Received: 19 December 2023 | Revised: 31 January 2024 | Accepted: 2 February 2024

DOI: 10.1002/jcc.27332

RESEARCH ARTICLE

Journal of COMPUTATIONAL CHEMISTRY WILEY

Thermal isomerization rates in retinal analogues using Ab-Initio molecular dynamics

Simon Ghysbrecht | Bettina G. Keller

Department of Biology, Chemistry and Pharmacy, Freie Universität Berlin, Berlin, Germany

Correspondence
Bettina G. Keller, Department of Biology, Chemistry and Pharmacy, Freie Universität Berlin, Arnimallee 22, 14195 Berlin, Germany.
Email: bettina.keller@fu-berlin.de

Funding information
Deutsche Forschungsgemeinschaft (DFG), Grant/Award Number: SFB 1114

[Correction added after first online publication, March 9, 2024: Equation (3a) has been revised.]

Abstract

For a detailed understanding of chemical processes in nature and industry, we need accurate models of chemical reactions in complex environments. While Eyring transition state theory is commonly used for modeling chemical reactions, it is most accurate for small molecules in the gas phase. A wide range of alternative rate theories exist that can better capture reactions involving complex molecules and environmental effects. However, they require that the chemical reaction is sampled by molecular dynamics simulations. This is a formidable challenge since the accessible simulation timescales are many orders of magnitude smaller than typical timescales of chemical reactions. To overcome these limitations, rare event methods involving enhanced molecular dynamics sampling are employed. In this work, thermal isomerization of retinal is studied using tight-binding density functional theory. Results from transition state theory are compared to those obtained from enhanced sampling. Rates obtained from dynamical reweighting using infrequent metadynamics simulations were in close agreement with those from transition state theory. Meanwhile, rates obtained from application of Kramers' rate equation to a sampled free energy profile along a torsional dihedral reaction coordinate were found to be up to three orders of magnitude higher. This discrepancy raises concerns about applying rate methods to one-dimensional reaction coordinates in chemical reactions.

KEYWORDS
DFT, DFTB, Kramers, metadynamics, rate theory, retinal, square-root approximation, umbrella sampling

1 | INTRODUCTION

Precise models of chemical reactions, encompassing reaction mechanisms and precise rate constants, are critical for a nuanced understanding of reactions occurring in nature, laboratory experiments, and industrial processes. Yet, computational models of chemical reactions remain challenging, because the transition state of a reaction has to be calculated using a highly accurate model of the Born-Oppenheimer potential energy surface (PES). This usually involves evaluating the electronic structure and explicitly calculating the electronic energy at each nuclear configuration. Thus, the current standard to calculate reaction rate constants remains Eyring transition state theory (TST),[1] which requires only calculations at a few select points along the PES. However, Eyring TST makes strong assumptions for the shape of the PES and the dynamics on this PES and is therefore limited to small to medium sized systems in the gas phase.

Eyring TST defines the transition state as a saddle point on the PES, assumes an equilibrium between reactant state *A* and transition

state *TS* and models the PES at *A* and *TS* by a harmonic approximation.[2] The dynamics of the nuclei are treated quantum mechanically. The most prominent error source is the accuracy of the energy barrier, which enters the equation for the rate exponentially. But the assumptions are easily violated if (*i*) the saddle point of the PES does not coincide with the bottleneck of the reaction (i.e., the free energy maximum along the optimal reaction coordinate), (*ii*) the PES at *A* or at *TS* is anharmonic, (*iii*) the reaction coordinate has a strong curvature in the configurational space, or (*iv*) the reactant state exhibits multiple minima.

These violations occur in systems with many degrees of freedom, in particular if these degrees of freedom are very mobile. Then the reactant state may comprise multiple molecular conformations and various vibrational modes may couple to the reactive vibrational mode. The situation is further complicated if the reaction occurs in solution or if the reactants otherwise strongly interacts with their environment, for example, in a catalysed reaction. To model reactions for these systems, one shifts to a classical description of the nuclear dynamics and samples the reaction using molecular dynamics (MD) simulations.[3]

The simplest estimator for a reaction rate constant from a MD simulation is to count the number of transition from reactant state *A* to product state *B*. However, since the accessible simulation times are orders of magnitude smaller than the mean first passage times even of very fast reactions, one uses enhanced sampling protocols to increase the statistics. A wide variety of methods to recover accurate rate constants and mechanisms from these sped-up simulations have been proposed.[2,4] They broadly fall into two categories: (*i*) dynamical reweighting methods[5] sample the reaction on a biased PES and reweight the transition count, (*ii*) reaction coordinate based methods define a one-dimensional reaction coordinate *s*[6] and calculate the rate constant from an effective dynamics on this reaction coordinate. Besides the definition of a reaction coordinate, the second approach involves the calculation of the free energy surface (FES)[7] and diffusion profile[8] via MD simulations.

Kramers' rate theory[9] is the most prominent example for this second approach. It assumes separation of timescales and models the FES at *A* and *TS* by a harmonic approximation. These two assumptions may be relaxed by using Pontryagin's rate theory.[10] In both cases, the dynamics are modelled by a stochastic classical equation of motion. Infrequent metadynamics[11] is an example for the first approach. The method assumes separation of timescales, but does not use a harmonic approximation. The dynamics are modelled and simulated in the full configurational space using either a deterministic or a stochastic classical equation of motion.

Owing to recent progress in electronic structure calculations[12–14] and quantum mechanics/molecular mechanics approaches,[15] the development of reactive force fields,[16] and the emergence of neural network potentials,[17] chemical reactions will increasingly be modeled through simulations rather than through Eyring TST. Thus, models of chemical reactions in large molecular systems with complex environments come within reach. However, moving from Eyring TST to sampling-based rate estimates involves a considerable reconstruction of the theoretical foundation through which the reaction is modelled. Most importantly, the quantum mechanical description of the nuclear degrees of freedom is replaced by a classical approximation. Furthermore, the search for a transition state *TS* is replaced by a statistical estimate of the transition count (first approach) or by the search for an optimal reaction coordinate. It is not obvious how these changes influence the accuracy with which the reaction rate constant can be determined.

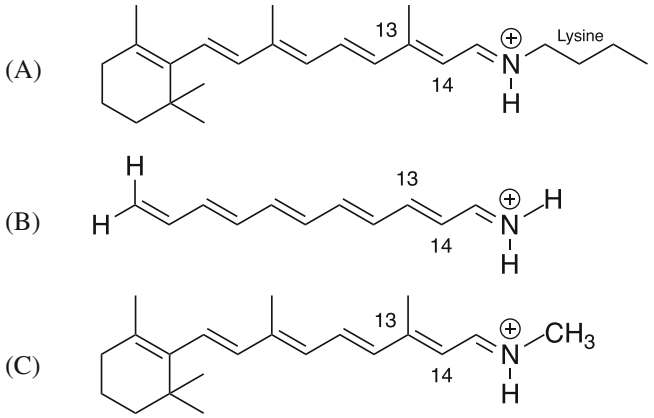

FIGURE 1 Structures of retinal compounds: (A) Retinal cofactor attached to lysine chain, (B) pSb5 and (C) pSb1.

The first aim of this study is to explore and to quantify the influence of these approximations on the estimate of a reaction rate constant. As test reactions we choose the thermal cis-trans isomerization in two analogues of retinal: pSb5 and pSb1 (Figure 1). When modelled in vacuum, the thermal cis-trans isomerization over a C=C double bond fulfills the assumptions of Eyring TST well. Additionally, it is an unimolecular reaction, so that the encounter complex of the reactants does not need to be modelled. Thus for this specific system, we expect that Eyring TST yields an accurate rate constant, which can serve as a reference for more approximate TST models or for sampling-based approaches. On the other hand, the reaction energy barrier is high and the two molecules are large enough for non-trivial coupling between vibrational modes, so that the two test systems pose a reasonable challenge for sampling-based approaches.

Extensive literature has addressed the precise modeling of the potential energy surface (PES) for the cis-trans isomerization in retinal,[18] as well as for retinal analogues.[19–21] We here model the PES by self-consistent-charge density-functional tight-binding method with a third-order expansion of the total energy around a reference density (DFTB3)[12,22] and include density functional theory (DFT)[23,24] as a reference. This allows us to explore the sensitivity of the reaction rate constant to variations in both the rate model and the underlying PES. Thus, as a second aim of the study, we ask whether the precision of the activation energy is indeed the most pivotal parameter when calculating a reaction rate constant.

2 | THEORY

We here summarize the rate theories used in this study. For a more in-depth exploration of rate theories, please refer to References 2, 4, and section I of the supplementary material.

2.1 | Eyring TST

The cis-trans isomerization around a C=C double bond is a unimolecular reaction

$$A \xrightarrow{k_{AB}} B, \tag{1}$$

which, according to the theory of the activated complex, is modeled as

$$A \rightleftharpoons AB^{\ddagger} \rightarrow B \tag{2}$$

where A is the reactant state, B is the product state and $AB^{\ddagger}$ is the activated complex. The critical assumption in Equation (2) is that reactant and transition state configurations are in equilibrium. Eyring TST[1] models this equilibrium by statistical thermodynamics and arrives at the following equation for the reaction rate:[2,4]

$$k_{AB}^{\mathrm{Eyr}} = \frac{RT}{h}\frac{\tilde{q}_{AB^{\ddagger}}}{q_A}\exp\left(-\frac{E_b}{RT}\right) \tag{3a}$$

$$= \frac{RT}{h}\exp\left(-\frac{\Delta F^{\ddagger}}{RT}\right) \tag{3b}$$

where R is the ideal gas constant, T is the temperature, and h is the Planck constant in molar units. See section I of the supplementary material.

The free energy difference $\Delta F^{\ddagger}$ between the $AB^{\ddagger}$ and A can be calculated from the molecular partition function at the transition state $\tilde{q}_{AB^{\ddagger}}$ and the molecular partition function at the reactant state q_A:

$$\Delta F^{\ddagger} = E_b - RT\ln\left(\frac{\tilde{q}_{AB^{\ddagger}}}{q_A}\right) \tag{4}$$

where E_b is the potential energy barrier, that is, the potential energy difference between the reactant minimum and the maximum of the energy barrier. The partition functions are calculated relative to the electronic ground state energy of the respective configurations. The tilde symbol in $\tilde{q}_{AB^{\ddagger}}$ indicates that, for $AB^{\ddagger}$, the vibrational contribution corresponding to the reaction coordinate is excluded in the free energy calculation.

We calculate and report potential and free energies in units of J mol^{-1}, correspondingly the thermal energy is also reported as a molar quantity: RT. If units of energy are used for potential and free energies, R should be replaced by the Boltzmann constant $k_B = R/N_A$ in Equations (3) and (4) and all of the following equations. N_A is the Avogadro constant.

The molecular partition functions are determined by separating their translational, rotational, vibrational and electronic degrees of freedom. Each part is treated using appropriate quantum mechanical models, that is, particle-in-a-box for translational, rigid rotor for rotational and harmonic oscillator for vibrational degrees of freedom. In the case of a unimolecular reaction, the contributions of the translational degrees of freedom to the free energy difference in Equation (4) will cancel. We assume $q_{\mathrm{el}} = 1$ for all situations, that is, the electronic ground state energy level is non-degenerate, and any contributions from higher electronic states can be ignored. With these approximations, the free energy difference (Equation 4) of the cis-trans isomerization can be decomposed as

$$\Delta F^{\ddagger} = E_b + \Delta F_{\mathrm{rot}} + \Delta F_{\mathrm{vib}} \tag{5}$$

where

$$\Delta F_{\mathrm{rot}} = -RT\ln\left[\frac{q_{AB^{\ddagger},\mathrm{rot}}}{q_{A,\mathrm{rot}}}\right] \tag{6a}$$

$$\Delta F_{\mathrm{vib}} = -RT\ln\left[\frac{\tilde{q}_{AB^{\ddagger},\mathrm{vib}}}{q_{A,\mathrm{vib}}}\right] \tag{6b}$$

define the rotational and vibrational free energy difference. $q_{A,\mathrm{rot}}$ and $q_{AB^{\ddagger},\mathrm{rot}}$ are the rotational partition functions of A and $AB^{\ddagger}$. $q_{A,\mathrm{vib}}$ and $\tilde{q}_{AB^{\ddagger},\mathrm{vib}}$ are the vibrational partition functions of A and $AB^{\ddagger}$, where the tilde symbol indicates that the reactive vibrational mode has been excluded from vibrational partition function of $AB^{\ddagger}$. See section I of the supplementary material.

2.2 | High-temperature TST

If the thermal energy RT is large compared to the energy difference of the vibrational states, the following high-temperature approximation to Eyring TST may be used

$$k_{AB}^{\mathrm{Eyr}} \approx k_{AB}^{\mathrm{ht}} = \frac{\prod_{k=1}^{3N-6}\nu_{A,k}}{\prod_{k=1,k\neq r}^{3N-6}\nu_{AB^{\ddagger},k}}\exp\left(-\frac{E_b}{RT}\right) \tag{7}$$

where the frequencies $\nu_{A,k}$ and $\nu_{AB^{\ddagger},k}$ correspond to the harmonic frequencies at the reactant and transition state respectively. Note that the frequency of the reactive vibrational mode $\nu_{AB^{\ddagger},r}$ is excluded from the product.

Equation (7) can be brought into the form of Equation (3b) by setting

$$\begin{aligned}\Delta F^{\ddagger} \approx \Delta F^{\ddagger,\mathrm{ht}} &= E_b - RT\ln\left(\frac{\tilde{q}^{\mathrm{ht}}_{AB^{\ddagger},\mathrm{vib}}}{q^{\mathrm{ht}}_{A,\mathrm{vib}}}\right)\\ &= E_b + \Delta F^{\mathrm{ht}}_{\mathrm{vib}}\end{aligned} \tag{8}$$

where $\tilde{q}^{\mathrm{ht}}_{AB^{\ddagger},\mathrm{vib}}$ and $q^{\mathrm{ht}}_{A,\mathrm{vib}}$ are the high-temperature approximations to the vibrational partition functions of A and $AB^{\ddagger}$. In deriving Equations (7) and (8), one assumes that the moments of inertia of the reactant and TS configuration are approximately the same and thus the rotational contribution to the $\Delta F^{\ddagger}$ is negligible. Additionally, one neglects the vibrational zero-point energy and takes the continuum

limit of the vibrational partition function. See section I of the supplementary material.

The high-temperature TST can also be derived by treating the partition functions in Equations (3a) and (4) classically and using a harmonic approximation for the PES. See section I of the supplementary material. Equations (7) and (8) thus constitute the link between a quantum mechanical and a classical treatment of the activated complex.

2.3 | Infrequent metadynamics

A statistical estimate for the reaction coordinate rate constant is obtained via the mean first-passage time τ_{AB}

$$k_{AB} = \frac{1}{\tau_{AB}}, \tag{9}$$

where τ_{AB} is the average time it takes for the system to reach the product state B from the reactant state A. The relation between k_{AB} and τ_{AB} stated in Equation (9) relies on a separation of timescales between the timescale of equilibration within A and the much slower timescale of equilibration between A and B. From MD simulations on the PES $V(\mathbf{x})$, where $\mathbf{x}$ is the molecular configuration, τ_{AB} can be calculated as the arithmetic mean of the first-passage times from A to B.[2] However, a better statistical accuracy is obtained by fitting a the cumulative distribution function of a Poisson process[25]

$$P(\tau_{AB,i}) = 1 - \exp\left(-\frac{\tau_{AB,i}}{\tau_{AB}}\right) \tag{10}$$

to the cumulative distribution histogram of these fist passage times. In Equation (10), $\tau_{AB,i}$ is the i first-passage time observed in the simulation and τ_{AB} is the MFPT and acts as a fitting parameter, which is inserted into Equation (9) to obtain the reaction rate.

Infrequent metadynamics[11] is a method to calculate transition times for systems in which the mean first-passage times is larger than the accessible simulation time. The molecular system is prepared in the reactant state A and a time dependent bias function $U(\mathbf{x},t)$ is introduced that increases in strength as the simulation proceeds and pushes the system over the barrier into state B. One terminates the simulation and records the biased transition time $\tau_{AB,i}^{\text{InMetaD}}$, where i is the index of the infrequent metadynamics simulation. Each accelerated first-passage time is then reweighted to the corresponding physical first-passage time by a discretized time-integral over the length of the trajectory[11,26–28]

$$\tau_{AB,i} = \Delta t \sum_{k=1}^{T_i} \exp\left(\frac{U(\mathbf{x}_{i,k}, k\Delta t)}{RT}\right), \tag{11}$$

where Δt is the time step of the trajectory, T_i is the total number of time steps in the ith trajectory, $\mathbf{x}_{i,k}$ is the kth configuration in this trajectory, and $t = k\Delta t$ is the corresponding time. This reweighting assumes that no bias has been deposited on the transition state, which is approximately ensured by the slow deposition of the infrequent metadynamics protocol.[29]

Equation (11) is derived from the Equation (3b), that is, the method assumes that the reaction proceeds via an activated complex. In contrast to Eyring TST, partition functions q_A and $q_{AB^\ddagger}$ are treated classically. The derivation considers a statistical estimate of $q_{AB^\ddagger}/q_A$ from MD simulation data, which has the advantage that no harmonic approximation is needed. See section I of the supplementary material.

2.4 | Reaction coordinate based rate theories

In reaction coordinate based rate theories, one assumes that the system evolves according to a diffusive dynamics along a reaction coordinate $s(\mathbf{x})$. This approach requires the free energy surface (FES) along $s(\mathbf{x})$, which is defined as:[2]

$$F(s) = -RT \ln \pi(s) \tag{12}$$

where $\pi(s)$ is the equilibrium distribution in s:

$$\pi(s) = Z_{\mathbf{x}}^{-1} \int_{\Gamma_{\mathbf{x}}} d\mathbf{x} \exp\left(-\frac{V(\mathbf{x})}{RT}\right) \delta(s(\mathbf{x}) - s). \tag{13}$$

$Z_{\mathbf{x}}$ the configurational partition function, $\Gamma_{\mathbf{x}}$ is the configurational space, and $\delta(s(\mathbf{x}) - s)$ is the Dirac delta function.

The interaction of the internal degrees of freedom with the effective dynamics along s is modelled as a thermal bath, that is, by a friction and random force which are balanced by the Einstein relation. The friction force can be scaled by a friction coefficient or collision rate ξ (with units time^{-1}).

The following two rate theories assume separation of timescales which can be quantified in terms of the FES as $F_{AB}^\ddagger \gg RT$, where $F_{AB}^\ddagger = F(s_{AB^\ddagger}) - F(s_A)$ is the difference of the FES between the FES minimum in the reactant state and the maximum of the free energy barrier. We remark that the location of the free energy maximum $s_{AB^\ddagger}$ does not necessarily coincide with a saddle point in the PES.

In Kramers' rate theory,[2,4,9] the reactant and product state, as well as the maximum of the FES are modelled using a harmonic approximation of the FES around these extrema. In the medium-to-high friction regime, one obtains the following analytical expression for the reaction rate constant

$$k_{AB}^{\text{Kra}} = \frac{\xi}{\omega_\ddagger}\left(\sqrt{\frac{1}{4} + \frac{\omega_\ddagger^2}{\xi^2}} - \frac{1}{2}\right)\frac{\omega_A}{2\pi}\exp\left(-\frac{F_{AB}^\ddagger}{RT}\right) \tag{14}$$

where ω_A is the angular frequency of the harmonic approximation in the reactant state A, and $\omega^\ddagger$ is the angular frequency of the harmonic approximation at the maximum of the free energy barrier.

By relaxing the harmonic approximation one obtains Pontryagin's expression for the rate constant:[10]

TABLE 1 Overview of the model assumptions for different reaction rate models.

	Equation	Activated complex	Separat. of timescales	Harmonic approx.	High T	QM versus CM	TS versus RC	Sampling	Further assumptions
Eyring TST	3b	✓	–	✓	–	QM	TS	–	
High T TST	7	✓	–	✓	✓	QM/CM	TS	–	
InMetaD	11	✓	✓	–	✓	CM	(RC)	✓	Poisson statistics
									No bias on TS
Pontryagin	15	–	✓	✓	✓	CM	RC	✓	High friction
Kramers	14	–	✓	–	✓	CM	RC	✓	Medium-to-high friction

Abbreviations: CM, classical mechanics; QM, quantum mechanics; RC, reaction coordinate; TS, transition state.

$$k_{AB}^{\mathrm{Pon}} = \left\{ \int_{s_A}^{s_B} \mathrm{d}s' \left[\frac{1}{D(s')} e^{\beta F(s')} \int_{-\infty}^{s'} \mathrm{d}s'' \, e^{-\beta F(s'')} \right] \right\}^{-1} \tag{15}$$

where $\beta = 1/RT$, and $D(s) = \frac{RT}{\mu_q \xi(s)}$. $D(s)$ is the position dependent diffusion profile, which arises from the position dependent friction coefficient $\xi(s)$. μ_q is a effective molar mass. $D(s)$ can be estimated form MD simulations following Reference 8. Note that, while Equation (14) is valid in the intermediate and in the high friction regime, Equation (15) is only valid in the high friction regime, where the effective dynamics can be modelled by overdamped Langevin dynamics. Equation (15) is often quite generically referred to as the formula for the mean first-passage time (MFPT) for diffusion over a barrier (which is inverted to get the rate) or the escape rate. For the sake of clarity, we shall refer to it as the Pontryagin rate equation after Reference 10.

The assumptions of the reaction rate models introduced in this section are summarized in Table 1. We remark that all sampling-based approaches use the high-temperature approximation, and that infrequent metadynamics needs a reaction coordinate to apply the bias, but not for the actual estimate for the rate constant.

3 | MODEL SYSTEMS AND POTENTIAL ENERGY SURFACE

pSb5 and pSb1 (Figure 1B,C) are model compounds for retinal. In proteins, retinal is covalently linked to a lysine side chain via a protonated Schiff base (Figure 1A). In pSb5 (naming following Reference 18), the β-ionone ring and methyl substituents as well as the lysine chain have been removed. In pSb1 (naming following Reference 18), the β-ionone ring and methyl substituents remain but the lysine chain has been replaced by a methyl group. Both compounds have been used as models for retinal in previous studies.[18–20,30–34]

Our goal is to evaluate various rate theories for two model compounds on a specific potential energy surface. Here, we outline our selection of the electronic structure method for PES calculation. Even though most computational models emphasize photo-isomerization in electronically excited states, our focus centers on the thermal isomerization within the electronic ground state.

A critical point in modelling the thermal isomerization is the highly correlated π-electron system along the retinal polyene chain, which allows for two possible pathways for the cis-trans isomerization. In the first pathway, the double bond is broken homolytically when the torsion angle reaches ca. 90 degrees, creating a transition state with diradical character. In the second pathway, cis-trans isomerization over the double bond occurs through charge transfer, with the electrons moving towards the protonated imine in the transition state.

From quantum chemical methods that account for dynamic electron correlation, there is little consensus as to whether cis-trans isomerization in molecules with three conjugated double bonds proceeds via a charge-transfer or a diradical mechanism.[21,35–37] However, DFT studies of retinal and related systems[18,20,32–34] conclude that the isomerization over double bonds in the polyene chains proceeds through a charge-transfer pathway if the Schiff-base is protonated. Since both pSb5 and pSb1 feature a protonated Schiff-base, the charge-transfer pathway seems to be a reasonable assumption for the isomerization of the C_{13}=C_{14} double bond in our model compounds. In a charge-transfer pathway, electrons stay paired (closed-shell) during isomerization, and we thus do not necessarily need an electronic structure method that models unpaired electrons.

Ab-initio MD simulations of the thermal isomerization in retinal at the level of DFT are limited to simulation times in the order 1 ns to 10 ns, which is not enough to converge a free energy surface. An alternative is the self-consistent-charge tight-binding density-functional method (DFTB),[12,22] whose computational cost is 2–3 orders of magnitude lower than DFT, thus giving access to much longer simulation timescales. DFTB is an approximation to DFT based on expansion of the total energy around a reference density,[22] where DFTB3[12] includes the third order of the expansion. Even though spin polarization has been introduced for DFTB,[38–40] most applications are based on restricted DFT and cannot model unpaired electrons.

For retinal compounds, DFTB-predicted structures are in good agreement with NMR experiments.[41] Relative to DFT, DFTB yields a reasonable description of the torsional properties of retinal not only in the gas phase,[34] but also in the protein environment.[18,42,43] Torsional barriers for the C_{13}=C_{14} bond in retinal compounds are slightly

underestimated (about 2 kcal mol^{-1}) when using DFTB as compared to DFT/B3LYP.

More approximate potential energy functions, including semi-empirical methods such as AM1 or PM3, overestimate the delocalization more dramatically than DFTB.[34] In empirical force fields, the delocalization can be modelled by imposing the bond lengths along the polyene chain. But since these potential energy functions use fixed partial atomic charges, they are not well suited to describe the charge-shift in the polyene chain during the isomerization, and consequently the isomerization is highly sensitive to the choice of these charges.

Since DFTB strikes a suitable balance between the accuracy of the potential energy function and the computational cost of conducting ab-initio MD simulations, we use will use it in our simulations. For rate theories that do not require sampling, we include calculations at the level of unrestricted DFT/B3LYP/6-31G* for comparison to a higher level of theory.

4 | RESULTS

4.1 | Free energy surface and diffusion profile

The potential energy functions of pSb5 and pSb1 are high-dimensional functions of 72 and 156 internal degrees of freedom, respectively. With MD simulations, one can characterize these high-dimensional energy functions in more manageable, lower-dimensional collective variable spaces using free energy surfaces (FES) and diffusion profiles.

Figure 2 shows the FES (Equation 12) along the C_{13}=C_{14} torsion angle φ, as estimated from umbrella sampling[44] (US) and well-tempered metadynamics[29,45] (MetaD) simulations using ab-initio MD with the DFTB3 method. The line thickness shows the statistical error in the estimated FES.

A full rotation around φ yields two barriers which, as expected, have the same absolute height. The vertical rotational barriers are $F^{\ddagger}_{t\to c} \approx 89\,\mathrm{kJ\,mol^{-1}}$ and $F^{\ddagger}_{c\to t} \approx 81\,\mathrm{kJ\,mol^{-1}}$ for pSb5 and $F^{\ddagger}_{t\to c} \approx 79\,\mathrm{kJ\,mol^{-1}}$ and $F^{\ddagger}_{c\to t} \approx 75\,\mathrm{kJ\,mol^{-1}}$ for pSb1 (both from MetaD1 in Figure 2). The rotational barrier in pSb5 is slightly higher than in pSb1, because the tertiary C_{13} in pSb1 stabilizes the positive charge at the charge-transfer transition state better than the secondary C_{13} in pSb5.[19,20] As usual for a carbon double bond, the trans state at $\varphi = \pi$ rad is slightly more stable than the cis state at $\varphi = 0$ rad, however the stabilization is larger in pSb5 ($9.43 \pm 1.20\,\mathrm{kJ\,mol^{-1}}$) than in pSb1 ($2.70 \pm 1.23\,\mathrm{kJ\,mol^{-1}}$). A possible explanation might be that in the trans state of pSb1, the methyl group at C_{13} sterically interacts with the hydrogens at C_{15}, which destabilizes this conformation.

The effective dynamics along a reaction coordinate are suitably modelled by stochastic dynamics with position dependent diffusion constant. The dependence of the diffusion constant on the collective variable is due to the dynamics in the orthogonal degrees of freedom and due to the curvature of the collective variable. The diffusion profile in Figure 2 show the form that is expected for a dihedral angle rotations. Note that the estimate of the diffusion constant fails in the barrier region, because of the sharpness of the barriers.

In pSb1, umbrella sampling and metadynamics yield essentially the same FES for various parameter settings (Figure 2B). By contrast, estimates of the FES for pSb5, and especially the relative stability of the cis state, depends on the method that is used to construct the FES and on the parameter settings (Figure 2A). Additionally, the metadynamics simulations converge much slower for pSb5 than for pSb1. Convergence of the metadynamics simulation can be checked by monitoring the estimated free-energy difference between cis and trans state ΔF as function of simulation time (Figure 3A), or by monitoring average errors in block analysis[46] as a function of block size (Figure 3B).

The kinks in the lines in Figure 3A correspond to transitions between cis and trans state during the metadynamics build-up. The larger and less frequent kinks in the simulations for pSb5 compared to those for pSb1 imply that the bias builds up within one state longer before moving to the other. Slow convergence can be caused by correlated motion in degrees of freedom orthogonal to the biased coordinate, which in this case is the C_{13}=C_{14} dihedral φ.

4.2 | Correlated degrees of freedom

Using MD simulations, correlation between various collective variables can be assessed. In the case of pSb1, previous research in Ref. 18 documented correlations between the C_{13}=C_{14} dihedral angle and adjacent dihedral angles along the minimum energy path, that is, at 0 K. In Figure 4, the correlations at 300 K are presented for both pSb5 and pSb1. These correlation plots were generated from US simulations, with each color in the plot representing a different umbrella potential. Crosses represent the minimum energy path calculated at unrestricted DFT/B3LYP/6-31G*.

The MD simulations at DFTB3 follow closely the B3LYP/6-31G* minimum energy path, which substantiates that DFTB faithfully represents the DFT-PES of these two molecules. However, there is considerable thermal fluctuations around the minimum energy path. For the single-bond dihedrals C_{12}-C_{13} and C_{14}-C_{15}, the spread is ± 0.4 rad ($\approx \pm 23$ degrees), whereas for the improper torsions the spread is ± 0.2 rad ($\approx \pm 11$ degrees). Overall, we find that the correlation extends to the neighboring single bond, but not to the improper dihedrals at C_{12} and C_{15}. We observed a certain level of correlation to the C_{15}-N double bond along the minimum energy paths. However, this correlation is obscured by thermal fluctuations at 300 K (see Figure S1 in the supplement).

The most remarkable feature of the correlations plots are the sudden jumps at the transition states ($\varphi = +\pi/2$ rad and $\varphi = -\pi/2$ rad). The improper dihedral angles at C_{13} and C_{14} represent the planarity at these sp^2-carbon atoms, with $\chi = 0$ rad representing a planar conformation. Consider the improper torsion at C_{14} when approaching the transition state at $\varphi = -\pi/2$ rad from the trans-state, the local conformation at C_{14} bends out of plane up to 20 degrees (0.35 rad, minimum energy value). At the transition state, it suddenly inverts to an

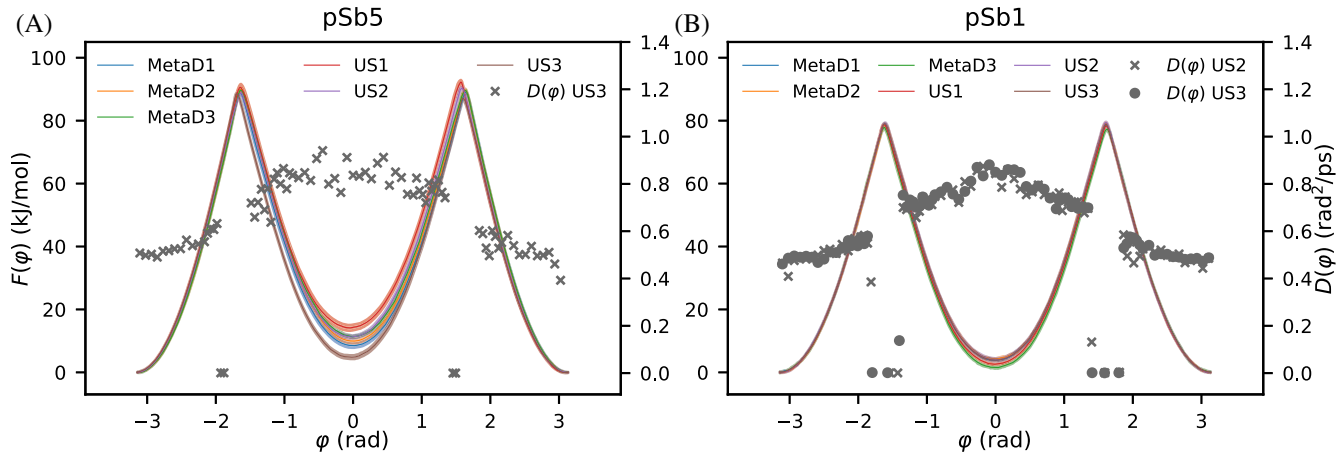

FIGURE 2 Free energy surfaces $F(\varphi)$ and diffusion profiles $D(\varphi)$ along C_{13}=C_{14} dihedral angle φ for (A) pSb5 and (B) pSb1 from metadynamics (MetaD) and umbrella sampling (US) using DFTB3. Parameters for MetaD and US simulations are reported in Tables 4 and 5. Free energy curves are filled between plus and minus one standard error.

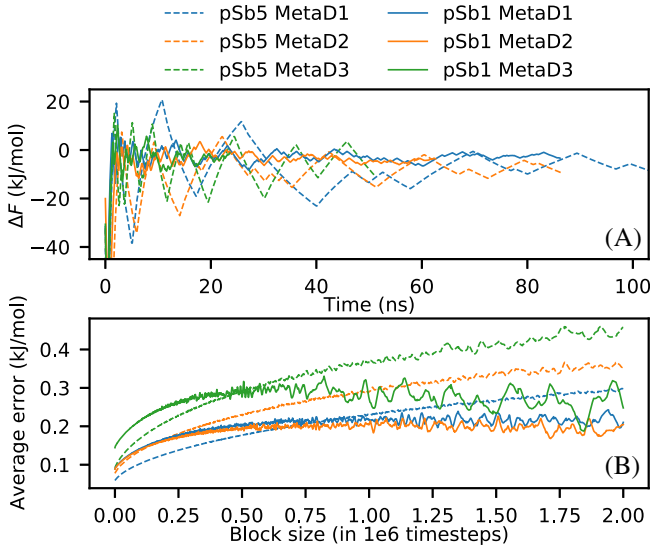

FIGURE 3 (A) Convergence of free energy difference $\Delta F = F_{cis} - F_{trans}$ from metadynamics bias as a function of simulation time for pSb5 (dashed) and pSb1 (full). (B) Convergence of the average errors from block averaging analysis as a function of block size for the same simulations as above.

out-of-plane distortion of −20 degrees. On top of the out-of-plane wagging at C_{14}, the substituent at N (H for pSb5 and CH_3 for pSb1) slightly rotates. The concerted motion is illustrated in Figure 5, where structures of pSb1 nearing the transition state ($\varphi \approx -90$ degrees) from cis (orange structure, $\varphi = -60$ degrees) and from trans (red structure, $\varphi = -120$ degrees) are aligned along the C_{13}, C_{15} and N atoms. Note that the out-of-plane wagging at C_{14} contributes to the correlation between C_{13}=C_{14} dihedral and C_{14}-C_{15} dihedral.

C_{13} shows a similar out-of-plane wagging as C_{14}. However, while at C_{14} we do not find any difference between pSb5 and pSb1, the correlation of the C_{13}=C_{14} dihedral to the improper dihedral at C_{13} and the C_{13}-C_{12} torsion is less pronounced in pSb1 than in pSb5. Presumably, the methyl substituent hinders the out-plane motion at C_{13} in pSb1 compared to C_{14} in the same molecule and thus interrupts the correlation.

4.3 | Rates for the DFTB3 potential energy surface

With a model of the free energy surface of pSb5 and pSb1 and good understanding of the reaction mechanism, we are ready to discuss the reaction rate constants for the thermal isomerization at the level of DFTB3 (Tables 2 and 3). Transition states for both compounds were optimized using the Nudged Elastic Band (NEB) method. NEB optimization converged well for pSb5, but was very sensitive to the choice of the NEB parameters (spring constants, maximal force, amount of nodes) for pSb1.

4.3.1 | Eyring TST

In pSb5, the potential energy barrier E_b for trans→cis reaction is 112.2 kJ mol^{-1}, which is in good agreement with the previously reported value of 27.5 kcal mol^{-1} = 115.1 kJ mol^{-1}.[34] The barrier for cis → trans reaction is about 7.7 kJ mol^{-1} lower, which implies that the cis reactant state is slightly higher in energy than the trans state. This aligns closely with the free energy difference of 8 to 10 kJ mol^{-1} between cis and trans states in pSb5 (Figure 2). For pSb1, the potential energy barriers have about equal height (93.7 and 91.1 kJ mol^{-1}) and are about 10 kJ mol^{-1} lower than for pSb5. Again, this aligns closely with the FES along φ for this molecule.

In each of the four reactions, the free energy difference $\Delta F^{\ddagger}$ at $T = 300$ K is about 8 to 12 kJ mol^{-1} lower than E_b due to the vibrational and rotational contribution to the free energy difference. For pSb5, the Eyring TST rates are $8.30 \cdot 10^{-6}$ s^{-1} for the trans → cis

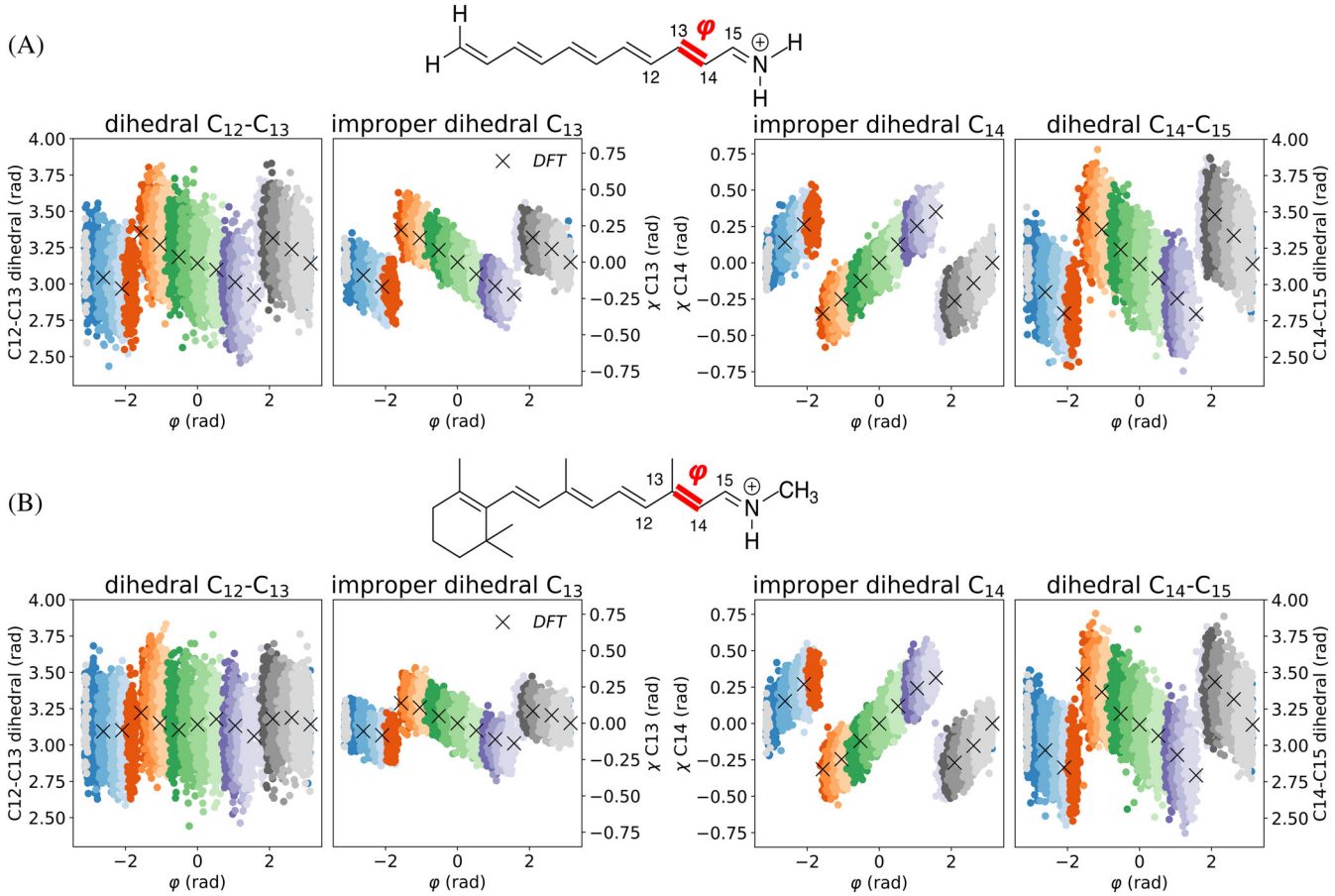

FIGURE 4 Correlations in pSb5 and pSb1. (A) Umbrella sampling simulations using DFTB3 (set US1) for pSb5 showing correlation between C_{13}=C_{14} dihedral φ and dihedral C_{12}-C_{13}, the improper dihedral on C_{13}, the dihedral C_{14}-C_{15} and the improper dihedral on C_{14}. Different colors represent different umbrella windows. Black crosses represent constrained optimizations along φ using unrestricted DFT/B3LYP/6-31G*. (B) Same analysis for pSb1.

reaction and $1.99 \cdot 10^{-4}\,s^{-1}$ for the reverse reaction. In pSb1, the lower energy barrier E_b leads to considerably faster rates, namely $1.42 \cdot 10^{-2}\,s^{-1}$ for the trans→cis transition and $2.06 \cdot 10^{-1}\,s^{-1}$ for the reverse reaction.

4.3.2 | High-temperature TST

The high-temperature approximation approximates the free energy contribution to the rates by neglecting the contribution due to the rotational degrees of freedom and by making a classical approximation for the harmonic vibrational partition function. Tables 2 and 3 show that for pSb5 and pSb1 the rotational contribution is less than 1 kJ mol^{-1}, and thus neglecting this contribution is well justified. In our systems, the vibrational component contributes negatively to the free energy difference, thereby reducing the overall free energy difference $\Delta F^{\ddagger}$ in comparison to the potential energy barrier E_b. This effect is slightly underestimated in the classical approximation. Consequently, $\Delta F^{\ddagger,\mathrm{ht}}$ in high-temperature TST appears higher than $\Delta F^{\ddagger}$ in Eyring TST, and the high-temperature TST rates are slower than Eyring TST rates. The effect amounts to about 3 kJ mol^{-1} which lowers the rate by about a factor of two. Thus, the high-temperature approximation is suitable for our two systems.

4.3.3 | Infrequent metadynamics

The high-temperature approximation constitutes the link between a quantum partition function and the classical partition functions. Models based on classical partition functions can be sampled by MD simulations to obtain a statistical estimate of the rate. One method to do this is infrequent metadynamics, in which Gaussian bias function are deposited in the potential energy well of the reactant state, and the enhanced reaction rate constant is subsequently reweighted to the unbiased reaction rate constant. Rate constants for two different deposition paces of the Gaussian bias functions are shown in Tables 2 and 3. The obtained rate constants were insensitive with regards to doubling the pace of deposition (Table S1), indicating that the deposition rate is infrequent enough for rate constants to be reliable. The infrequent metadynamics simulations passed the Kolmogorov-Smirnoff test[25] which serves as indication whether the assumptions of TST are violated. The rate constants from infrequent metadynamics

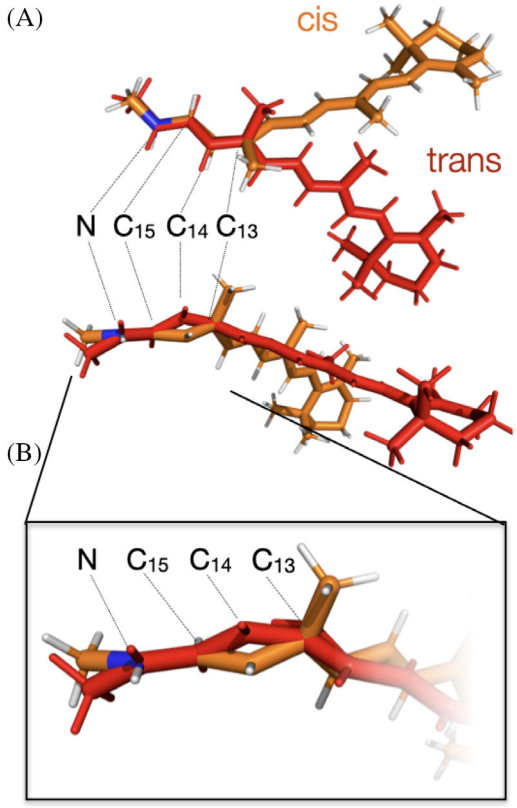

FIGURE 5 (A) Minimum energy structures from constrained optimizations using unrestricted DFT/B3LYP/6-31G* calculations on pSb1 at $\varphi = -60$ degrees (cis, orange structure) and $\varphi = -120$ degrees (trans, red structure). Structures are aligned along the N, C_{15} and C_{13} atoms. (B) Zoom on the reaction center.

are in very good agreement with the rate constants from high-temperature TST. Only in pSb1, the rate constant for the cis → trans reaction is slightly underestimated by infrequent metadynamics.

Infrequent metadynamics and high-temperature TST are based on very similar assumptions (Table 1). In this sense, infrequent metadynamics may be regarded as a sampling-based analogue to high-temperature TST. Because the results from infrequent metadynamics and high-temperature TST are in excellent agreement, we may conclude that sampling-based approaches are suitable for the calculation of reaction rates of chemical reaction and the required MD simulation times are accessible when using DFTB3 to calculate the PES.

4.3.4 | Reaction coordinate based rate theories

We calculated the rate constants from Kramers' (Equation 14) and Pontryagin's (Equation 15) rate theories using the free energy surfaces and diffusion profiles in Figure 2 (Tables 2 and 3). Values for the parameters in Equation (14) are reported in Tables S2 and S3 in the supplement. Surprisingly, these rates are orders of magnitude higher than rates from TST or from infrequent metadynamics. In pSb5, Kramers' rate constant is three orders of magnitude higher then the rate constant from high-temperature TST. This can be traced back to the low free energy barriers $F^{\ddagger}_{AB}$ in the FES compared to the energy difference $\Delta F^{\ddagger}$, as calculated within TST.

Kramers' rate theory can treat reactions that violate the TST assumptions. Therefore, differences between high-temperature TST and Kramers' rate theory might be explained by a complicated reactant state or by violations of the harmonic approximation. However, the choice of our test system, the slow convergence of the metadynamics simulations for the FES in pSb5, as well as the sensitivity of the FES to parameters collectively raise concerns about the optimality of the chosen reaction coordinate. With a sub-optimal reaction coordinate, the free-energy barriers are underestimated and thus Kramers' rate theory overestimates the rate constants. Although φ appears an intuitive choice for the reaction coordinate, the correlated motions in orthogonal degrees of freedom described above suggests that these motions need to be taken into account to construct a sufficiently accurate reaction coordinate.

On the other hand, while pSb1 exhibits the same correlated motions as pSb5, the discrepancy between $F^{\ddagger}_{AB}$ and $\Delta F^{\ddagger}$ is much smaller. Consequently Kramers' rate theory overestimates the rate constants from high-temperature TST for pSb1 only by a factor of 40 for the trans → cis reaction and by a factor of 17 for the reverse reaction. Additionally, the calculation of the FES converges quickly for pSb1. It is not obvious, why the C_{13}=C_{14} dihedral angle φ would be a poor reaction coordinate for pSb5 but a reasonably accurate reaction coordinate in pSb1.

Pontryagin's rate theory yields even higher rate constants than Kramers' rate theory and this points to a second effect that might be at play. Pontryagin's rate theory assumes overdamped Langevin dynamics along the reaction coordinate and would overestimate the rates if the effective dynamics actually falls into the intermediate or weak friction regime. In the weak friction regime, also Kramers' rate theory for intermediate friction (Equation 14) would overestimate the rate constant. The friction regime is in part determined by the "sharpness" of the free energy barriers as measured by $\omega_{\ddagger}$, the angular frequency of the harmonic approximation of the FES maximum. Both systems in fact exhibit very sharp barriers and thus high values of $\omega_{\ddagger}$, which might shift the effective dynamics into the weak-to-intermediate friction regime.

4.4 | Comparison across different PES

Tables 2 and 3 compare the energy barriers and the rotational and vibrational contribution to the free energy differences at the level of DFTB3 to those at the unrestricted DFT/B3LYP/6-31G* level, abbreviated DFT/B3LYP in the following. The potential energy barriers E_b from DFT/B3LYP calculations closely aligning with literature-reported values.[2,20,21]

Compared to unrestricted DFT/B3LYP, DFTB3 tends to underestimate the barrier heights, as has been reported previously.[34,42,43] Energies of constrained optimizations along φ for DFT/B3LYP and DFTB3 are shown in Figure S2 in section III of the supplementary

TABLE 2 Rates for thermal cis-trans isomerization around the C_{13}=C_{14} double bond in **pSb5** using **DFTB3** as well as DFT/B3LYP/6-31G* without (**DFT**) and with (**DFT-D3**) Grimme's dispersion correction.

		DFTB3		DFT		DFT-D3	
	Equation	trans → cis	cis → trans	trans → cis	cis → trans	trans → cis	cis → trans
Free energy difference between *A* and $AB^{\ddagger}$ in Eyring TST [kJ mol^{-1}]							
E_b		112.2	104.5	123.6	108.5	123.6	109.3
ΔF_{rot}	6a	−0.5	0.2	−0.4	0.1	−0.4	0.1
ΔF_{vib}	6a	−7.3	−8.3	−6.2	−6.5	−6.5	−6.6
$\Delta F^{\ddagger}$	5	104.4	96.5	117.0	102.2	116.8	102.8
Free energy difference betweeen *A* and $AB^{\ddagger}$ in high-temperature TST [kJ mol^{-1}]							
E_b		112.2	104.5	123.6	108.5	123.6	109.3
ΔF_{vib}^{ht}	8	−4.6	−5.2	−3.6	−3.6	−3.7	−3.7
$\Delta F^{\ddagger,ht}$	8	107.6	99.3	120.0	104.8	119.9	105.6
Free energy barrier $F_{AB}^{\ddagger}$ [kJ mol^{-1}]							
$F_{AB}^{\ddagger}$		89.0	80.5	n/a	n/a	n/a	n/a
Rates [s^{-1}]							
Eyring TST	3	8.30×10^{-6}	1.99×10^{-4}	5.40×10^{-8}	2.04×10^{-5}	5.81×10^{-8}	1.56×10^{-5}
high *T* TST	7	2.28×10^{-6}	6.39×10^{-5}	1.62×10^{-8}	7.00×10^{-6}	1.65×10^{-8}	5.19×10^{-6}
InMetaD 1	11	1.94×10^{-6}	1.04×10^{-4}	n/a	n/a	n/a	n/a
InMetaD 2	11	2.63×10^{-6}	9.22×10^{-5}	n/a	n/a	n/a	n/a
Kramers	14	4.06×10^{-3}	1.74×10^{-1}	n/a	n/a	n/a	n/a
Pontryagin	15	1.14×10^{-2}	2.74×10^{-1}	n/a	n/a	n/a	n/a

material. The discrepancy between DFT/B3LYP and DFTB3 is larger for the trans → cis transitions than in the cis → trans transitions, with a discrepancy as high as 11.4 kJ mol^{-1} in pSb5. The exception to this trend is the cis → trans transition in pSb1, for which the DFT/B3LYP is 1 kJ mol^{-1} lower than the DFTB3 barrier.

The rotational contributions to total free energy difference are nearly identical at DFT/B3LYP and at DFTB3, which can be attributed to the rigid molecular scaffold. The vibrational motion lowers the total free energy barrier in all four reactions. However, the effect is smaller at the level of DFT/B3LYP than with DFTB3. The discrepancy between DFT/B3LYP and DFTB3 for the vibrational contribution can be as large as the discrepancy for the potential energy barrier E_b (see e.g., cis → trans reaction in pSb1). This highlights the need to consider not only the potential energy barrier but also the vibrational free energy of the reactant and transition state when comparing different PES.

Overall, we find that the total free energy difference $\Delta F^{\ddagger}$ is 6 to 13 kJ mol^{-1} larger in DFT/B3LYP than in DFTB3. Consequently, the Eyring TST rate constants at the level of DFT/B3LYP are one to two orders of magnitude lower than at the level of DFTB3. As with DFTB3, the high-temperature rate constant is slightly lower than the Eyring TST rate constant, because the reduction of the total free energy difference due to the vibrational contribution is underestimated when using the high-temperature limit. Rates from infrequent metadynamics or reaction-coordinate based methods are not available, because they require ab-initio MD simulations. The necessary simulation time to converge these rate estimates is challenging to attain at this level of theory.

The last two columns in Tables 2 and 3 report the influence of the the D3 dispersion correction for DFT[47] on the energy barriers and the free energy contributions to the rates. The effect is less than 1 kJ mol^{-1} (only exception: ΔF_{vib} for trans → cis in pSb1). The difference in the energies for constrained optimizations along φ (Figure S2 in the supplement) is equally small. As a result, there is a minimal difference in the rates when calculated with and without D3 correction. We suspect that the small influence of the D3 correction on the PES stems from the linear and rigid structure of the two molecules. Dispersion is a strongly distance-dependent pairwise interaction. Due to the linearity, the molecules likely have a small intramolecular dispersion overall. Due to the rigid polyene scaffold, most pairwise distances do not change during the reaction. Even though rotation from trans to cis shortens the distance between the Schiff base and the β-ionone ring, the distance remains so large that the two groups have minimal dispersion interaction.

5 | METHODS

Calculations for DFTB3 were carried out with the DFTB+ software package[48] using the 3ob-3-1 Slater-Koster parameter set.[49] Energy minimizations, constrained optimizations and Nudged Elastic Band

TABLE 3 Rates for thermal cis-trans isomerization around the $C_{13}=C_{14}$ double bond in **pSb1** using **DFTB3** as well as DFT/B3LYP/6-31G* without (**DFT**) and with (**DFT-D3**) Grimme's dispersion correction.

		DFTB3		DFT		DFT-D3	
	Equation	trans → cis	cis → trans	trans → cis	cis → trans	trans → cis	cis → trans
Free energy difference between A and $AB^{\ddagger}$ in Eyring TST [kJ mol^{-1}]							
E_b		93.7	91.1	98.0	90.1	98.2	90.4
ΔF_{rot}	6a	−0.1	0.0	−0.1	−0.0	−0.1	−0.0
ΔF_{vib}	6a	−7.7	−12.0	−2.4	−3.8	−4.1	−3.7
$\Delta F^{\ddagger}$	5	85.8	79.2	95.4	86.2	94.0	86.7
Free energy difference$\Delta F^{\ddagger}$ between A and $AB^{\ddagger}$ in high-temperature TST [kJ mol^{-1}]							
E_b		93.7	91.1	98.0	90.1	98.2	90.4
ΔF_{vib}^{ht}	8	−5.1	−9.4	0.5	−1.1	−1.1	−0.9
$\Delta F^{\ddagger,ht}$	8	88.5	81.7	98.4	89.0	97.0	89.5
Free energy barrier$F_{AB}^{\ddagger}$ [kJ mol^{-1}]							
$F_{AB}^{\ddagger}$		78.9	75.0	n/a	n/a	n/a	n/a
Rates [s^{-1}]							
Eyring TST	3	1.42×10^{-2}	2.06×10^{-1}	3.09×10^{-4}	1.24×10^{-2}	5.35×10^{-4}	9.92×10^{-3}
high T TST	7	4.81×10^{-3}	7.44×10^{-2}	9.21×10^{-5}	4.12×10^{-3}	1.58×10^{-4}	3.31×10^{-3}
InMetaD1	11	2.97×10^{-3}	9.30×10^{-3}	n/a	n/a	n/a	n/a
InMetaD2	11	3.09×10^{-3}	1.22×10^{-2}	n/a	n/a	n/a	n/a
Kramers	14	1.92×10^{-1}	1.30×10^{0}	n/a	n/a	n/a	n/a
Pontryagin	15	5.43×10^{-1}	2.07×10^{0}	n/a	n/a	n/a	n/a

(NEB) calculations were done by interfacing DFTB+ with the Atomic Simulation Environment (ASE)[50] and using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm[51] for numerical optimization. Vibrational analysis of the optimized structures was done using DFTB + to obtain the vibrational frequencies, while rotational moments of inertia were calculated by entering the optimized configuration into the Gaussian 16 software.[52] From these data we calculated rates for Eyring TST and high-temperature TST.

Ab-initio MD simulations were performed using DFTB+ using the velocity-Verlet integrator with a time step of 1 fs. Before simulations, energy minimization was done, followed by temperature equilibration at 300 K in two steps. In a first equilibration run, the Berendsen thermostat[53] with a coupling time of 2 ps is employed, while in a second equilibration run a Nosé-Hoover chain setup[54–56] of coupling time 2 ps and chain length 3 is used. For production runs, the same thermostat setup was used as for the second equilibration runs.

Well-tempered metadynamics[29] and umbrella sampling[44] were carried out by interfacing the PLUMED[57] software package with DFTB+. Parameter sets for metadynamics and umbrella sampling sets can be found in Tables 4 and 5 respectively. Sets of runs for infrequent metadynamics were set up by equilibrating in the reactant state, after which a metadynamics bias is applied until a transition is registered. The transition times were reweighted using the acceleration factor which was directly calculated by PLUMED. The set of reweighted transition times was fitted to the theoretical cumulative distribution function of a Poisson distribution (Equation 10) to obtain a mean first-passage time and corresponding rate.

Diffusion profiles were calculated using the method from Reference 8. Effective masses of the reactant states were calculated by measuring the average squared velocity along the dihedral angle and using the equipartition theorem. Frequencies of the harmonic approximations of the reactant wells and transition state barriers were calculated from spring constants obtained by harmonically fitting the corresponding wells or barriers. Free energy barriers $F_{AB}^{\ddagger}$ are measured from the FES directly. One-dimensional rate methods (Kramers and Pontryagin) can then be applied straightforwardly.

Calculations at the DFT level were performed using the Gaussian 16 software[52] using unrestricted DFT with the B3LYP functional[23,24] and the 6-31G* basis set.[58] Full geometry optimizations as well as constrained optimizations were done using the Berny optimization algorithm[59] as implemented in Gaussian. Transition state search was performed using the Synchronous Transit-guided Quasi-Newton (STQN) method[60,61] as implemented in Gaussian, where the reactant and product state input configurations were chosen to be the geometry optimized structures in the trans and cis states. Gaussian performs a full thermochemical analysis including calculation of the translational, rotational and

TABLE 4 Parameters for metadynamics and infrequent metadynamics simulations for pSb5 and pSb1 using DFTB3.

	Height [kJ mol^{-1}]	Width [rad]	Pace [ps]	Bias factor	Time [ns]	Runs
pSb5						
MetaD1	1.3	0.15	0.5	16	152	n/a
MetaD2	1.3	0.10	0.5	16	87	n/a
MetaD3	0.75	0.075	0.25	25	51	n/a
InMetaD1	1.3	0.05	5.0	16	n/a	25
InMetaD2	1.3	0.05	10.0	16	n/a	30
pSb1						
MetaD1	1.3	0.10	0.5	16	86	n/a
MetaD2	1.3	0.05	0.5	16	63	n/a
MetaD3	1.3	0.10	0.5	16	24	n/a
InMetaD1	1.3	0.05	5.0	16	n/a	25
InMetaD2	1.3	0.05	10.0	16	n/a	30

TABLE 5 Parameters for umbrella sampling using DFTB3.

Windows	Biased region [rad]	Interval [rad]	Force constant [kJ mol^{-1} rad^{-2}]
63	$[-3.1, +3.1]$	0.1	500
10	$[-1.95, -1.05]$	0.1	500
10	$[+1.05, +1.95]$	0.1	500

Note: Each umbrella sampling set was run with 83 windows positioned as shown here. In total, three sets were run for pSb5 and three for pSb1 (Figure 2).

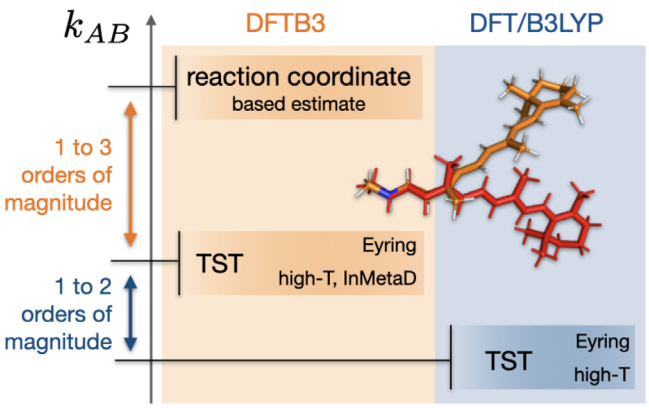

FIGURE 6 Effect of rate model and PES on the estimated reaction rate constant.

vibrational partition functions and corresponding energies and entropies.[62]

This allows for straightforward calculation of rates for Eyring TST. Vibrational frequencies were obtained from Gaussian separately[63] and used to calculate rates for high-temperature TST.

A complete overview of the computational details is given in section II of the supplementary information.

6 | CONCLUSIONS

We studied the thermal cis-trans isomerization in two retinal analogues at 300 K in the gas phase. This reaction falls well within the approximation of TST, and thus accurate values for the reaction rate constant can be obtained from this theoretical framework. However, reactions in molecules with numerous flexible degrees of freedom or reactions in complex environments may not be accurately modeled by TST. We therefore explored whether accurate estimates of the reaction rate constant can be obtained from MD simulations. The impact of changing the theoretical framework for modelling the reaction rate must be assessed relative to the error in the potential energy surface (PES), which is often considered to be the primary error source in rate modeling. To gauge the effect of changing the PES, we compared TST rate constants at the level of at the DFTB3 and the unrestricted DFT/B3LYP/6-31G* level. Figure 6 summarizes our results.

Reaction rate constants at DFTB3 are larger than those at DFT, with a difference of one to two orders magnitude. However only in pSb5 the increase in reaction rate can be mainly attributed to a lower potential energy barrier. In pSb1, the change of the vibrational free energy has an equally strong (trans → cis) or even larger (cis → trans) contribution to the increase of the reaction rate constant. Thus, reducing the comparison of different PES to the height of the potential energy barrier and neglecting entropic effects may be misleading.

It is important to note that we only compared PES that are ultimately derived from DFT. DFT and wavefunction-based methods tend to differ in delocalization of the conjugated π-electron system, with DFT typically overdelocalizing the π-electrons.[64] Overdelocalization reduces the double-bond character in the C_{13}=C_{14} double bond, and thus decreases the reaction barrier. In fact, for cis-trans isomerization in polyene chains, DFT/B3LYP/6-31G* underestimates the torsional barrier compared to CASSCF[65] as well as compared to experimental data.[66–68] Thus, the true reaction rate constant might be even lower than our DFT estimates.

An important assumption when estimating reaction rate constants from MD simulations is that classical mechanics are used to

model the dynamics of the nuclei and to approximate the partition functions. The classical limit is well justified for rugged potential energy landscapes with broad minima and relatively low energy barriers, but might not be appropriate for modelling a chemical reaction. In both of our molecules, the classical limit (high-temperature TST) yields a lower rate constant than Eyring TST, but the effect is less than an order of magnitude. Thus, the error due to the classical limit is small compared to the uncertainty due to the model of the PES.

Since the classical limit is justified for our systems, one should in principle be able to estimate reaction rate constants from MD simulations. Indeed, the infrequent metadynamics results are in excellent agreement with results from high-temperature TST. The length of the simulations (on a biased PES) were in the nanosecond regime, the mean-first passage times $\tau_{AB} = 1/k_{AB}$ are in the regime of hundreds to hundred thousands of seconds. This is a enormous speedup, with the largest acceleration factors being of the order of 10^{14}.

By contrast, our results from Kramers and Pontryagin's rate theory overestimate the rate constant by multiple orders of magnitude. These two methods rely on a reaction coordinate – in our case the C_{13}=C_{14} torsion angle φ - but the FES can be very sensitive to the choice of this reaction coordinate. In fact, we found that the improper dihedrals of the substituents on the C_{13} and C_{14} atoms correlate with the reaction coordinate φ. We hypothesized that this was indicative of a isomerization mechanism consisting of a concerted motion where the C_{14} atom temporarily nods out of the polyene plain, before the isomerization is completed. Thus, even though the torsion angle φ is a very intuitive reaction coordinate, it might not be optimal enough to yield accurate results for Kramers and Pontryagin's rate theory.

This places us in a challenging position. Both MD based-approaches, infrequent metadynamics and reaction coordinate based rate models, come with a high computational cost. Our simulations required approximately 10 to 100 ns for each rate estimation. However, since infrequent metadynamics is derived from TST, it is particularly suitable for chemical reactions that align well with Eyring TST. By contrast, reaction coordinate based rate theories have the potential to model systems deviating from the harmonic approximation and the separation of timescales. However, their robustness is compromised due to sensitivity to the reaction coordinate and the friction regime.

Fortunately, several alternatives exist. Within reaction coordinate-based rate models, one can optimize the reaction coordinate,[69,70] and we will explore this approach in subsequent work.[71] Interestingly, neural networks[17,72] can be used to optimize reaction coordinates. Another avenue extends these models to encompass effective dynamics within multidimensional collective variable spaces.[73,74] Alternatively, one can opt for transition path sampling[75–77] and leverage dynamical reweighting techniques that are not based on TST.[5,78,79] In summary, while sampling chemical reactions in complex systems poses a formidable challenge, there is optimism that this wide variety of ideas will allow us to solve this task.

ACKNOWLEDGMENTS

This research has been funded by Deutsche Forschungsgemeinschaft (DFG) through grant SFB 1114 "Scaling Cascades in Complex Systems"–project number 235221301. Simon Ghysbrecht acknowledges funding by the Einstein Center of Catalysis/BIG-NSE. Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Simon Ghysbrecht https://orcid.org/0009-0003-7556-9592

REFERENCES

[1] H. Eyring, *J Chem Phys* **1935**, *3*, 107.
[2] B. Peters, *Reaction rate theory and rare events*, Elsevier, Amsterdam **2017**.
[3] D. Frenkel, B. Smit, *Understanding molecular simulation: from algorithms to applications*, Elsevier, Amsterdam **2023**.
[4] P. Hänggi, P. Talkner, M. Borkovec, *Rev. Mod. Phys.* **1990**, *62*, 251.
[5] S. Kieninger, L. Donati, B. G. Keller, *Curr. Opin. Struct. Biol.* **2020**, *61*, 124.
[6] B. Peters, *Annu. Rev. Phys. Chem.* **2016**, *67*, 669.
[7] J. Hénin, T. Lelièvre, M. Shirts, O. Valsson, L. Delemotte, *Living J Comput Mol Sci* **2022**, *4*, 1583.
[8] G. Hummer, *New J. Phys.* **2005**, *7*, 34.
[9] H. A. Kramers, *Physica* **1940**, *7*, 284.
[10] L. Pontryagin, A. Andronov, A. Vitt, *J. Exp. Theor. Phys.* **1933**, *3*, 165.
[11] P. Tiwary, M. Parrinello, *Phys. Rev. Lett.* **2013**, *111*, 230602.
[12] M. Gaus, Q. Cui, M. Elstner, *J. Chem. Theory Comput.* **2011**, *7*, 931.
[13] P. Pracht, F. Bohle, S. Grimme, *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169.
[14] R. Schade, T. Kenter, H. Elgabarty, M. Lass, T. D. Kühne, C. Plessl, *Int J High Perform Comput Appl* **2023**, *37*, 530.
[15] H. M. Senn, W. Thiel, *Angew Chem Int Ed* **2009**, *48*, 1198.
[16] T. P. Senftle, S. Hong, M. M. Islam, S. B. Kylasa, Y. Zheng, Y. K. Shin, C. Junkermeier, R. Engel-Herbert, M. J. Janik, H. M. Aktulga, *Npj Comput. Mater.* **2016**, *2*, 1.
[17] P. Gkeka, G. Stoltz, A. Barati Farimani, Z. Belkacemi, M. Ceriotti, J. D. Chodera, A. R. Dinner, A. L. Ferguson, J.-B. Maillet, H. Minoux, *J. Chem. Theory Comput.* **2020**, *16*, 4757.
[18] A.-N. Bondar, M. Knapp-Mohammady, S. Suhai, S. Fischer, J. C. Smith, *Theor. Chem. Acc.* **2011**, *130*, 1169.
[19] P. Tavan, K. Schulten, D. Oesterhelt, *Biophys. J.* **1985**, *47*, 415.
[20] E. Tajkhorshid, S. Suhai, *J Phys Chem B* **1999**, *103*, 5581.
[21] S. Gozem, M. Huntress, I. Schapiro, R. Lindh, A. A. Granovsky, C. Angeli, M. Olivucci, *J. Chem. Theory Comput.* **2012**, *8*, 4069.
[22] M. Elstner, *Theor. Chem. Acc.* **2006**, *116*, 316.
[23] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785.
[24] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098.
[25] M. Salvalaglio, P. Tiwary, M. Parrinello, *J. Chem. Theory Comput.* **2014**, *10*, 1420.
[26] H. Grubmüller, *Phys Rev E* **1995**, *52*, 2893.
[27] A. F. Voter, *Phys. Rev. Lett.* **1997**, *78*, 3908.
[28] S. A. Khan, B. M. Dickson, B. Peters, *J Chem Phys* **2020**, *153*, 054125.
[29] A. Barducci, G. Bussi, M. Parrinello, *Phys. Rev. Lett.* **2008**, *100*, 020603.
[30] J. Baudry, S. Crouzy, B. Roux, J. C. Smith, *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1018.
[31] J. Baudry, S. Crouzy, B. Roux, J. C. Smith, *Biophys. J.* **1999**, *76*, 1909.
[32] E. Tajkhorshid, S. Suhai, *Chem. Phys. Lett.* **1999**, *299*, 457.
[33] E. Tajkhorshid, B. Paizs, S. Suhai, *J Phys Chem B* **1999**, *103*, 4518.
[34] H. Zhou, E. Tajkhorshid, T. Frauenheim, S. Suhai, M. Elstner, *Chem. Phys.* **2002**, *277*, 91.
[35] L. De Vico, C. S. Page, M. Garavelli, F. Bernardi, R. Basosi, M. Olivucci, *J. Am. Chem. Soc.* **2002**, *124*, 4124.

[36] S. Gozem, H. L. Luk, I. Schapiro, M. Olivucci, *Chem. Rev.* **2017**, *117*, 13502.
[37] A. Zen, E. Coccia, S. Gozem, M. Olivucci, L. Guidoni, *J. Chem. Theory Comput.* **2015**, *11*, 992.
[38] B. Aradi, B. Hourahine, T. Frauenheim, *J Phys Chem A* **2007**, *111*, 5678.
[39] B. Hourahine, S. Sanna, B. Aradi, C. Köhler, T. Niehaus, T. Frauenheim, *J Phys Chem A* **2007**, *111*, 5671.
[40] P. Melix, A. F. Oliveira, R. Rüger, T. Heine, *Theor. Chem. Acc.* **2016**, *135*, 1.
[41] M. Sugihara, V. Buss, P. Entel, M. Elstner, T. Frauenheim, *Biochemistry* **2002**, *41*, 15259.
[42] A.-N. Bondar, M. Elstner, S. Suhai, J. C. Smith, S. Fischer, *Structure* **2004**, *12*, 1281.
[43] N. Elghobashi-Meinhardt, P. Phatak, A.-N. Bondar, M. Elstner, J. C. Smith, *J Membrane Biol* **2018**, *251*, 315.
[44] G. M. Torrie, J. P. Valleau, *J. Comput. Phys.* **1977**, *23*, 187.
[45] A. Laio, M. Parrinello, *Proc Natl Acad Sci* **2002**, *99*, 12562.
[46] G. Bussi, G. A. Tribello, in *Biomolecular Simulations. Methods in Molecular Biology* (Eds: M. Bonomi, C. Camilloni), Vol. 2022, Humana, New York, NY **2019**.
[47] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
[48] B. Hourahine, B. Aradi, V. Blum, F. Bonafé, A. Buccheri, C. Camacho, C. Cevallos, M. Deshaye, T. Dumitrică, A. Dominguez, *J. Chem. Phys.* **2020**, *152*, 124101.
[49] M. Gaus, A. Goez, M. Elstner, *J. Chem. Theory Comput.* **2013**, *9*, 338.
[50] A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, *J. Phys: Condens. Matter* **2017**, *29*, 273002.
[51] R. Fletcher, *Practical methods of optimization*, John Wiley & Sons, Hoboken, NJ, USA **2000**.
[52] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, Gaussian 16 Revision C.01. **2016**.
[53] H. J. Berendsen, J. Postma, W. F. Van Gunsteren, A. DiNola, J. R. Haak, *J Chem Phys* **1984**, *81*, 3684.
[54] S. Nosé, *J Chem Phys* **1984**, *81*, 511.
[55] W. G. Hoover, *Phys. Rev. A* **1985**, *31*, 1695.
[56] G. J. Martyna, M. L. Klein, M. Tuckerman, *J Chem Phys* **1992**, *97*, 2635.
[57] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, *Comput. Phys. Commun.* **2014**, *185*, 604.
[58] W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.* **1972**, *56*, 2257.
[59] H. B. Schlegel, *J. Comput. Chem.* **1982**, *3*, 214.
[60] C. Peng, H. Bernhard Schlegel, *Isr. J. Chem.* **1993**, *33*, 449.
[61] C. Peng, P. Y. Ayala, H. B. Schlegel, M. J. Frisch, *J. Comput. Chem.* **1996**, *17*, 49.
[62] J. W. Ochterski, Gaussian Inc. **2000**.
[63] J. W. Ochterski, Gaussian Inc. **1999**.
[64] M. Wanko, M. Hoffmann, P. Strodel, A. Koslowski, W. Thiel, F. Neese, T. Frauenheim, M. Elstner, *J Phys Chem B* **2005**, *109*, 3606.
[65] B. Paizs, E. Tajkhorshid, S. Suhai, *J Phys Chem B* **1999**, *103*, 5388.
[66] T. Okada, M. Sugihara, A.-N. Bondar, M. Elstner, P. Entel, V. Buss, *J. Mol. Biol.* **2004**, *342*, 571.
[67] F. Buda, P. Giannozzi, F. Mauri, *J Phys Chem B* **2000**, *104*, 9048.
[68] A.-N. Bondar, J. C. Smith, M. Elstner, *Theor. Chem. Acc.* **2010**, *125*, 353.
[69] E. Vanden-Eijnden, *Annu. Rev. Phys. Chem.* **2010**, *61*, 391.
[70] B. Roux, *J Phys Chem A* **2021**, *125*, 7558.
[71] S. Ghysbrecht, L. Donati, B. G. Keller, arXiv preprint arXiv: 2312.12948. **2023**.
[72] H. Jung, R. Covino, A. Arjun, C. Leitold, C. Dellago, P. G. Bolhuis, G. Hummer, *Nat Comput Sci* **2023**, *3*, 334.
[73] H. C. Lie, K. Fackeldey, M. Weber, *SIAM J Matrix Anal Appl* **2013**, *34*, 738.
[74] L. Donati, M. Weber, B. G. Keller, *J. Phys: Condens. Matter* **2021**, *33*, 115902.
[75] P. G. Bolhuis, D. Chandler, C. Dellago, P. L. Geissler, *Annu. Rev. Phys. Chem.* **2002**, *53*, 291.
[76] D. M. Zuckerman, L. T. Chong, *Annu. Rev. Biophys.* **2017**, *46*, 43.
[77] K. Palacio-Rodriguez, F. Pietrucci, *J. Chem. Theory Comput.* **2022**, *18*, 4639.
[78] L. Donati, B. G. Keller, *J Chem Phys* **2018**, *149*, 072335.
[79] K. Shmilovich, A. L. Ferguson, *J. Phys. Chem. A* **2023**, *127*, 3497.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: S. Ghysbrecht, B. G. Keller, *J. Comput. Chem.* **2024**, *45*(16), 1390. https://doi.org/10.1002/jcc.27332