Dynamical Re-crossing in the Intercalation Process of the Anticancer Agent Proflavine into DNA

V. M Hridya, James T Hynes, Arnab Mukherjee

Dynamical Recrossing in the Intercalation Process of the Anticancer Agent Proflavine into DNA

V. M. Hridya[†], James T. Hynes[§,‡,*], and Arnab Mukherjee[†,*]

[†]Department of Chemistry, Indian Institute of Science Education and Research Pune-411008, India

[§]Department of Chemistry, University of Colorado, Boulder, CO 80309-0215 USA

[‡]Ecole Normale Supérieure-PSL Research University, Chemistry Department, Sorbonne Universités-UPMC University Paris 06, CNRS UMR 8640 Pasteur, 24 rue Lhomond, 75005 Paris, FR

ABSTRACT: Intercalation into DNA is the interaction mode of some anthracycline antibiotics. Recently, this process's molecular mechanism was explored using the static free energy landscape. Here we explore the dynamical effects in the intercalation of proflavine into DNA by calculating the transmission coefficient κ -- providing a measure of the departure from Transition State Theory for the reaction rate constant -- by examination of the recrossing events at the transition state. For that purpose, we first found the accurate transition state of this complex system -- as judged by a committor analysis -- using a set of all-atom simulations of total length 6.3 millisecond. In a subsequent calculation of the transmission coefficient κ in another extensive set of simulations, the small value $\kappa=0.1$ was found, indicating a significant departure from TST. Comparison of this result with Grote-Hynes and Kramers theories shows that neither theory is able to capture this complex system's recrossing events; the source of this striking failure is discussed, as are related aspects of the mechanism. This study suggests that, for biomolecular processes similar to this, dynamical effects essential for the process are complex in nature and require novel approaches for their elucidation.

INTRODUCTION

Molecular recognition, a specific non-covalent interaction between two or more molecules, is an important process governing many biological activities. Intercalation is a specific type of molecular recognition where the planar aromatic moiety of a small molecule inserts between two consecutive base pairs of a DNA strand[1-3]. Intercalation causes structural changes to the strand and can inhibit replication and transcription of DNA, subsequently leading to cell death[4]. Hence, these small molecules are potential drugs for cancer therapeutics.

Proflavine is one such small aromatic molecule which can intercalate into DNA. Due to its toxicity and nonspecificity[5], it is not currently used as an anticancer drug. However, modifications are underway for better specificity and druggability[5]. From the kinetic studies of proflavine-DNA interaction, it was found that the intercalation process comprises at least two steps: formation of a fast outside bound state (minor groove-bound drug or major groove-bound drug), followed by a slow insertion process[6]. It was proposed earlier[6] that the insertion proceeds by either of two mechanisms: (i) one where the drug waits outside the DNA and inserts itself between the base pairs when the cavity becomes large enough due to fluctuations and (ii) the second where drug induces a cavity between the base pairs and forces itself into the DNA. Recently, some of us calculated intercalation pathways of proflavine from major and minor grooves using well-tempered metadynamics and provided evidence in favor of the drug-induced cavity formation mechanism[7-8]. The constructed minimum free energy path also revealed the process's mechanistic details through its static free energy landscape[8].

The dynamical aspects of the intercalation process are far less clear than its current thermodynamical understanding. Transition state theory (TST)[9] is widely used to calculate the rate constant (k_{TST}) of a reaction from its static free energy profile. However, TST neglects any dynamical effect such as system's recrossing of the transition state surface and thus overestimates the reaction rate constant. In the absence of tunneling, the correct rate constant is provided by multiplying the TST rate by the transmission coefficient (TC). It is therefore a correction factor, often denoted by , with a value of unity or less. It has been shown that incorporation of the dynamical effects in the reaction via a transmission coefficient in TST rate improves the rate prediction.[10-11] The exploration of the recrossing or dynamical effects thus provides the "true" description of the process. For intercalation system, this knowledge may well be relevant for designing better intercalators and, therefore, better anticancer drugs.

The importance of recrossing has been illustrated in some biomolecular systems. For example, it has been observed that the amount of recrossing is less in the active site environment of an enzyme-catalyzed reaction compared to its uncatalyzed aqueous solution counterpart. Therefore, beyond decreasing the reaction barrier, the enzyme increases the TC (less recrossing) to increase the rate of the reaction, although to a smaller extent compared to the former.[12-14] The k_{TST} and the TC values depend on the choice of the reaction coordinate (RC). But the actual rate constant k of the reaction (which is related to a steady-state flux) is independent of that choice. An improper RC or a non-optimal dividing surface can automatically result in a very poor TC estimate. However, Peters *et al.* showed that even with an optimal dividing surface choice, the recrossing phenomenon can remain unavoidable, as illustrated for an ion-pair dissociation.[15] In the presence of a supposedly perfect dividing surface, the coupling of the environment degrees of freedom with the RC motion can cause recrossing.

The first attempt to include the effect of environment-induced recrossing on the rate of a reaction was given by Kramers in 1940.[16] He considered the escape of a Brownian particle in a well

across a barrier, with a solvent as environment; he used the ordinary Langevin equation (with continuum solvent description and a markovian, memory-less friction) to describe its motion in a potential in the presence of a frictional force and a random force. The Langevin equation employs a static friction constant, which -- it is now appreciated -- assumes that the correlation time of the solvent frictional force is small on the time scale of the barrier crossing. In many cases, however, the solvent forces will often not relax fast on the time scale of barrier crossing. Then the time correlations or memory effects of the solvent forces need to be included via a friction kernel. This was realized by Grote and Hynes, who focused on dynamics in the transition state region and employed the generalized Langevin equation (GLE) to describe the motion of the reacting system across the barrier.[17-18] The GLE

for the motion of the reaction coordinate $x(t)$ in a parabolic potential of $G[(x, t)]$ is given by

$$m\ddot{x}(t) = -\frac{\partial G[(x,t)]}{\partial x(t)} - \int_0^t m\,\xi(\tau)\,\dot{x}(t-\tau)\,d\tau + \delta F_R(t), \quad (1)$$

Here the first term on the right-hand side represents the situation when the solvent molecules equilibrate to the RC motion, i.e., it describes the equilibrium solvation force due to the potential of mean force (PMF). The non-equilibrium solvation force, in which the solvent molecules do not have enough time to equilibrate, is given by the second term, which incorporates a friction kernel accounting for both the short and long-time correlations of the solvent forces. The last term represents the random fluctuating forces.
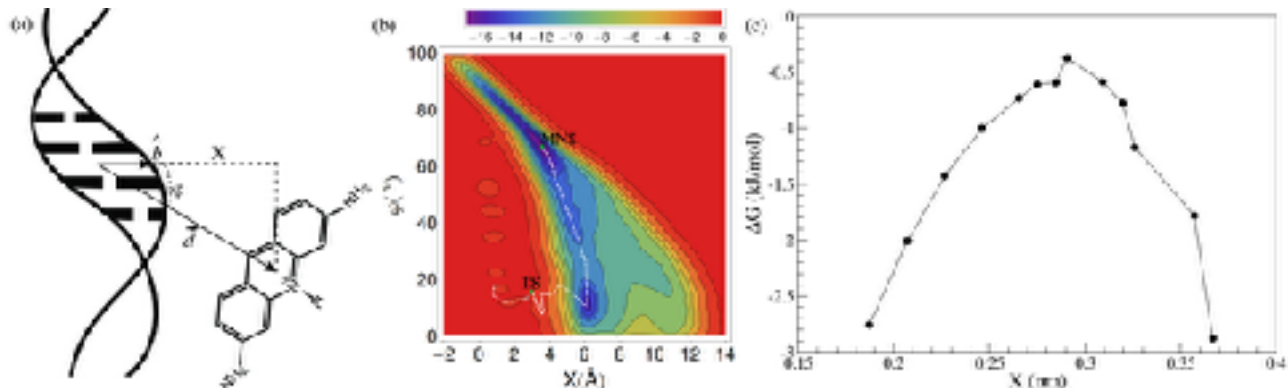

Figure 1. (a) Schematic representation of the reaction coordinates X and φ. $\vec{d}$ is the vector from the center of mass of the intercalating base pairs to the center of mass of the proflavine. $\hat{b}$ is the unit vector from the center of mass of the same base pairs to the center of mass of the two sugars. $X = \hat{b} \cdot \vec{d}$ and $\varphi = \cos^{-1}\left(\hat{b} \cdot \vec{d} / |\vec{d}|\right)$. (b) The two-dimensional free energy surface of the intercalation process from minor groove bound state (MNS) against X and φ.[7-8] The FES does not have a well-sampled intercalated state and therefore it is not shown. The color bar on the top indicates the free energy values. The dotted white line represents the minimum free energy path. (c) Free energy values around the TS along the minimum free energy path, i.e., the white line shown in (b), plotted against the collective variable X, with the intercalated product state having a small X value. Improvement of the latter calculation is described later.

Using the GLE, combined with the Stable States Picture (SSP) involving the definition of reactant and product regions away from the transition state[19], Grote Hynes theory (GHT) calculates the rate constant via the time correlation function of the reactive flux. The overall rate is obtained from the product of TST rate and the transmission coefficient (TC), $k = \kappa_{GH}\,k_{TST}$, where the transmission coefficient is given by the ratio of the reactive frequency ω_r and the equilibrium barrier frequency ω_{eq} as $\kappa_{GH} = \dfrac{\omega_r}{\omega_{eq}}$.

Here ω_{eq} is related to the magnitude of curvature of the equilibrium free energy at the transition state, while ω_r is the effective frequency for the reaction system's motion across the barrier. In the GHT approach, the effective friction felt by the reaction coordinate during the barrier passage at the TS can be much less than the full friction assumed by the Kramers theory (KT). The resulting GHT transmission coefficient can thus be considerably closer to unity (less recrossing due to less effective friction) than that predicted by KT.[12, 20-21]

GHT has been applied successfully in many reaction model studies, including S_N1 and S_N2 reactions[22-23], ion-pair association[24-25], ion-transfer reaction[21], electron transfer reaction[26], proton transfer

reaction[27], etc. Several transition state recrossing studies using GHT have also been used to understand the role of the environment in enhancing reaction rates in enzyme catalyses; examples include Michael addition by chalcone isomerase[12], methyl transfer by catechol O-methyltransferase[13] and glycine N-methyltransferase[28], reductive methylation by thymidylate synthase,[20] etc. However, studies of dynamical recrossing remain uncommon in all-atom molecular dynamics simulations for complex biomolecular recognition processes such as the one presented here; this is due to the dual difficulty of finding an accurate description of the TS and identifying the appropriate reaction coordinate near the TS.

Here we study the dynamical recrossing effect in the intercalation process of the anticancer agent proflavine through the minor groove of DNA in aqueous solution. With the aid of our previous detailed study of the intercalation pathway of proflavine[7], we have searched for the accurate transition state using committor probability calculations with several probable TS candidates. Subsequently, we performed reactive flux formalism simulations to estimate the intercalation's transmission coefficient (TC); the result was then compared with the theoretical estimates using GHT and KT. We find that both GHT (surprisingly, given its ex-

tensive success indicated above) and KT are in considerable disagreement with the reactive flux simulation result. We will discuss reasons for the discrepancy in terms of trajectory and mechanistic details, closing with a discussion of some implications of the dynamical effect in this complex biophysical process.

METHODS

(a) Simulation details. The simulated system consists of a twelve base pair DNA d(GCGCTCGAGCGC)2 and a protonated proflavine (constituting 784 atoms), all solvated with 11763 TIP3P[29] water molecules in a cubic box. Twenty-two sodium ions and one chloride ion were added to neutralize the negative charges on the DNA and the positive charge on the proflavine, respectively. The AMBER99/parmbsc0[30-31] force field was used for DNA and the general Amber force field (GAFF)[32] was used for proflavine. Constant particle, volume, and temperature (NVT) simulations were carried out using the Nose-Hoover thermostat[33-34] with a coupling constant of 0.4 ps to maintain the temperature at 300 K. The time step of each simulation was 2 fs. The electrostatic interaction was treated using Particle Mesh Ewald[35] with a cutoff at 10 Å, and the van der Waals cutoff was taken to be 10 Å. All the simulations were carried out with GROMACS[36].

(b) Design of the reaction coordinate. Reaction coordinates used for monitoring the intercalation simulations were the distance and angular coordinates X[37-38] and φ[7], respectively (see Fig. 1a). While X defines the position of the drug with respect to the intercalation base pairs, the collective variable φ denotes the position of the drug along the helical axis of the DNA. As discussed in Ref. [7], the coordinates X and φ together can distinguish between the reactant, product, and intermediate states of the intercalation. As we show within, X is more important near the TS, and accordingly we chose X as the reaction coordinate for estimating recrossing and calculating the TC value. Apart from these two, a few other coordinates such as DNA base step parameters (Rise, Roll, etc.) and number of water molecules around DNA and proflavine seem to be relevant for the recrossing effects for the reaction, as discussed within.

(c) Transition state (TS) determination. In principle, recrossing can be studied by starting a trajectory from the reactant. However, the free energy barrier for proflavine intercalation is estimated to be 16.9kcal/mol[7], implying a time scale of milliseconds for the process. Several such simulations for a millisecond-long process are currently beyond computational reach. Hence, we adopt the common strategy of starting a simulation from the TS, following it forward and backward in time until it forms reactant and/or product. This requires far less time (by many orders of magnitude), since falling from the TS to the reactant or product basins is a process downhill in free energy. One characterization of a TS would be that it is a point (more generally a surface) which would produce statistically equal proportion of reactants and products in unbiased simulations -- this is the prescription for a committor analysis.[39-40] Naturally, a precise and accurate TS structure needs to be first found for such calculations, now described.

In this context, we sought assistance from our previous study on the detailed free energy landscape (Fig. 1b) of proflavine interca-

lation, starting from the minor groove and following the process through to the insertion into DNA.[7] In a one dimensional perspective, the point in the minimum free energy path with the highest free energy -- shown in Fig. 1c involving one of the reaction coordinates X -- can be regarded as the first guess for the probable TS. This guess is however insufficient. Accurate calculation of the transmission coefficient requires an accurate transition state, but there are errors associated with the quoted metadynamics free energy calculation. In fact, we will show that the TS obtained from these calculations -- although structurally and free energetically close to the actual TS -- is not sufficiently accurate for a committor probability calculation. Nevertheless, it does provide an initial guess in the search for a more accurate TS.

To obtain a more precise TS, we first performed umbrella sampling simulations to calculate free energy along X around the highest free energy value obtained from the earlier metadynamics study[7]. Next, we chose a few configurations with the reaction coordinate value close to the highest value in the free energy profile obtained from umbrella sampling. We then calculated the committor probability of each of these TS candidates by carrying out ~1000 simulations for 300 ps, each with different initial velocity distributions; for these we counted whether these configurations landed in the product (intercalated state with $X \leq 1\,\mathring{A}$) or reactant (away from DNA with $X \geq 6\,\mathring{A}$) state. A true TS will result in 50% probability for both reactant and product.[41] The reactant and the product Stable States' boundaries were decided based on the distribution of the reaction coordinate values at the endpoint (300th ps) of each trajectory, as shown in SI sec. I.

We point out that the state we consider as reactant in this study is not precisely the minor-groove bound state, but is rather an intermediate separated state that is free energetically close to that state. [8] This intermediate state can be seen in Fig 1b as the minimum located at X=0.6 nm and $\varphi = 10°$. If a particular trajectory did not reach any of the Stable States during the given simulation time, it was discarded. Time reversibility was achieved by using a negative time step in the velocity Verlet algorithm[42] used to integrate the equations of the motion.

(d) Evaluation of transmission coefficient from MD study. Once the TS was identified, we created many configurations by fixing the reaction coordinate's value and allowing other degrees of freedom to equilibrate. Subsequently, we chose ~ 1000 configurations randomly from this collection and performed simulations both forward and backward in time. Since the RC was fixed, we chose the RC velocity randomly from the Maxwell-Boltzmann distribution. If the system reached either of the Stable States, they were classified as reactant or product according to the value of RC; otherwise the trajectory was discarded.

The trajectories starting from the TS were classified as RP (reactant→product) if the forward part of a trajectory (positive time) resulted in the product and the reverse part of the trajectory (negative time) ended up in the reactant basin (see SI sec. II). If both forward and backward parts of a trajectory led to the reactant basin, it was categorized as RR (reactant→reactant). Similarly, if both parts led to the product, it was labeled as PP (product→product) trajectory. The last possibility for a trajectory is to end in the product in backward propagation and reactant basin in forward propagation. These were labeled as PR (product→reactant) trajec-

tories. The RR and the PP trajectories correspond to the recrossed trajectories. Once these statistics were obtained, the transmission coefficient κ was calculated by reactive flux formalism[22, 43], which defines κ as the ratio of the actual rate to the rate predicted by TST as

$$\kappa = \frac{\sum_{i,+}^{N} p_i |v_i| Q_i}{\sum_{i,+}^{N} p_i |v_i|}, \quad (2)$$

where "+" represents trajectories with an initial positive flux, p_i is the probability to have the ith initial configuration and v is the velocity of the reaction coordinate $\dot{X}$ (calculated by forward difference method[15]). Since the value of X decreases from reactant to product in the present system, a negative RC velocity obtained from the forward difference method is considered as positive flux for our calculations. The factor Q_i is given by

$$Q_i = \begin{array}{ll} 1 & \text{if RP} \\ 0 & \text{if RR/PP} \\ -1 & \text{if PR} \end{array} \quad (3)$$

(e) Estimation of transmission coefficient (TC) from Grote-Hynes theory (GHT). As mentioned in the Introduction, GHT expresses the TC κ as the ratio of the reactive frequency ω_r to the equilibrium barrier frequency ω_{eq} [17-18]

$$\kappa_{GH} = \frac{\omega_r}{\omega_{eq}}. \quad (4)$$

The reactive frequency is the effective frequency of the motion across the TS barrier along the RC for successful reaction. It depends on both the equilibrium barrier frequency and the effective friction kernel and is obtained by solving the self-consistent equation

$$\omega_r^2 - \omega_{eq}^2 + \omega_r \int_0^\infty e^{-\omega_r t} \xi_{TS}(t) dt = 0. \quad (5)$$

$$\xi(t)_{TS} = \frac{\langle F_{RC}(0) F_{RC}(t) \rangle}{\mu_{RC} k_B T} \quad (6)$$

The time-dependent friction kernel $\xi_{TS}(t)$ measures the non-equilibrium coupling of the RC to other degrees of freedom at the TS; it is obtained from the time correlation function of the forces[44] $F_{RC}(t)$ exerted by all other degrees of freedom on the fixed RC, evaluated at the TS; μ_{RC} is the RC's reduced mass whose determination is described below, T is the temperature, and k_B is the Boltzmann constant. In order to calculate ω_{eq}, the barrier frequency when equilibrium solvation is maintained, the reaction's potential of mean force (PMF) obtained along the RC near the TS, using umbrella sampling simulations mentioned before, was fit to a parabolic function as

$$\Delta PMF(X) = -\frac{1}{2} K_{eq} \left(X - X^\# \right)^2. \quad (7)$$

Here $\Delta PMF(X)$ is the PMF's change in the RC X with respect to the maximum, located at $X^\#$. From the magnitude of the negative force constant thus obtained, we calculated the equilibrium barrier frequency (the imaginary barrier frequency's absolute magnitude) as

$$\omega_{eq} = \frac{1}{2\pi c} \sqrt{\frac{K_{eq}}{\mu_{RC}}}. \quad (8)$$

Finally, the average square of the RC velocity, v_{RC} at the TS provides the reduced mass via the equipartition theorem,

$$\frac{1}{2} \mu_{RC} \langle v_{RC, TS}^2 \rangle = \frac{1}{2} k_B T.[45]$$

RESULTS

(a) Finding the transition state. Starting from an accurate TS, the system has equal probabilities to end up as products (intercalated state) and reactants (minor groove-bound state). As the first TS guess, we chose the configuration corresponding to the meta-dynamics-generated (see Methods) free energy surface's highest point (referred as CN1 in Fig. 3 which shows the location and structure of different TS candidates) obtained from the previous metadynamics calculation (Fig. 1b). With this choice of configuration, we then calculated the committor probability. While we had expected close to 50% product formation, we found instead that out of 1000 simulations, 93% formed products.

Since this selected point obviously cannot be the actual TS, it might initially be regarded as surprising that the MFEP's highest point obtained from metadynamics deviates so strongly from the expected committor probability. The likely reason is that the non-equilibrium sampling metadynamics method – such as employed by Sasikala et al [7] -- may not sample the high energy TS region adequately unless both the reactant and product are completely sampled. The free energy surface metadynamics exploration in Ref. 7 did not require adequate sampling of the intercalation state product, resulting in an inaccurate TS estimate. Nonetheless, we can infer from the 93% product formation result that this CN1 point lies only 1.18 kcal/mol towards the product. Thus, despite the extreme difficulty -- in a system with more than 30,000 atoms -- to precisely locate the TS (surface) providing equal probability for generation of reactant and product, our prior metadynamics simulation can aid us to narrow our TS search.
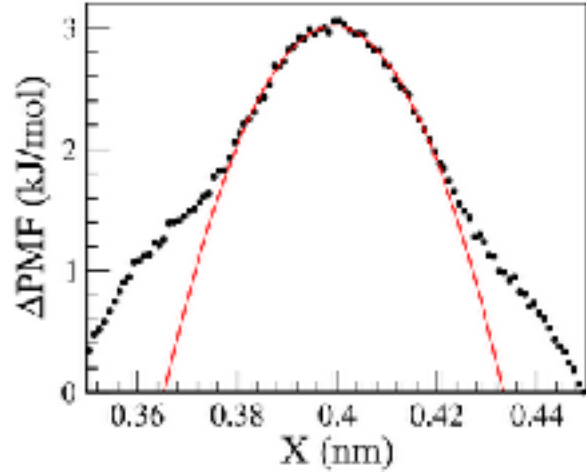

Figure 2. The black circles represent the free energy obtained from umbrella sampling simulations along the reaction coordinate X. The red line is the parabolic fit to the potential of mean force near the TS (X in the range from 0.38 nm to 0.42 nm) to obtain the equilibrium barrier frequency ω_{eq} via Eqs. 7 and 8. Rapid departure of the parabolic approximation away from the TS is apparent.

Accordingly, we recalculated the free energy profile along X in the neighborhood of the metadynamics TS. Umbrella sampling simulations[46] were performed to both locate the highest free energy point and obtain the free energy profile's curvature. This gave a free energy surface with $X=0.4$ nm as the highest point (Fig. 2). Thus, twenty more configurations (CN2-CN21), most having the value of X in the range 0.4±0.05 nm, were chosen as TS candi-

4

dates to be verified through committor probability calculation. Even though these structures have X values within the neighborhood of 0.4 nm, they differ in terms of various other parameters, as shown in Table S1 of SI sec. III. Figure 3 shows the locations of some of the representative TS candidates in the free energy surface (FES).
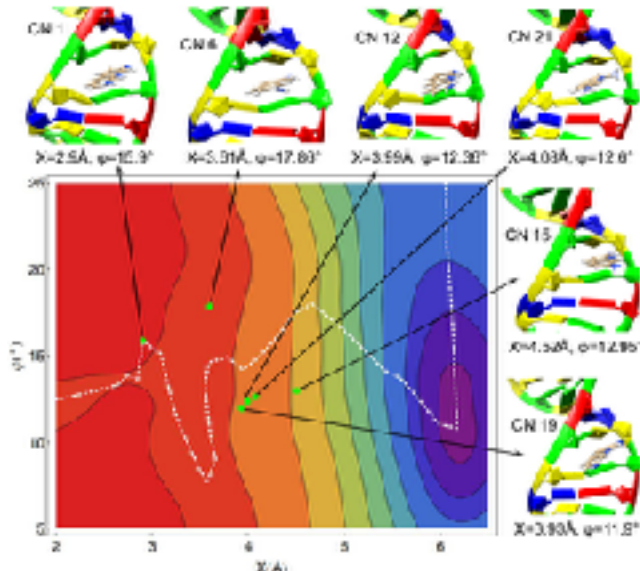

Figure 3. The free energy surface (FES) in Figure 1b is zoomed in near the transition state. The dotted line represents the minimum free energy path. The green circles in the FES represent some of the transition state candidates chosen for committor probability calculation. The reaction coordinate values are indicated below each structure. The detailed structural comparison of the selected TS and that obtained from metadynamics is given in Table S1 of SI.

For every TS candidate, we have run multiple simulations of 300 ps each to obtain a statistically converged probability of Stable product formation (the probability of Stable reactant formation is complementary). The probability of product formation is just the ratio of the trajectories that reach the product by this total number. We counted the number of trajectories reaching either the reactant or the product within 300 ps. As the number of simulations increases, the probability estimate finally converges to a particular value. We have plotted this percentage of product formation against the number of trajectories in Fig. 4 for each TS candidate. After the initial fluctuations, the probability converges to a particular value beyond ~600 trajectories. The number of simulations was extended to 1000 to test the convergence. Thus, we have performed 1000 simulations, each 300 ps long, for each of the 21 configurations, i.e., a total of 6.3 millisecond simulation to find the most probable candidate for the TS. Figure 4 shows that product formation from CN21 (X = 4.1 Å) is closest to 50% (51.6%); it is therefore taken as the most accurate TS. We point out here that although we used a single structure to determine the committor probability, an ensemble of TS configurations could be more appropriate. In fact, we have used an ensemble of TS configurations for determining the transmission coefficient, as we later discuss.
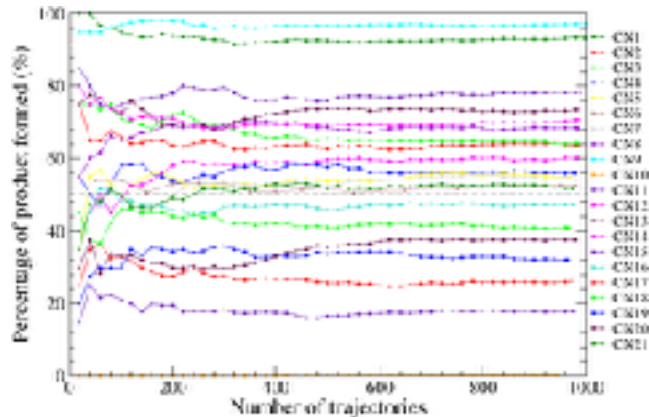

Figure 4. Percentage of Stable product formation is plotted against the number of trajectories. The CN # corresponds to the different initial configurations. The black dotted line at 50% is to guide the eye.

(b) MD-simulated reactive flux and transmission coefficient. Once CN21 was identified as the TS using committor analysis just described, we ran a MD simulation with CN21 as the initial structure to sample the TS configurations by freezing the intercalating base pairs and proflavine coordinates to both maintain the TS value $X = 0.4\ nm$ and prevent any significant variation of other DNA parameters such as Rise, Roll etc. Configurations randomly picked from this simulation were then used as the initial configurations for the 417 simulations to calculate the transmission coefficient; these were carried out forward and backward in time with different Maxwell-Boltzmann distributed initial velocities for 300 ps each. 24 trajectories were discarded since they reached neither the reactant nor the product state during the simulation time.

The transmission coefficient calculated from this data is 0.1. Figure 5a shows the same as a function of simulation time, indicating that the plateau value converges beyond 150 ps (a remarkably long time). This value -- a result far below the TST predicted value of unity -- indicates significant recrossing in proflavine's intercalation into DNA through the minor groove pathway. We next turn to the Grote-Hynes and Kramers predictions for this transmission coefficient.
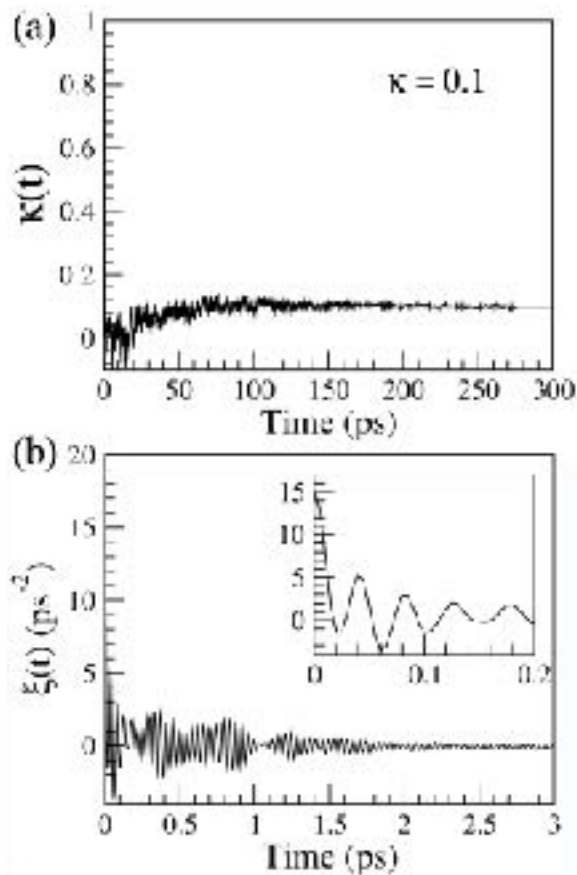
Figure 5. (a) The time-dependent transmission coefficient $\kappa(t)$ obtained from MD simulations using the reactive flux formalism. $\kappa(t)$ plateaus to the value $\kappa = 0.1$. (b) The friction kernel for the RC fixed at the transition state, with the inset showing the initial decay. The friction kernel does not contribute to the transmission coefficient beyond 3 ps, so motions occurring on timescales longer than this do not play a role in the coefficient's value.

Table 1. Various intercalation reaction values[a]

Equilibrium barrier frequency (ω_{eq})	**4.63 ps^{-1}**
Initial friction coefficient ($\xi(t=0)$)	15.19 ps^{-2}
Reactive frequency (ω_r)	4.52 ps^{-1}
Kramers' limit of reactive frequency (ω_{Kr})	4.51 ps^{-1}
Friction constant ($\int_0^\infty \xi(t)dt$)	0.25 ps^{-1}
Reactive time scale (τ_r)	0.22 ps
Environment time scale (τ_{env})	0.016 ps
MD Transmission Coefficient (κ_{MD})	0.1
Grote-Hynes TC (κ_{GH})	0.98
Kramers TC (κ_{Kr})	0.97

a – calculated via friction kernel integration up to 3 ps (see Fig. 5b)

(c) Grote-Hynes Theory. We compare our numerical result for the intercalation reaction's TC with GHT, which provides an estimate κ_{GH} as the ratio of the reactive frequency and barrier frequency (see Method for details). The barrier frequency was obtained to be 4.63 ps^{-1} via Eqs. 7 and 8 (Table 2) after fitting the PMF obtained from the umbrella sampling simulations (Fig. 2) to an inverted parabola near the TS.

To obtain the reactive frequency, we first calculated the friction kernel $\xi(t)$ in Eq. 6 after constraining the system at the transition state. $\xi(t)$, shown in Fig. 5b, is obtained as an average of ~300 simulations of 10 ps each (a method selected since multiple short simulations with different initial conditions give a better statistical average than a single long simulation)[47]; it displays a fast, initial decay followed by a much more slowly decaying small-amplitude oscillations around the mean value zero. The initial friction (zero time) value is 15.19 ps^{-2} and the total, integrated, friction constant ξ estimate is 0.25 ps^{-1}. The reactive frequency ω_r, calculated by solving the self-consistent Grote-Hynes equation Eq. 5 was found to be 4.52 ps^{-1}. Thus, the GHT estimate of the transmission coefficient κ_{GH}, calculated via Eq. 4, is 0.98; this value, which is close to the TST result of κ=1, is an order of magnitude larger than the MD result (see Table 1).

If instead the KT approximation was adopted and the $\xi(t)$'s full-time integral ξ was inserted into the GH Eq. 5, the resulting reactive frequency would be 4.51 ps^{-1}; and the Kramers TC, κ_{Kr}, is 0.97 , effectively the same result as the GHT, thus also incorrectly predicting that TST is quite accurate for the intercalation(see Table 2). A model system calculation supports this κ value estimate (SI sec. IV).

These considerable underestimations of the actual extensive recrossing indicate that there are more complexities involved in the drug intercalation system than can be captured by a frictional approach, even with the GHT time-dependent friction. We discuss these difficulties in the following section.

DISCUSSION

(a) Difficulties for GHT and KT with Reactive Flux Results for Drug Intercalation. Both Grote-Hynes and Kramers theories predict very little recrossing; the respective transmission coefficients 0.98±0.001 and 0.97±0.01, which are the same within the error bars, are extremely close to the TST value of unity. Why this conclusion follows from the theories and why they give the same prediction is detailed in SI sec. V. Here we discuss the marked discrepancy with the simulation κ result 0.1.

The key assumptions of GHT require that all recrossing events, which determine the κ value, occur on a fairly short time scale in the TS barrier top's close vicinity; in this region, the free energy is taken to be locally parabolic, and the time-dependent friction kernel is evaluated fixed at the TS (KT can be regarded as a special case of GHT in which a simple friction constant description suffices[48]). While these assumptions evidently hold in quite a wide range of reaction systems[12-13, 28, 49-50], they are very clearly violated here. A number of trajectories recross the TS after wandering rather far from the TS, as now recounted.
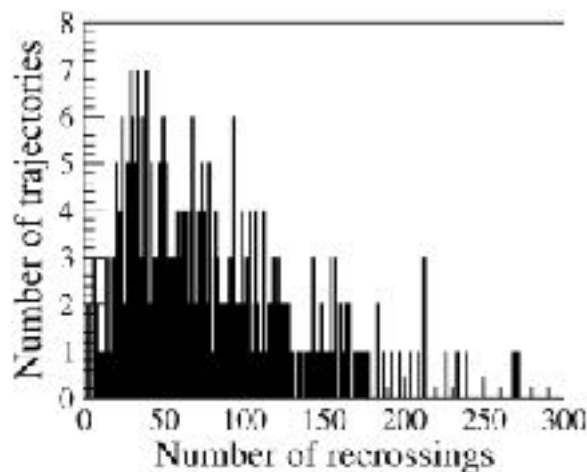
Figure 6: The histogram of proflavine intercalation process trajectories characterized by a given number of TS recrossings.

First, in Fig. 6, we characterize in more detail the TS barrier top recrossings of the proflavine intercalation process. This histogram shows the calculated number of recrossings in each trajectory, determined by observing how many times it crosses the TS reaction coordinate value, $X = 4.1$ Å. Clearly there are numerous recrossings, in contrast to their incorrect description as minimal in GHT and KT. Next, we display in Fig. 7a-c three types of trajectories (with Table 2 indicating the occurrence frequency); these are respectively a successful RP trajectory, an RR trajectory where the Stable intercalated product is not reached, and a PP trajectory, in which the Stable product is ultimately reached; in fact the trajectory originated from the P side rather than the R side. A simplified discussion on this type of classification is given in SI sec. II. All these panels and their insets (except for Fig. 7a) show considerable recrossing of the TS involving large excursions in X away from the TS before recrossing (some with considerable time delay), in clear violation of the GHT assumption. This phenomenon has been previously seen in only a few other reaction studies. [48, 51-52]
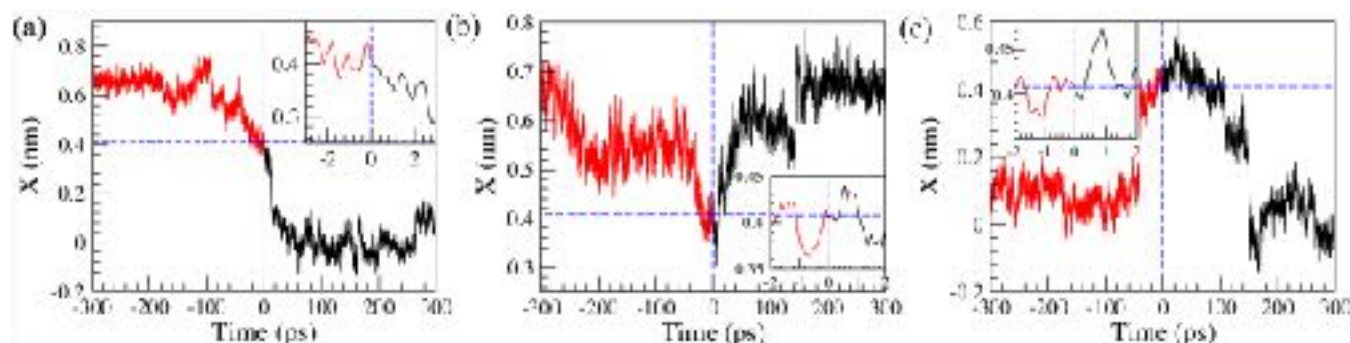

Figure 7: Three representative trajectory results following trajectories that crossed the TS from the Stable reactant R to the Stable product P side at t=0 (see also SI sec. II for trajectory classification). The vertical dashed blue line denotes time t=0 and the horizontal dashed blue line denotes the reaction coordinate X's value at t=0. Black solid line trajectories result from the forward simulation and the red solid line trajectories are their continuation simulation backwards in time. (a) A reactive RP trajectory, with a reasonably straightforward overall successful pathway from Stable reactant R to Stable product P. The inset shows recrossings near the TS, with recrossings evident prior to t=0. (b) A non-reactive trajectory of the RR variety that after the t=0 crossing of the TS moves towards the product P Stable State in the first few ps, but later recrosses to ultimately form reactant R. Multiple recrossings, including large amplitude ones, before and after t=0 are evident in the trajectory and in the short timescale inset, where large amplitude excursions are also apparent. (c) Another non-reactive trajectory, here of the PP variety, so termed because even though the Stable product P is ultimately formed after crossing the TS at t=0, the history before t=0 shows that the trajectory ultimately originated from the product P and not the Stable reactant R. Large amplitiude recrossings are evident in the trajectory and the inset. Figure 2 can be consulted concerning the parabolic free energy assumption.

Table 2. Statistics of RR, RP, PR and PP trajectories

No. of trajectories*	RP	RR	PP	PR	κ
393	87	156	82	68	0.1

*Note that there are more RR trajectories than PP trajectories; we speculate that this might be due to e.g. a potential hindrance in the product side, but in any event this requires further investigation.

We can further characterize the recrossing events by estimating the time taken for recrossing. To this end, we define and plot a histogram in Fig. 8a of $\tau_{recross}$, the time taken between consecutive recrossings. The main plot shows that most recrossings occur in a very short time, with the inset showing that there is only a small number of recrossings taking more than 20 ps. But this perspective -- while belying the GHT assumption that only very short time events are key for the transmission coefficient -- does not tell us how far the trajectory wanders from the TS before recrossing. To address this key aspect, we calculated the farthest value of X away from the TS attained by a trajectory before recrossing; its histogram is plotted in Fig 8b. It is seen that majority of the recrossing occurs near or perhaps not inordinately far from the TS. But a very significant number of trajectories recross after wandering far away from the TS region. The fundamental assumptions of GHT (and *mutatis mutandis* KT) are violated.
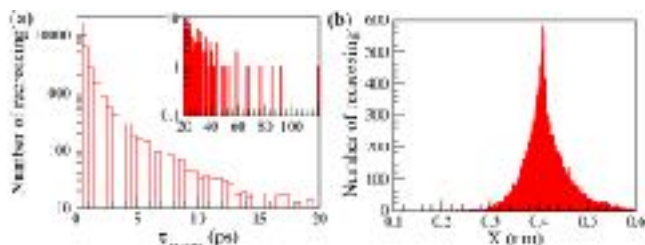
7

Figure 8. (a) A histogram of the time taken between consecutive recrossings up to 20 ps. The inset shows the corresponding results for the times larger than 20 ps. (b) The histogram of the largest magnitude value of X reached away from TS by a trajectory before recrossing.

(b) Exploration of the intercalation mechanism. Figures 6-8 and their discussion above show that the Proflavine intercalation process is a complex one with multiple recrossings -- often of large amplitude excursions -- of the selected X reaction coordinate from the TS; and it is noteworthy that this is despite the TS's selection via a committor analysis. It seems clear that a more extended molecular level view of the process will be required in constructing a proper rate description. Here we take a first step in that process by a detailed analysis of the overall intercalation process in terms of an extended set of drug, DNA and environmental variables.

Since intercalation is a millisecond long process, it is not possible to obtain the intercalation's molecular mechanism with unbiased simulations. However, our approximate TS determination, followed by the reactive flux approach of running forward and backward simulations, enables us to at least probe some important aspects of the unbiased molecular mechanism of intercalation. In an effort to shed some light on the intercalation mechanism's molecular level details, we have focused on the successful trajectories RP, and have examined the behavior of several quantities in addition to the selected reaction coordinate X (The results for all trajectory types are given in the SI sec. VI Fig. S5). These are the coordinates d, φ (see Fig. 1a and subsection (b) of Methods); the intercalating base pairs (IBP) variables Rise, Roll, Buckle, Shift, Slide, Twist[7, 53] (see SI sec. VI Fig. S4 for these motions' character) and finally several variables related to water in the IBP neighborhood: the number of hydrogen bonds between water molecules and the drug, as well as the number of waters around the drug and the IBP within 0.34 nm.

Figure 9 displays plots of the simulation averages of all these variables versus time -- both before and after the TS passage for the successful RP trajectories --; each starts from CN21 as the TS (see Fig. 3 and subsection (b) of the Results section). A global picture of aspects of the mechanism in terms of these variables is provided by the Figure's results during 300 ps preceding and 300 ps after the trajectory initiation, since within this time frame, Stable reactant and product can be attained. We now examine details of these variables' behavior.

The distance parameters X and d are similar if the angle φ between the drug and the DNA changes little, as is the case very near the TS, which we adopt as our central reference point in this discussion. They both decrease from their Stable reactant (R) value by about a third as the TS is approached, and ultimately each decreases the remaining ~2/3 fraction, to ~0.1 nm in the Stable product. The angle φ for R (~20°) and for the TS (~10°) are very similar, so this variable changes little as the TS is approached. But then it climbs to ~90° in the product (P), since after

intercalation, a small intercalated drug displacement changes φ from 0° to 180° producing an average of 90°.
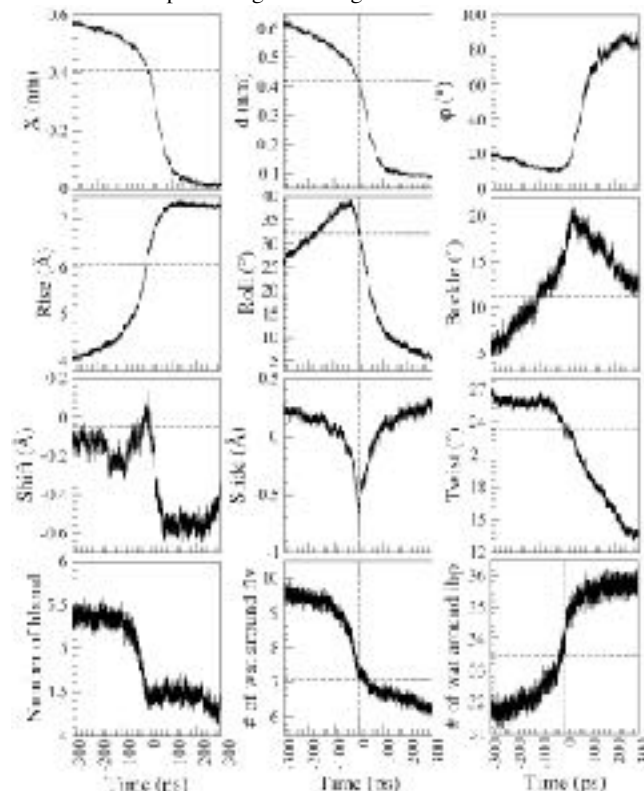

Figure 9. The average successful RP reaction trajectory values, versus time before and after the TS, of the coordinates X, d, ; the IBP variables Rise, Roll, Buckle, Shift, Slide, Twist (see SI sec. VI for these motion's character); and the water-related variables the number of hydrogen bonds between water and drug, the number of waters around the drug and the IBP within 0.34 nm. The vertical dashed lines indicate t=0. The horizontal dashed lines indicate the variable's value at t=0. The changes in some of these profiles are not monotonic in nature.

Turning to the IBP variables, most of the significant Rise increase from R occurs on the way to reaching the TS in order to allow the drug access to the intercalation region, with a smaller further subsequent increase in the intercalated P state. The Roll first increases to allow drug access, but then decreases as the TS is approached, continuing thereafter a large decrease towards P. This decrease when the proflavine completely intercalates is consistent with the previous observation by Sasikala *et al*. and others,[7, 37-38] that the Roll is high in the TS compared to a relaxed DNA. Buckle (a base variable) being similar to Roll (base pair variable) shows similar behavior.

We now turn to the behavior of the Shift, Slide and Twist motions in Fig. 9. Compared to R, the TS is characterized by high Shift and low Slide values. The Shift increases appropriately from its normal DNA value of 0 Å in R as the TS is reached; it then rapidly drops significantly to the negative intercalated P state value, indicating that the lower BPs are shifted towards the major groove more than the upper BPs. The Slide rapidly and significantly drops from its small positive value in R to a negative value in the TS and then rapidly and symmetrically returns to its P state value which is the same as that in R. This change is consistent with the intercalating drug being closer to one of the asymmetric

8

DNA strands (one $5'\,to\,3'$, the other $3'\,to\,5'$), followed by relaxation. Finally the R state is characterized by an almost normal twist angle, which only decreases (i.e., untwisting) somewhat as the TS is approached, with the major untwisting occurring after the TS is passed and intercalation is concluded; indeed untwisting is known for intercalated DNA[54-55]. All these variables have significant changes in the neighborhood of the selected TS (and beyond), and presumably require inclusion in an improved description.

Finally, we consider the water variables in the RP trajectory evolution. It is convenient to first discuss the numbers of waters around the drug in the separated R state and around the IBP. These both change just before the TS is reached, with the former decreasing and the latter increasing, by a similar, but not identical, number of water molecules; smaller changes in the same directions follow as the drug is completely intercalated. This part of the mechanism involves the number of waters around proflavine decreasing as it departs the separated R state with its large solvent exposed surface area, and approaches the intercalated state where the drug is buried in the DNA. From the IBP perspective, the number of water molecule increases, due to the larger volume in the intercalated state. Correspondingly, the number of water-proflavine hydrogen bonds remains same in the product, but increases for the reactant state. Since all these variables have a significant change in the selected TS neighborhood and beyond, their participation in an improved description seems appropriate.

Figure 9 and its description have shown that the successful RP intercalation mechanism involves many variable changes in and beyond the spatial and temporal neighborhood of the TS adopted in the present work. Our discussion above has provided some hints as to which additional variables might be included in an improved -- albeit a still reduced -- description of the reaction coordinate. An additional source of information here could be to analyze the energy flow in the proflavine-DNA-water system subsequent to and in response to the flux across the current TS in terms of the power/work formalism, which allows molecular level identification and characterization of the energy flow in complex systems.[56-58] With a key set of variables selected -- and still within the framework of GHT -- the frequency analysis of the time-dependent friction involving the forces for these variables could be used to construct the detailed, improved reaction coordinate.[12] Further studies are undergoing to understand the influence of the variables mentioned in Fig 9 in the recrossing dynamics using machine learning approach.

CONCLUSION

We have investigated, via extensive molecular dynamics simulations, the presence of recrossing in a large and complex system, the proflavine drug intercalation into DNA in aqueous solution. Using an earlier metadynamics free energy surface, we could narrow down the transition state region. Umbrella sampling simulations followed by committor probability calculations were then employed to find the most accurate transition state (TS) for the intercalation process from the minor groove side.

This TS identification enabled the investigation of the dynamical recrossing at the TS region, which reflects the departure of the rate constant from its Transition State Theory (TST) value; the significant recrossing found leads to a very low value of the transmission coefficient (TC), indicating that TST significantly overestimates the rate. In contrast to experience in many other reaction systems, Grote-Hynes theory (GHT) fails to account for this TC predicting insignificant recrossing and a TC close to unity (the same prediction and failure occur for Kramers theory). Detailed trajectory investigation shows that many recrossings occur far from the TS in both space and time, thus violating the basic assumption of the GHT.

We have also monitored the change of different variables involved in the intercalation process in reactive (RP) trajectories, which portray the unbiased mechanism of the intercalation process. We find that the various variables such as Rise, Roll, number of water molecules and hydrogen bonds involved, etc. undergo significant changes during intercalation in a concerted manner. Hence, inclusion and adequate sampling of these variables in an improved reaction coordinate description seem worthy of investigation. However, the overall mechanism is consistent with that found in our previous equilibrium study[7]. Overall, our study shows that dynamical effects in this bio-molecular system is complex in nature and require multifaceted approaches for its elucidation.

ASSOCIATED CONTENT

Supporting Information. Figure for stable state determination, schematic figure for different types of trajectories in reactive flux simulations, table with the properties of the transition state candidates, friction kernel of Lennard Jones particle under a potential, discussion of GHT and KT, and average behavior of RR, PR and PP trjaectories. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Arnab Mukherjee, Tel:+91 2590 8051;Fax:+91 20 2589 9790; Email: arnab.mukherjee@iiserpune.ac.in
*J. T. Hynes (chynes43@gmail.com)

Author Contributions

The manuscript was written through contributions of all authors.

ABBREVIATIONS

GHT, Grote Hynes Theory; GLE, Generalized Langevin Equation; IBP, Intercalating Base Pairs; KT, Kramers Theory; MD, molecular dynamics; PMF, Potential of Mean Force; RC, Reaction Coordinate; SSP, Stable State Picture; TC, Transmission Coefficient; TS, Transition State.

REFERENCES

1. Lerman, L. S. Structural considerations in the interaction of DNA and acridines. *J. Mol. Biol.* **1961,** *3*, 18-IN14.

2. Chaires, J. B. A thermodynamic signature for drug–DNA binding mode. *Arch. of Biochem. Biophys.* **2006,** *453*, 26-31.

3. Chaires, J. B.; Dattagupta, N.; Crothers, D. M. Studies on interaction of anthracycline antibiotics and deoxyribonucleic acid: equilibrium binding studies on the interaction of daunomycin with deoxyribonucleic acid. *Biochemistry* **1982,** *21*, 3933-3940.

4. Brana, M. F.; Cacho, M.; Gradillas, A.; De Pascual-Teresa, B.; Ramos, A. Intercalators as anticancer drugs. *Curr. Pharm. Des.* **2001,** *7*, 1745-1780.

5. Baruah, H.; Bierbach, U. Unusual intercalation of acridin−9−ylthiourea into the 5′−GA/TC DNA base step from the minor groove: implications for the covalent DNA adduct profile of a novel platinum–intercalator conjugate. *Nucleic Acids Res.* **2003,** *31*, 4138-4146.

6. Li, H. J.; Crothers, D. M. Relaxation studies of the proflavine-DNA complex: The kinetics of an intercalation reaction. *J. Mol. Biol.* **1969,** *39*, 461-477.

7. Sasikala, W. D.; Mukherjee, A. Molecular mechanism of direct proflavine–DNA intercalation: Evidence for drug-induced minimum base-stacking penalty pathway. *J. Phys. Chem. B* **2012,** *116*, 12208-12212.

8. Sasikala, W. D.; Mukherjee, A. Intercalation and de-intercalation pathway of proflavine through the minor and major grooves of DNA: roles of water and entropy. *Phys. Chem. Chem. Phys.* **2013,** *15*, 6446-6455.

9. Eyring, H. The activated complex in chemical reactions. *J. Chem. Phys.* **1935,** *3*, 107-115.

10. Roca, M.; Pascual-Ahuir, J.-L.; Tuñón, I. Reversibility and diffusion in mandelythiamin decarboxylation. Searching dynamical effects in decarboxylation reactions. *J. Am. Chem. Soc.* **2012,** *134*, 10509-10514.

11. Hynes, J. T. Crossing the transition state in solution. In *solvent effects and chemical reactivity*, Tapia, O.; Bertrán, J., Eds. Springer Netherlands: Dordrecht, 2002; pp 231-258.

12. Ruiz-Pernía, J. J.; Tuñón, I.; Moliner, V.; Hynes, J. T.; Roca, M. Dynamic effects on reaction rates in a michael addition catalyzed by chalcone isomerase. beyond the frozen environment approach. *J. Am. Chem. Soc.* **2008,** *130*, 7477-7488.

13. Roca, M.; Moliner, V.; Tuñón, I.; Hynes, J. T. Coupling between protein and reaction dynamics in enzymatic processes: application of Grote–Hynes theory to catechol O-methyltransferase. *J. Am. Chem. Soc.* **2006,** *128*, 6186-6193.

14. Roca, M.; Andrés, J.; Moliner, V.; Tuñón, I.; Bertrán, J. On the nature of the transition state in catechol O-methyltransferase. A complementary study based on molecular dynamics and potential energy surface explorations. *J. Am. Chem. Soc.* **2005,** *127*, 10648-10655.

15. Mullen, R. G.; Shea, J.-E.; Peters, B. Transmission coefficients, committors, and solvent coordinates in ion-pair dissociation. *J. Chem. Theor. Comput.* **2014,** *10*, 659-667.

16. Kramers, H. A. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **1940,** *7*, 284-304.

17. Grote, R. F.; Hynes, J. T. The stable states picture of chemical reactions. II. Rate constants for condensed and gas phase reaction models. *J. Chem. Phys.* **1980,** *73*, 2715-2732.

18. Grote, R. F.; Hynes, J. T. Saddle point model for atom transfer reactions in solution. *J. Chem. Phys.* **1981,** *75*, 2191-2198.

19. Northrup, S. H.; Hynes, J. T. The stable states picture of chemical reactions. I. Formulation for rate constants and initial condition effects. *J. Chem. Phys.* **1980,** *73*, 2700-2714.

20. Kanaan, N.; Roca, M.; Tuñón, I.; Martí, S.; Moliner, V. Application of Grote–Hynes theory to the reaction catalyzed by thymidylate synthase. *J. Phys. Chem. B* **2010,** *114*, 13593-13600.

21. Pecina, O.; Schmickler, W. On the dynamics of electrochemical ion-transfer reactions. *J. Electroanal. Chem.* **1998,** *450*, 303-311.

22. Gertner, B. J.; Wilson, K. R.; Hynes, J. T. Nonequilibrium solvation effects on reaction rates for model SN2 reactions in water. *J. Chem. Phys.* **1989,** *90*, 3537-3558.

23. Keirstead, W. P.; Wilson, K. R.; Hynes, J. T. Molecular dynamics of a model SN1 reaction in water. *J. Chem. Phys.* **1991,** *95*, 5256-5267.

24. Das, A. K.; Madhusoodanan, M.; Tembe, B. L. Dynamics of Na+−Cl-, Na+−Na+, and Cl−−Cl- ion pairs in dimethyl sulfoxide: friction kernels and transmission coefficients. *J. Phys. Chem. A* **1997,** *101*, 2862-2872.

25. Sese, G.; Guardia, E.; Padro, J. A. Ionic association of Na+-Cl-, Na+-Na+ and Cl--Cl- in methanol: mean force potentials and friction kernels. *J. Phys. Chem.* **1995,** *99*, 12647-12654.

26. Smith, B. B.; Hynes, J. T. Electronic friction and electron transfer rates at metallic electrodes. *J. Chem. Phys.* **1993,** *99*, 6517-6530.

27. Staib, A.; Borgis, D.; Hynes, J. T. Proton transfer in hydrogen−bonded acid–base complexes in polar solvents. *J. Chem. Phys.* **1995,** *102*, 2487-2505.

28. Castillo, R.; Roca, M.; Soriano, A.; Moliner, V.; Tuñón, I. Using Grote–Hynes theory to quantify dynamical effects on the reaction rate of enzymatic processes. The case of methyltransferases. *J. Phys. Chem. B* **2008,** *112*, 529-534.

29. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79*, 926-935.

30. Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000,** *21*, 1049-1074.

31. Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham Iii, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophys. J.* **2007,** *92*, 3817-3829.

32. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004,** *25*, 1157-1174.

33. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984,** *52*, 255-268.

34. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985,** *31*, 1695-1697.

35. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98*, 10089-10092.

36. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theor. Comput.* **2008,** *4*, 435-447.

37.   Mukherjee, A.; Lavery, R.; Bagchi, B.; Hynes, J. T. On the molecular mechanism of drug intercalation into DNA: A simulation study of the intercalation pathway, free energy, and DNA structural changes. *J. Am. Chem. Soc.* **2008,** *130*, 9747-9755.

38.   Wilhelm, M.; Mukherjee, A.; Bouvier, B.; Zakrzewska, K.; Hynes, J. T.; Lavery, R. Multistep drug intercalation: Molecular dynamics and free energy studies of the binding of daunomycin to DNA. *J. Am. Chem. Soc.* **2012,** *134*, 8588-8596.

39.   Onsager, L. Initial recombination of ions. *Phys. Rev.* **1938,** *54*, 554-557.

40.   Király, P.; Kiss, D. J.; Tóth, G. Committor of elementary reactions on multistate systems. *J. Chem. Phys.* **2018,** *148*, 134107.

41.   Du, R.; Pande, V. S.; Grosberg, A. Y.; Tanaka, T.; Shakhnovich, E. S. On the transition coordinate for protein folding. *J. Chem. Phys.* **1998,** *108*, 334-350.

42.   Martys, N. S.; Mountain, R. D. Velocity Verlet algorithm for dissipative-particle-dynamics-based models of suspensions. *Phys. Rev. E* **1999,** *59*, 3733-3736.

43.   Bergsma, J. P.; Reimers, J. R.; Wilson, K. R.; Hynes, J. T. Molecular dynamics of the A+BC reaction in rare gas solution. *J. Chem. Phys.* **1986,** *85*, 5625-5643.

44.   Peters, B. Chapter 17 - Grote-Hynes theory. In r*eaction rate theory and rare events simulations*, Peters, B., Ed. Elsevier: Amsterdam, 2017; pp 451-471.

45.   Best, R. B.; Hummer, G. Diffusive model of protein folding dynamics with Kramers turnover in rate. *Phys. Rev. Lett.* **2006,** *96*, 228104.

46.   Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977,** *23*, 187-199.

47.   Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids*. Clarendon Press: 1989; p 385.

48.   Spångberg, D.; Rey, R.; Hynes, J. T.; Hermansson, K. Rate and mechanisms for water exchange around Li+(aq) from MD simulations. *J. Phys. Chem. B* **2003,** *107*, 4470-4477.

49.   Tolokh, I. S.; White, G. W. N.; Goldman, S.; Gray, C. G. Prediction of ion channel transport from Grote—Hynes and Kramers theories. *Mol. Phys.* **2002,** *100*, 2351-2359.

50.   Roux, B.; Karplus, M. Ion transport in a gramicidin-like channel: dynamics and mobility. *J. Phys. Chem.* **1991,** *95*, 4856-4868.

51.   Rey, R.; Hynes, J. T. Hydration shell exchange kinetics: An MD study for Na+(aq). *J. Phys. Chem.* **1996,** *100*, 5611-5615.

52.   Rey, R.; Hynes, J. T. Hydration shell exchange dynamics for in water. *J. Phys. Condens. Matter* **1996,** *8*, 9411-9416.

53.   Dickerson, R. E. Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Res.* **1989,** *17*, 1797-1803.

54.   Günther, K.; Mertig, M.; Seidel, R. Mechanical and structural properties of YOYO-1 complexed DNA. *Nucleic Acids Res.* **2010,** *38*, 6526-6532.

55.   Celedon, A.; Wirtz, D.; Sun, S. Torsional mechanics of DNA are regulated by small-molecule intercalation. *J. Phys. Chem. B* **2010,** *114*, 16929-16935.

56.   Rey, R.; Ingrosso, F.; Elsaesser, T.; Hynes, J. T. Pathways for H_2O bend vibrational relaxation in liquid water. *J. Phys. Chem. A* **2009,** *113*, 8949-8962.

57.   Gertner, B. J.; Whitnell, R. M.; Wilson, K. R.; Hynes, J. T. Activation to the transition state: reactant and solvent energy flow for a model SN2 reaction in water. *J. Am. Chem. Soc.* **1991,** *113*, 74-87.

58.   Whitnell, R. M.; Wilson, K. R.; Hynes, J. T. Vibrational relaxation of a dipolar molecule in water. *J. Chem. Phys.* **1992,** *96*, 5354-5369.

11

TOC Graphic

Transition state

Stable
Reactant

Stable
Product

12
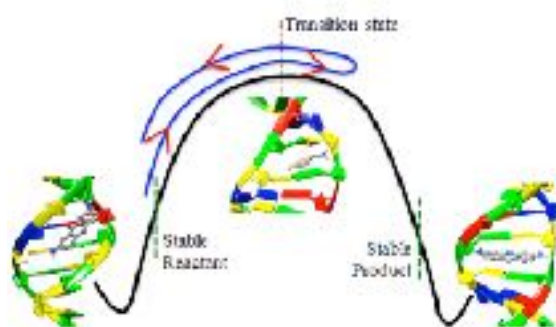
Supporting Information

Dynamical Re-crossing in the Intercalation Process of the Anticancer Agent Proflavine into DNA

V. M. Hridya[†], James. T. Hynes[§, ‡, *], and Arnab Mukherjee[†, *]

[†]Department of Chemistry, Indian Institute of Science Education and Research Pune-411008, India
[§]Department of Chemistry, University of Colorado, Boulder, CO 80309-0215, USA
[‡] PASTEUR, Department of Chemistry, École normale supérieure, PSL University, Sorbonne Université, CNRS, 75005 Paris, France

I. Stable States determination

In Fig. S1 we display the distribution of reaction coordinates at the 300th ps of the trajectories starting from the transition state (TS), used for the definition of the reactant and product Stable States[1]. At the TS, the value of the separation coordinate X is 4.08 Å and the angle φ is 12.6° (see main text Fig. 3). The reactant distribution peaks at $X \sim 5$ Å and the X value for the product always remains less than 2 Å. Since φ can take all the allowed values in the product, we did not use φ as a distinguishing parameter for the reactant and product separations. We adopted a strict criteria for the Stable States, defining the reactant Stable State R at $X > 6$ Å, and the product Stable State P at $X < 1$ Å.
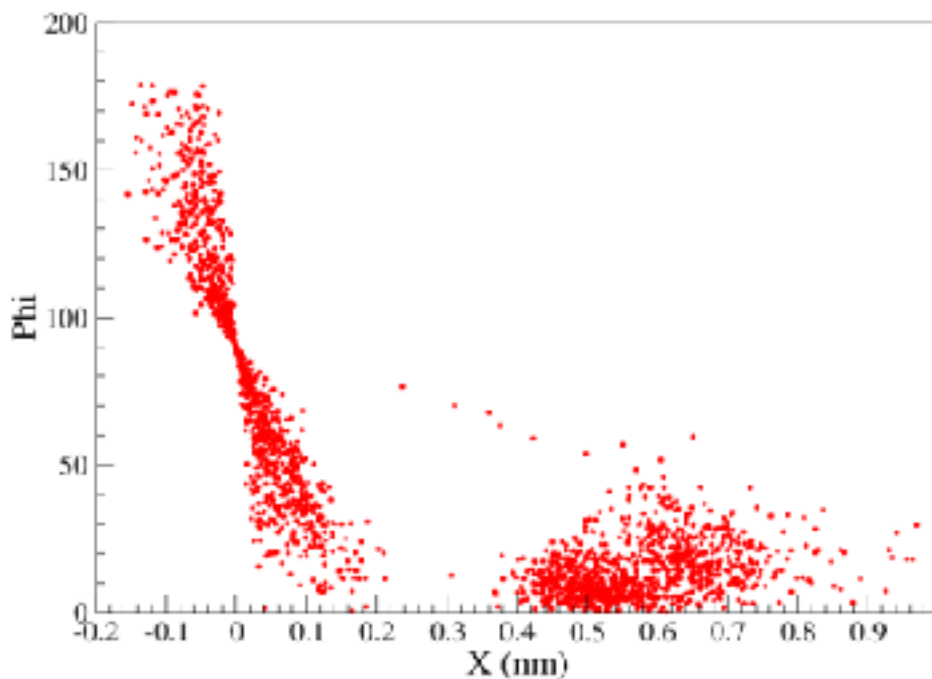

Figure S1: Distribution of the reaction coordinates, the separation coordinate X and the angle φ, at 300th ps of trajectories started from the proflavine transition state.

II. Different types of trajectories in reactive flux simulations

Here we provide in Fig. S2 schematic illustrations of the different TS recrossing patterns considered in the discussion of the proflavine intercalation process simulation results
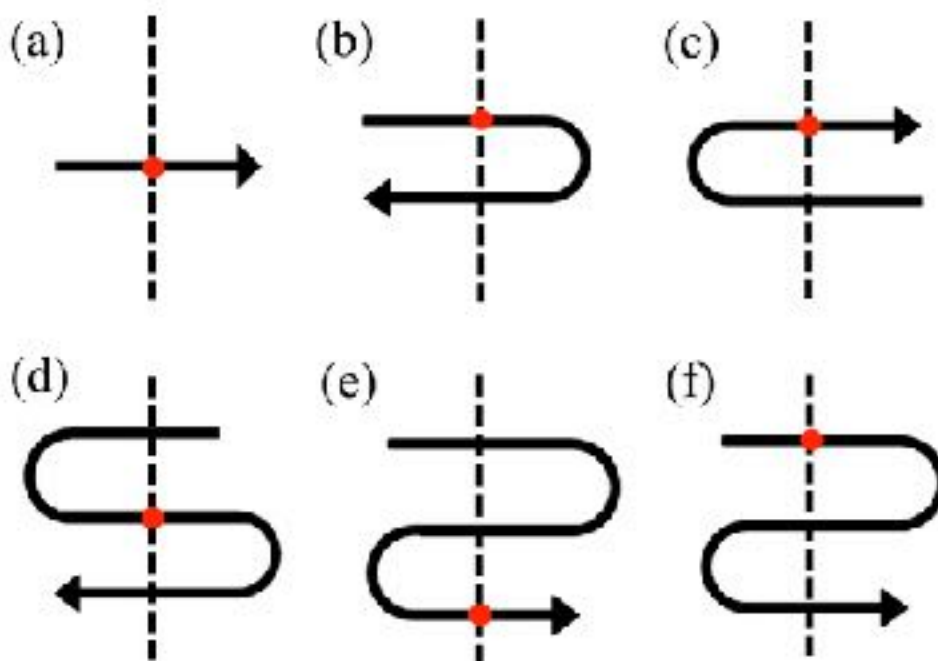

Figure S2: Different recrossing trajectory types. We consider a trajectory with positive flux at the TS at t=0; this is indicated by a red dot in the figure within the vertical dashed line indicating the TS, and represents a crossing of the TS from the stable reactant (R) side to the stable product (P) side. (a) If the number of crossings of barrier is one (before reaching a stable reactant or product region), then there is only one type of trajectory possible -- an RP trajectory. When there are two crossings of the TS barrier: (b) if one crossing occurs after the t=0 crossing, then it is a RR trajectory; (c) if the crossing occurs before the t=0 crossing, then it is a PP trajectory. If there are three crossings of the barrier: (d) one crossing occurs prior to t=0 crossing and other occurs after the t=0 crossing, then it is a PR trajectory; (e) when both the crossings occur prior to the t=0 crossing, then it is a RR trajectory; (f) when both the crossings occur after the t=0 crossing, then it is a PP trajectory. One can extend this type of scenario to multiple crossings at the TS (for a trajectory with positive flux at t=0 at the TS). Then we infer that (i) if the number of crossing is

1, then it is definitely an RP trajectory; (ii) if the number of crossings is an even number, then it is always an RR or PP trajectory; (iii) if the number of crossings is an odd number > 1, then it is always an RP or PR trajectory. This description applies irrespective of the location of R and P stable states.

III . Values of different variables for the different transition state candidates.

Table S1. Various reaction, DNA and water solvent properties for the initial guess points (CN #) employed to understand the role of these properties in determining the proflavine intercalation transition state. The Table is broken into two parts for tabulating all properties.

Property	CN1	CN2	CN3	CN4	CN5	CN6	CN7	CN8	CN9	CN10	CN11
Product (%)	**93**	63.6	64.1	55.9	55	**52.4**	54.2	68.4	97.2	0.2	78.4
X (Å)	2.9	3.7	3.6	3.5	3.5	3.6	3.7	4.0	3.0	5.00	4.1
φ (degree)	15.9	9.0	10.5	13.2	17.2	17.9	17.3	12.4	12.8	12.2	12.5
Rise (Å)	7.3	6.4	6.5	6.6	6.7	6.8	6.2	6.3	6.5	4.1	6.2
Roll (degree)	33.3	31.2	33.2	35.1	38.8	29.1	27.2	33.0	29.2	40.3	28.9
Shift (Å)	-0.4	0.0	0.1	0.1	0.0	0.1	-0.1	-0.1	0.1	-2.3	-0.5
Slide (Å)	-1.0	-0.4	-0.4	-0.7	-0.2	-0.7	-1.1	-1.2	-0.7	1.9	-0.7
Buckle (degree)	34.3	13.1	19.3	24.5	29.8	24.4	17.5	14.8	7.3	27	12.9
Twist (degree)	19.8	23.8	23.7	24.6	18.7	23.3	22.9	20.7	15.8	23.1	27.9
Water around flv	5	6	7	7	6	5	6	6	3	13	6

Water around ibp	31	35	37	39	40	36	37	35	34	34	40
Hbond	3	4	3	3	3	3	3	3	2	5	2

Table S1 continued.

Property	CN12	CN13	CN14	CN15	CN16	CN17	CN18	CN19	CN20	CN21
Product (%)	60	73.4	70.7	17.9	47.3	26.2	41.0	32.1	37.6	**51.6**
x (Å)	4.0	4.1	4.0	4.5	4.3	4.2	4.0	4.0	4.0	**4.1**
φ (degree)	12.4	12.6	12.4	12.9	12.1	12.4	12.5	11.9	12.2	**12.6**
Rise (Å)	6.2	6.3	6.1	5.9	6.1	6.1	5.9	6.1	6.1	**6.1**
Roll (degree)	28.2	30.5	26.0	29.0	26.3	31.5	28.8	29.6	26.6	**30.6**
Shift (Å)	0.0	0.5	0.1	-0.2	-0.4	-0.1	-0.0	-0.7	-0.3	**-0.1**
Slide (Å)	-1	-0.8	-0.8	0.3	-0.7	-0.2	-0.3	-0.5	0.2	**-0.8**
Buckle (degree)	22.8	18.0	11.4	15.5	1.2	17.9	14.9	17.7	19.8	**13.3**
Twist (degree)	26.1	19.1	23.8	23.4	22.7	25.8	27.7	27.4	21.3	**23.3**
Water around flv	5	6	7	5	6	6	6	11	6	**8**

Water around ibp	31	37	37	33	38	35	35	33	33	**33**
Hbond	3	3	4	3	3	3	3	5	4	**6**

IV. Friction kernel of Lennard Jones particle in a double well potential

We have used a model system to assess our proflavine intercalation system friction kernel calculation in the main text and to place its impact in perspective. We have taken a Lennard-Jones particle in water to estimate the friction kernel; the particle is assigned the mass of the RC used in the text and moves under an external potential that has similar curvature 4.6 ps^{-1}. MD simulation of the fixed force time correlation function (see Eq. 6 of the text) produced a friction kernel (see Fig. S3) characterized as a weak friction similar to the proflavine intercalation result in Fig. 5b in the manuscript (but without an oscillatory behavior); calculation of the associated transmission coefficients in the GHT and KT perspective yields $\kappa_{GH} = 0.90$ and $_{Kr} = 0.88$; these values are similar to those in the main text.
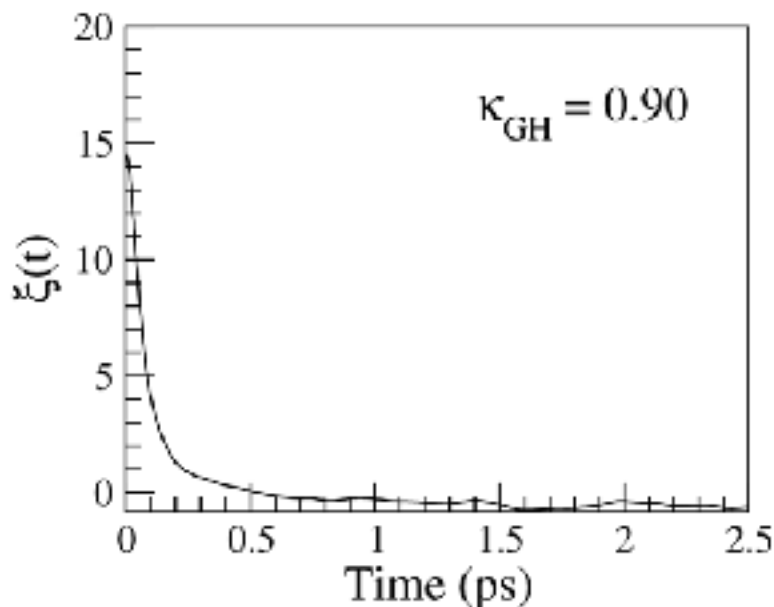
Figure S3. Friction kernel along the RC (x-coordinate) of the Lennard Jones particle in water system experiencing an external double well potential along its x-coordinate. The particle experiences weak friction with a short memory.

V. Discussion of Grote-Hynes and Kramers Theories.

As presented in the text, the Grote-Hynes and Kramers theories (GHT and KT respectively) predict proflavine intercalation simulation transmission coefficient TC values close to unity, i.e. very little recrossing at the TS reaction barrier, which indicates that transition state theory (TST) is essentially perfect for the intercalation reaction. Here we indicate why this conclusion follows from the theories and why they give the same prediction.

The predicted TC for these theories will be close to unity if the effective friction inducing the recrossing of the TS is small compared to the equilibrium barrier frequency ω_{eq}. In GHT, this ratio reflects the power of the effective friction to act on the reactive time scale, $\tau_r \approx \omega_{eq}^{-1}$. On this time scale, the reacting system is within $\sim k_B T$ of the barrier top; in this transition state region, the effective friction forces can induce recrossing of the barrier before the free energetic forces are sufficiently great to safely drive the system to the product region. In GHT, this effective friction is in effect a short-time friction which acts on this time scale; in contrast, in KT, this time scale feature is ignored and the full long time scale friction constant (time integral formula) is employed.[2] Typically, the equilibrium barrier frequency is sufficiently high – and the reaction time scale is sufficiently short – such that these two types of friction (and the resulting TCs) can differ significantly[3-5], but in the present case, they are almost identical. This can be seen from main text Fig. 5c and Table 2. The reaction time scale from GHT is $\tau_r \approx \omega_{eq}^{-1} = 0.22$ ps; this is an order of magnitude longer than the environmental time scale, (the effective time scale for the environmental friction function in Fig. 5c to vanish), so that the full friction constant value is established; modulo the small amplitude, though slowly decaying, oscillations which have little impact, this time is ~ 0.02 ps. Therefore the GHT effective friction and the KT friction

constant for the intercalation reaction are the same; 'memory' effects are negligible. But the friction is very weak -- the key ratio of the effective friction to the barrier frequency is very small, ~ 0.06, and the predicted neglible frictional recrossing is the same in both theories. As discussed in text, the assumptions underlying GHT (and *mutatis mutandis* KT) of recrossing events limited in space and time to the TS region is strongly violated for the proflavine intercalation problem, and the GHT and KT predictions are far from the simulation result (TC = κ = 0.1).

VI. DNA base pair parameters and the average trajectory behavior of assorted Coordinate, DNA and water variables for RR, RP, PR and PP trajectories
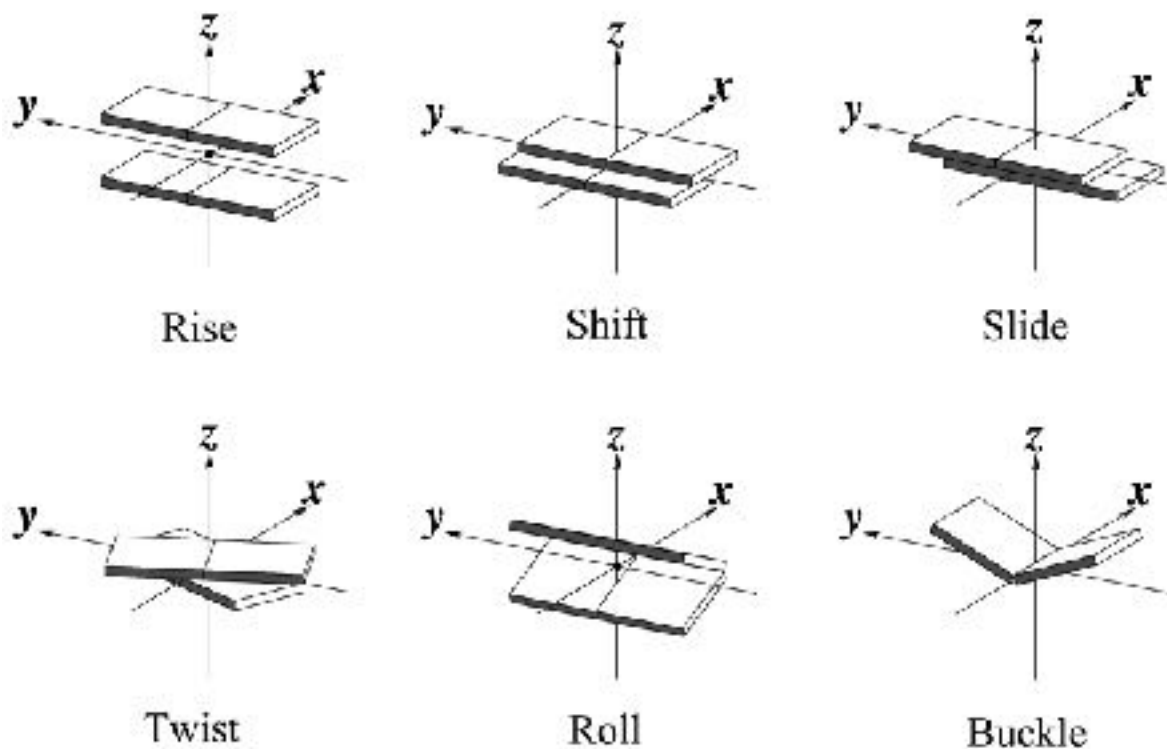
Figure S4. The different translational base pair parameters: Rise, Shift and Slide and the rotational base pair parameters: Twist and Roll. Buckle is a rotational base step parameter. The figure[6] is adapted from Ref. 6. We have studied variation of these parameters in addition to other variables for different types of trajectories as shown below.

Figure S5 complements the RP trajectory results of Fig. 9 of the main text by providing the full complement of trajectory results for the three remaining reaction outcome possibilities.
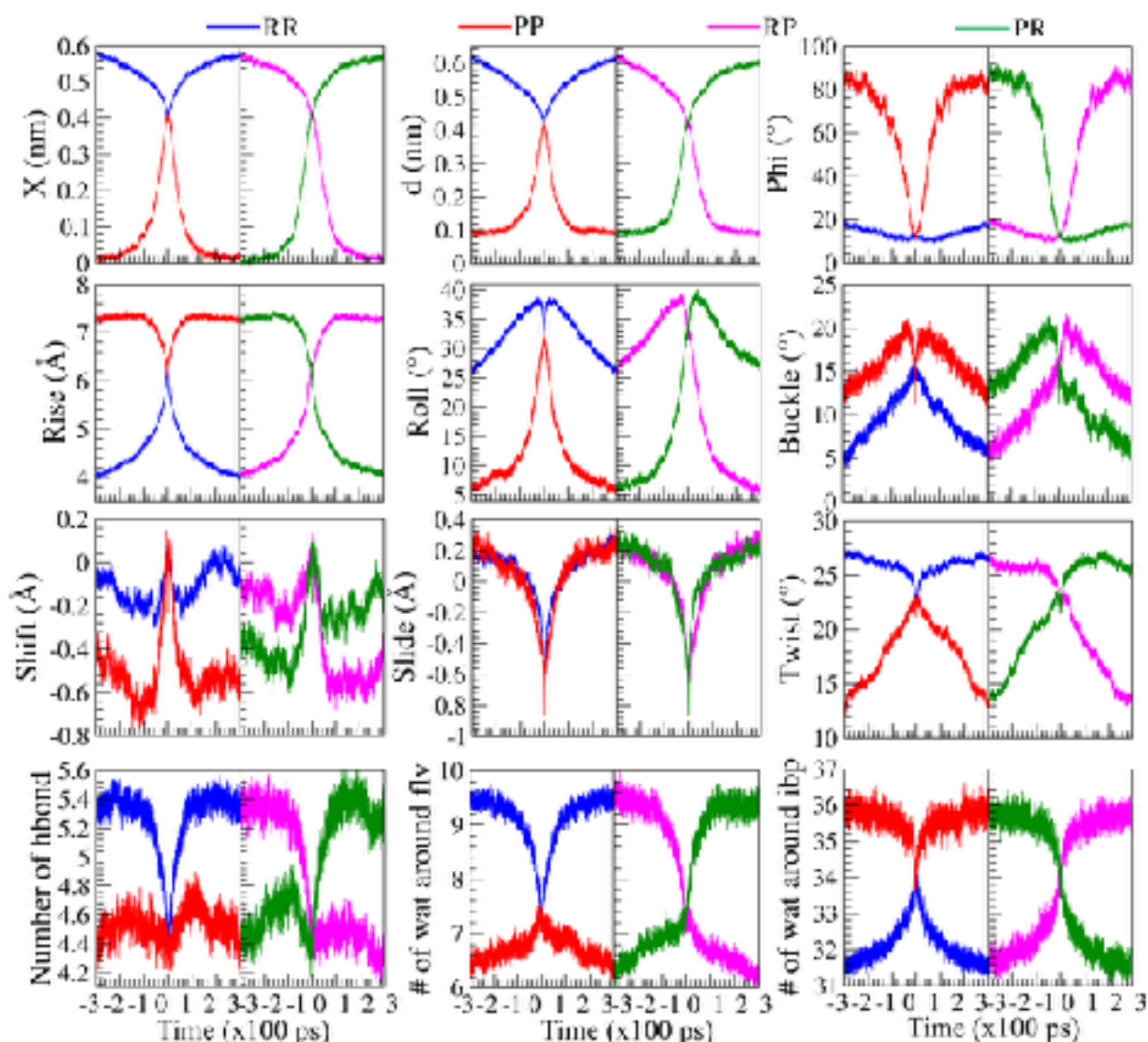
Figure S5. The average trajectory of different variables along time is plotted for the RR, RP, PR and PP trajectories. The variables plotted are *X*, *d*, φ; Rise, Roll, Buckle, Shift, Slide, Twist of the IBP; the number of hydrogen bonds between water molecules and proflavine, the number of water around the drug and around the IBP within 0.34 nm.

References

1.   Northrup, S. H.; Hynes, J. T., The stable states picture of chemical reactions. I. Formulation for rate constants and initial condition effects. *J. Chem. Phys.* **1980,** *73* (6), 2700-2714.

2.   Hynes, J. T., Crossing the Transition State in Solution. In *Solvent Effects and Chemical Reactivity*, Tapia, O.; Bertrán, J., Eds. Springer Netherlands: Dordrecht, 2002; pp 231-258.

3.   Roca, M.; Moliner, V.; Tuñón, I.; Hynes, J. T., Coupling between Protein and Reaction Dynamics in Enzymatic Processes: Application of Grote−Hynes Theory to Catechol O-Methyltransferase. *J. Am. Chem. Soc.* **2006,** *128* (18), 6186-6193.

4.   Pecina, O.; Schmickler, W., On the dynamics of electrochemical ion-transfer reactions. *J. Electroanal. Chem.* **1998,** *450* (2), 303-311.

5.   Gertner, B. J.; Wilson, K. R.; Hynes, J. T., Nonequilibrium solvation effects on reaction rates for model S_N2 reactions in water. *J. Chem. Phys.* **1989,** *90* (7), 3537-3558.

6.   Lu, X. J.; Olson, W. K., 3DNA: a software package for the analysis, rebuilding and visualization of three−dimensional nucleic acid structures. *Nucleic Acids Res.* **2003,** *31* (17), 5108-5121.