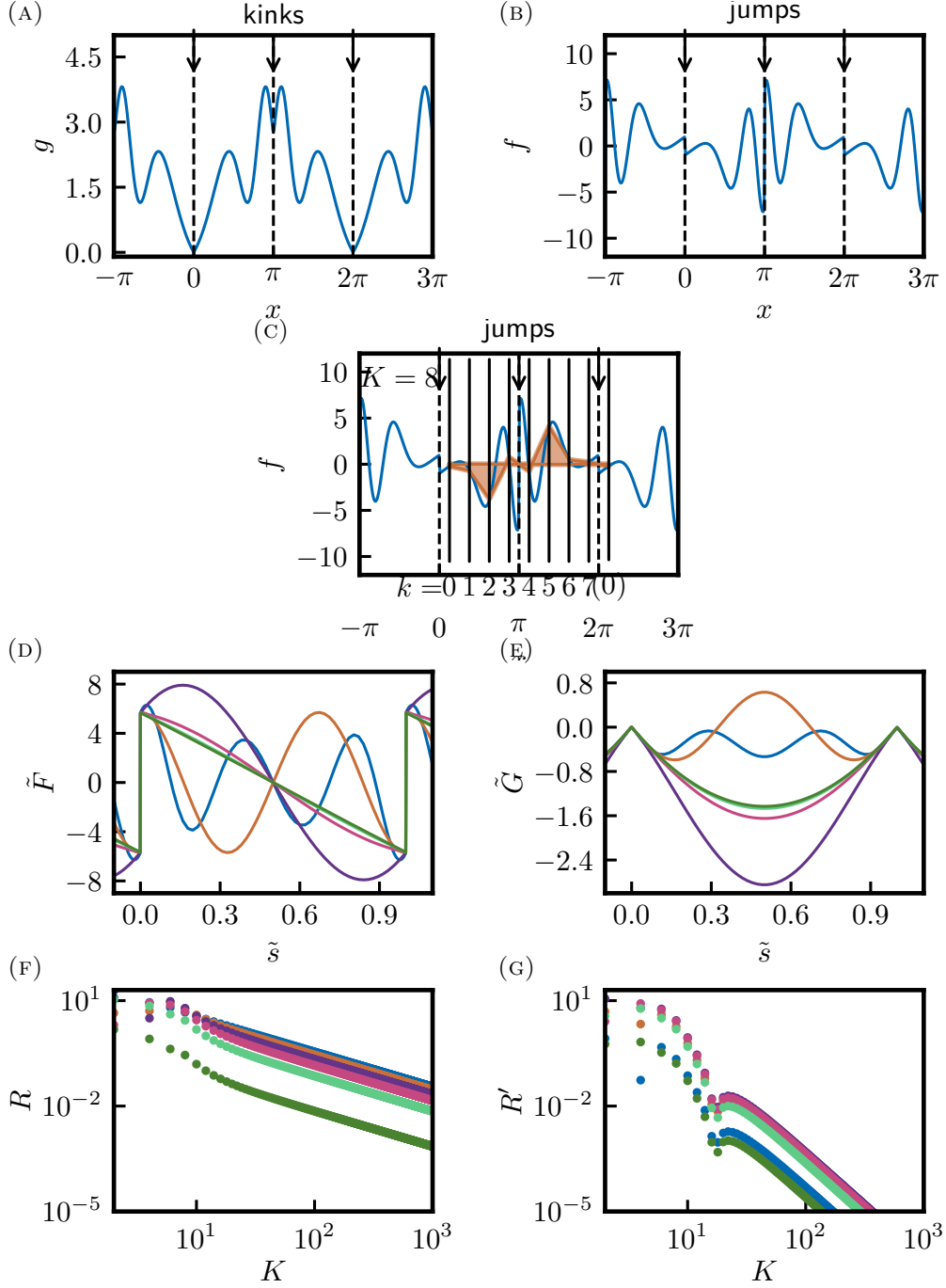higher-level quadrature, *e.g.* Simpson or even Romberg).

FIGURE 1.A.2: *Illustrative functions supporting the discussion in Sect. 1.A.* Panel (a) shows the illustrative function $g(x)$ defined over the reference period $[0; 2\pi[$ as $x + \sin(x^2)$ if $x < \pi$ or $(2\pi - x) + \sin((2\pi - x)^2)$ if $x \geq \pi$. Panel (b) shows the negative derivative $f(x)$ of $g(x)$. Panel (c) illustrates the trapezoidal integration of $f(x)$ (blue) as a shaded area (orange) with $K = 8$ and $s = 0.4$. The vertical lines show the positions of the carriages at values x_m. The trapezoids are seen to represent a poor approximation at the discontinuities. Panel (d) illustrates the total force $\tilde{F}(\tilde{s})$ of Eq. 1.A.4 for different values of K. Panel (e) illustrates the potential energy $\tilde{G}(\tilde{s})$ for different values of K. which is the running integral of $\tilde{F}$ for different numbers of replicas K. Panel (f) shows in a log-log form the residual integral $R(\tilde{s})$ of Eq. 1.A.8 as a function of K for different values of $\tilde{s}$. Panel (g) shows in a log-log form the corrected residual integral $R'(\tilde{s})$ of Eq. 1.A.10. Note that the logarithm of the absolute value of R or R' is displayed. The cusp in Panel (g) is explained by a change of sign.

1.B Appendix CHOICE OF THE CBTI
PARAMETERS

Here, we explore the influence of the CBTI mass parameter m_Λ and thermostat coupling time τ_Λ considering different numbers K of replicas.

A first set of simulations involves the choices $m_\Lambda = 16, 160, 800, 1600$ or $3200\,\mathrm{u\,nm}^2$, along with $K = 16$ replicas in the absence of separate thermostat coupling for Λ, *i.e.* $\tau_\Lambda \to \infty$ (10 ns simulations after 0.2 ns equilibration). The results are shown graphically in Figs. 1.F.1 - 1.F.5 and reported numerically in Tab. 1.1 (entries 1-5). In the absence of explicit thermostating for the Λ-variable, the average temperature T_Λ of this variable is controlled exclusively by its coupling to the conformational degrees of freedom, themselves thermostated by the application of SD with a reference temperature of $298.15\,\mathrm{K}$ and an effective coupling time of $0.1\,\mathrm{ps}$. All choices of m_Λ considered lead to average temperatures T_Λ close to $298.15\,\mathrm{K}$, suggesting an appropriate kinetic-energy exchange. The mass-parameter choices of $m_\Lambda = 16\,\mathrm{u\,nm}^2$ and $1600\,\mathrm{u\,nm}^2$ present the largest deviations of about $10\,\mathrm{K}$ and $6\,\mathrm{K}$, respectively, while the deviation is at most $2\,\mathrm{K}$ for the other choices.

Expectedly, increasing m_Λ tendentially leads to a decrease in the diffusion coefficient D_Λ. This results from a smaller width $\sigma_{\dot\Lambda}$ of the $\dot\Lambda$ distribution (Eq. 1.25; see also Fig. 1.5a where m_Λ increases from left to right), *i.e.* a lower average magnitude of the velocity along Λ. However, D_Λ is also affected by the presence of residual variations in the free-energy profile $G_\Lambda(\Lambda)$, which may be more easily overcome at higher m_Λ (more inertia). This is reflected in the increase of the $\dot\Lambda$ autocorrelation time $\tau_{\dot\Lambda}$ upon increasing m_Λ (see also Fig. 1.4 for the corresponding normalized autocorrelation functions). This second effect is likely to have more influence for small K, where the $G_\Lambda(\Lambda)$ variations are more pronounced. In the present case with $K = 16$, these opposite effects probably explain why the trend in D_Λ upon increasing m_Λ is not strictly monotonic, with a comparatively high D_Λ for $m_\Lambda = 1600\,\mathrm{u\,nm}^2$ (see also Fig. 1.5c where the diffusion trend is non-monotonic upon increasing K

when m_Λ is made proportional to $K^{1/2}$).

In terms of the calculated free-energy change ΔG, all simulations provide the same values within the statistical error, with a slightly larger deviation for the simulation involving the lowest mass $m_\Lambda = 16 \, \mathrm{u \, nm^2}$, probably due to the comparatively large T_Λ deviation. Because it leads to accurate T_Λ and ΔG values while presenting a high D_Λ, the second-to-lowest mass $m_\Lambda = 160 \, \mathrm{u \, nm^2}$ was chosen for all following CBTI simulations employing $K = 16$ replicas.

A second set of simulations involves a separate coupling of the Λ variable to a Nosé-Hoover chain thermostat with the choices $\tau_\Lambda = 0.05, 0.1, 0.5, 1$ or $2 \, \mathrm{ps}$ along with $K = 16$ replicas and $m_\Lambda = 160 \, \mathrm{u \, nm^2}$ (10 ns simulations after 0.2 ns equilibration). The results are shown graphically in Figs. 1.F.6 - 1.F.10 and reported numerically in Tab. 1.1 (entries 6-10). Except for the two shortest coupling times, the average temperature T_Λ is very close to the reference temperature, with a maximal deviation of about 2 K. The larger deviations for $\tau_\Lambda = 0.05$ or $0.1 \, \mathrm{ps}$ suggest that the coupling of Λ to its thermostat should be made less tight than that of the conformational degrees of freedom to their thermostat (here SD with 0.1 ps). The coupling parameter τ_Λ has no influence on the width $\sigma_{\dot{\Lambda}}$ of the $\dot{\Lambda}$ distribution, which is the same for all five simulations and very close to the Maxwell-Boltzmann one (Eq. 1.25). On the other hand, a lower τ_Λ, *i.e.* a tighter coupling, reduces the inertia of the conveyor belt, which results in a decrease of the autocorrelation time $\tau_{\dot{\Lambda}}$. This leads to a tendential decrease of D_Λ upon decreasing τ_Λ. In terms of the calculated ΔG, all simulations provide consistent values within the statistical error, with a slightly larger deviation for the simulation involving $\tau_\Lambda = 0.05 \, \mathrm{ps}$, probably again due to the comparatively large T_Λ deviation. In comparison to the first set of simulations with $\tau_\Lambda \to \infty$, the ΔG values are slightly higher ($0.1 - 0.3 \, \mathrm{kJ \, mol^{-1}}$), which could be due to the fact that the sampled $\dot{\Lambda}$ distribution $P_{\dot{\Lambda}}$ is even closer to the Maxwell-Boltzmann distribution (Eq. 1.25), also for shorter time intervals. For $\tau_\Lambda \geq 0.5 \, \mathrm{ps}$, the choice of τ_Λ has little influence on the calculation, and a coupling time $\tau_\Lambda = 0.5 \, \mathrm{ps}$ was chosen for all following CBTI simulations. Note, however, that a looser coupling could be of advantage by leading to a higher D_Λ.

A third set of simulations considers the choices $K = 8, 16, 32, 64$ or 128 along with $m_\Lambda = 40 K^{1/2} \, \mathrm{u} \, \mathrm{nm}^2$ and $\tau_\Lambda = 0.5 \, \mathrm{ps}$ ($256 K^{-1}$ ns simulations after 0.2 ns equilibration). The results are shown graphically in Figs. 1.F.11 - 1.F.15 and reported in Tab. 1.1 (entries 11-15). Simulations 11, 13 and 15 are discussed in details in Sect. 1.4 (see Figs. 1.3 and 1.5). For the five simulations, the average temperature T_Λ is close to the target temperature, with a maximal deviation of about $6 \, \mathrm{K}$. The free-energy differences ΔG are consistent within the statistical error, except for the simulation employing $K = 8$ replicas (entry 11), which is due to the non-uniform sampling of the λ-range as discussed in Sect. 1.4.

The rationale for making the mass m_Λ proportional to the square-root of K, an arbitrary parameter choice, is the following. If one wishes the CB variable Λ to evolve dynamically on comparable timescales irrespective of the number K of replicas attached to it, one should ensure that its acceleration $\ddot{\Lambda}$ depends only weakly on K. Assuming that the forces exerted by the K replicas are essentially uncorrelated, their sum (*i.e.* the net force on the CB) will scale as $K^{1/2}$. Thus, an identical scaling of m_Λ is required to preserve an approximately constant $\ddot{\Lambda}$. The time series and distributions of $\ddot{\Lambda}$ for different choices of K when using this scaling for m_Λ (see Fig. 1.5b) are indeed similar, supporting the above considerations.

In summary, this appendix shows that when selected within reasonable ranges, the parameters m_Λ and τ_Λ have only a limited influence on the kinetic-energy exchange between Λ and the conformational degrees of freedom, on the temperature T_Λ, on the diffusion constant D_Λ, and on the calculated free energies ΔG. In Sect. 1.4, working choice $m_\Lambda = 40 K^{1/2} \, \mathrm{u} \, \mathrm{nm}^2$ and $\tau_\Lambda = 0.5 \, \mathrm{ps}$ is selected for all CBTI calculations.

1.C Appendix SIMPLIFIED FREE-ENERGY ESTIMATORS

The free-energy estimator employed here for CBTI is given by Eq. 1.22. The approximation involved corresponds to a simple forward rectangular quadrature, where the Hamiltonian derivative is averaged over J successive bins considering all replicas simultaneously. If K replicas sample L configurations each, and $p(\lambda)$ is close to homogeneous, the number of data points per bin will be close to KL/J, with limited variations across bins. In this case, one may consider a simpler alternative to Eq. 1.22 that does not involve the specification of a number of bins.

Considering all the K replicas simultaneously, the KL pairs of λ-values and associated Hamiltonian derivatives are sorted in ascending order for λ (index $i = 0, ..., KL - 1$), One then calculates

$$\Delta G_{\mathrm{alt}} = (KL)^{-1} \sum_{i=0}^{KL-1} \frac{\lambda_{i+1} - \lambda_{i-1}}{2} \left(\frac{\partial \mathcal{H}}{\partial \lambda} \right)_i \tag{1.C.1}$$

using $\lambda_{-1} = \lambda_{KL} = 0$. This alternative estimate, noted ΔG_{alt}, considers that each sample i defines its own single-point bin, of width determined by contact to the next lower and higher data points $i - 1$ and $i + 1$, respectively, thereby accounting for a possible heterogeneity in the λ-sampling.

In the limit where the sampling becomes sufficiently close to homogeneous, Eqs. 1.22 or 1.C.1 can be replaced by an even simpler approximate expression, namely

$$\Delta G_{\mathrm{app}} \approx K^{-1} \sum_{k=0}^{K-1} \left\langle \frac{\partial \mathcal{H}(\mathbf{x}_k; \lambda_k)}{\partial \lambda_k} \right\rangle^{\dagger\star}, \tag{1.C.2}$$

which corresponds to averaging the Hamiltonian derivative over all replicas and over the entire CBTI simulation. This approximate estimate, noted ΔG_{app}, assumes that the KL data points are on average equispaced along λ, which only holds in the limit of homogeneous sampling.

The accuracy of the estimate ΔG of Eq. 1.22 with $J = 500$ (except for entry 11, $J = 200$) is compared in Tab. 1.1 to those of the simpler expressions ΔG_{alt} and ΔG_{app}, which do not require the specification of a number of bins. The estimate ΔG_{alt} is very close to ΔG (within error bar), suggesting that Eq. 1.C.1, which assumes that each sampled λ-value is at the center of its own bin, is a viable parameter-free alternative to Eq. 1.22. On the other hand, the estimate ΔG_{app}, which assumes that each bin encompasses the same number of sampled λ-points irrespective of J, is more approximate. As expected from the involved assumption, ΔG_{app} is only a good approximation to ΔG when the CBTI sampling is indeed close to uniform along λ. Considering entries 11-17, this is essentially the case for the unbiased simulations with $K \geq 32$. However, for the unbiased simulation with $K = 8$ and the two biased simulations, the sampling is not sufficiently homogeneous and the discrepancy can be significant (about $1 - 5\,\mathrm{kJ\,mol}^{-1}$).

1.D Appendix AUTOCORRELATION OF THE HAMILTONIAN DERIVATIVE
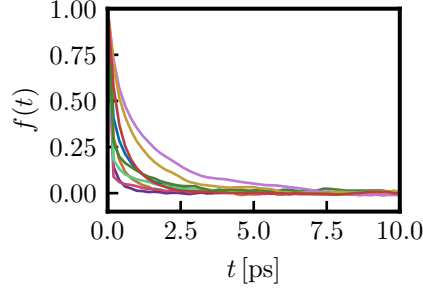
FIGURE 1.D.1: *Normalized autocorrelation function $f(t)$ of the Hamiltonian derivative for the aqueous methanol-to-dummy mutation at* $298.15\,\mathrm{K}$ *and* $1\,\mathrm{bar}$ *with the 2016H66 force field.* These functions are based on the TI calculation relying on $K_{\mathrm{TI}} = 17$ λ-points. The functions displayed correspond to every second λ-point. The associated autocorrelation times τ_f are reported in Tab. 1.D.1.

TABLE 1.D.1: *Autocorrelation times τ_f of the Hamiltonian derivative for the aqueous methanol-to-dummy mutation at* $298.15\,\mathrm{K}$ *and* $1\,\mathrm{bar}$ *with the 2016H66 force field.* These times are based on the TI calculation relying on $K_{\mathrm{TI}} = 17$ λ-points. The associated autocorrelation functions $f(t)$ are displayed in Fig. 1.D.1 for every second λ-point. The autocorrelation times were derived by performing an expontential fit to the $f(t)$ curve.

λ	τ_f [ps]	λ	τ_f [ps]
0.000	0.339	0.562	0.161
0.062	0.297	0.625	0.224
0.125	0.207	0.688	0.377
0.188	0.150	0.750	1.080
0.250	0.109	0.812	1.119
0.312	0.085	0.875	0.799
0.375	0.087	0.938	0.555
0.438	0.112	1.000	0.442
0.500	0.139		

1.E Appendix REFERENCE TI CALCULATIONS
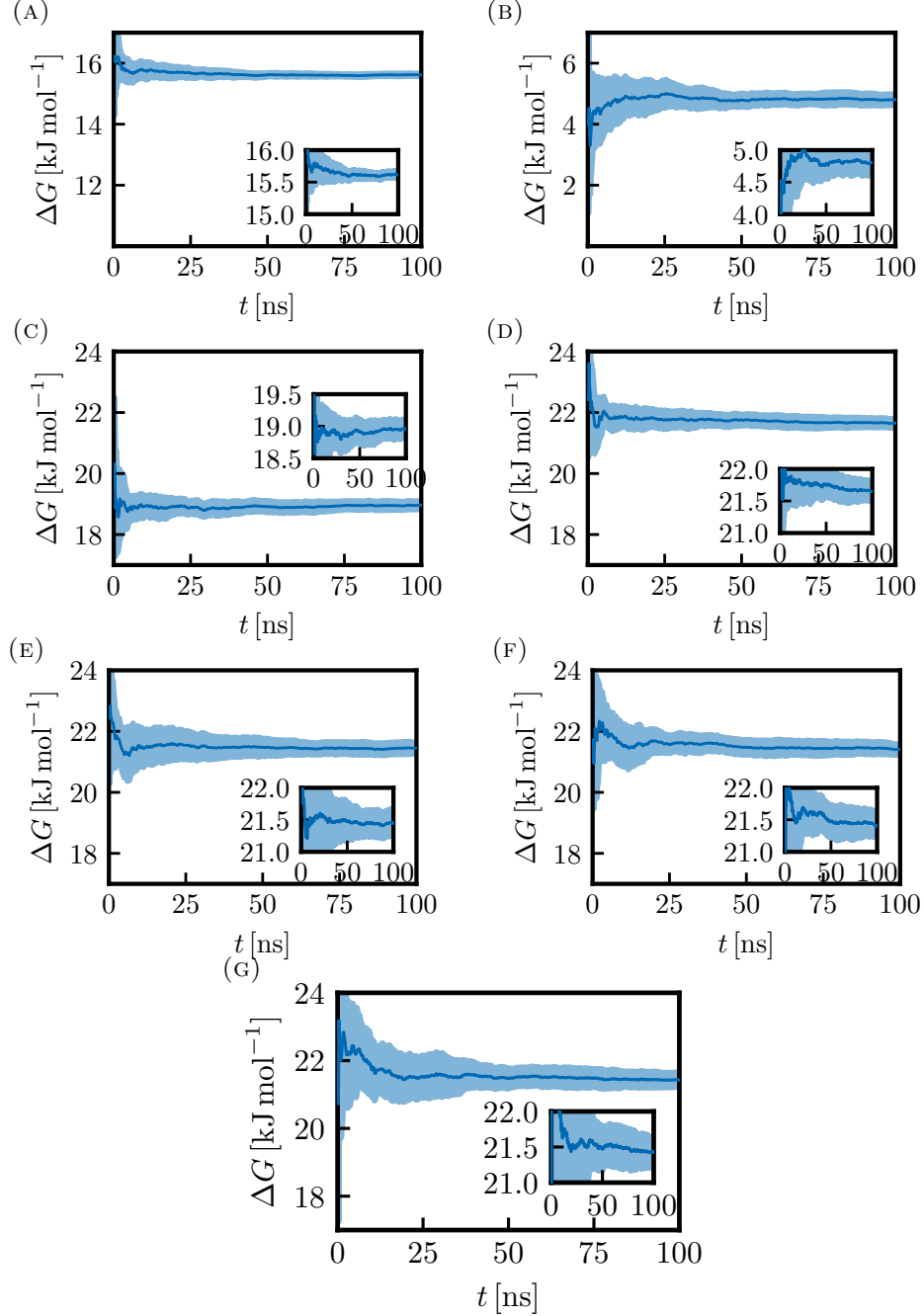
FIGURE 1.E.1: *Convergence properties of the TI calculations for the aqueous methanol-to-dummy mutation at* 298.15 K *and* 1 bar *with the 2016H66 force field.* This figure complements Fig. 1.2b by also including the TI calculations involving fewer than 129 λ-points. Here, the running ΔG estimates are shown for calculations involving $K_{\mathrm{TI}} = 2^n + 1$ equidistant λ-points with $n = 1, 2, .., 7$ resulting in $K_{\mathrm{TI}} = 3$ a, 5 b, 9 c, 17 d, 33 e, 65 f and 129 g. Note that the vertical ranges in Panels a and b differ from the range used in the other graphs. The corresponding numerical values at full sampling time are reported in Tabs. 1.2 and 1.E.1.

TABLE 1.E.1: *Repeat results over TI calculations for the aqueous methanol-to-dummy mutation at* $298.15\,\mathrm{K}$ *and* $1\,\mathrm{bar}$ *with the 2016H66 force field.* The free-energy change ΔG is reported for TI calculations involving $K_{\mathrm{TI}} = 2^n + 1$ λ-points with $n = 1, 2, .., 7$. For each choice of K_{TI} the results of ten repeats (successive rows) are listed, all involving a total sampling time of $100\,\mathrm{ns}$ equally distributed over the λ-points. For each repeat, the calculated ΔG (Simpson quadrature) is reported along with a bootstrap error estimate (first parenthesis) and an error propagated from the bootstrap errors of the Hamiltonian derivative at each λ-point (second parenthesis). These two error estimates do not include a Student t-factor. Repeat statistics based on this data can be found in Tab. 1.2.

	$\Delta G\,[\mathrm{kJ\,mol^{-1}}]$						
n	1	2	3	4	5	6	7
K_{TI}	3	5	9	17	33	65	129
1	15.68 (0.08) (0.10)	4.82 (0.18) (0.15)	19.22 (0.18) (0.16)	21.87 (0.17) (0.15)	21.48 (0.20) (0.15)	21.18 (0.24) (0.15)	21.45 (0.30) (0.14)
2	15.52 (0.11) (0.10)	4.32 (0.18) (0.14)	18.72 (0.17) (0.16)	21.83 (0.18) (0.14)	21.59 (0.19) (0.14)	21.72 (0.22) (0.14)	21.87 (0.29) (0.14)
3	15.64 (0.09) (0.09)	4.50 (0.22) (0.16)	18.92 (0.17) (0.14)	21.57 (0.20) (0.14)	21.72 (0.17) (0.14)	21.62 (0.20) (0.14)	21.60 (0.26) (0.14)
4	15.58 (0.10) (0.09)	4.98 (0.17) (0.14)	19.18 (0.17) (0.14)	21.88 (0.18) (0.14)	21.85 (0.19) (0.14)	21.66 (0.22) (0.14)	21.67 (0.31) (0.14)
5	15.71 (0.10) (0.09)	4.97 (0.18) (0.15)	18.77 (0.17) (0.15)	21.37 (0.18) (0.15)	21.37 (0.22) (0.15)	21.29 (0.25) (0.14)	21.30 (0.26) (0.14)
6	15.62 (0.10) (0.09)	5.11 (0.18) (0.16)	18.99 (0.17) (0.15)	21.54 (0.21) (0.15)	21.19 (0.17) (0.15)	21.66 (0.22) (0.14)	21.76 (0.27) (0.14)
7	15.66 (0.10) (0.08)	4.75 (0.19) (0.15)	18.92 (0.14) (0.15)	21.67 (0.14) (0.15)	21.71 (0.21) (0.14)	21.29 (0.22) (0.14)	21.39 (0.29) (0.14)
8	15.48 (0.09) (0.09)	4.79 (0.18) (0.15)	18.77 (0.16) (0.15)	21.37 (0.17) (0.15)	21.03 (0.19) (0.15)	21.34 (0.25) (0.14)	20.99 (0.33) (0.14)
9	15.58 (0.09) (0.08)	4.77 (0.19) (0.16)	18.98 (0.19) (0.15)	21.49 (0.18) (0.15)	21.12 (0.18) (0.14)	21.02 (0.23) (0.14)	21.16 (0.32) (0.14)
10	15.65 (0.10) (0.09)	5.06 (0.19) (0.14)	19.17 (0.17) (0.15)	21.95 (0.19) (0.15)	21.45 (0.18) (0.15)	21.24 (0.22) (0.14)	21.09 (0.26) (0.14)

1.F Appendix UNBIASED CBTI SIMULATIONS

1.F.1 EXPLORATION OF THE INFLUENCE OF m_Λ IN CBTI

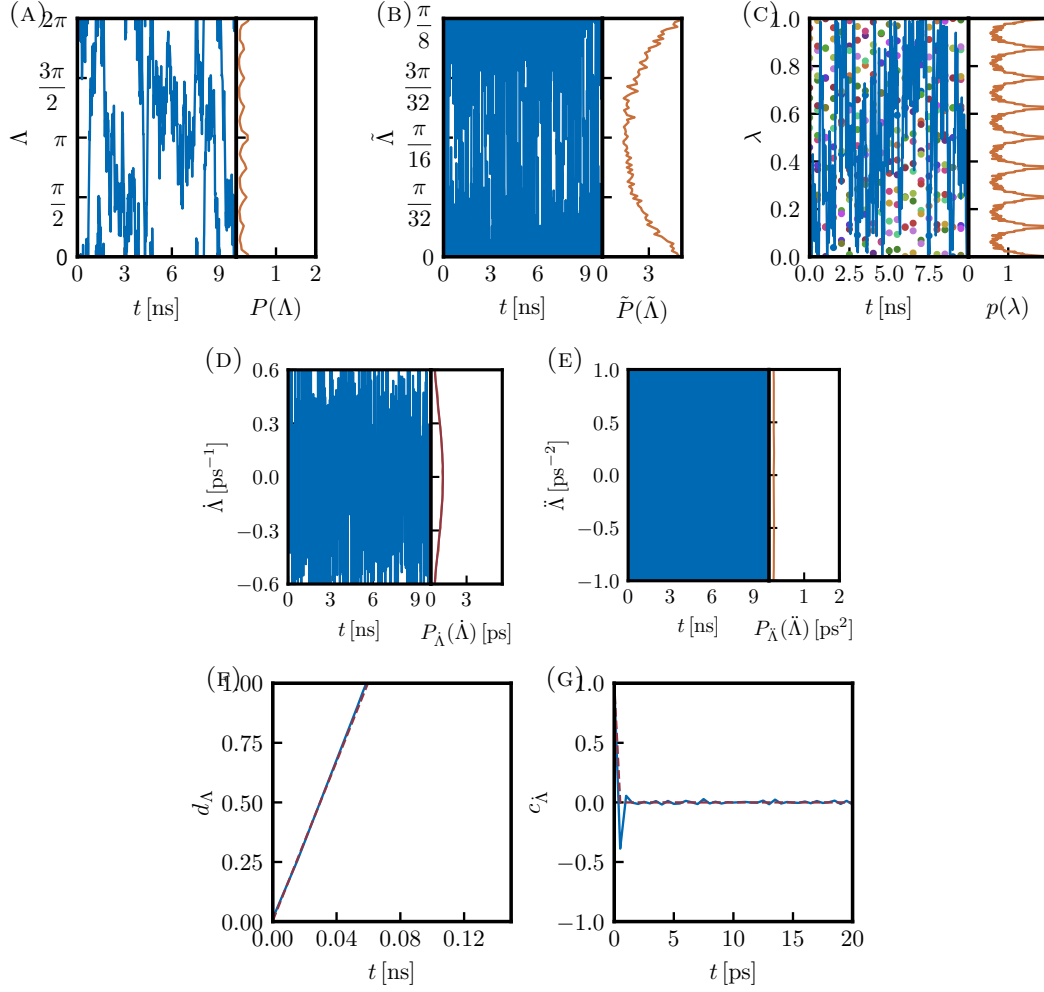**Simulation #1, $m_\Lambda = 16\,\mathrm{u\,nm^2}$**

FIGURE 1.F.1: *Results from the CBTI simulation of* $10\,\mathrm{ns}$ *employing* $K = 16$ *replicas with a mass-parameter* $m_\Lambda = 16\,\mathrm{u\,nm^2}$ *and no thermostat coupling of the* Λ*-variable* ($\tau_\Lambda \to \infty$). This simulation corresponds to entry 1 in Tab. 1.1. Panel a time series (blue) and distribution (orange) of Λ. Panel b time series (blue) and distribution (orange) of $\tilde{\Lambda}$. Panel c time series of λ for replica $k = 0$ (blue) along with the λ-values for the $K - 1$ other replicas (colored dots) as well as distribution of λ considering all replicas (orange). Panel d time series (blue) and distribution (blue) of the velocity $\dot{\Lambda}$ along with the analytical Maxwell-Boltzmann distribution (orange). Panel e time series (blue) and distribution (blue) of the acceleration $\ddot{\Lambda}$. Panel f time series of the mean-square displacement of Λ (blue) along with a linear least-square fit (dashed brown). Panel g autocorrelation function of $\dot{\Lambda}$ (blue) along with an exponential fit (dashed brown).
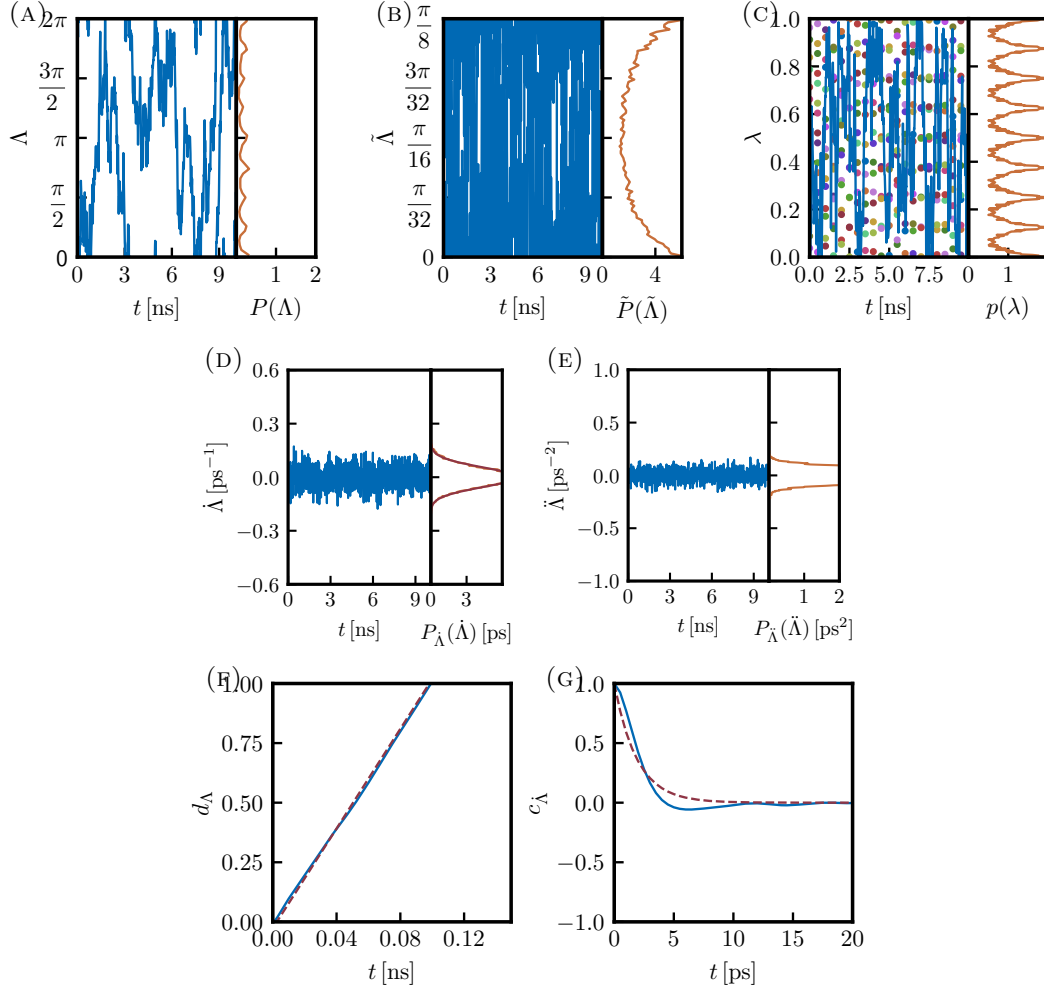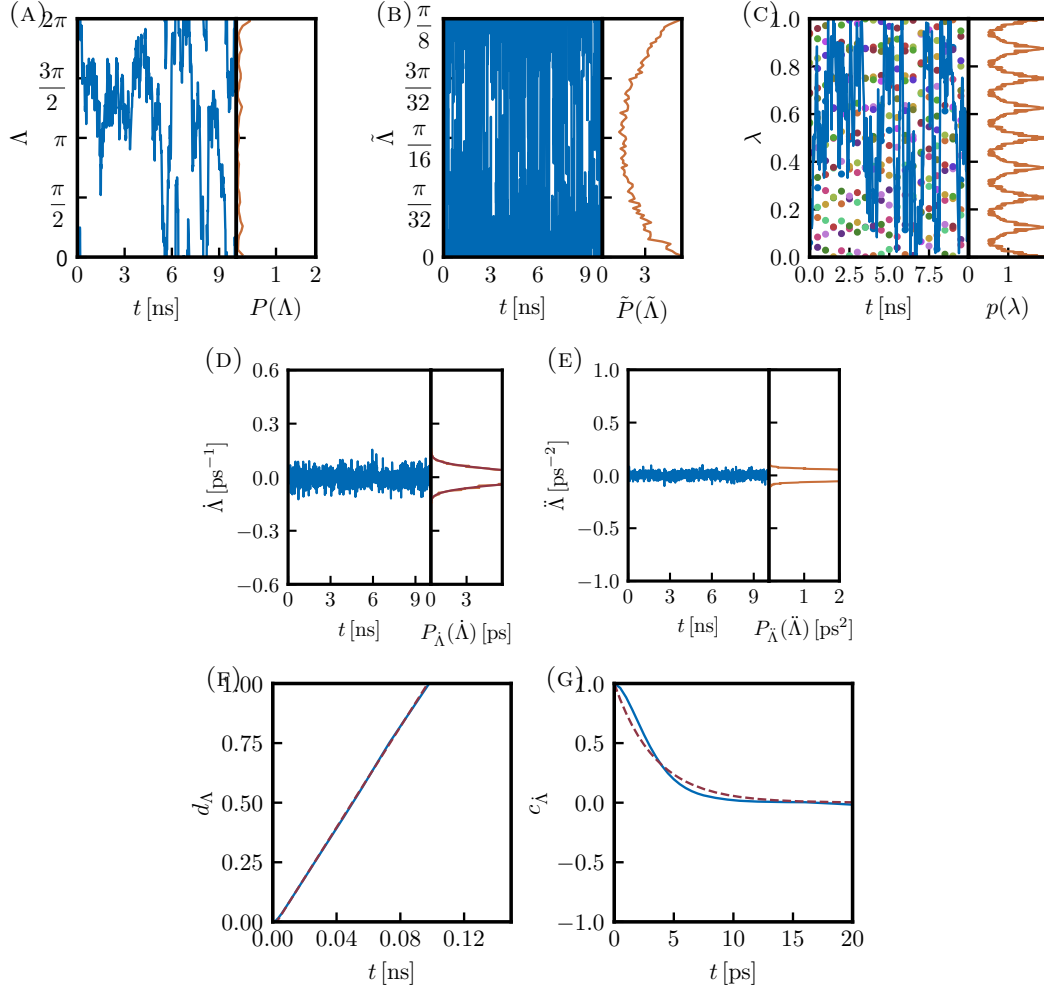
Simulation #2, $m_\Lambda = 160\,\mathrm{u\,nm}^2$
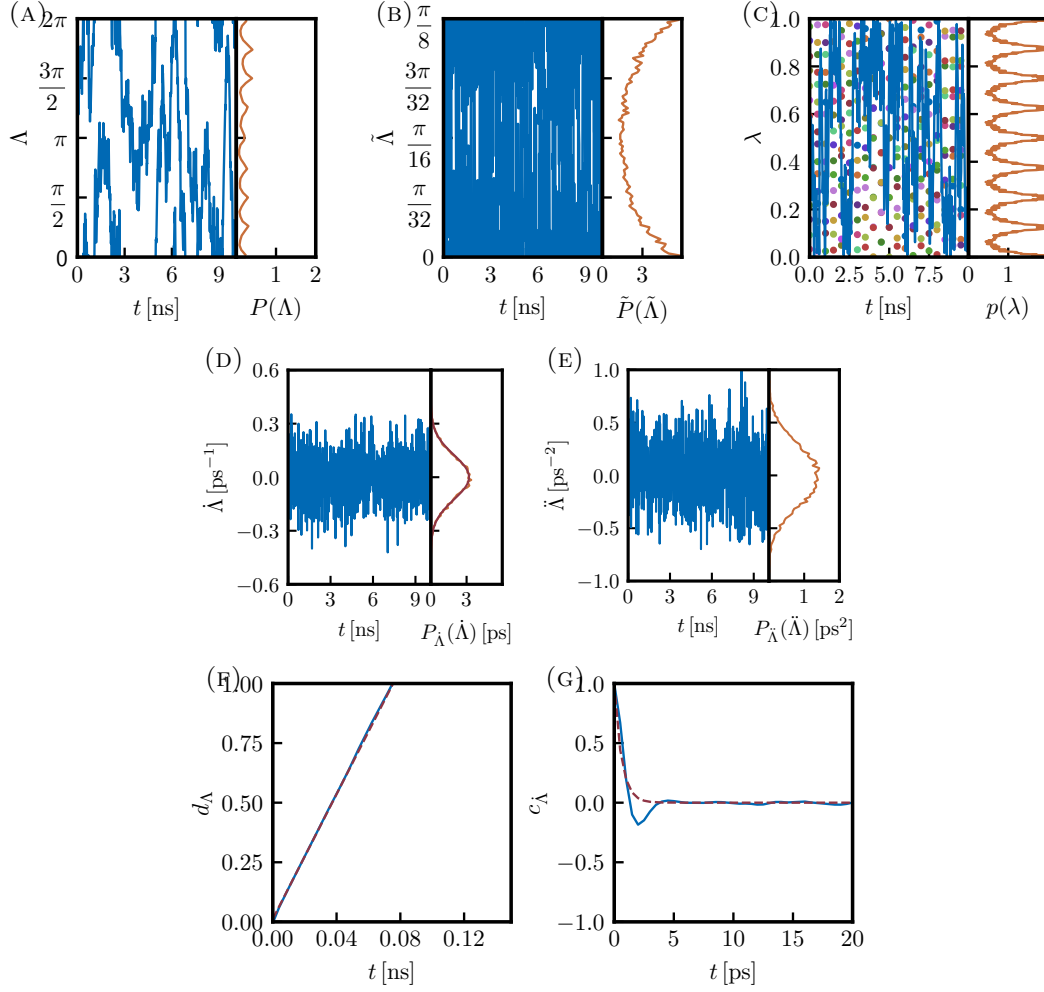
FIGURE 1.F.2: *Results from the CBTI simulation of* $10\,\mathrm{ns}$ *employing* $K = 16$ *replicas with a mass-parameter* $m_\Lambda = 160\,\mathrm{u\,nm}^2$ *and no thermostat coupling of the Λ-variable ($\tau_\Lambda \to \infty$).* This simulation corresponds to entry 2 in Tab. 1.1. Panel a time series (blue) and distribution (orange) of Λ. Panel b time series (blue) and distribution (orange) of $\tilde{\Lambda}$. Panel c time series of λ for replica $k = 0$ (blue) along with the λ-values for the $K - 1$ other replicas (colored dots) as well as distribution of λ considering all replicas (orange). Panel d time series (blue) and distribution (blue) of the velocity $\dot{\Lambda}$ along with the analytical Maxwell-Boltzmann distribution (orange). Panel e time series (blue) and distribution (blue) of the acceleration $\ddot{\Lambda}$. Panel f time series of the mean-square displacement of Λ (blue) along with a linear least-square fit (dashed brown). Panel g autocorrelation function of $\dot{\Lambda}$ (blue) along with an exponential fit (dashed brown).

66

Simulation #3, $m_\Lambda = 800\,\mathrm{u\,nm^2}$

FIGURE 1.F.3: *Results from the CBTI simulation of* $10\,\mathrm{ns}$ *employing* $K = 16$ *replicas with a mass-parameter* $m_\Lambda = 800\,\mathrm{u\,nm^2}$ *and no thermostat coupling of the Λ-variable ($\tau_\Lambda \to \infty$).* This simulation corresponds to entry 3 in Tab. 1.1. Panel a time series (blue) and distribution (orange) of Λ. Panel b time series (blue) and distribution (orange) of $\tilde{\Lambda}$. Panel c time series of λ for replica $k = 0$ (blue) along with the λ-values for the $K - 1$ other replicas (colored dots) as well as distribution of λ considering all replicas (orange). Panel d time series (blue) and distribution (blue) of the velocity $\dot{\Lambda}$ along with the analytical Maxwell-Boltzmann distribution (orange). Panel e time series (blue) and distribution (blue) of the acceleration $\ddot{\Lambda}$. Panel f time series of the mean-square displacement of Λ (blue) along with a linear least-square fit (dashed brown). Panel g autocorrelation function of $\dot{\Lambda}$ (blue) along with an exponential fit (dashed brown).

Simulation #4, $m_\Lambda = 1600\,\mathrm{u\,nm}^2$
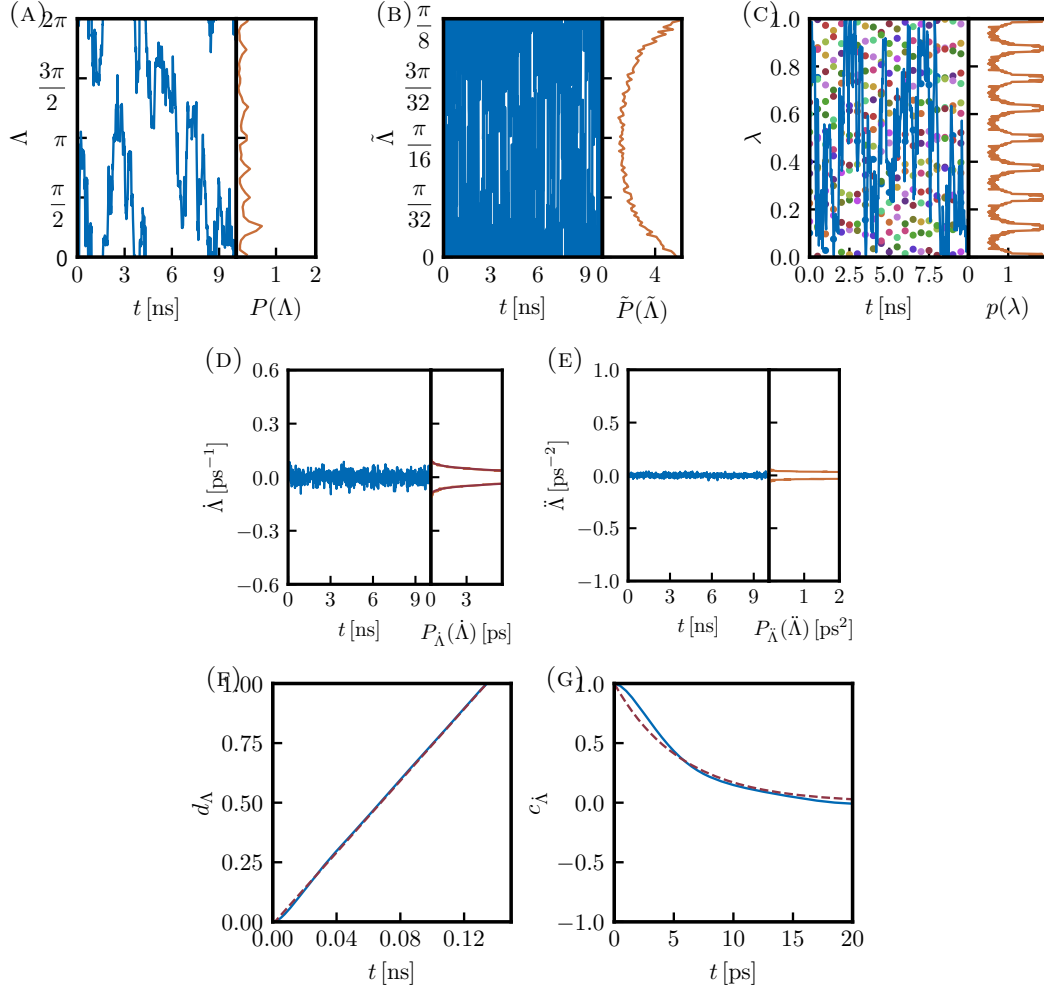

FIGURE 1.F.4: *Results from the CBTI simulation of $10\,\mathrm{ns}$ employing $K = 16$ replicas with a mass-parameter $m_\Lambda = 1600\,\mathrm{u\,nm}^2$ and no thermostat coupling of the Λ-variable ($\tau_\Lambda \to \infty$).* This simulation corresponds to entry 4 in Tab. 1.1. Panel a time series (blue) and distribution (orange) of Λ. Panel b time series (blue) and distribution (orange) of $\tilde{\Lambda}$. Panel c time series of λ for replica $k = 0$ (blue) along with the λ-values for the $K - 1$ other replicas (colored dots) as well as distribution of λ considering all replicas (orange). Panel d time series (blue) and distribution (blue) of the velocity $\dot{\Lambda}$ along with the analytical Maxwell-Boltzmann distribution (orange). Panel e time series (blue) and distribution (blue) of the acceleration $\ddot{\Lambda}$. Panel f time series of the mean-square displacement of Λ (blue) along with a linear least-square fit (dashed brown). Panel g autocorrelation function of $\dot{\Lambda}$ (blue) along with an exponential fit (dashed brown).
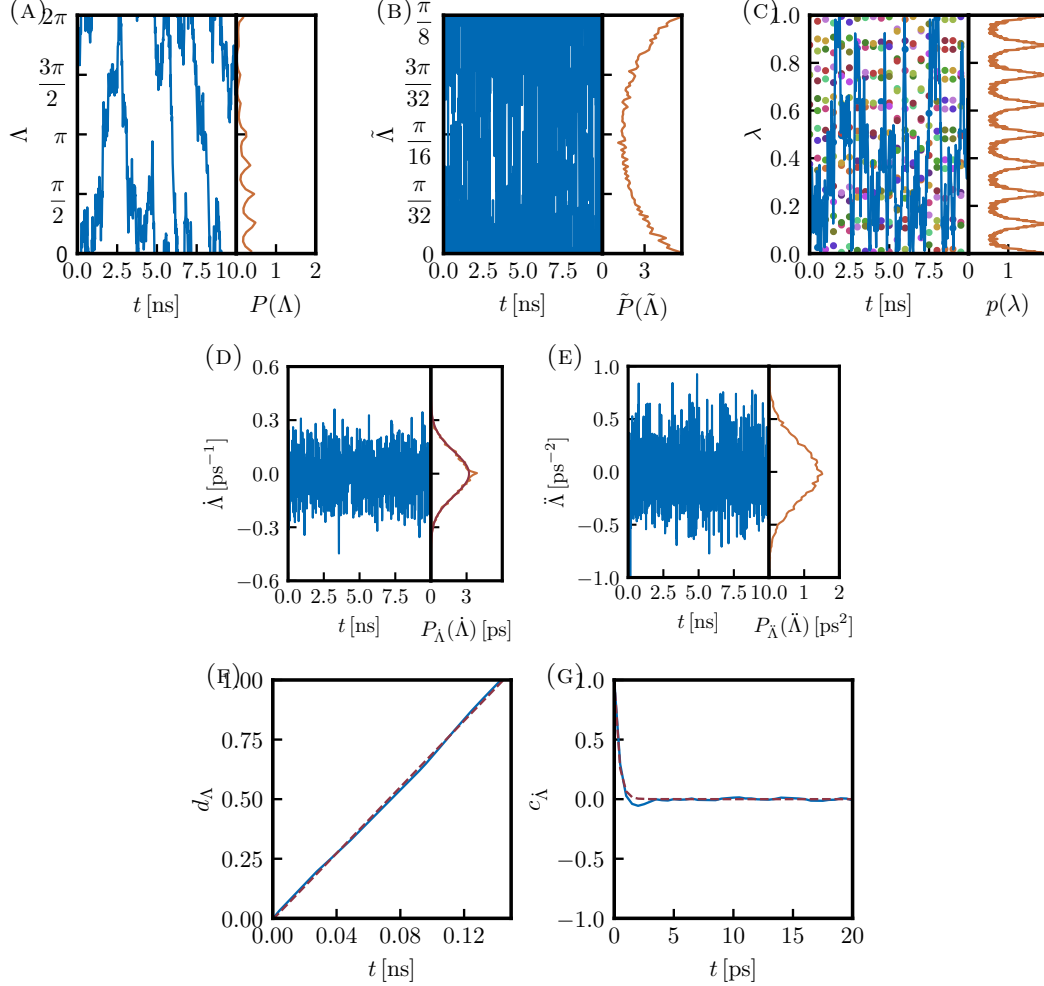
Simulation #5, $m_\Lambda = 3200\,\mathrm{u\,nm}^2$

FIGURE 1.F.5: *Results from the CBTI simulation of* $10\,\mathrm{ns}$ *employing* $K = 16$ *replicas with a mass-parameter* $m_\Lambda = 3200\,\mathrm{u\,nm}^2$ *and no thermostat coupling of the* Λ*-variable* ($\tau_\Lambda \to \infty$). This simulation corresponds to entry 5 in Tab. 1.1. Panel a time series (blue) and distribution (orange) of Λ. Panel b time series (blue) and distribution (orange) of $\tilde{\Lambda}$. Panel c time series of λ for replica $k = 0$ (blue) along with the λ-values for the $K - 1$ other replicas (colored dots) as well as distribution of λ considering all replicas (orange). Panel d time series (blue) and distribution (blue) of the velocity $\dot{\Lambda}$ along with the analytical Maxwell-Boltzmann distribution (orange). Panel e time series (blue) and distribution (blue) of the acceleration $\ddot{\Lambda}$. Panel f time series of the mean-square displacement of Λ (blue) along with a linear least-square fit (dashed brown). Panel g autocorrelation function of $\dot{\Lambda}$ (blue) along with an exponential fit (dashed brown).
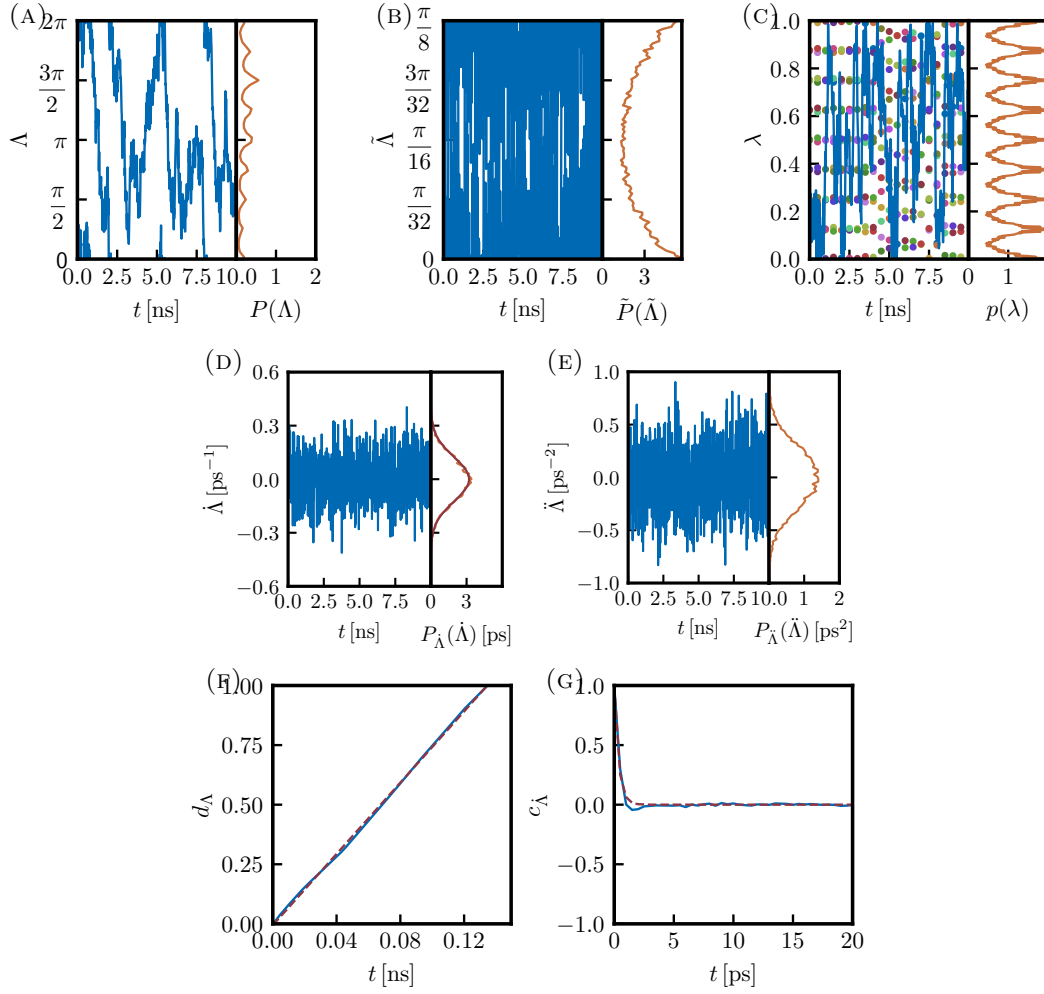
1.F.2 Exploration of the Influence of τ_Λ in CBTI

Simulation #6, $\tau_\Lambda = 0.05\,\mathrm{ps}$

FIGURE 1.F.6: *Results from the CBTI simulation of* $10\,\mathrm{ns}$ *employing* $K = 16$ *replicas with a mass-parameter* $m_\Lambda = 160\,\mathrm{u\,nm}^2$ *and thermostat coupling of the Λ-variable with* $\tau_\Lambda = 0.05\,\mathrm{ps}$. This simulation corresponds to entry 6 in Tab. 1.1. Panel a time series (blue) and distribution (orange) of Λ. Panel b time series (blue) and distribution (orange) of $\tilde\Lambda$. Panel c time series of λ for replica $k = 0$ (blue) along with the λ-values for the $K - 1$ other replicas (colored dots) as well as distribution of λ considering all replicas (orange). Panel d time series (blue) and distribution (blue) of the velocity $\dot\Lambda$ along with the analytical Maxwell-Boltzmann distribution (orange). Panel e time series (blue) and distribution (blue) of the acceleration $\ddot\Lambda$. Panel f time series of the mean-square displacement of Λ (blue) along with a linear least-square fit (dashed brown). Panel g autocorrelation function of $\dot\Lambda$ (blue) along with an exponential fit (dashed brown).
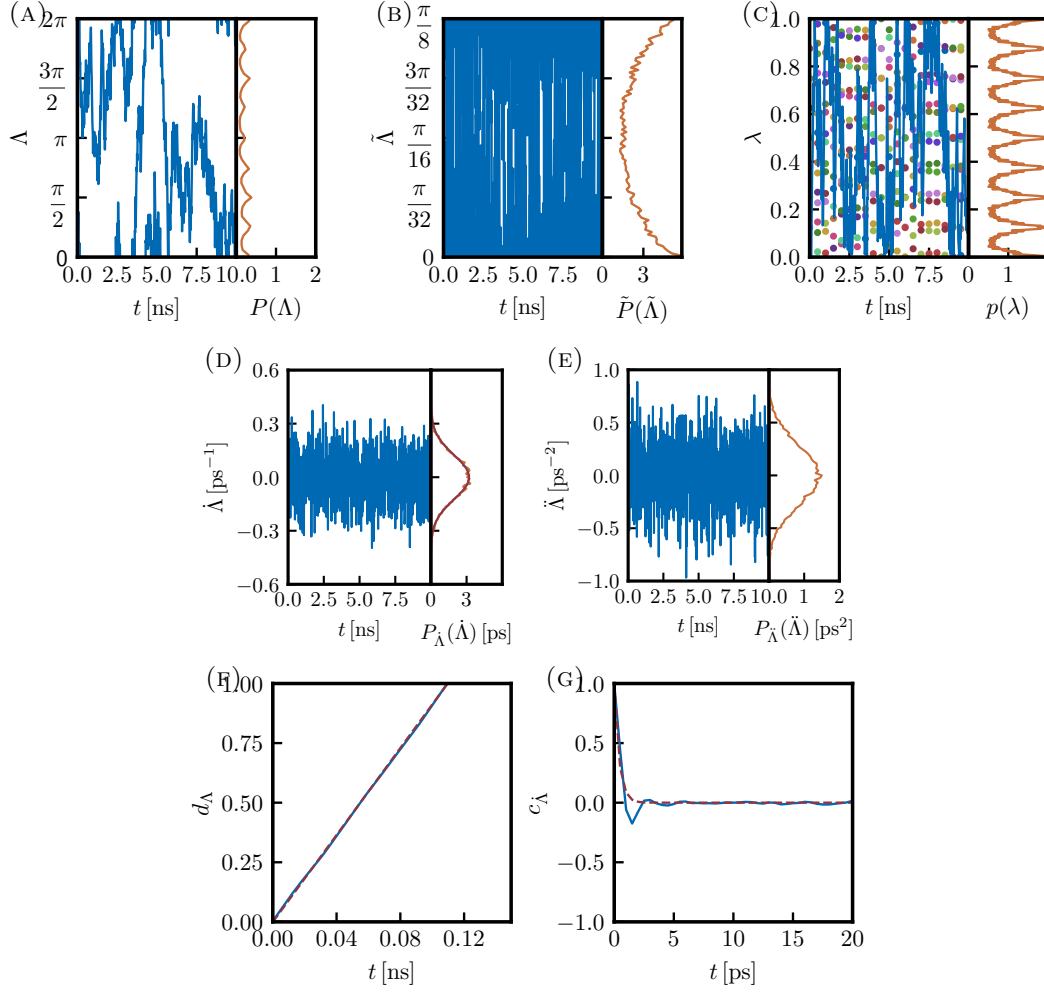
Simulation #7, $\tau_\Lambda = 0.1\,\mathrm{ps}$

FIGURE 1.F.7: *Results from the CBTI simulation of* $10\,\mathrm{ns}$ *employing* $K = 16$ *replicas with a mass-parameter* $m_\Lambda = 160\,\mathrm{u\,nm^2}$ *and thermostat coupling of the Λ-variable with* $\tau_\Lambda = 0.1\,\mathrm{ps}$. This simulation corresponds to entry 7 in Tab. 1.1. Panel a time series (blue) and distribution (orange) of Λ. Panel b time series (blue) and distribution (orange) of $\tilde{\Lambda}$. Panel c time series of λ for replica $k = 0$ (blue) along with the λ-values for the $K - 1$ other replicas (colored dots) as well as distribution of λ considering all replicas (orange). Panel d time series (blue) and distribution (blue) of the velocity $\dot{\Lambda}$ along with the analytical Maxwell-Boltzmann distribution (orange). Panel e time series (blue) and distribution (blue) of the acceleration $\ddot{\Lambda}$. Panel f time series of the mean-square displacement of Λ (blue) along with a linear least-square fit (dashed brown). Panel g autocorrelation function of $\dot{\Lambda}$ (blue) along with an exponential fit (dashed brown).
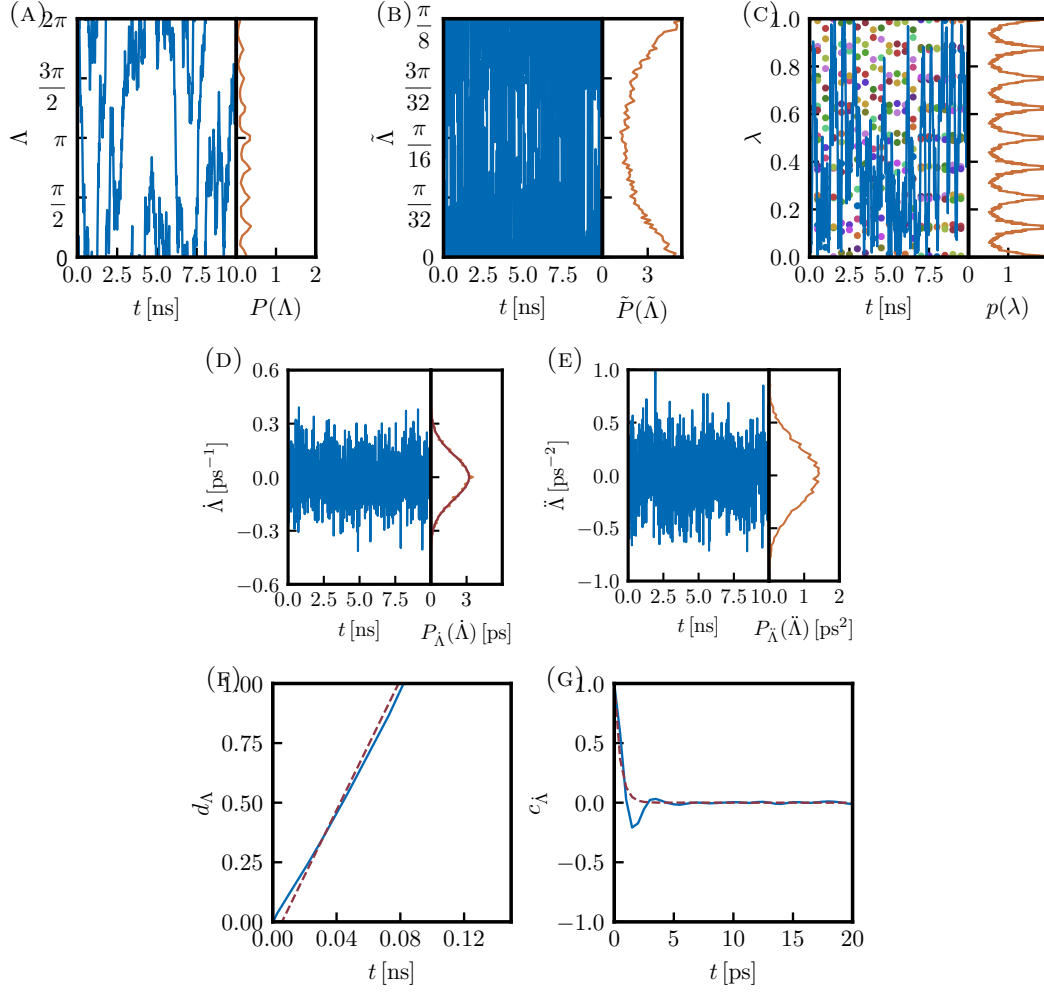
Simulation #8, $\tau_\Lambda = 0.5\,\mathrm{ps}$

FIGURE 1.F.8: *Results from the CBTI simulation of* $10\,\mathrm{ns}$ *employing* $K = 16$ *replicas with a mass-parameter* $m_\Lambda = 160\,\mathrm{u\,nm^2}$ *and thermostat coupling of the Λ-variable with* $\tau_\Lambda = 0.5\,\mathrm{ps}$. This simulation corresponds to entry 8 in Tab. 1.1. Panel a time series (blue) and distribution (orange) of Λ. Panel b time series (blue) and distribution (orange) of $\tilde{\Lambda}$. Panel c time series of λ for replica $k = 0$ (blue) along with the λ-values for the $K - 1$ other replicas (colored dots) as well as distribution of λ considering all replicas (orange). Panel d time series (blue) and distribution (blue) of the velocity $\dot{\Lambda}$ along with the analytical Maxwell-Boltzmann distribution (orange). Panel e time series (blue) and distribution (blue) of the acceleration $\ddot{\Lambda}$. Panel f time series of the mean-square displacement of Λ (blue) along with a linear least-square fit (dashed brown). Panel g autocorrelation function of $\dot{\Lambda}$ (blue) along with an exponential fit (dashed brown).
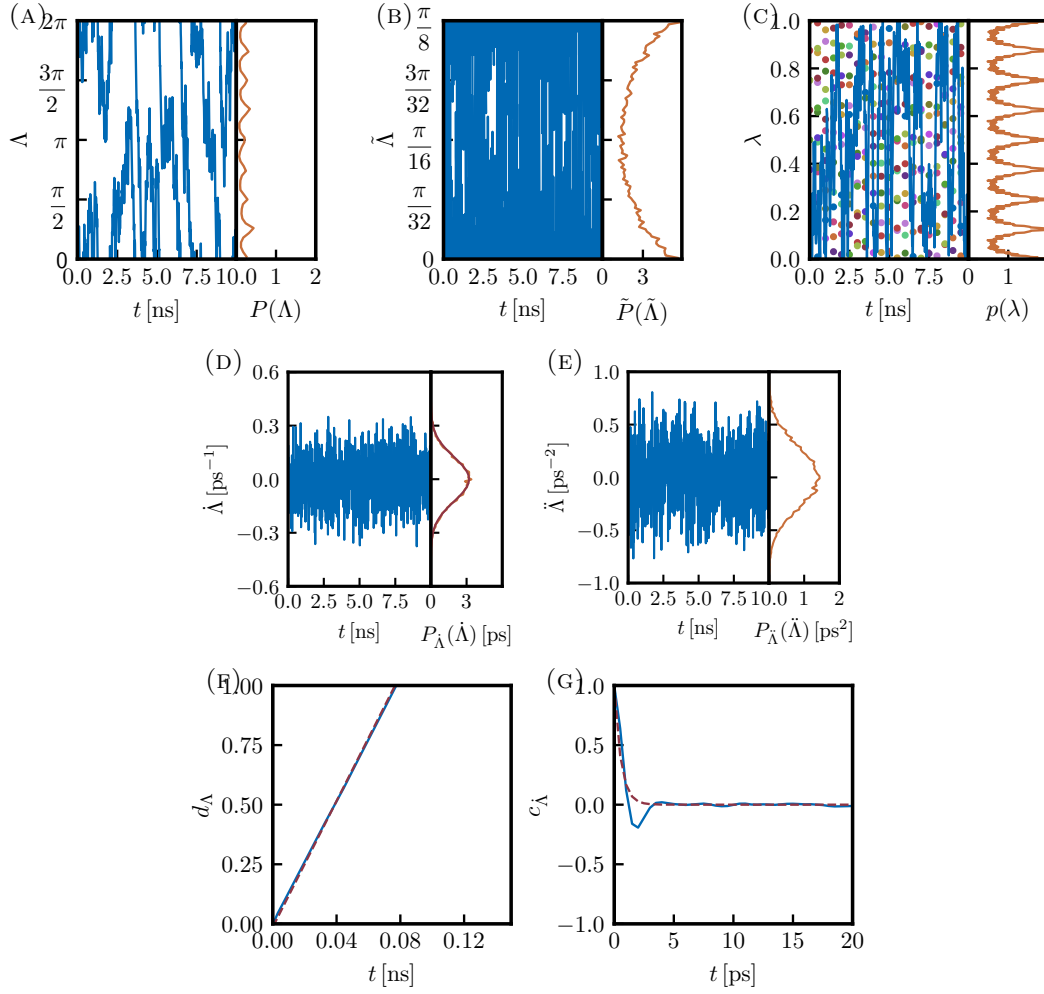
Simulation #9, $\tau_\Lambda = 1\,\mathrm{ps}$

FIGURE 1.F.9: *Results from the CBTI simulation of* $10\,\mathrm{ns}$ *employing* $K = 16$ *replicas with a mass-parameter* $m_\Lambda = 160\,\mathrm{u\,nm^2}$ *and thermostat coupling of the Λ-variable with* $\tau_\Lambda = 1\,\mathrm{ps}$. This simulation corresponds to entry 9 in Tab. 1.1. Panel a time series (blue) and distribution (orange) of Λ. Panel b time series (blue) and distribution (orange) of $\tilde{\Lambda}$. Panel c time series of λ for replica $k = 0$ (blue) along with the λ-values for the $K - 1$ other replicas (colored dots) as well as distribution of λ considering all replicas (orange). Panel d time series (blue) and distribution (blue) of the velocity $\dot{\Lambda}$ along with the analytical Maxwell-Boltzmann distribution (orange). Panel e time series (blue) and distribution (blue) of the acceleration $\ddot{\Lambda}$. Panel f time series of the mean-square displacement of Λ (blue) along with a linear least-square fit (dashed brown). Panel g autocorrelation function of $\dot{\Lambda}$ (blue) along with an exponential fit (dashed brown).
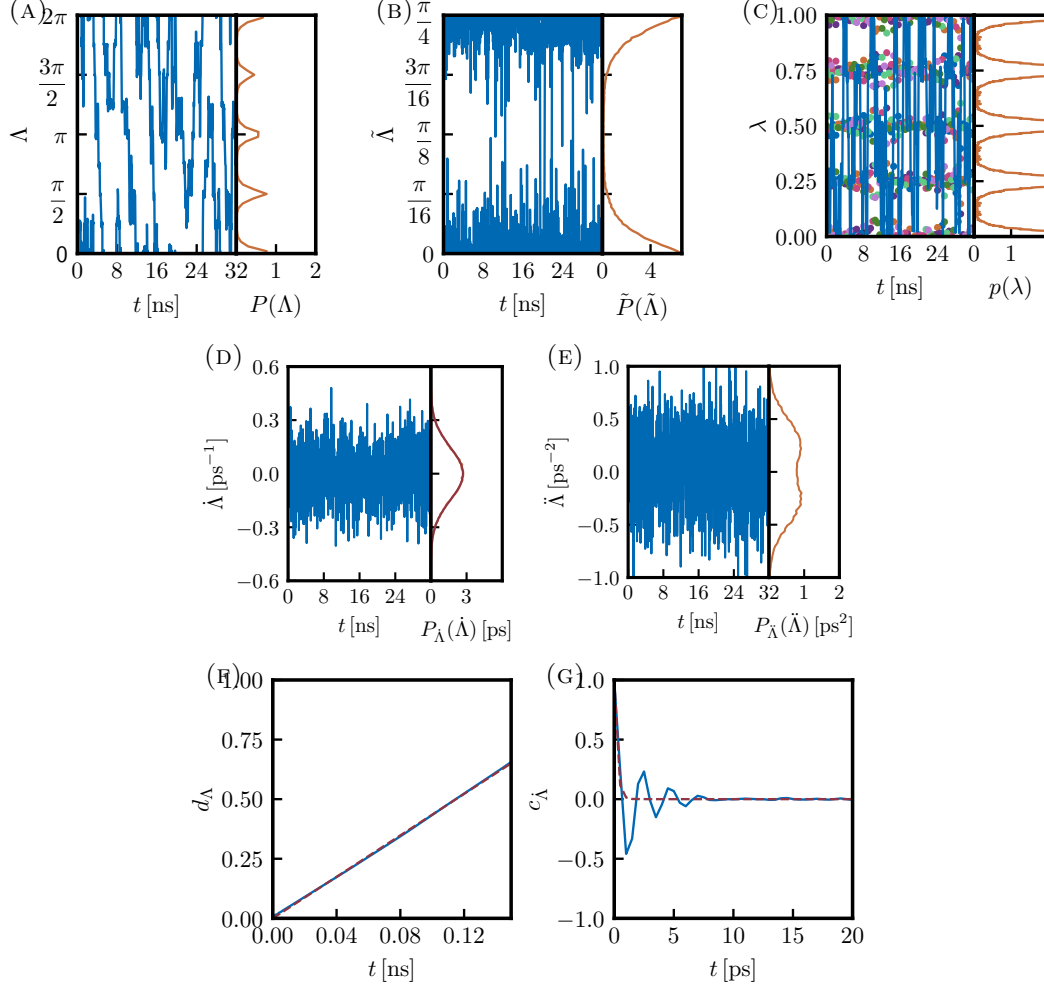
Simulation #10, $\tau_\Lambda = 2\,\mathrm{ps}$

FIGURE 1.F.10: *Results from the CBTI simulation of* $10\,\mathrm{ns}$ *employing* $K = 16$ *replicas with a mass-parameter* $m_\Lambda = 160\,\mathrm{u\,nm^2}$ *and thermostat coupling of the* Λ-*variable with* $\tau_\Lambda = 2\,\mathrm{ps}$. This simulation corresponds to entry 10 in Tab. 1.1. Panel a time series (blue) and distribution (orange) of Λ. Panel b time series (blue) and distribution (orange) of $\tilde{\Lambda}$. Panel c time series of λ for replica $k = 0$ (blue) along with the λ-values for the $K - 1$ other replicas (colored dots) as well as distribution of λ considering all replicas (orange). Panel d time series (blue) and distribution (blue) of the velocity $\dot{\Lambda}$ along with the analytical Maxwell-Boltzmann distribution (orange). Panel e time series (blue) and distribution (blue) of the acceleration $\ddot{\Lambda}$. Panel f time series of the mean-square displacement of Λ (blue) along with a linear least-square fit (dashed brown). Panel g autocorrelation function of $\dot{\Lambda}$ (blue) along with an exponential fit (dashed brown).
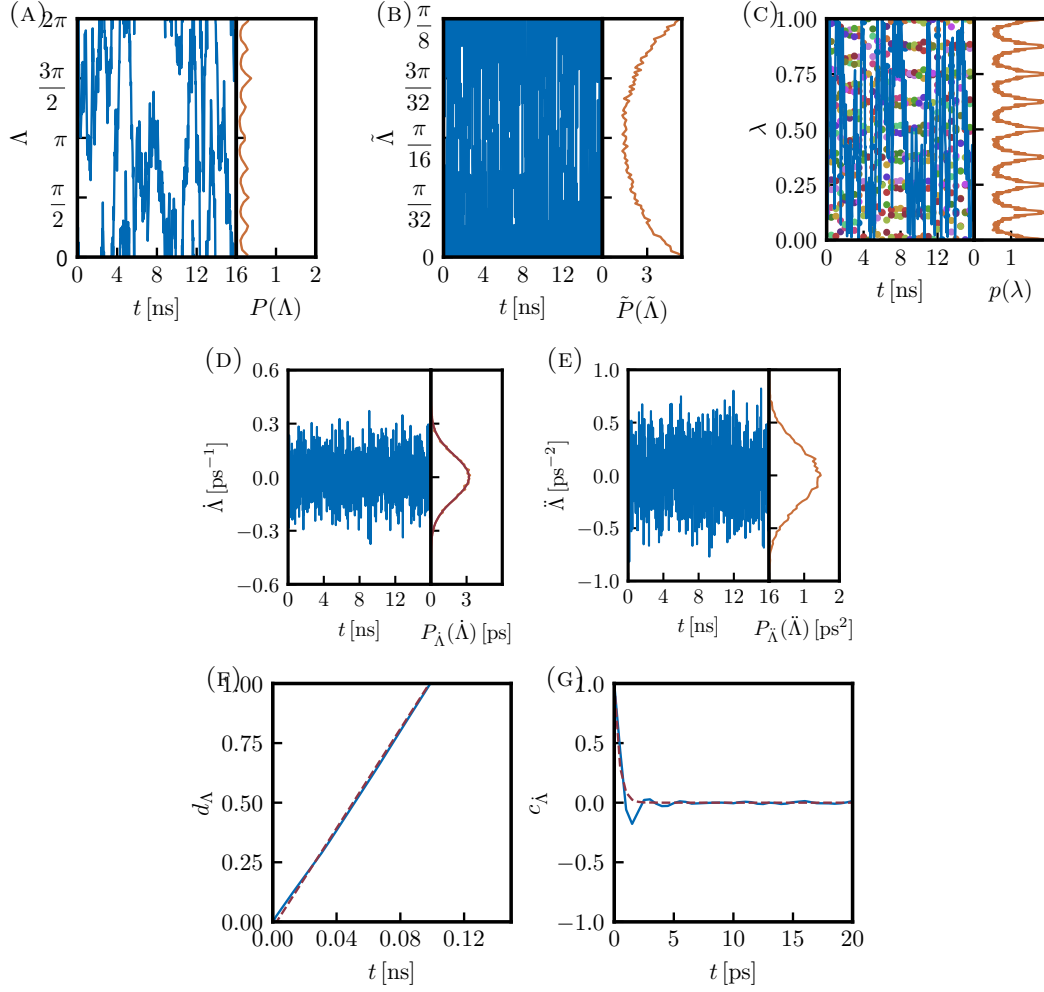
1.F.3 Exploration of the Influence of K in CBTI

Simulation #11, $K = 8$

FIGURE 1.F.11: *Results from the CBTI simulation of* $32\,\mathrm{ns}$ *employing* $K = 8$ *replicas with a mass-parameter* $m_\Lambda = 113\,\mathrm{u\,nm^2}$ *and thermostat coupling of the* Λ*-variable with* $\tau_\Lambda = 0.5\,\mathrm{ps}$. This simulation corresponds to entry 11 in Tab. 1.1. Panel a time series (blue) and distribution (orange) of Λ. Panel b time series (blue) and distribution (orange) of $\tilde{\Lambda}$. Panel c time series of λ for replica $k = 0$ (blue) along with the λ-values for the $K - 1$ other replicas (colored dots) as well as distribution of λ considering all replicas (orange). Panel d time series (blue) and distribution (blue) of the velocity $\dot{\Lambda}$ along with the analytical Maxwell-Boltzmann distribution (orange). Panel e time series (blue) and distribution (blue) of the acceleration $\ddot{\Lambda}$. Panel f time series of the mean-square displacement of Λ (blue) along with a linear least-square fit (dashed brown). Panel g autocorrelation function of $\dot{\Lambda}$ (blue) along with an exponential fit (dashed brown).
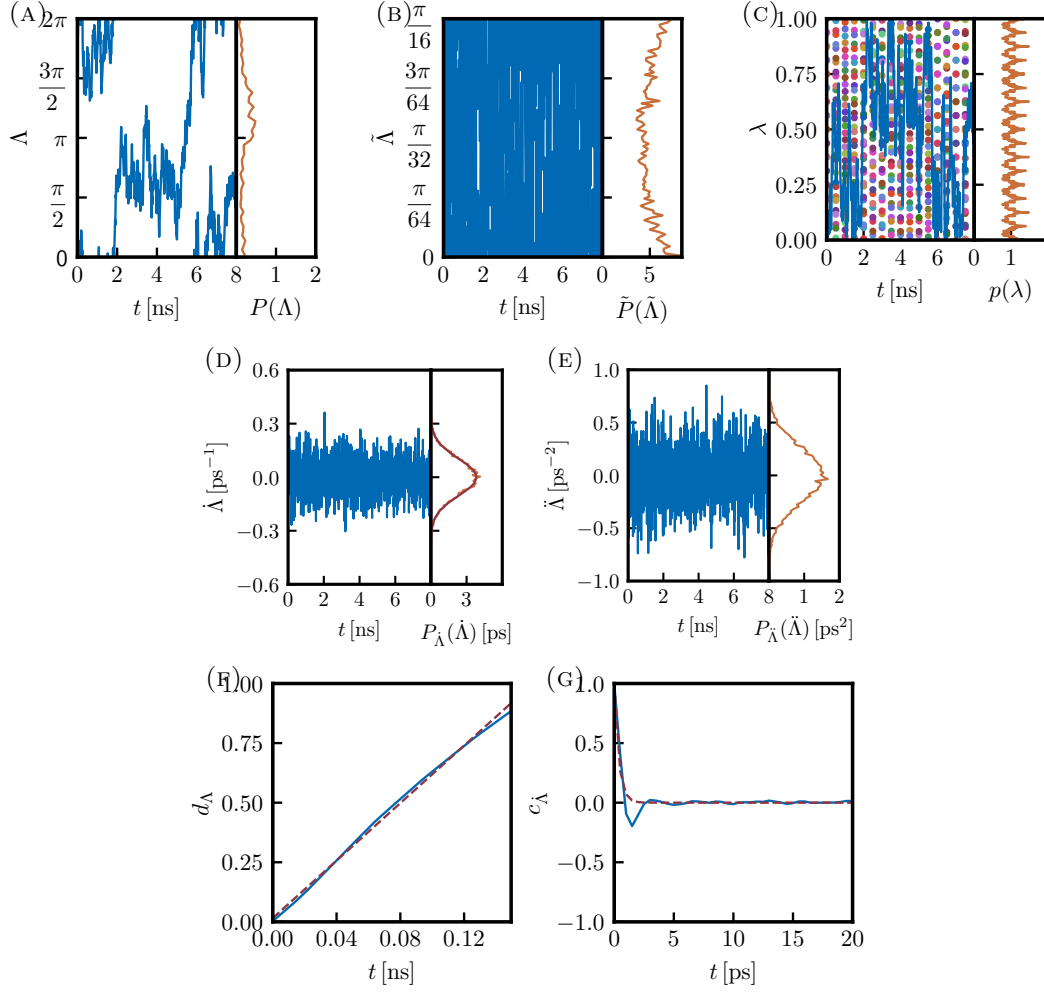
Simulation #12, $K = 16$

FIGURE 1.F.12: *Results from the CBTI simulation of* 16 ns *employing $K = 16$ replicas with a mass-parameter $m_\Lambda = 160$ u nm^2 and thermostat coupling of the Λ-variable with $\tau_\Lambda = 0.5$ ps.* This simulation corresponds to entry 12 in Tab. 1.1. Panel a time series (blue) and distribution (orange) of Λ. Panel b time series (blue) and distribution (orange) of $\tilde{\Lambda}$. Panel c time series of λ for replica $k = 0$ (blue) along with the λ-values for the $K - 1$ other replicas (colored dots) as well as distribution of λ considering all replicas (orange). Panel d time series (blue) and distribution (blue) of the velocity $\dot{\Lambda}$ along with the analytical Maxwell-Boltzmann distribution (orange). Panel e time series (blue) and distribution (blue) of the acceleration $\ddot{\Lambda}$. Panel f time series of the mean-square displacement of Λ (blue) along with a linear least-square fit (dashed brown). Panel g autocorrelation function of $\dot{\Lambda}$ (blue) along with an exponential fit (dashed brown).
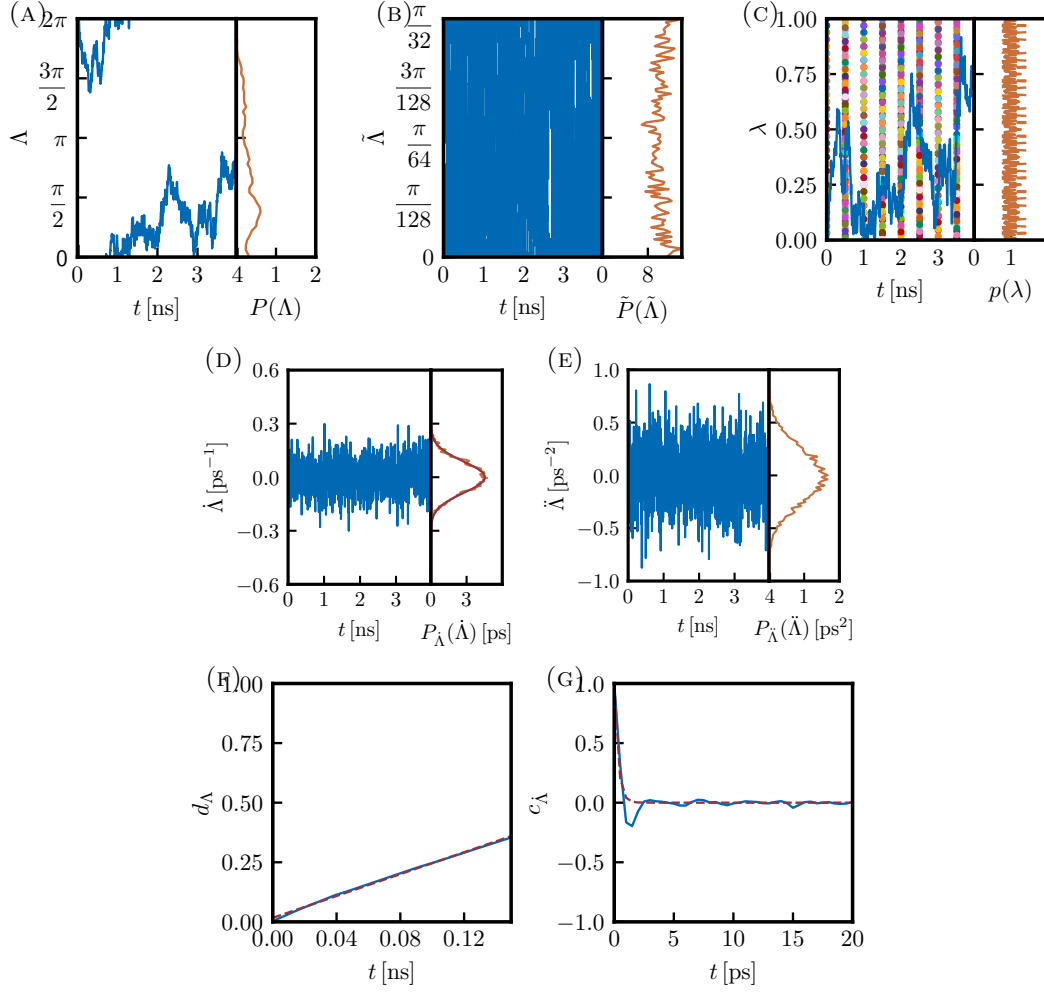
Simulation #13, $K = 32$

FIGURE 1.F.13: *Results from the CBTI simulation of* $8\,\mathrm{ns}$ *employing* $K = 32$ *replicas with a mass-parameter* $m_\Lambda = 226\,\mathrm{u\,nm^2}$ *and thermostat coupling of the* Λ-*variable with* $\tau_\Lambda = 0.5\,\mathrm{ps}$. This simulation corresponds to entry 13 in Tab. 1.1. Panel a time series (blue) and distribution (orange) of Λ. Panel b time series (blue) and distribution (orange) of $\tilde{\Lambda}$. Panel c time series of λ for replica $k = 0$ (blue) along with the λ-values for the $K - 1$ other replicas (colored dots) as well as distribution of λ considering all replicas (orange). Panel d time series (blue) and distribution (blue) of the velocity $\dot{\Lambda}$ along with the analytical Maxwell-Boltzmann distribution (orange). Panel e time series (blue) and distribution (blue) of the acceleration $\ddot{\Lambda}$. Panel f time series of the mean-square displacement of Λ (blue) along with a linear least-square fit (dashed brown). Panel g autocorrelation function of $\dot{\Lambda}$ (blue) along with an exponential fit (dashed brown).

Simulation #14, $K = 64$

FIGURE 1.F.14: *Results from the CBTI simulation of* 4 ns *employing* $K = 64$ *replicas with a mass-parameter* $m_\Lambda = 320$ u nm² *and thermostat coupling of the* Λ-*variable with* $\tau_\Lambda = 0.5$ ps. This simulation corresponds to entry 14 in Tab. 1.1. Panel a time series (blue) and distribution (orange) of Λ. Panel b time series (blue) and distribution (orange) of $\tilde{\Lambda}$. Panel c time series of λ for replica $k = 0$ (blue) along with the λ-values for the $K - 1$ other replicas (colored dots) as well as distribution of λ considering all replicas (orange). Panel d time series (blue) and distribution (blue) of the velocity $\dot{\Lambda}$ along with the analytical Maxwell-Boltzmann distribution (orange). Panel e time series (blue) and distribution (blue) of the acceleration $\ddot{\Lambda}$. Panel f time series of the mean-square displacement of Λ (blue) along with a linear least-square fit (dashed brown). Panel g autocorrelation function of $\dot{\Lambda}$ (blue) along with an exponential fit (dashed brown).
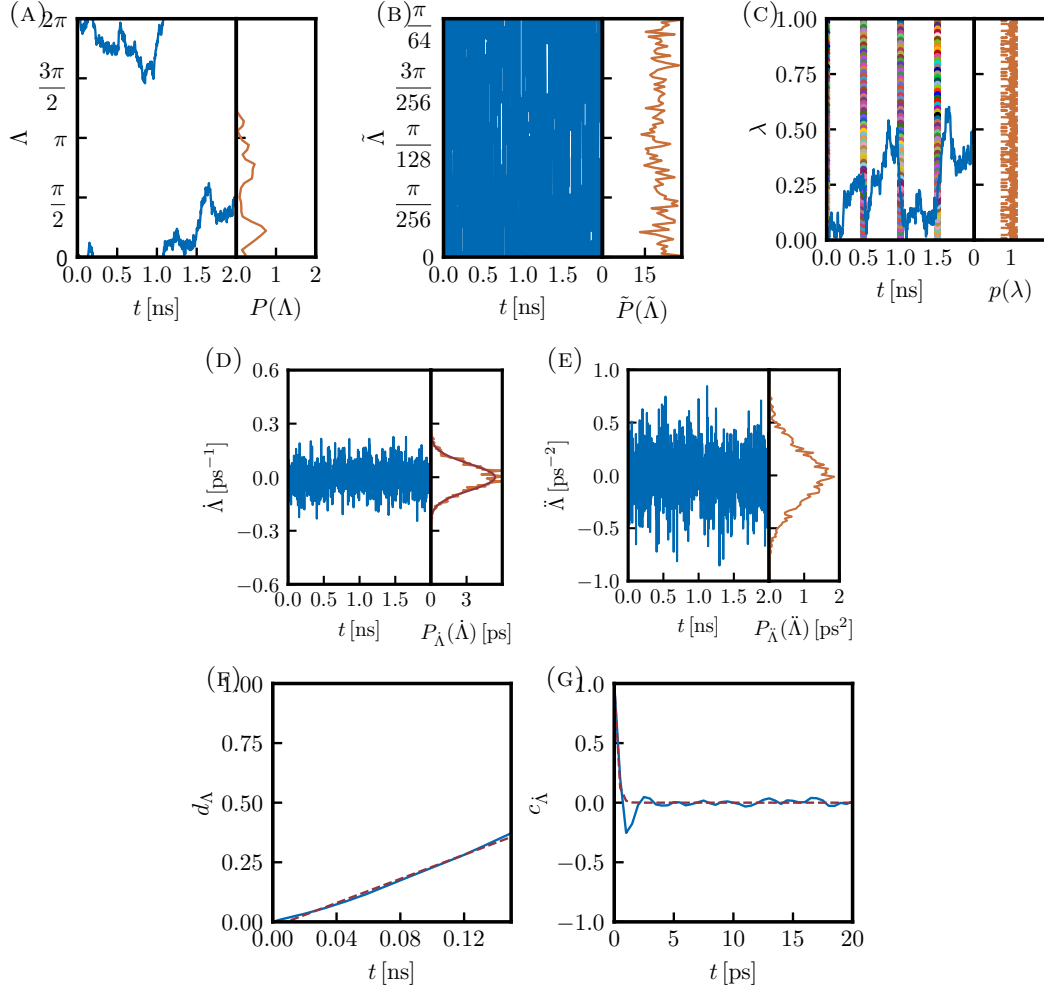
Simulation #15, $K = 128$

FIGURE 1.F.15: *Results from the CBTI simulation of $2\,\mathrm{ns}$ employing $K = 128$ replicas with a mass-parameter $m_\Lambda = 452\,\mathrm{u\,nm^2}$ and thermostat coupling of the Λ-variable with $\tau_\Lambda = 0.5\,\mathrm{ps}$. This simulation corresponds to entry 15 in Tab. 1.1. Panel a time series (blue) and distribution (orange) of Λ. Panel b time series (blue) and distribution (orange) of $\tilde{\Lambda}$. Panel c time series of λ for replica $k = 0$ (blue) along with the λ-values for the $K - 1$ other replicas (colored dots) as well as distribution of λ considering all replicas (orange). Panel d time series (blue) and distribution (blue) of the velocity $\dot{\Lambda}$ along with the analytical Maxwell-Boltzmann distribution (orange). Panel e time series (blue) and distribution (blue) of the acceleration $\ddot{\Lambda}$. Panel f time series of the mean-square displacement of Λ (blue) along with a linear least-square fit (dashed brown). Panel g autocorrelation function of $\dot{\Lambda}$ (blue) along with an exponential fit (dashed brown).*

1.G Appendix REPEATS OF A CBTI SIMULATION

TABLE 1.G.1: *Repeat results of a CBTI calculation for the aqueous methanol-to-dummy mutation at* 298.15 K *and* 1 bar *using the 2061H66 force field.* The free-energy change ΔG is reported for the CBTI calculation involving $K = 16$ replicas along with $m_\Lambda = 160 \, \mathrm{u \, nm^2}$ and $\tau_\Lambda = 0.5 \, \mathrm{ps}$. The results of ten repeats (successive rows) are listed, all of 6.25 ns duration (total single-system sampling time of 100 ns). For each repeat, the calculated ΔG (Eq. 1.22 with $J = 500$) is reported along with a bootstrap error estimate. This error does not include a Student t-factor. Repeat statistics based on this data can be found in Tab. 1.2.

	$\Delta G \, [\mathrm{kJ \, mol^{-1}}]$
1	21.42 (0.16)
2	21.33 (0.16)
3	21.44 (0.17)
4	21.38 (0.15)
5	21.08 (0.14)
6	21.43 (0.15)
7	21.59 (0.15)
8	21.54 (0.15)
9	21.31 (0.15)
10	21.42 (0.16)

1.H Appendix FREE-ENERGY PROFILES $G_{\tilde{\Lambda}}(\tilde{\Lambda})$ ALONG $\tilde{\Lambda}$
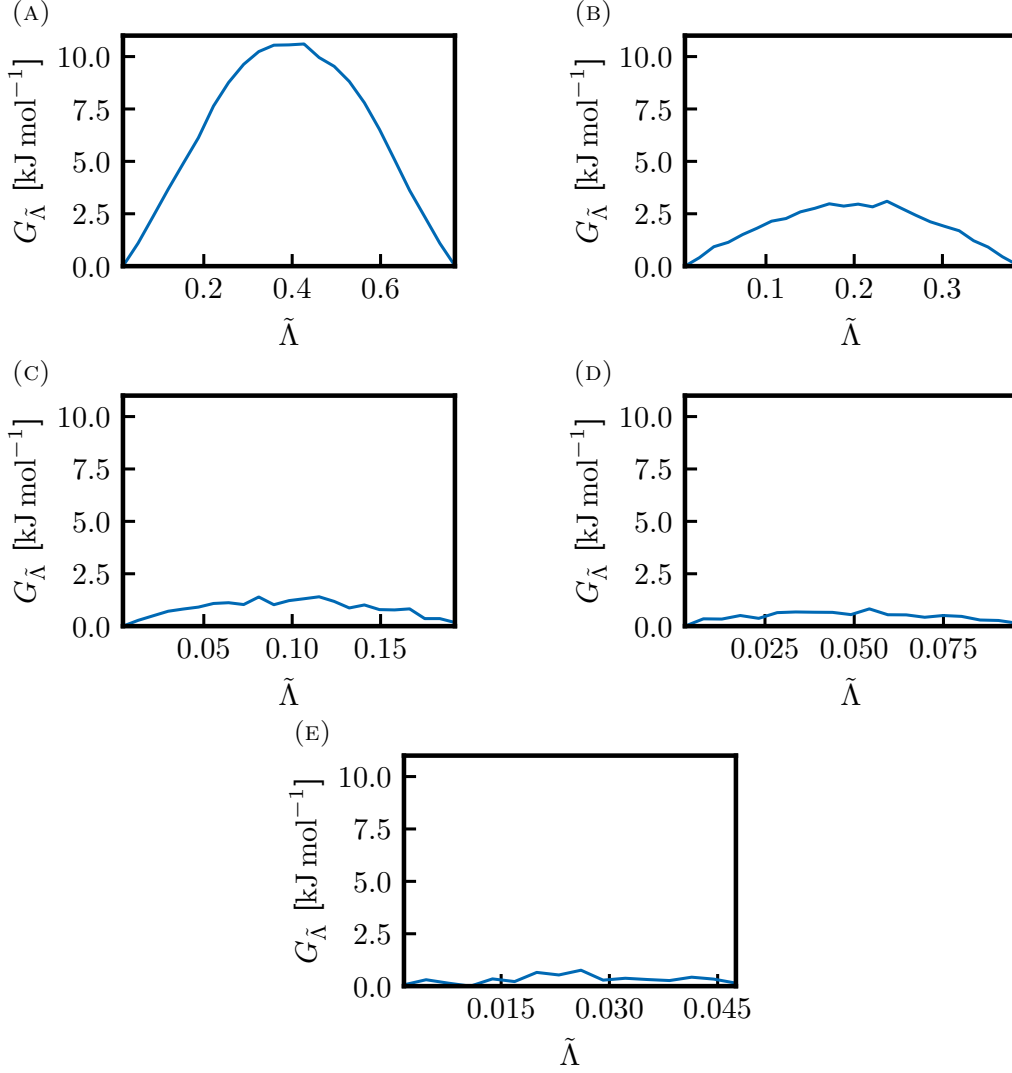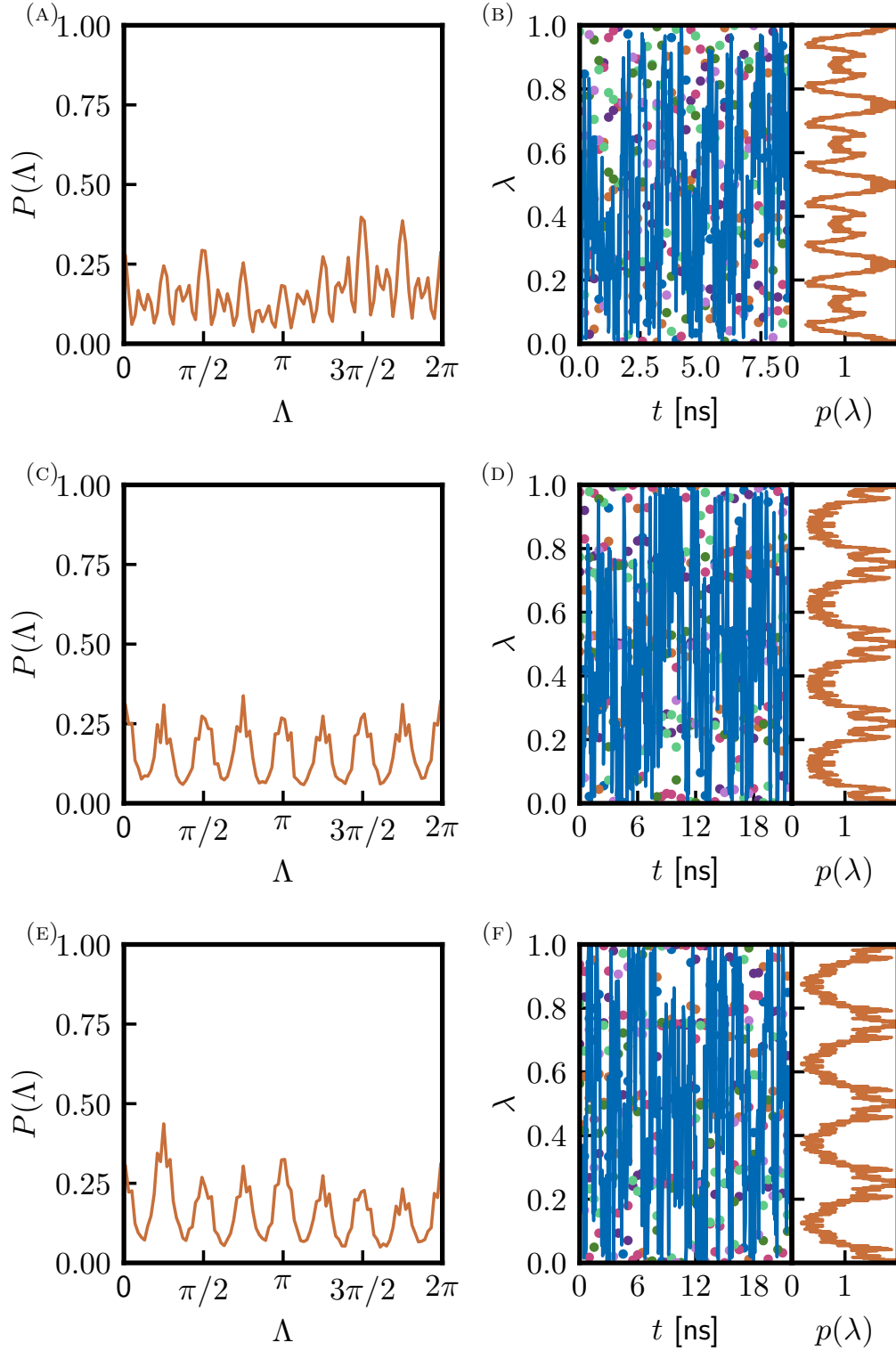

FIGURE 1.H.1: *Free-energy profiles $G_{\tilde{\Lambda}}(\tilde{\Lambda})$ in unbiased CBTI simulations of the aqueous methanol-to-dummy mutation at* 298.15 K *and* 1 bar *with the 2016H66 force field.* The simulations relied on $K = 8$ a, $K = 16$ b, $K = 32$ c, $K = 64$ d, and $K = 128$ e replicas, a mass-parameter $m_\Lambda = 40K^{1/2}$ u nm^2 and thermostat coupling of the Λ-variable with $\tau_\Lambda = 0.5$ ps. The free-energy profiles were calculated as $G_{\tilde{\Lambda}}(\Lambda) = -\beta^{-1} \ln P_{\tilde{\Lambda}}(\tilde{\Lambda})$, where $P_{\tilde{\Lambda}}(\tilde{\Lambda})$ is the normalized probability distribution of $\tilde{\Lambda}$, and anchored to zero at their minimum. The value at the maximum corresponds to $G_{\tilde{\Lambda}}^{\star}$, shown graphically in Fig. 1.6.

1.I Appendix CBTI SIMULATIONS WITH BIASING POTENTIAL

TABLE 1.I.1: *Parameters and results of biased CBTI simulations of the aqueous methanol-to-dummy mutation at* $298.15\,\mathrm{K}$ *and* $1\,\mathrm{bar}$ *with the 2016H66 force field.* The successive entries are the index of the simulation (sim), the number K of replicas, the number N_{gp} of gridpoints for $\tilde{\Lambda}$ in the range $[0; 2\pi/K]$, the build-up force constant c_{LE}, the reduction factor f_{red}, the LE build-up time t_{LE} for the replica system, the number N_{ds} of double-sweeps over the $\tilde{\Lambda}$ range during the build-up, the US umbrella sampling time t_{US} for the replica system, and the free-energy difference ΔG calculated using Eq. 1.22 with $J = 500$. The corresponding results are illustrated graphically in Figs. 1.I.2 and 1.I.4. Only entries 2 and 8 are discussed in Sect. 1.4.

| sim	K	N_{gp}	$c_{\mathrm{LE}}\,[\mathrm{kJ\,mol}^{-1}]$	f_{red}	$t_{\mathrm{LE}}\,[\mathrm{ns}]$	N_{ds}	$t_{\mathrm{US}}\,[\mathrm{ns}]$	$\Delta G\,[\mathrm{kJ\,mol}^{-1}]$
1	8	10	0.01	0.1	0.050	3	22	21.11 ± 0.15
2	8	34	0.001	0.1	0.150	3	22	21.48 ± 0.18
3	8	34	0.001	0.1	0.200	6	22	21.29 ± 0.17
4	8	34	0.001	0.8	0.054	3	22	21.54 ± 0.22
5	8	34	0.001	0.8	0.128	6	22	21.38 ± 0.17
6	8	34	0.001	0.8	0.156	9	22	21.46 ± 0.16
7	16	6	0.01	0.1	0.050	3	22	21.23 ± 0.22
8	16	18	0.001	0.1	0.070	3	22	21.30 ± 0.13
9	16	18	0.001	0.1	0.083	6	22	21.48 ± 0.16
10	16	18	0.001	0.8	0.030	3	10	21.00 ± 0.14
11	16	18	0.001	0.8	0.050	6	20	21.57 ± 0.14
12	16	18	0.001	0.8	0.084	9	20	21.21 ± 0.17

Continued next page.
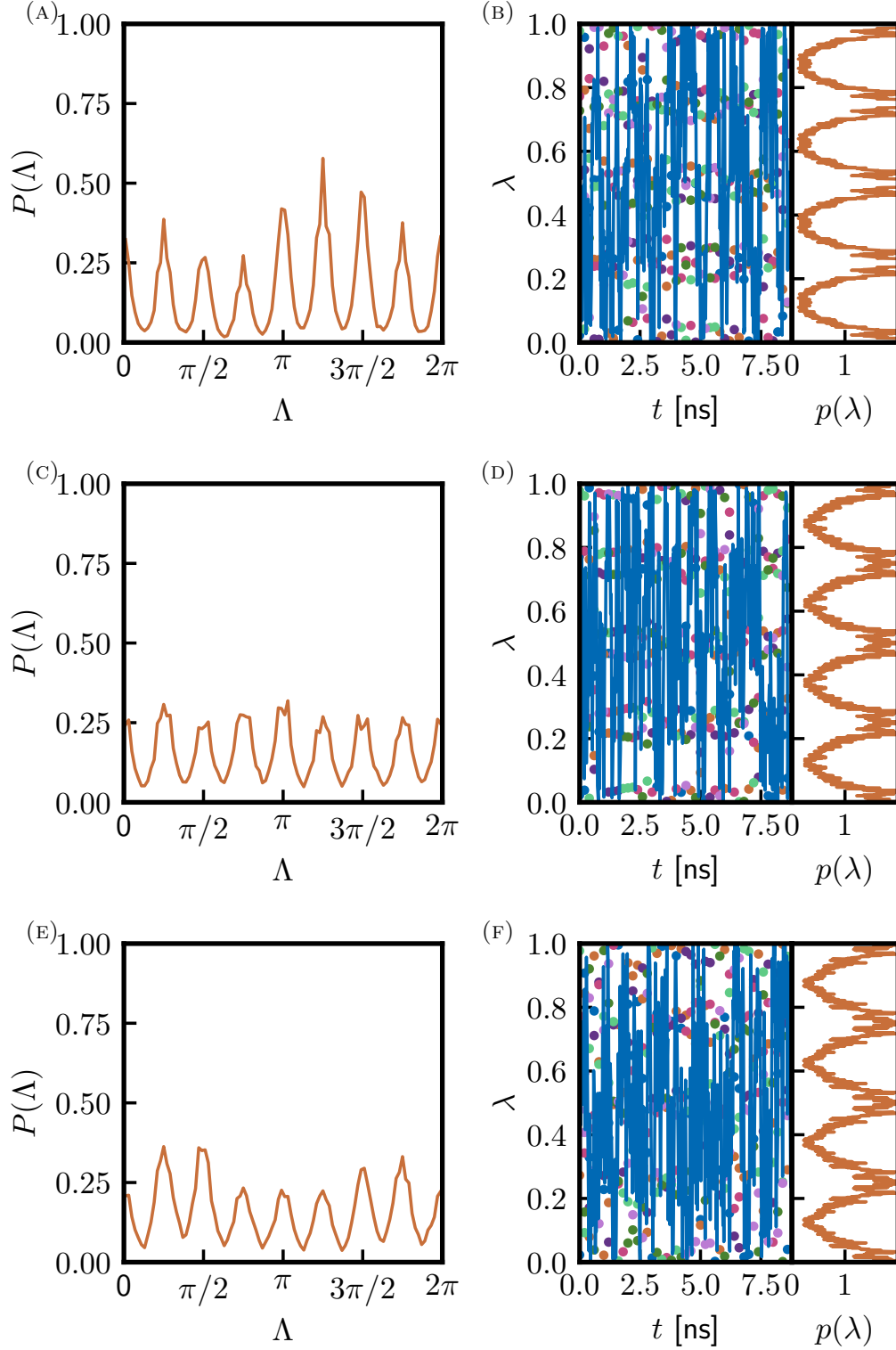
FIGURE 1.I.2: *Time series and probability distributions of the relevant CB-variables in biased CBTI calculations of the aqueous methanol-to-dummy mutation at 298.15 K and 1 bar with the 2016H66 force field.* The calculations rely on $K = 8$ replicas. Each row shows the probability distribution $P(\Lambda)$ of the CB advance variable Λ (left) and the time series $\lambda(t)$ and probability distribution $p(\lambda)$ of the coupling variable λ for all replicas (right). The successive panels correspond to different protocol settings listed in Tab. 1.I.1. Panels (a,b) entry 1. Panels (c,d) entry 2. Panels (e,f) entry 3. Panels (g,h) entry 4. Panels (i,j) entry 5. and Panels (k,l) entry 6.
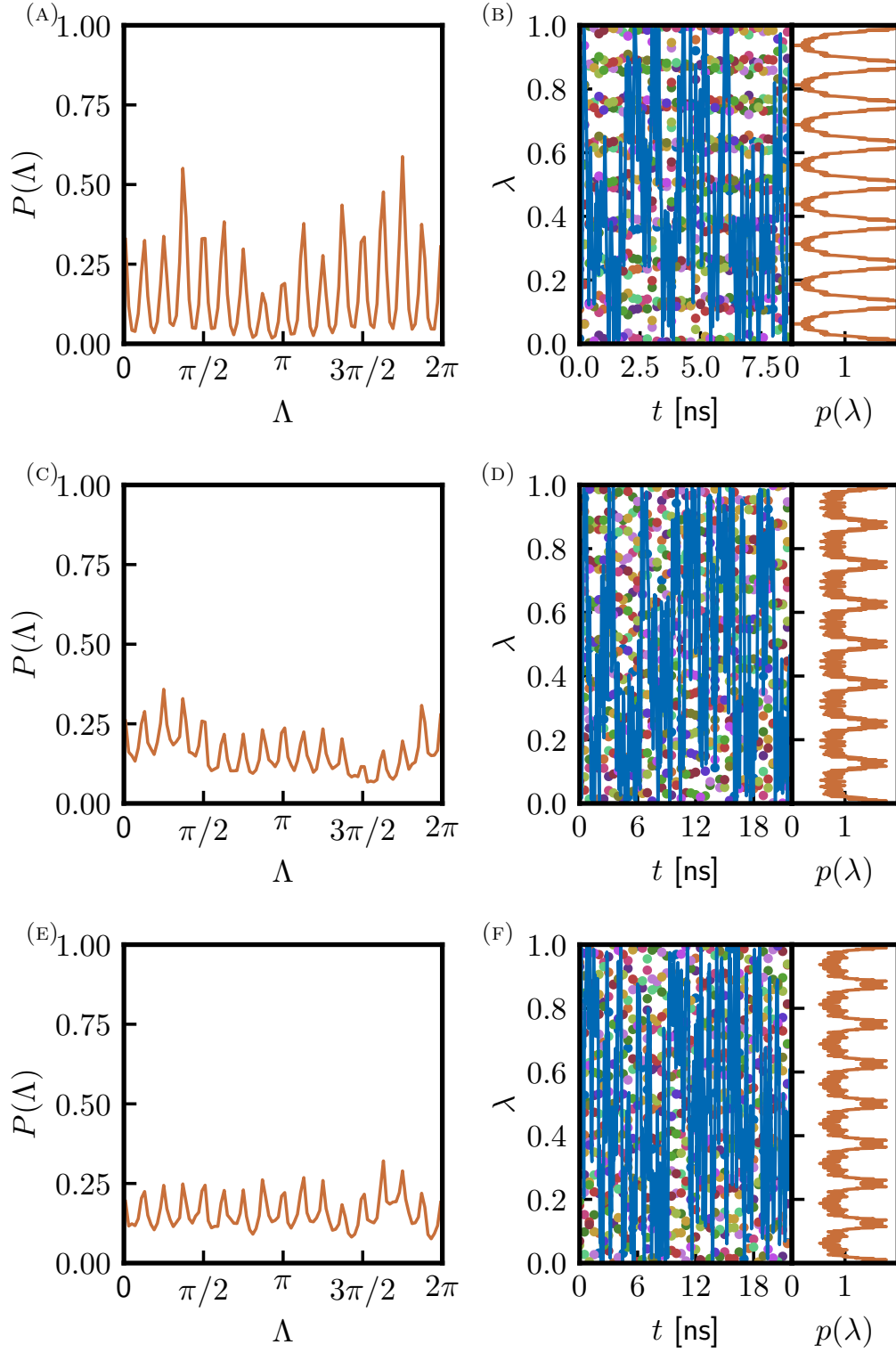
Continued next page.
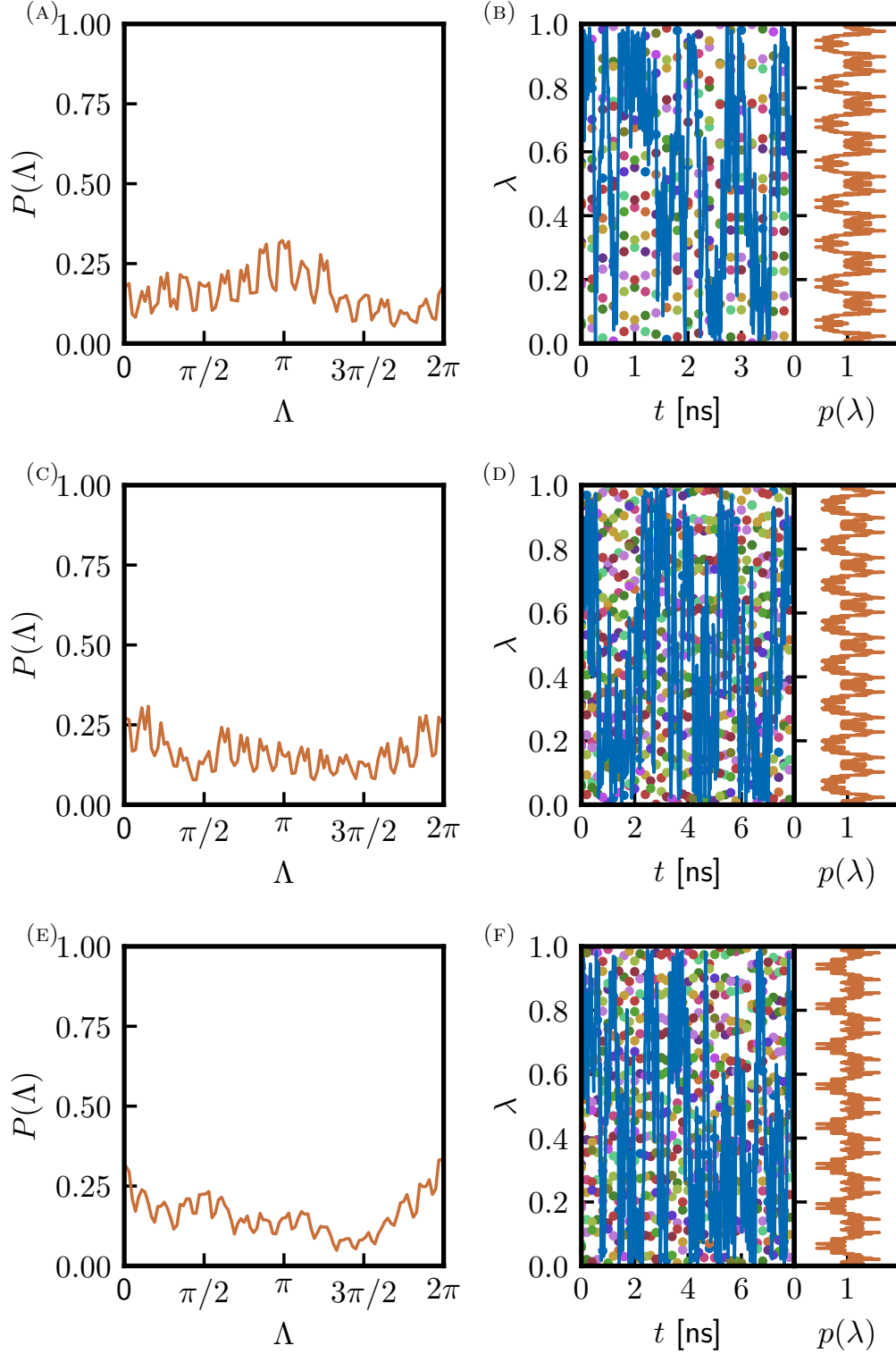
FIGURE 1.I.4: *Time series and probability distributions of the relevant CB-variables in biased CBTI calculations of the aqueous methanol-to-dummy mutation at 298.15 K and 1 bar with the 2016H66 force field.* The calculations rely on $K = 16$ replicas. Each row shows the probability distribution $P(\Lambda)$ of the CB advance variable Λ (left) and the time series $\lambda(t)$ and probability distribution $p(\lambda)$ of the coupling variable λ for all replicas (right). The successive panels correspond to different protocol settings listed in Tab. 1.I.1. Panels (a,b) entry 7. Panels (c,d) entry 8. Panels (e,f) entry 9. Panels (g,h) entry 10. Panels (i,j) entry 11. and Panels (k,l) entry 12.
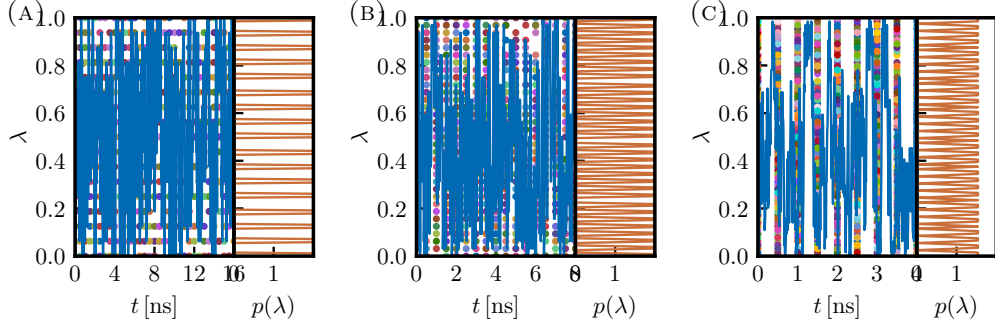
1.J Appendix HRE SIMULATIONS

FIGURE 1.J.1: *Time series $\lambda(t)$ and probability distributions $p(\lambda)$ of the coupling variable λ in HRE simulations of the aqueous methanol-to-dummy mutation at 298.15 K and 1 bar with the 2016H66 force field.* The time series is shown as a blue curve for replica $k = 0$, and as individual colored points at 0.5 ns interval for the $K_{\mathrm{HRE}} - 1$ other replicas. Panel a: $K_{\mathrm{HRE}} = 17$. Panel b: $K_{\mathrm{HRE}} = 33$. Panel c: $K_{\mathrm{HRE}} = 65$.
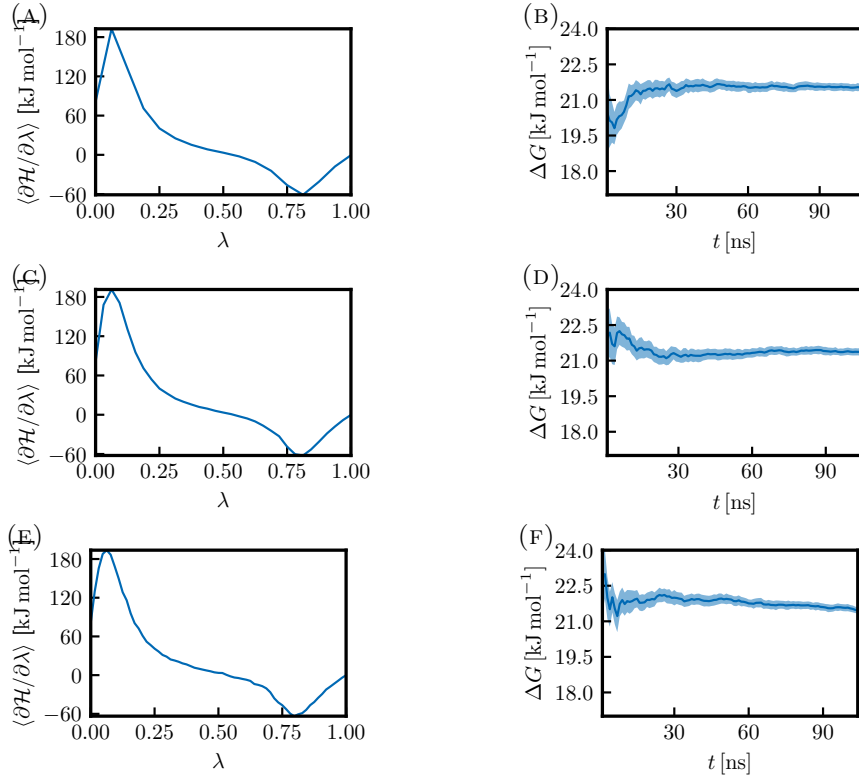
FIGURE 1.J.2: *Relevant results from HRE calulations of the aqueous methanol-to-dummy mutation at 298.15 K and 1 bar with the 2016H66 force field.* Panels a, c and e show the Hamiltonian derivative curve considering a total single-system sampling time of 100 ns. Panel a: $K_{\mathrm{HRE}} = 17$. Panel c: $K_{\mathrm{HRE}} = 33$. Panel e: $K_{\mathrm{HRE}} = 65$. Panels b, d and f show the Hamiltonian derivative curve considering a total single-system sampling time of 100 ns. Panel b: $K_{\mathrm{HRE}} = 17$. Panel d: $K_{\mathrm{HRE}} = 33$. Panel f: $K_{\mathrm{HRE}} = 65$.
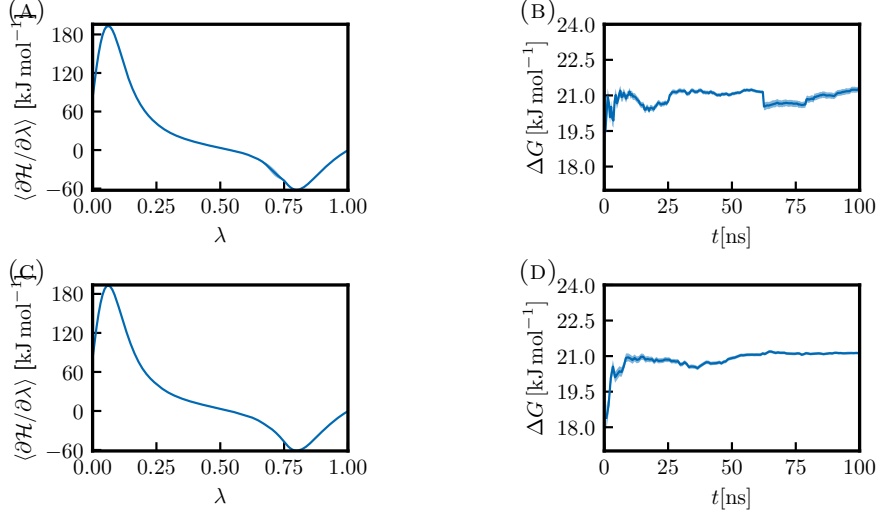
1.K Appendix TI/EXTI CALCULATIONS

FIGURE 1.K.1: *Relevant results from TI/EXTI calulations of the aqueous methanol-to-dummy mutation at 298.15 K and 1 bar with the 2016H66 force field.*Panels a and c show the Hamiltonian derivative curve, which was predicted at 129 λ-points. Panel a: $K_{\text{TI}} = 9$. Panel c: $K_{\text{TI}} = 17$.Panels b and d show the convergence of the ΔG value dependent on the total single-system sampling time. Panel b: $K_{\text{TI}} = 9$. Panel d: $K_{\text{TI}} = 17$.
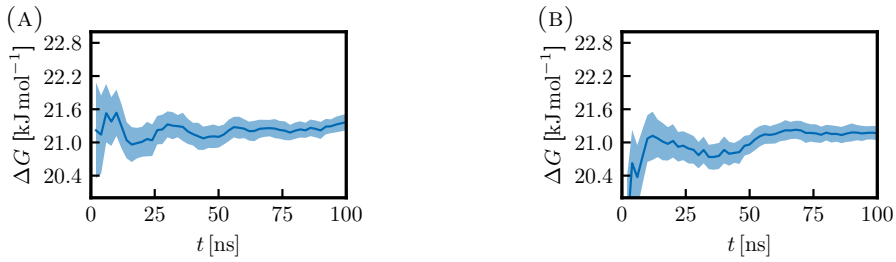
1.L Appendix TI/MBAR CALCULATIONS

FIGURE 1.L.1: *Convergence properties of TI/MBAR calulations for the aqueous methanol-to-dummy mutation at 298.15 K and 1 bar with the 2016H66 force field.* Panels b and d show the convergence of the ΔG value dependent on the total single-system sampling time. Panel a: $K_{\text{TI}} = 9$. Panel b: $K_{\text{TI}} = 17$.