DYNAMICAL RECROSSING, INTERNAL FRICTION AND MEMORY EFFECTS: INVESTIGATING MODEL SYSTEMS AND DRUG-DNA INTERCALATION PROCESS

A THESIS

SUBMITTED IN PARTIAL FULFILMENT OF THE

REQUIREMENTS

OF THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

HRIDYA V M

20122028

DEPARTMENT OF CHEMISTRY

INDIAN INSTITUTE OF SCIENCE EDUCATION AND

RESEARCH, PUNE – 411 008

Dedicated to

My Teachers, Family and Friends

भारतीय बिज्ञान शिक्षा एबं अनुसंधान सँस्थान, पुणे
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH (IISER), PUNE
(An Autonomous Institution, Ministry of Human Resource Development, Govt. of India)
Dr Homi Bhabha Rd, Ward No. 8, NCL Colony, Pashan, Pune, Maharashtra 411008, India.

CERTIFICATE

Certified that, the work incorporated in the thesis entitled, "**Dynamical Recrossing, Internal Friction and Memory Effects: Investigating Model Systems and Drug-DNA Intercalation Process**" submitted by ***Hridya V M*** was carried out by the candidate, under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or institution.

Date: 30th August 2019
Pune (MH), India

Dr. Arnab Mukherjee
(Thesis Supervisor)

DECLARATION

I declare that this written submission represents my ideas in my own words and where others' ideas have been included; I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Date: 30th August 2019 **Hridya V M**

Pune (MH), India **Reg: No: 20122028**

ACKNOWLEDGMENTS

I take this opportunity to mention the people who deserve my deepest gratitude and without whose support this thesis would have not been possible. First and foremost, I sincerely thank my mentor, Dr. Arnab Mukherjee for his invaluable guidance and continuous support during the course of my PhD. He has given me enough freedom, and his untiring optimism has fuelled this thesis work. His enthusiasm to tackle any research problem is quite uplifting. All the long scientific discussions and the systematic analysis of papers and concepts taught during group meetings have molded us to form better researchers. His motivational talks and story-telling sessions would be fondly remembered.

I owe a debt of gratitude to the RAC members Dr. Anirban Hazra and Dr. Srabanti Chaudhury. The questions asked and the constructive feedbacks and motivation given during the RAC meetings have immensely helped my research.

I would like to express my gratitude to my collaborator Prof. Danith Ly who gave me an opportunity to explore a different field in addition to my thesis work. His life story and his pursuit to solve far-reaching research problems have been inspiring. I would also like to acknowledge the opportunity to work with Prof. James T. Hynes though it has been limited to e-mail communications. His meticulous and critical examination of research work has taught me a few lessons.

I would like to thank my labmates Wilbee, Mandar, Reman, Debasis, Abhijit, Rahul, Amal, Fatehjeet, Amitosh, and Atul for the stimulating discussions and the lighter moments we had shared in the lab. They all have been very co-operative and extremely helpful. I would like to especially thank Wilbee for giving me the courage to take up this field as my research career and patiently teaching me the basics of computational chemistry. Mandar bhaiya's infectious laugh, Reman's helping nature, Amitosh's and Atul's encouragements will not be forgotten. Though I spent little time with Amal, Fatehjeet and Rahul, their support was indelible.

I would like to mention here my friends Anjusha, Prachi, Ajith, Giri, Dibyata, Krithika and Sumit who have made my life in IISER the most enjoyable. The exhilarating trips that we went together, the entertaining long dinner table chats, the fun games and quizzes we played in weekends will all be deeply missed. You all, especially Anjusha, have

been my emotional support in IISER and thank you all for listening, offering me advice and being there when I needed you.

I would like to thank my iPhD batchmates Meghna, Sneha, Jerrin, Aditi, Abhishek, Amogh and Rahi. The time we spent together in the classes and outside; and the joy we shared after each exam and presentation are delightful memories.

A special mention to Minal and family who have always welcomed me like kin, Gregor with whom I often had very interesting conversations (especially for the career advices), Bhagya, and my computational chemistry friends.

I am grateful to the IT department especially Neeta and Suresh who always solved our computer issues and the Admin and Accounts staff especially Tushar, Anjali, Sayalee, Prabhas and Mayuresh who made paperwork less cumbersome.

A very special gratitude goes to IISER for the fellowship and the admirable facilities provided for a seamless research experience.

My heart-felt thanks to the people who mean the world to me, my parents and my sister for encouraging me in all my pursuits. I would like to thank my relatives, especially my uncle, and friends for their thoughts, well-wishes, phone-calls and visits. I would also like to expand my gratitude to those who have directly or indirectly helped me in this journey.

List of Publications

1) **Hridya, V. M**.; Mukherjee, A. Probing the Viscosity Dependence of Rate: Internal Friction or the Lack of Friction? ***J. Phys. Chem. B*** **2018,** *122* (39), 9081-9086 (Discussed in Chapter 3)
2) **Hridya, V. M**.; Mukherjee, A. Separating memory effects from internal friction in viscosity dependence of rate (Manuscript is under preparation) (Discussed in Chapter 4)
3) **Hridya, V. M**.; Hynes, J. T.; Mukherjee, A. Dynamical Re-crossing in the Intercalation Process of the Anticancer Agent Proflavine into DNA ***J. Phys. Chem. B*** **2019,** *123* (51), 10904-10914 (Discussed in Chapter 5)
4) Thadke, S. A.; Perera, J. D. R.; **Hridya, V. M**.; Bhatt, K; Shaikh, A. Y.; Hsieh, W; Chen, M; Gayathri, C; Gil, R. R.; Rule, G. S.; Mukherjee, A; Thornton, C. A.; Ly, D. H. Design of Bivalent Nucleic Acid Ligands for Recognition of RNA-Repeated Expansion Associated with Huntington's Diseases ***Biochemistry*** **2018,** *57* (14) 2094-2108 (Discussed in Appendix)
5) Thadke, S. A.; **Hridya, V. M**.; Perera, J. D. R.; Gil, R. R.; Mukherjee, A.; Ly, D. H. Shape Selective Bifacial Recognition of Double Helical DNA. ***Communications Chemistry*** **2018,** *1* (1), 79 (Discussed in Appendix)

CONTENTS

Synopsis

Title: Dynamical Recrossing, Internal Friction and Memory Effects: Investigating Model Systems and Drug-DNA Intercalation Process

Transition state theory is widely used to estimate reaction rates in condensed phase systems. The theory assumes that all the reactants that achieve the activation energy will definitely form products. However, this is not always true. The system, after crossing the transition state surface towards product, may recross and fall back into the reactant valley reducing the overall rate. Thus, transition state theory always gives an upper bound to the rate. First attempt to include recrossing effects in the reaction rate was given by Kramers in 1940. Kramers illustrated that in the spatial diffusion regime, the correction factor to the transition state theory due to recrossing is given by, $\kappa = \omega_b/\gamma$, where ω_b is the curvature of the transition state surface and γ is the friction felt by the system. Though Kramers theory was able to describe several reactions, for some others it was found that the rate did not exhibit an inverse dependence on the viscosity (the macroscopic equivalent of the friction). Furthermore, addition of a frictional term to the solvent viscosity was found to explain the rate dependence on the viscosity. Consequently, this additional frictional term was coined internal friction since it was thought to arise from the system itself. Several experimental and computational studies thereafter attributed the deviation from the Kramers' inverse viscosity dependence of rate to the presence of internal friction. The origin of internal friction in proteins has been ascribed to ruggedness on the energy surface, dihedral rotations, hydrogen bonds, non-native salt bridges etc. In this thesis, we show that deviation from inverse viscosity dependence of rate *does not necessarily imply* the presence of internal friction. We demonstrate using a simple model system, a Lennard Jones particle in water, that memory effects arising from the solvent motions can result in a non-linear dependence of rate on

viscosity which can be accounted by Grote-Hynes theory (GHT). Further, this study was extended to a model system and a simple realistic system to understand the interplay between memory friction and internal friction.

To further our understanding on the dynamical effects of rate on a complex biomolecular process, we chose intercalation of drug into DNA due to its potential to obstruct the growth of cancerous cells. Quite a few studies explored the molecular mechanism of the process using the static free energy landscape. In this thesis, we investigate the dynamical effects in the intercalation process of an anticancer agent proflavine into DNA by estimating transmission coefficient that gives a measure of the extent of recrossing occurring at the transition state surface. For that, we used both reactive flux approach and compared the result with both Kramers theory and GHT. We found significant recrossing in the process and showed that both the theories cannot capture the true recrossing event due to the breakdown of local parabolic approximation assumed in the above theories.

Chapter 1: Introduction. This chapter gives an overview of the different reaction rate theories, namely, transition state theory, Kramers theory and Grote-Hynes theory. We discuss the origin of these theories and examine their limitations. Further, we analyze the use of these theories in literature to understand the rate of reaction process in solvent medium. We have also discussed in detail the dynamical effect in barrier crossing process, recrossing, and internal friction -- the focus of the thesis.

Chapter 2: Methodology. In this chapter, we discuss the general molecular dynamics method employed to carry out all the studies in this thesis. In addition, we discuss umbrella sampling, an accelerated sampling technique, used to sample high energy conformations. We then explain the reactive flux formalism, the method used to determine transmission coefficient in numerical simulations. We also discuss the use of Kramers theory and Grote-Hynes theory to estimate recrossing in the system. Finally, we describe the method applied to

vary the viscosity of the solvent in simulations without affecting the energy surface of the reaction.

Chapter 3: Probing viscosity dependence of rate: Internal friction or Lack of friction? In this chapter we calculated recrossing to capture the dynamical aspects of the rate in a simple model and thereby investigated the origin of a highly debated topic of internal friction. Deviation from the Kramers's inverse viscosity dependence of rate, $k \propto 1/\eta$, is often attributed to the presence of internal friction in proteins after Ansari *et al.* in 1992 showed that the folding rate could fit the equation, $k \propto 1/(\eta + \sigma)$, where σ is considered as internal friction. In this study, we have used a simple model system (one LJ particle under external potential) to show that memory effects can result in a non-zero value of σ in molecular dynamics with explicit solvent which may be misinterpreted as the presence of internal friction. We show that the same system follow Kramers behavior when simulated using Langevin dynamics confirming that the deviation is due to memory effects arising from the solvent motions. We thus demonstrate that the current methods employed to estimate internal friction in different systems cannot distinguish between the memory effects and internal friction. Additionally, we have systematically investigated the effect of barrier curvature, barrier height of the potential, size of the LJ solute, and solute-solvent coupling on σ. This study demonstrated that it is the relative deviation of rate from Kramers prediction at different viscosities that determines the value of σ.

Chapter 4: Separating memory effects from internal friction in viscosity dependence of rate. Here we have extended the previous study to investigate internal friction in a model system, a diatomic system, interacting under an external potential and a real system, dihedral rotation in a butane molecule. Given that these are systems with more than one atom, the system may have internal friction. At the same time, it may also exhibit memory effects. Here, we decouple both the effects and realize in the process that *neither of the systems*

exhibit internal friction. We demonstrate by adding an external potential which mimics internal friction that the presence of internal friction always decreases the rate while also resulting in deviation from Kramers theory. This study therefore portrays two competing physical phenomena giving rise to similar behavior of rate dependence on viscosity.

Chapter 5: Dynamical recrossing in the intercalation process of the anticancer agent proflavine into DNA. Intercalation is a specific type of molecular recognition where the drug inserts itself between the base pairs of a DNA strand. The dynamical aspects of intercalation process are far less clear than its current thermodynamical understanding. Here we explore the dynamical recrossing in the intercalation of proflavine into the DNA through the minor groove. We first identity the transition state to estimate transmission coefficient and using reactive flux formalism we show that system undergoes significant recrossing at the transition state surface. We then compare the numerical results with Grote-Hynes theory and find that the theory is inadequate to address recrossing in such a complex system. We also discuss the possible reasons for the discrepancy in the results obtained from numerical simulations and the Grote-Hynes theory. This chapter was the first step towards our future goal of estimating internal friction in intercalation process by calculating transmission coefficient at different viscosities.

Addendum: The addendum furnishes the studies on topics not directly related to recrossing and internal friction. It is related to a targeted therapy for neurodegenerative diseases such as myotonic dystrophy, Huntington's disease, amyotrophic lateral sclerosis (ALC) etc, which occur as a result of unstable repeat expansion in the nucleic acid. A challenging solution is to design molecules that would bind such repeat targets with high affinity, specificity, and selectivity. To this purpose, a set of peptide nucleic acid (PNA) capable of forming bifacial H-bonding interactions with nucleobases on the repeat nucleic acids were designed by our collaborator Dr. Danith Ly from Carnegie Mellon University. We have successfully

demonstrated the proof of concept followed by illustration of application on CAG repeats in RNA which cause Huntington's disease using molecular dynamics simulations. Initially, the triple helix bound structure of the PNA to the RNA was created since no crystal structure was available. The force field for the PNA ligands was then generated and MD simulations were performed to monitor the stability of the PNA-RNA bound complex. Both complexes exhibited stability even after 500 ns of simulation thus establishing this method to a potential mode to target RNA-repeat expansion diseases. This corroborated with the experimental finding that ligands are able to bind the repeats in a sequence specific manner.

Chapter 1

Introduction

1.1 Rate of Reaction

Reactions occur around us all the time in gas phase, in variety of solutions, at interfaces etc. Some of these reactions occur in femtoseconds whereas some others take years. It is imperative to study the rate of a reaction since it also helps us understand the mechanism of the reaction. It is now known that rate processes are rare events and the thermal fluctuations in the system leads to an instantaneous flow of energy to the reactant facilitating the barrier crossing process. Rate is influenced by different factors like temperature, pressure, concentration and nature of the reactants, the solvent, and the presence of catalyst.

There are quite a few experimental tools to measure the rate of the reaction. The technique employed to measure the rate is dependent on the timescale of the reaction. For processes ranging from microseconds to seconds, generally temperature jump, pressure jump, Nuclear Magnetic Resonance, flash photolysis, pulse radiolysis etc. are used. If the timescale associated with the reaction is even smaller, then acoustical methods, electrical pulse, fluorescence decay etc. are more convenient methods to estimate rate. Theoretical advances in the calculation of rate as we know today began with the findings of van't Hoff and Arrhenius in the nineteenth century. Since then the reaction rate theory has been fostered by developments in other fields like chemical kinetics, diffusion theories, etc. In this chapter, I will briefly discuss a few landmark theories in rate calculation.

1.2 Reactions in solution

Given that many chemical and biological processes take place in aqueous environment, it is important to understand the effect of solvent on the rate constants. There are three steps involved in a solution phase reaction: (i) the encounter of the reactant molecules, (ii) the chemical transformation converting the reactants to products, (iii) the diffusion of products away from each other. When the barrier is significant, the second step is the rate determining step. Nonetheless, solvent has several effects on the dynamics of the reaction. Solvent viscosity affects the diffusion of the molecules in the solution. Further, solvation free energies of the reactant and the product can contribute to the free energy of the overall reaction significantly. Crucially, collisions with the solvent molecules regulate the energy flow to and from the solute degrees of freedom. Hence, the rate of a reaction in the condensed phase will be much different than that in the gas phase. The first noteworthy rate theory developed that is applicable for reactions in the condensed phase is the transition state theory[1].

1.2.1 Transition State Theory

Arrhenius, known as the father of the rate theory, in 1889 investigated temperature dependence of the rate of inversion of sugar in presence of acid.[2] He proposed that the concentration of the activated complex is exponentially dependent on the temperature ($k \propto e^{-A/RT}$). However, the pre-exponential factor for this familiar exponential rate expression was undetermined. The prefactor was later identified in the Transition state theory (TST) which was simultaneously reported by Henry Eyring[1] and by M. G Evans and M Polanyi[3] in 1935. TST predicts that the rate depends on the probability to find the system at the transition state region and the frequency by which the region is crossed. Thus TST gives the rate as $k^{TST} = \frac{k_B T}{h} e^{-\Delta G^{\#}/RT}$, where $k_B T/h$ is the frequency factor and $\Delta G^{\#}$ is the activation free energy. TST has been valuable in interpreting a wide range of reactions, including

complicated ones, in a qualitative way; but it often fails to find the absolute reaction rates owing to two reasons: (i) calculation of absolute reaction rate requires accurate knowledge of the potential energy surfaces (ii) chemical systems often violate the assumptions made in TST.

Transition state theory assumes the existence of a dividing surface separating the reactants and products, the crossing of which is followed by thermalization of the system in the product valley before crossing the surface again. This assumption that the trajectories originating from the reactant valley after crossing the dividing surface do not recross it before getting stabilized in the product region is the no recrossing assumption. TST also makes several other assumptions. It assumes that (i) the reactant's energy follow a Boltzmann distribution, (ii) the concentration of the transition state complex can be calculated using equilibrium theory even when the reaction is not completely in equilibrium, (iii) the motion of the reaction coordinate can be separated from other degrees of freedom at the barrier top, and (iv) the passage across the barrier is entirely a classical motion, ignoring any quantum effects. However, the no-recrossing assumption adopted in TST is often violated in classical reactive systems.

1.2.2 Dynamical recrossing in rates

Recrossing occurs in chemical systems mainly due to following two reasons: (i) if the forces due to bath degrees of freedom are large enough to cause change in the direction of the motion of the reaction coordinate, then the system can recross back to reactant valley, and (ii) the specific shape of the potential energy surface at the barrier top – e.g. when the surface is extremely flat or if there is a small minimum -- can cause the system to undergo several motions before escaping from the basin. This may result in the loss of memory of its initial momentum and then the trajectory has an equal probability to reach either reactant or product basin. A schematic figure of recrossed trajectory is shown in Fig.1. The presence of

recrossing in the system leads to an overestimation of the rate by TST. [4-9] Hence, TST prediction can be considered as the upper bound to the rate. The exact rate can be found by incorporating a correction factor called transmission coefficient κ, which accounts for the recrossing effects, $k = \kappa k^{TST}$. In the absence of recrossing, $\kappa = 1$ and TST is exact. In the presence of recrossing, κ falls below unity.

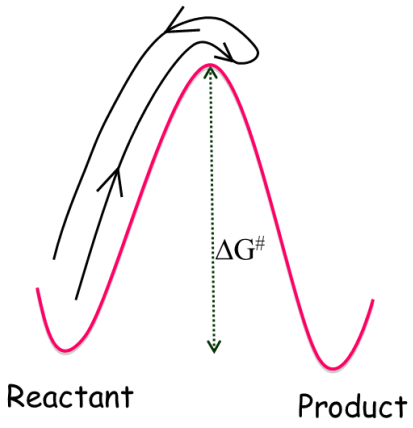

Figure 1. *A schematic diagram of recrossing at the transition state.*

Recrossings are most common in reactions whose barrier is flat near the transition state region. They are not always dynamical in nature. A poor choice of the dividing surface can lead to more recrossings. The method of variationally changing the dividing surface to minimize the recrossing is called variational TST, which is an improvement to the conventional TST.[10-12] However, it has been demonstrated that even with an optimal dividing surface choice, the recrossing phenomenon cannot be eliminated.[13] It is important to note here that the major contributing factor to the rate is the exponential term determined by the activation energy and the impact of dynamical correction to rate due to recrossing is much less in magnitude compared to the exponential factor.

1.3 Kramers theory

First attempt to quantity recrossing effects through the calculation of transmission coefficient was by Kramers in 1940.[14] He formulated a theory to calculate the escape rate of particle from a metastable state. Kramers theory was based on the ordinary Langevin equation where, in addition to Newton's equation of motion, a random force coupled to the reaction coordinate motion and a damping force proportional to the velocity of the particle with an appropriate friction constant are used to represent the effect of the thermal environment. The chemical reaction equivalent of the theory is that the motion of a particle can be considered as the motion of the reaction coordinate. All other degrees of freedom (including the other molecular and the solvent degrees of freedom) constitute the bath and are represented by the random force and the damping force. Kramers recast the Langevin equation for the motion of the particle into a Fokker-Planck equation and used a flux-over-population method to calculate the rate of barrier passage. Kramers identified three regimes based on the friction felt by the system. (1) First is the low friction regime where the reaction coordinate is very weakly coupled to the bath. Hence, the flow of energy into and out of RC is very limited and infrequent. Thus, an equilibrium energy distribution cannot be maintained. This leads to rare activation of the reactant and deactivation of the product, rendering a low rate in this regime. Kramers thus found that rate is proportional to friction in this weak bath coupling regime. (2) Second is the intermediate friction regime where enough collisions of the RC with the bath occur to maintain an equilibrium energy distribution of the barrier passage. However, the collisions are not too frequent to obstruct the motion across the barrier. Hence, there is a negligible recrossing in this regime. Therefore, in this regime, TST is mostly valid. Kramers rate equation in the intermediate to strong friction regime is written as, $k = \frac{1}{\omega_b}\left(\frac{-\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + \omega_b^2}\right) k^{TST}$ where ω_b is the equilibrium barrier frequency and γ is the friction. (3) Finally, in the high friction regime, the collisions of the RC with bath are too often that they

impede the passage of the system across the barrier in this regime. These collisions lead to recrossing before the formation of stable product and thereby result in a rate estimate much lower than the TST prediction. Here the rate becomes inversely proportional to the friction and the reaction is essentially a diffusion-controlled passage across the barrier. Kramers rate equation in the strong friction regime is $k = \frac{\omega_b}{\gamma} k^{TST}$. A schematic figure demonstrating the dependence of rate on friction in these three regimes is shown in Fig. 2.

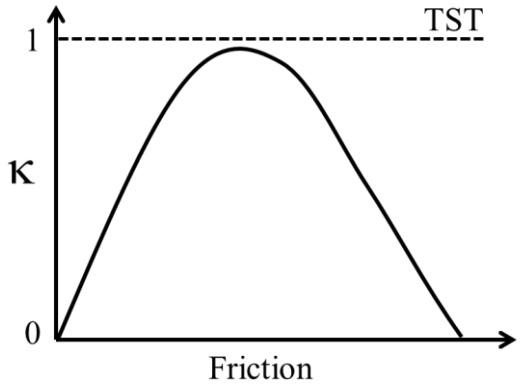

Figure 2. *Schematic figure of dependence of the transmission coefficient on the friction (damping) felt by the system. At lower damping, κ increases with friction whereas at large damping, κ is inversely proportional to the friction. The exact shape of the curve may depend on the system.*

Kramers theory has been successful in explaining reaction rates of several systems. However, in some studies it was observed that the rate did not follow an exact inverse dependence on the viscosity (macroscopic equivalent of the friction) of the solvent even in the intermediate to high friction regime as suggested by Kramers theory.[15-17]

1.3.1 Internal Friction

Although the deviation from Kramers theory was attributed to the limitations in the highly simplified model of Kramers theory.[18], a different explanation for the deviation was given by Ansari and co-workers in 1992.[19] They studied the rate of conformational changes of myoglobin protein at different viscosities shown in Fig. 3. In the rate versus viscosity plot,

they obtained a regime which did not follow Kramers prediction of the inverse dependence of rate on viscosity. They found that the plot instead fits to the equation $k = C/(\eta + \sigma)$ where the term σ was appended to the solvent viscosity η. Since the friction arising from the solvent is already incorporated in η, σ was thought to arise from the friction imparted due to the protein degrees of freedom. Hence, σ was coined internal friction. Several other studies also observed a fractional dependence of rate on viscosity ($k = C/\eta^{\alpha}$ with $\alpha < 1$ whereas in Kramers theory $\alpha = 1$) and attributed the phenomena to the presence of internal friction.[20-24] A non-zero intercept obtained after extrapolating the time taken for reaction versus viscosity plot to zero viscosity is also ascribed to the presence of internal friction in the system.[22, 24-26]

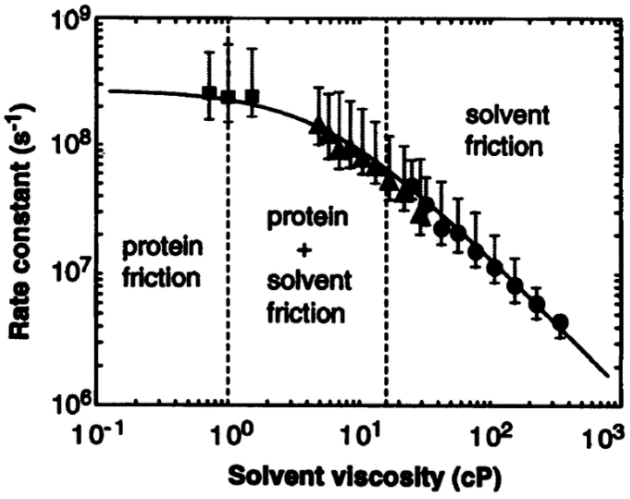

Figure 3. *Rate constant of myoglobin conformational change as a function of solvent viscosity obtained by Ansari et al. In the middle regime, both the solvent friction and the protein friction contributes to the decrease in the reaction rate constant. Reprinted with permission from Ref. 19, Copyright (1992), AAAS.*

Several experimental and theoretical studies ensued after Ansari's work attempting to unearth the origin of internal friction. Internal friction was attributed to the presence of ruggedness in the free energy landscape following the theory of Zwanzig[27] who had showed that the diffusion of particle in a rugged potential is slower than that in a smooth potential.[28] Schulz *et al.* identified that α-helices are characterized by larger internal friction than β-hairpins from the forced unfolding studies of polypeptides.[29] They suggest that this difference is caused by

the different number of intrapeptide hydrogen bonds in the α-helix and β-hairpin configurations. Best and co-workers state that the insensitivity of dihedral angle changes to solvent viscosity is the source of internal friction.[22] They also stated that the internal friction is not affected by changes in global folding barrier heights while increase in local dihedral angle barrier leads to an increased internal friction.[21] Concurrently, internal friction in the unfolding dynamics of cold shock protein was attributed to dihedral rotations belonging to a correlated move by Echeverria *et al.*[25] Eaton and co-workers employed both single molecule fluorescence studies and molecular dynamics simulations to investigate the presence of internal friction in a designed α helical protein.[30] Their study claims that the non-native salt bridge formation between helices gives rise to internal friction. Sashi *et al.* investigated the ring flipping motions of phenylalanine and tyrosine in cytochrome-C at different viscosities.[31] They concluded that with the increase in viscosity, molecular compressibility of the protein reduced and argued that the increased number of dispersion interaction between the non-bonded atoms is responsible for internal friction. Thus, the origin of internal friction is a highly debated topic and we will try to address some issues related to its origin in Chapter 3 and 4.

1.3.2 Limitations of Kramers theory

As mentioned before, the failure of Kramers theory could be attributed to a highly simplified model used to calculate the rate. The theory makes several assumptions. First of all, Kramers calculated the rate for a one-dimensional motion of the particle which may not be generally valid for multi-dimensional reaction coordinate motion. Secondly, Kramers used Langevin equation to describe the motion of the particle which assumes that the system-bath coupling can be characterized by a fixed parameter (friction coefficient) γ. Use of Langevin equation to describe the system-bath coupling is also justified only when the interactions between the

system and the bath are linear in nature. Therefore, Langevin equation is not appropriate for non-linear interactions between the system and the bath. Further, in Langevin equation the environment is represented using a Markovian friction which implies that the response of the bath to the motion of the reaction coordinate is immediate and does not have any memory. However, often it is found that the environment takes time to respond to the solute motions. For example, solvent molecules take time to rearrange their hydrogen bond network, reorient the polar solvent dipoles etc. for a change in the solute configuration. In cases where there is a lag created in the solvent configuration due to solute motion, one has to take into account the time-dependent friction or memory friction instead of employing the full Markovian static friction as discussed below.

1.4 Grote Hynes Theory

Though Kramers theory is the first to provide a description for non-equilibrium solvent effects on reaction rates, it often breaks down due to the non-Markovian nature of the friction.[5, 17, 32] A better description for rate processes incorporating memory effects present in the system was developed by Grote and Hynes in 1980.[33] Grote and Hynes realized that the friction may not act instantaneously on the solute motions, and hence used generalized Langevin equation (GLE) -- which takes into account the non-Markovian effect by means of a time-dependent friction kernel -- to describe the motion of the particle. They also used the Stable States Picture previously introduced by Northrup and Hynes.[34] A Stable State reactant (SSR) and Stable State product (SSP) are located away from the TS region and forbids recrossing of any trajectories that reach either of the states without thermalizing (absorbing or equilibrating) in these regions. Rate is then calculated as the correlation function of the flux originating from the SSR, crossing the TS, and traversing the SSP at a later time. Grote-Hynes theory (GHT) yields transmission coefficient as the ratio of reactive frequency λ_r to

equilibrium barrier frequency ω_{eq}. ω_{eq} is the real magnitude of the frequency at the transition state when an equilibrium solvation is assumed. λ_r is the modified barrier frequency in the presence of memory friction. The reactive frequency is a positive root in the interval $0 < \lambda_r < \omega_b$ and can be found by solving the self-consistent equation, $\lambda_r^2 - \omega_{eq}^2 + \lambda_r \int_0^\infty e^{-\lambda_r t} \xi_{TS}(t) dt = 0$ where $\xi_{TS}(t)$ is the friction kernel – a measure of the friction felt by the RC at the TS. If the friction is very low, then $\lambda_r \approx \omega_b$ and consequently $\kappa \approx 1$. The reactive frequency decides the timescale of the events that will cause recrossing at the TS and thus ensures that only the part of the friction kernel relevant for the barrier crossing is only taken into account unlike the Kramers theory which takes into account the entire spectrum of the friction kernel. The modes of the bath with frequency close to the reactive frequency λ_r are mostly known to affect κ.[35-37]

1.4.1 Limiting cases of GHT

Grote Hynes theory is a generalized theory. One can arrive at different special cases of Grote-Hynes theory,[38] based on the friction strength and the memory time of the friction kernel which is defined as, $\tau_M = \xi / \xi(t = 0)$. Here, ξ is total friction (zero frequency friction) obtained from the integral of the friction kernel ($\xi = \int_0^\infty \xi(t) dt$ is the total friction used in the Kramers theory) and $\xi(t = 0)$ is the initial value of the friction kernel. Based on these quantities, the four limiting cases of GHT arise as mentioned below.

(i) Kramers theory: If $\tau_M \ll \omega_b^{-1}$, then the friction becomes Markovian for the system under study. GHT rate essentially then reduces to Kramers prediction and $\lambda_{r,Kr}^2 - \omega_b^2 + \int_0^\infty \xi(t) dt = 0$. (ii) Non-adiabatic limit: If $\tau_M \gg \omega_b^{-1}$, the solvent responds very slowly to the motion of the solute and thus can appear frozen in comparison to the solute motion. In such cases, $\xi(t)$ can be replaced with its initial value, $\xi(t = 0)$ in the GHT equation to obtain

$\lambda_{r,NA}^2 - \omega_b^2 + \xi(t=0) = 0$. Here the solute does not move along the equilibrium barrier since the solvent molecules do not equilibrate to solute motion; the solute moves along a different barrier $\lambda_{r,NA}$ that depends on the initial solvent configurations.

(iii) Caging-Regime: If $\lambda_{r,NA}$ becomes imaginary, then it implies that the solute is trapped in a cage by the solvent at the barrier top. Only after the relaxation of solvent, the solute is able to move. Solvent motions are critical here since the time required for the solvent to relax determines the number of recrossing and thus the value of κ.

(iv) Weak solvation limit: If the friction is weak even when the solvent is not in equilibrium, recrossings are very few resulting in $\kappa \approx 1$.

1.4.2 Applications of GHT

GHT has been successfully applied in a wide variety of reactions. Some of the earlier demonstrations of the applicability of GHT are in model atom transfer A+BC reaction in rare gas solvents[39], model S_N2 reaction between chloride ion and methyl chloride in aqueous solution[4-5] and S_N1 reaction in water[32]. In the S_N2 and S_N1 reaction models, Kramers' prediction was found to be much lower than that of GHT, thus illustrating the importance of incorporating memory effects in rate calculation. The validity of GHT has been established in several other cases: the dynamics of Na-Cl ion pairs in dimethyl-sulfoxide[40], model electrochemical ion transfer reaction[41], potassium permeation in the selective potassium IRK1 channel[42], sodium permeation in a gramicidin like channel[9], Na-Cl ion pair association-dissociation in water[43], electron transfer at electrode-water interface[44], etc. GHT has also been effectively employed to compare the reactions in aqueous solution and in the enzyme counterpart; a few examples include methyl transfer from S-adenosylmethionine to catecholate both in aqueous solution and in the catechol O-methyltransferase active site[35], Michael addition reaction leading from 6-deoxychalcone to the corresponding flavanone in

aqueous solution and the reaction catalyzed by chalcone isomerase[36], and methyl transfer by glycine N-methyltransferase[45]. A few cases where failure of GHT has been reported include hydration shell kinetics around Na^+ and Li^+ ions.[6-7] The failure of the theory has been traced to the breakdown in the theory's assumptions.

1.4.3 Shortcomings of GHT

GHT uses generalized Langevin equation (GLE) to describe the motion of the particle. Though GLE is more general than the ordinary Langevin equation (LE), it is not universally applicable to describe all chemical reactions. As in the case of LE, the potential terms involving the bath are quadratic or bilinear in GLE, and thus the theory will fail if there are non-linear interactions between bath and the solute. GHT also assumes an infinite parabolic barrier in the rate calculation where local harmonic approximation holds at the TST. Hence, several of the failures of GHT occur due to anharmonicity at the barrier top. In other cases of failures of GHT, the friction is dependent on the value of the reaction coordinate where one has to employ coordinate-dependent friction models. GHT also ignores any quantum effects in the rate calculation.

There have been several efforts to improve both Kramers theory and Grote-Hynes theory by including quantum effects,[46] barrier anharmonicity,[47-48] spatially dependent friction[49] etc. However both Kramers theory and Grote-Hynes theory have remained popular among the community due to their ability to explain different rate-related concepts and their ease of application to different chemical systems.

1.5 Outline of the Thesis

In this thesis, we have focused on the dynamical effects in rate process like recrossing, memory effects, and internal friction. It is challenging to study non-equilibrium effects like

recrossing using experimental techniques. For example, changing the solvent to vary the friction to study the role of friction in recrossing will affect the free energy in experiments. Therefore computational methods are essential to understand the exact nature of these phenomena mentioned above. In Chapter 2, we summarize the methods used in the thesis to investigate these three phenomena. In Chapter 3 and 4, we study the highly debated topic of internal friction using model systems. In Chapter 5, we investigate the recrossing effects in the intercalation of proflavine into DNA. This is followed by the summary of the thesis and a few future directions to augment this thesis work. The addendum to the thesis provides details of molecular dynamics studies of design of ligands developed to target repeat expansions in genetic material, which is a work not directly related to the thesis.

References

1. Eyring, H., The Activated Complex in Chemical Reactions. *The Journal of Chemical Physics* **1935,** *3* (2), 107-115.
2. Arrhenius, S., Über die Reaktionsgeschwindigkeit bei der Inversion von Rohrzucker durch Säuren. In *Zeitschrift für Physikalische Chemie*, 1889; Vol. 4U, p 226.
3. Evans, M. G.; Polanyi, M., Some applications of the transition state method to the calculation of reaction velocities, especially in solution. *Transactions of the Faraday Society* **1935,** *31* (0), 875-894.
4. Bergsma, J. P.; Gertner, B. J.; Wilson, K. R.; Hynes, J. T., Molecular dynamics of a model SN2 reaction in water. *The Journal of Chemical Physics* **1987,** *86* (3), 1356-1376.
5. Gertner, B. J.; Wilson, K. R.; Hynes, J. T., Nonequilibrium solvation effects on reaction rates for model SN2 reactions in water. *The Journal of Chemical Physics* **1989,** *90* (7), 3537-3558.
6. Rey, R.; Hynes, J. T., Hydration Shell Exchange Kinetics: An MD Study for Na+(aq). *The Journal of Physical Chemistry* **1996,** *100* (14), 5611-5615.
7. Spångberg, D.; Rey, R.; Hynes, J. T.; Hermansson, K., Rate and Mechanisms for Water Exchange around Li+(aq) from MD Simulations. *The Journal of Physical Chemistry B* **2003,** *107* (18), 4470-4477.
8. Neria, E.; Karplus, M., A position dependent friction model for solution reactions in the high friction regime: Proton transfer in triosephosphate isomerase (TIM). *The Journal of Chemical Physics* **1996,** *105* (24), 10812-10818.
9. Roux, B.; Karplus, M., Ion transport in a gramicidin-like channel: dynamics and mobility. *The Journal of Physical Chemistry* **1991,** *95* (12), 4856-4868.
10. Wigner, E., Calculation of the Rate of Elementary Association Reactions. *The Journal of Chemical Physics* **1937,** *5* (9), 720-725.
11. Horiuti, J., On the Statistical Mechanical Treatment of the Absolute Rate of Chemical Reaction. *Bulletin of the Chemical Society of Japan* **1938,** *13* (1), 210-216.

12. Evans, M. G., Thermodynamical treatment of transition state. *Transactions of the Faraday Society* **1938,** *34* (0), 49-57.
13. Mullen, R. G.; Shea, J.-E.; Peters, B., Transmission Coefficients, Committors, and Solvent Coordinates in Ion-Pair Dissociation. *Journal of Chemical Theory and Computation* **2014,** *10* (2), 659-667.
14. Kramers, H. A., Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **1940,** *7* (4), 284-304.
15. Goldenberg, M.; Emert, J.; Morawetz, H., Intramolecular excimer study of rates of conformational transitions. Dependence on molecular structure and the viscosity of the medium. *Journal of the American Chemical Society* **1978,** *100* (23), 7171-7177.
16. Velsko, S. P.; Fleming, G. R., Photochemical isomerization in solution. Photophysics of diphenyl butadiene. *The Journal of Chemical Physics* **1982,** *76* (7), 3553-3562.
17. Velsko, S. P.; Waldeck, D. H.; Fleming, G. R., Breakdown of Kramers theory description of photochemical isomerization and the possible involvement of frequency dependent friction. *The Journal of Chemical Physics* **1983,** *78* (1), 249-258.
18. Bagchi, B.; Oxtoby, D. W., The effect of frequency dependent friction on isomerization dynamics in solution. *The Journal of Chemical Physics* **1983,** *78* (5), 2735-2741.
19. Ansari, A.; Jones, C. M.; Henry, E. R.; Hofrichter, J.; Eaton, W. A., The role of solvent viscosity in the dynamics of protein conformational changes. *Science* **1992,** *256* (5065), 1796-1798.
20. Zheng, W.; De Sancho, D.; Hoppe, T.; Best, R. B., Dependence of Internal Friction on Folding Mechanism. *Journal of the American Chemical Society* **2015,** *137* (9), 3283-3290.
21. Zheng, W.; de Sancho, D.; Best, R. B., Modulation of Folding Internal Friction by Local and Global Barrier Heights. *The Journal of Physical Chemistry Letters* **2016,** *7* (6), 1028-1034.
22. de Sancho, D.; Sirur, A.; Best, R. B., Molecular origins of internal friction effects on protein-folding rates. *Nat Commun* **2014,** *5*.
23. Qiu, L.; Hagen, S. J., A Limiting Speed for Protein Folding at Low Solvent Viscosity. *Journal of the American Chemical Society* **2004,** *126* (11), 3398-3399.
24. Schulz, J. C. F.; Schmidt, L.; Best, R. B.; Dzubiella, J.; Netz, R. R., Peptide Chain Dynamics in Light and Heavy Water: Zooming in on Internal Friction. *Journal of the American Chemical Society* **2012,** *134* (14), 6273-6279.
25. Echeverria, I.; Makarov, D. E.; Papoian, G. A., Concerted Dihedral Rotations Give Rise to Internal Friction in Unfolded Proteins. *Journal of the American Chemical Society* **2014,** *136* (24), 8708-8713.
26. Soranno, A.; Buchli, B.; Nettels, D.; Cheng, R. R.; Müller-Späth, S.; Pfeil, S. H.; Hoffmann, A.; Lipman, E. A.; Makarov, D. E.; Schuler, B., Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proceedings of the National Academy of Sciences* **2012,** *109* (44), 17800.
27. Zwanzig, R., Diffusion in a rough potential. *Proceedings of the National Academy of Sciences of the United States of America* **1988,** *85* (7), 2029-2030.
28. Wensley, B. G.; Batey, S.; Bone, F. A. C.; Chan, Z. M.; Tumelty, N. R.; Steward, A.; Kwa, L. G.; Borgia, A.; Clarke, J., Experimental evidence for a frustrated energy landscape in a three-helix-bundle protein family. *Nature* **2010,** *463* (7281), 685-688.
29. Schulz, J. C. F.; Miettinen, M. S.; Netz, R. R., Unfolding and Folding Internal Friction of β-Hairpins Is Smaller than That of α-Helices. *The Journal of Physical Chemistry B* **2015,** *119* (13), 4565-4574.
30. Chung, H. S.; Piana-Agostinetti, S.; Shaw, D. E.; Eaton, W. A., Structural origin of slow diffusion in protein folding. *Science* **2015,** *349* (6255), 1504-1510.

31. Sashi, P.; Ramakrishna, D.; Bhuyan, A. K., Dispersion Forces and the Molecular Origin of Internal Friction in Protein. *Biochemistry* **2016,** *55* (33), 4595-4602.
32. Keirstead, W. P.; Wilson, K. R.; Hynes, J. T., Molecular dynamics of a model SN1 reaction in water. *The Journal of Chemical Physics* **1991,** *95* (7), 5256-5267.
33. Grote, R. F.; Hynes, J. T., The stable states picture of chemical reactions. II. Rate constants for condensed and gas phase reaction models. *The Journal of Chemical Physics* **1980,** *73* (6), 2715-2732.
34. Northrup, S. H.; Hynes, J. T., The stable states picture of chemical reactions. I. Formulation for rate constants and initial condition effects. *The Journal of Chemical Physics* **1980,** *73* (6), 2700-2714.
35. 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. *Journal of the American Chemical Society* **2006,** *128* (18), 6186-6193.
36. 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. *Journal of the American Chemical Society* **2008,** *130* (23), 7477-7488.
37. Kanaan, N.; Roca, M.; Tuñón, I.; Martí, S.; Moliner, V., Application of Grote−Hynes Theory to the Reaction Catalyzed by Thymidylate Synthase. *The Journal of Physical Chemistry B* **2010,** *114* (42), 13593-13600.
38. 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.
39. Bergsma, J. P.; Reimers, J. R.; Wilson, K. R.; Hynes, J. T., Molecular dynamics of the A+BC reaction in rare gas solution. *The Journal of Chemical Physics* **1986,** *85* (10), 5625-5643.
40. 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. *The Journal of Physical Chemistry A* **1997,** *101* (15), 2862-2872.
41. Pecina, O.; Schmickler, W., On the dynamics of electrochemical ion-transfer reactions. *Journal of Electroanalytical Chemistry* **1998,** *450* (2), 303-311.
42. Tolokh, I. S.; White, G. W. N.; Goldman, S.; Gray, C. G., Prediction of ion channel transport from Grote—Hynes and Kramers theories. *Molecular Physics* **2002,** *100* (14), 2351-2359.
43. Rey, R.; Guardia, E., Dynamical aspects of the sodium(1+)-chloride ion pair association in water. *The Journal of Physical Chemistry* **1992,** *96* (11), 4712-4718.
44. Rose, D. A.; Benjamin, I., Molecular dynamics of adiabatic and nonadiabatic electron transfer at the metal–water interface. *The Journal of Chemical Physics* **1994,** *100* (5), 3545-3555.
45. 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†. *The Journal of Physical Chemistry B* **2008,** *112* (2), 529-534.
46. Wolynes, P. G., Quantum Theory of Activated Events in Condensed Phases. *Physical Review Letters* **1981,** *47* (13), 968-971.
47. Lee, S.; Hynes, J. T., Solution reaction path Hamiltonian for reactions in polar solvents. I. Formulation. *The Journal of Chemical Physics* **1988,** *88* (11), 6853-6862.
48. Lee, S.; Hynes, J. T., Solution reaction path Hamiltonian for reactions in polar solvents. II. Applications. *The Journal of Chemical Physics* **1988,** *88* (11), 6863-6869.

49. Voth, G. A., A theory for treating spatially-dependent friction in classical activated rate processes. *The Journal of Chemical Physics* **1992,** *97* (8), 5908-5910.

Chapter 2

Methodology

2.1 Classical Simulations

Macroscopic properties are often determined by the behavior of the system at the microscopic level, i.e., atomic level. Classical simulations can mimic these microscopic behaviors to give us qualitative and quantitative information about the macroscopic properties. In my thesis, I have used two types of classical simulations, molecular dynamics which is deterministic simulation and Langevin dynamics which is stochastic in nature. Both molecular dynamics and Langevin dynamics propagate the motion of the atoms following Newtonian dynamics where the net force on the atoms is calculated at every instant. While in molecular dynamics this net force arise only from interaction with other molecules, in Langevin dynamics, in addition to the Newtonian force, a frictional force and a random force determine the atomic motion. Inter-atomic interactions are determined by potential functions that constitute the force field and dictate the interactions between each atom. The general form of a force field is given below.

$$U = \sum_{bonds} \frac{k_{bond}}{2}\left(r - r_{eq}\right)^2 + \sum_{angle} \frac{k_{angle}}{2}\left(\theta - \theta_{eq}\right)^2 + \sum_{torsions} \frac{k_{dihedral}}{2}[1 + \cos(n\emptyset - \gamma)] + \sum_{i>j}\left[\left\{\frac{A_{ij}}{{R_{ij}}^{12}} - \frac{B_{ij}}{{R_{ij}}^{6}}\right\} + \frac{q_i q_j}{\varepsilon R_{ij}}\right], \quad Eq\ 2.1$$

where k_{bond}, k_{angle} and $k_{dihedral}$ are force constants for the bond, angle, and dihedral potentials, respectively. r_{eq} and θ_{eq} represent the equilibrium values for a given bond and angle, respectively. n is the multiplicity and γ is the phase angle for dihedral angles. The

first three terms thus constitute the potential for bonded interactions. The last term in the equation represents the non-bonded interaction between atoms and is a combination of the Lennard-Jones potential and the Coulomb potential. A_{ij} and B_{ij} are atom-specific parameters for a pair of atoms i and j in the Lennard-Jones potential and q_i and q_j are partial charges on atoms i and j, respectively. ε is the permittivity of medium.

Both molecular dynamics and Langevin dynamics use the force field to determine the interatomic interactions. In molecular dynamics, interactions with solvent molecules are also precisely calculated using force field. When the solvent degrees of freedom are not directly involved in the reaction, but rather serve as a medium for the reaction to occur, then those degrees of freedom can be considered as bath and need not be treated explicitly. Hence, in Langevin dynamics these degrees of freedom are omitted. The average effect of these solvent degrees of freedom are mimicked by a frictional force term and a random force term in the Langevin equation as given below.

$$m_i \frac{d^2 r}{dt^2} = -\nabla \mathrm{U(r)} - M\gamma \frac{dr}{dt} + R_i(t) \qquad Eq\ 2.2$$

$$\langle R(t)R(t')\rangle = 6\gamma k_B T M \delta(t - t') \qquad Eq\ 2.3$$

The first term in $Eq.(2.2)$ represents the force due to the force field potentials, the second term represents the frictional drag caused by the solvent viscosity, and the third term mimics the random forces associated with the thermal motions of the solvent molecule. Here, γ represents the friction coefficient which determines the interaction strength of the solvent and $R_i(t)$ is a delta-correlated stationary Gaussian process with zero mean ($\langle R(t)\rangle = 0$). Since friction opposes motion, the second term is proportional to the particle's velocity and is oppositely directed. The friction coefficient, γ, is related to the fluctuations of the random

forces by the fluctuation-dissipation theorem given in Eq (2.3). The delta function assures that the random force is completely uncorrelated at different times.

Molecular biologists use classical simulations to mimic real systems. However, the timescale of different bio-molecular processes vary from picoseconds to seconds. Classical simulations, typically with an integration time of femtoseconds, may not adequately sample the processes that have a timescale larger than microseconds. In such cases, one can use enhanced sampling techniques such as umbrella sampling[1], metadynamics[2], parallel tempering,[3-4] etc. In my thesis, for the intercalation of drug into DNA, which is a millisecond process[5], I have used umbrella sampling technique to sufficiently sample the transition state region.

2.2 Umbrella Sampling

Umbrella sampling is a technique to sample phase space where ergodicity is hindered by the potential energy of the system. If one can find a reaction coordinate, $s(r)$, which defines the states to be sampled, a bias potential can be applied along the reaction coordinate to sample those states effectively. In general umbrella sampling scheme, multiple windows are made, each with a different value of reaction coordinate (see next subsection) to be sampled with the help of the bias potential. The bias potential can be of any form. However, it is generally taken to be a harmonic potential of the form $V_i^{umb}\big(s(r)\big) = \frac{k_i}{2}(s(r) - s(r)_i^{ref})^2$, where k_i is the force constant for the i^{th} window, $s(r)_i^{ref}$ is the reaction coordinate value to be sampled in the i^{th} window and $s(r)$ is the dynamical value of the reaction coordinate. Each window is constrained to sample a part of the phase space along the reaction coordinate such that there is sufficient overlap between the distributions of the reaction coordinate values in adjacent windows. Note that, the distribution obtained in each window using the above method is a biased distribution. If the biasing potential is strictly a function of $s(r)$, then the unbiased

free energy profile can be obtained by substracting the biasing potential from the biased free energy profile, i.e., $F_i(s) = -\frac{1}{\beta} ln\, P_i(s) - V_i^{umb}\big(s(r)\big)$ where P_i is the population of the i^{th} state. Finally, methods like WHAM[6] (Weighted histogram analysis method) can be used to recover the total free energy surface of the process along the reaction coordinate.

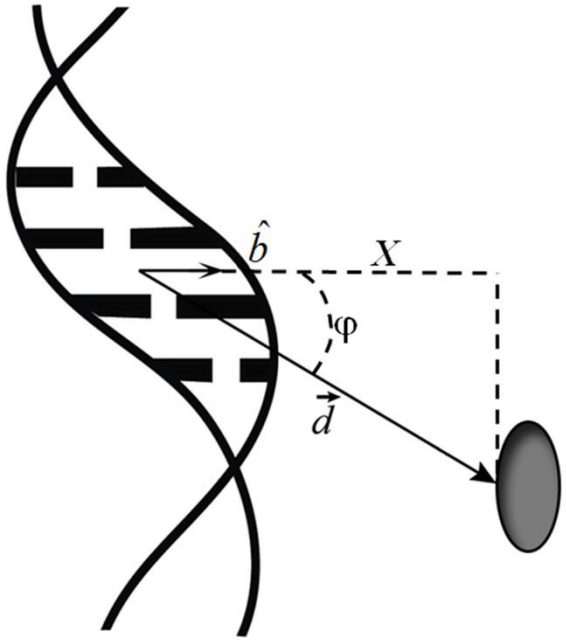

Figure 1: *The reaction coordinate X used in the umbrella sampling simulation in Chapter 5. See text for details.*

2.3 Reaction Coordinate

Almost all the enhanced sampling techniques require identification of a collective variable called reaction coordinate along which the external bias is to be applied. The choice of this collective variable is often heuristic and decided based on measuring the progress of the reaction. In this thesis (chapter 5), for the umbrella sampling simulations used in intercalation of drug into DNA, we use a coordinate named X[7-8] (see Fig 1). The two base pairs between which the drug inserts itself is called intercalating base pairs (IBP). $\vec{d}$ is defined as the vector from the center of mass of IBP to the center of mass of proflavine. $\hat{b}$ is the unit vector from the center of mass of IBP to the center of mass of two 5' sugars which lie more towards the

minor groove of IBP. X is defined as the dot product of the two vectors $\hat{b}$ and $\vec{d}$. Thus, X defines the distance of the drug from the DNA perpendicular to the DNA axis. X is close to zero when drug intercalates into the DNA.

2.4 Transmission coefficient calculation

The amount of recrossing in a system is quantified using the factor called transmission coefficient which, as already mentioned, is used as a correction to the transition state theory rate estimate. A transmission coefficient close to unity indicates very few recrossing while a decrease in transmission coefficient from unity corresponds to an increase in recrossing in the system. Transmission coefficient can be calculated using both MD simulations (reactive flux method) and using rates theories such as Kramers theory and Grote-Hynes theory. The chapters 3 and 5 of the thesis involve calculation of transmission coefficient using both simulation and theoretical approaches. The methods are discussed below.

2.4.1 Reactive flux method

Reactive flux method was initially developed by Chandler[9] as a technique to calculate rate for a high barrier process since otherwise the trajectory would most of the time wander in the reactant well. This method was used for improving the TST by counting explicitly the number of recrossings at the optimal dividing surface. Nearly, all of the numerical simulation studies of rate against which the TST is compared employ reactive flux method[10-11].

The TST demands that all the trajectories that have a positive flux towards the product at the transition state will form products, while trajectories with initial negative momentum will form reactants. Since there is no recrossing assumption in TST, the rate is thus given by $k^{TST} = \langle j_+ \rangle_R$, where j_+ denote the positive flux trajectories and $\langle\ \rangle_R$ denotes an average over equilibrium distribution, normalized by the reactant partition function. However, in an actual

reaction, trajectories with either positive or negative flux may later end up as products or reactants. Hence we need to follow these trajectories in time to find in which stable states -- reactant or product -- they finally fall into. Let us consider x to be the reaction coordinate whose value at the transition state is zero. We can define the step function $\theta[x(t)]$ which equals zero on the reactant side and equals one on the product side. In the presence of recrossing, the rate constant can be represented as $k = \langle j\theta[x(t)] \rangle_R$. This can be split into components from initially positive flux and negative flux as $k = \langle j_+\theta[x(t)] \rangle_R + \langle j_-\theta[x(t)] \rangle_R$. Time reversal converts j_- to $-j_+$ in the rate equation. Hence the rate constant can now be written as $k = \langle j_+\theta[x(t)] \rangle_R - \langle j_+\theta[x(-t)] \rangle_R$. Now the transmission coefficient, given as the ratio of actual rate to the rate given by TST, can be written as[10]

$$\kappa = \frac{k}{k^{TST}} = \frac{\langle j_+\theta[x(t)] \rangle_R - \langle j_+\theta[x(-t)] \rangle_R}{\langle j_+ \rangle_R} \qquad Eq\ 2.4$$

This can be rewritten as

$$\kappa = \frac{\sum_{i,+}^{N} \omega_i |v_i| Q_i}{\sum_{i,+}^{N} \omega_i |v_i|} \qquad Eq\ 2.5$$

where $+$ indicates all trajectories with initial positive flux, N is the total number of trajectories, ω_i is the probability of the initial configuration and v_i is the initial velocity of the trajectory. The value of Q_i depends on the initial and final states of the i^{th} trajectory and is given by $Q_i = \begin{cases} +1 & if\ reactant \rightarrow product \\ 0 & if\ reactant \rightarrow reactant\ or\ product \rightarrow product. \\ -1 & if\ product \rightarrow reactant \end{cases}$

Hence to calculate κ, several trajectories must be run forward and backward in time starting from the transition state to determine the value of Q_i. If the forward trajectory forms product and backward trajectory forms reactant, then it is effectively a $reactant \rightarrow product$

trajectory. Similarly, one obtains $reactant \rightarrow reactant$, $product \rightarrow reactant$ and $product \rightarrow product$ trajectories. We have used Eq (2.5) to calculate κ in Chapter 5. If the reaction is symmetric, then the statistics can be enhanced by including the negative flux trajectories. In such case, Eq (2.5) can be rewritten to include the negative flux trajectories as

$$\kappa = \frac{\sum_{i,+}^{N} \omega_i |v_i| Q_i - \sum_{i,-}^{N} \omega_i |v_i| Q_i}{\sum_{i,\pm}^{N} \omega_i |v_i|} \qquad Eq\ 2.6$$

where – denotes trajectories with initial negative momentum at the transition state. We have used this method to calculate transmission coefficient in chapter 3 since it is a symmetric reaction.

2.4.2 Evaluating TC from theories

The Grote-Hynes theory (GHT)[12-13] analytically solves the GLE to obtain the transmission coefficient as the ratio of reactive frequency to equilibrium barrier frequency, $\kappa_{GH} = \frac{\omega_r}{\omega_{eq}}$. The equilibrium barrier frequency can be obtained by fitting potential mean force of the reaction, within a small range of reaction coordinate (RC) values near the transition state region, to the parabolic function, $\Delta PMF = -\frac{1}{2} K_{eq} (RC - RC^{\#})^2$, where ΔPMF is the potential of mean force difference with respect to the maximum and $RC^{\#}$ is the value of RC at the transition state. From the equilibrium force constant, K_{eq}, so obtained one can calculate the equilibrium barrier frequency as $\omega_{eq} = \frac{1}{2\pi c} \sqrt{\frac{K_{eq}}{\mu_{RC}}}$.

The reactive frequency, which is the modified equilibrium barrier frequency in the presence of solvent friction, can be obtained by solving the following self-consistent equation,

$$\omega_r^2 - \omega_{eq}^2 + \omega_r \int_0^{\infty} e^{-\omega_r t} \xi_{TS}(t) dt = 0 \qquad Eq\ 2.7$$

The friction kernel, $\xi(t)$ given in the above equation gauges the coupling of reaction coordinate to other degrees of freedom and can be obtained from the time correlation function of the forces, $F(t)$ exerted on the RC at the transition state $\xi(t) = \frac{\langle F_{RC}(0)F_{RC}(t)\rangle}{\mu_{RC}k_BT}$ where μ_{RC} is the reduced mass of the reaction coordinate, T is the temperature and k_B is the Boltzmann constant. To obtain the friction kernel, the RC value must be constrained at the transition state using a harmonic potential to calculate the forces acting along the RC.

If the friction acting on the reaction coordinate is Markovian, then one can substitute the frequency-dependent friction in $Eq\ (2.7)$ with the zero frequency friction as

$$\omega_{Kr}^2 - \omega_{eq}^2 + \omega_{Kr}\int_0^\infty \xi_{TS}(t)dt = 0 \qquad Eq\ 2.8$$

This gives the reactive frequency in Kramers limit of GHT equation. Therefore, the Kramers estimate of transmission coefficient can be obtained as $\kappa_{Kr} = \frac{\omega_{Kr}}{\omega_{eq}}$.

Thus, in this thesis, we calculated the transmission coefficient numerically and also using GHT and Kramers theory for different systems. However, to find internal friction according to the prescribed protocol from literature, one has to calculate transmission coefficient or rate at multiple solvent viscosities. Below, we describe the technique employed to change the viscosity in simulations.

2.5 Changing viscosity in simulation

In protein folding experiments, the viscosity is scaled by addition of small-molecule viscogens like ethylene glycol, glucose, etc. to the system[14-17]; however they tend to increase the stability of proteins. Hence, a small quantity of chemical denaturant is added to keep the free energy surface unchanged.[16] Addition of these co-solvents can influence the protein dynamics and modify the protein-solvent interactions. Further, viscosity cannot be lowered,

but only increased by the addition of viscogenic cosolutes in experiments. On the other hand, simulation provides a novel way to overcome these shortcomings. In simulations, the viscosity can be scaled by scaling the mass of the solvent, without affecting its thermodynamic properties.[18-19] Scaling the mass of the solvent by a factor λ will change the forces on the solvent and also the kinetic energy of the solvent. Therefore, to keep the forces invariant and maintain the kinetic energy of the system, the timescale of the simulation has to be scaled by a factor $\sqrt{\lambda}$ (i.e., $F = \lambda m \frac{d^2r}{d(\sqrt{\lambda}t)^2}$ and $E_{kin} = \frac{1}{2}\lambda m \left(\frac{dr}{d(\sqrt{\lambda}t)}\right)^2$). This scaling of time by a factor of $\sqrt{\lambda}$, will scale the diffusion by a factor of $1/\sqrt{\lambda}$ since the latter is inversely proportional to time. Stokes law expresses viscosity of a liquid in terms of the mass m, the friction coefficient γ, and diameter d of the particles in the liquid as $\eta = \frac{m\gamma}{6\pi d}$. The diffusion coefficient, D, is related to the friction coefficient γ via Einstein's expression $D = \frac{k_B T}{m\gamma}$. Combining Stokes' law and Einstein's relation, one obtains $\eta = \frac{k_B T}{D 6\pi d}$. Thus when mass is scaled by a factor λ and time is scaled by a factor of $\sqrt{\lambda}$, effectively diffusion is scaled by a factor of $1/\sqrt{\lambda}$, thereby scaling viscosity, η, by a factor of $\sqrt{\lambda}$. This method does not alter the equilibrium properties or free energy surfaces; merely the transport properties are accelerated by the scaling. This method has been previously used to accelerate the equilibration of protein simulations in explicit solvent. [20-22] Here, in the thesis, we employ this method to vary the viscosity of the water that is used as the solvent to study rate dependence on the viscosity. We scale the masses of oxygen and hydrogen atoms in the water molecule and correspondingly scale the integration time step in the simulations. The values of time step and mass scaling used in the simulations are given in the respective chapters.

References

1. Torrie, G. M.; Valleau, J. P., Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **1977,** *23* (2), 187-199.
2. Barducci, A.; Bussi, G.; Parrinello, M., Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Physical Review Letters* **2008,** *100* (2), 020603.
3. Swendsen, R. H.; Wang, J.-S., Replica Monte Carlo Simulation of Spin-Glasses. *Physical Review Letters* **1986,** *57* (21), 2607-2609.
4. Earl, D. J.; Deem, M. W., Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics* **2005,** *7* (23), 3910-3916.
5. Li, H. J.; Crothers, D. M., Relaxation studies of the proflavine-DNA complex: The kinetics of an intercalation reaction. *Journal of Molecular Biology* **1969,** *39* (3), 461-477.
6. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A., The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry* **1992,** *13* (8), 1011-1021.
7. 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. *Journal of the American Chemical Society* **2008,** *130* (30), 9747-9755.
8. Sasikala, W. D.; Mukherjee, A., Molecular Mechanism of Direct Proflavine–DNA Intercalation: Evidence for Drug-Induced Minimum Base-Stacking Penalty Pathway. *The Journal of Physical Chemistry B* **2012,** *116* (40), 12208-12212.
9. Chandler, D., Statistical mechanics of isomerization dynamics in liquids and the transition state approximation. *The Journal of Chemical Physics* **1978,** *68* (6), 2959-2970.
10. Bergsma, J. P.; Reimers, J. R.; Wilson, K. R.; Hynes, J. T., Molecular dynamics of the A+BC reaction in rare gas solution. *The Journal of Chemical Physics* **1986,** *85* (10), 5625-5643.
11. Bergsma, J. P.; Gertner, B. J.; Wilson, K. R.; Hynes, J. T., Molecular dynamics of a model SN2 reaction in water. *The Journal of Chemical Physics* **1987,** *86* (3), 1356-1376.
12. Northrup, S. H.; Hynes, J. T., The stable states picture of chemical reactions. I. Formulation for rate constants and initial condition effects. *The Journal of Chemical Physics* **1980,** *73* (6), 2700-2714.
13. Grote, R. F.; Hynes, J. T., The stable states picture of chemical reactions. II. Rate constants for condensed and gas phase reaction models. *The Journal of Chemical Physics* **1980,** *73* (6), 2715-2732.
14. Wensley, B. G.; Kwa, L. G.; Shammas, S. L.; Rogers, J. M.; Clarke, J., Protein Folding: Adding a Nucleus to Guide Helix Docking Reduces Landscape Roughness. *Journal of Molecular Biology* **2012,** *423* (3), 273-283.
15. Cellmer, T.; Henry, E. R.; Hofrichter, J.; Eaton, W. A., Measuring internal friction of an ultrafast-folding protein. *Proceedings of the National Academy of Sciences* **2008,** *105* (47), 18320-18325.
16. Qiu, L.; Hagen, S. J., Internal friction in the ultrafast folding of the tryptophan cage. *Chemical Physics* **2005,** *312* (1–3), 327-333.
17. Wensley, B. G.; Batey, S.; Bone, F. A. C.; Chan, Z. M.; Tumelty, N. R.; Steward, A.; Kwa, L. G.; Borgia, A.; Clarke, J., Experimental evidence for a frustrated energy landscape in a three-helix-bundle protein family. *Nature* **2010,** *463* (7281), 685-688.
18. Walser, R.; Mark, A. E.; van Gunsteren, W. F., On the validity of Stokes' law at the molecular level. *Chemical Physics Letters* **1999,** *303* (5), 583-586.

19. Walser, R.; Hess, B.; Mark, A. E.; van Gunsteren, W. F., Further investigation on the validity of Stokes–Einstein behaviour at the molecular level. *Chemical Physics Letters* **2001,** *334* (4–6), 337-342.
20. Walser, R.; van Gunsteren, W. F., Viscosity dependence of protein dynamics. *Proteins: Structure, Function, and Bioinformatics* **2001,** *42* (3), 414-421.
21. Nguyen, P. H., Replica exchange simulation method using temperature and solvent viscosity. *The Journal of Chemical Physics* **2010,** *132* (14), 144109.
22. Lin, I. C.; Tuckerman, M. E., Enhanced Conformational Sampling of Peptides via Reduced Side-Chain and Solvent Masses. *The Journal of Physical Chemistry B* **2010,** *114* (48), 15935-15940.

Chapter 3

Probing Viscosity Dependence of Rate: Internal Friction or Lack of Friction?

3.1 Overview

Deviation from the Kramers' inverse viscosity dependence of rate, $k \propto 1/\eta$, is often attributed to the presence of internal friction in proteins after Ansari *et al.* in 1992 showed that the folding rate could fit the equation, $k \propto 1/(\eta + \sigma)$, where σ is considered as internal friction. Several experimental and computational studies thereafter used fits to Ansari's equation or extrapolated the rate to η=0 to estimate internal friction in proteins and attributed its origin to various internal interactions such as ruggedness, dihedral rotation, salt bridges, etc. Here, we show that the above method to calculate internal friction is incorrect since the rate in a simple model system without any internal friction yields a non-zero σ. We demonstrate that the same system in Langevin bath follows Kramers prediction of inverse dependence of rate on viscosity, indicating that the need to use a non-zero σ in the system with all-atom water models must originate from the microscopic motion of the solvent molecules, rather than some friction which is internal to the system. We have then systematically investigated the effect of barrier curvature, barrier height of the one-dimensional potential, size of the LJ solute, and solute-solvent coupling on the amount of σ. Our results show that σ increases with increase in barrier curvature, but is indifferent to changes in barrier height. Most interestingly, increase in the interaction with the solvent and solute size decreases σ. Further investigation reveals that σ correlates with the relative deviation from Kramers rate at different viscosities, where the deviation itself is caused due to absence of full solvent friction rather than the presence of internal friction.

3.2 Introduction

A biomolecular conformational change such as protein folding is essentially a classical barrier crossing problem. Transition state theory (TST) is widely used to calculate the rate of such barrier crossing processes where, $k_{TST} = \frac{\omega_R}{2\pi} e^{-E_b/k_B T}$; here E_b is the energy barrier, ω_R is the reactant well frequency, k_B is the Boltzmann constant, and T is the temperature. However, k_{TST} tends to overestimate the rate since it does not account for the dynamical effects of recrossing at the transition state. In 1940, Kramers derived a rate theory that provided a solution to the barrier crossing problem which modifies the TST rate by a correction factor called transmission coefficient (TC), κ_{KR}, as $k_{KR} = \kappa_{KR}\, k_{TST}$.[1] In the high friction limit, TC simplifies to $\kappa_{KR} = \frac{\omega_b}{\zeta}$. Here ω_b is the barrier curvature and ζ is the friction. Since k_{TST} does not depend on friction, rate is thus inversely proportional to friction. Note that, although Kramers solved the problem of barrier crossing in one dimension where the particle encounters a constant friction ζ, in realistic experimental setup the calculable quantity is often the viscosity, η (since friction is microscopic in nature). Therefore, following the standard Stokes-Einstein relation, Kramers rate in the high viscosity limit is often expressed as, $k(\eta) \propto 1/\eta$ assuming the barrier curvature to be independent of solvent viscosity.

Although, the incorporation of TC in TST improves the rate prediction,[2] failure in Kramers prediction was observed as early as 1982 in the viscosity-dependent isomerization rate of some organic molecules by Fleming and co-workers.[3-4] Many protein folding studies have also shown that Kramers theory (KT) often fails to explain the viscosity dependence of the rate process.[5-9] A fractional power law viscosity dependence of the rate on viscosity, $k \propto \eta^{-\alpha} (\alpha < 1)$, instead of a Kramers inverse dependence (i.e., α=1) seems to explain the observation.[5-9] Also, in 1980, Grote and Hynes proposed a modification to Kramers theory

where a frequency-dependent friction $\zeta(\lambda_r)$, instead of the static friction ζ, appears in the expression of TC, $\kappa_{GH} = 1/\sqrt{1 + (\zeta(\lambda_r)/\lambda_r)}$. [10-11] λ_r is the reactive frequency, i.e., frequency at which a system crosses the barrier. It was shown by Bagchi and Oxtoby subsequently that replacing the total friction by the frequency-dependent friction in the observation of Fleming and coworkers, discussed above, could explain the fractional dependence of the rate on solvent viscosity.[12]

In 1992, Ansari *et al.* measured the rate of conformational changes in myoglobin at different solvent viscosities.[13] At high solvent viscosity, they observed that the folding rate was inversely proportional to the solvent viscosity. However, at intermediate viscosities the rate had an unusual non-linear solvent viscosity dependence and could fit well to the following equation (hereafter referred as Ansari's equation), $k \propto 1/(\eta + \sigma)$, where η is the solvent viscosity and σ is a constant fit parameter. Assuming the source of this term to be internal to the system, σ was coined as internal friction. The internal friction σ, thus introduced, had stirred renewed interest in the community, and a series of studies followed measuring the amount of the internal friction and its origin.[14-15] These studies used either σ from the fit to Ansari's equation or extrapolated the rate to zero viscosity[14, 16-18] to estimate the internal friction. Also, fractional viscosity dependence (power law fit, $k \propto {}^{1}/_{\eta^{\alpha}}$ with $\alpha < 1$) was attributed to the presence of internal friction.[14-15, 18-20]

Internal friction is also attributed to the presence of ruggedness in the energy landscape following a theory of Robert Zwanzig who has showed in 1988 that the diffusion of a particle in a rugged landscape is slower than that in a smooth landscape of the same barrier height.[21] It is believed that internal motions of the solute that hinder the progress of the reaction manifest as ruggedness in the energy landscape. Any local hindrance that gives rise to a small energy penalty reveals as the ruggedness or roughness in the energy profile.

Wensley *et al*. measured viscosity dependence of rate of three α-spectrin domain proteins and attributed the ruggedness in the energy landscape as the reason for ~3000 times slower folding rate of R16 and R17 domains compared to R15 domain.[5, 22-23]

Apart from ruggedness, non-linear viscosity dependence of rate found in several other studies have attributed the origin of internal friction to several other phenomena. Best and coworkers argued that the insensitivity of the dihedral motion to the solvent viscosity is the source of internal friction in protein folding.[14] He has also demonstrated that when changes in torsion angles were an important part of the folding mechanism, they contributed to the internal friction[18], and increasing local torsion barriers led to increased internal friction[15]. They also attributed the presence of ruggedness in the overall free energy landscape of protein folding to these dihedral motions. A study on cold-shock protein by Papoian *et al.* argued that concerted dihedral rotations give rise to internal friction.[16] Chung *et al.* has shown that for a designed alpha helical protein, non-native salt bridges between the helices are the source of internal friction.[24] Conversely, studies by Netz *et al.* claimed that the formation of hydrogen bonds upon folding is connected to the presence of internal friction.[19, 25] Bhuyan *et al.*, after investigating the flipping motion of aromatic rings in cytochrome C, concluded that dispersion interactions and van der Waals forces between non-bonded atoms are the major contributors to internal friction.[6, 26] To reiterate, all studies mentioned above used either fractional viscosity dependence.[14-15, 18-20], or Ansari *et al.*'s equation[5, 13, 27], or extrapolation to zero viscosity[14, 16-18] to estimate internal friction.

In this study, we show that the additional friction term σ in Ansari's equation does not always represent internal friction in the literal sense; rather it is the consequence of relative deviation of rate from Kramers theory at different viscosities due to time-dependent (or frequency-dependent) friction arising out of correlated motion of the solvent molecules. Note that, we

do not claim that internal friction is not present in any system. In fact, we use a simple model system to study internal friction whereas the protein systems are much more complex harboring different types of interactions. However, we show that the method of calculating internal friction and subsequent attribution to its origin may not be correct. Therefore, our study calls for a new method to estimate internal friction in proteins and other systems.

3.3 Methods

(a) **System Design.** We used a simple system consisting of a Lennard Jones particle similar to a carbon atom without any charge under the influence of a one dimensional potential,[14]

$$V = V_0 \cos(2\pi n x/L) \qquad Eq.3.1$$

Here V_0 is half the barrier height, n is the periodicity of the potential and L is the length of the simulation box. This function mimics a double well potential when n = 2 (Fig.1). The particle was solvated in a box of length 2 nm using TIP3P[28] water molecules. This model was used by Best *et al.* in connection with their investigation of internal friction in proteins where they calculated the rate using mean first passage time and investigated the effect of barrier curvature on the internal friction. Using the same model, we have used a different approach, reactive flux formalism[29], to calculate the effect of not only barrier curvature, but also other parameters such as barrier height, solvent coupling, and solute size on σ.

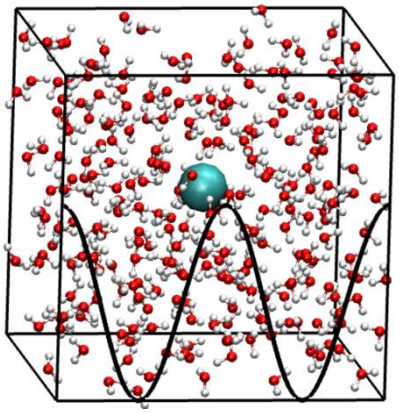

Figure 1: *Snapshot of the LJ system in water bath is shown. The schematic energy profile indicates the barrier that the LJ particle needs to cross for completing the reaction. LJ particle and the barrier profile constitute the system in the present study. Reprinted with permission, Copyright (2018) American Chemical Society.*

While objective of study by Best and co-workers was to investigate the origin of internal friction in which this model served as a prototype of rotational barriers constituting the ruggedness in the overall protein folding process, we have adopted this model to mimic general barrier driven processes (different heights) of different systems (different solute size, curvature and solute-solvent interaction). The details of the different systems used are given in Table 1.

Table 1. *We have varied barrier curvature, barrier height, solute-solvent coupling and solute size independently to study their effect on σ. The details are summarized here.*

Barrier Curvature (ps^{-1}) E_b = 50 kJ/mol $\varepsilon = \varepsilon_0$ $s = s_0$	9.06	18.13	27.19	36.26	
Barrier Height (kJ/mol) ω_b = 13.5 ps^{-1} $\varepsilon = \varepsilon_0$ $s = s_0$	100	25.54	11.76	6.96	
Solvent Coupling (ε_0) ω_b = 18.13 ps^{-1} E_b = 50 kJ/mol $s = s_0$	$0.1\varepsilon_0$	ε_0	$2\varepsilon_0$	$5\varepsilon_0$	$10\varepsilon_0$
Size of Solute (s_0) ω_b = 18.13 ps^{-1} E_b = 50 kJ/mol $\varepsilon = \varepsilon_0$	$0.2s_0$	$0.5s_0$	$1s_0$	$2s_0$	

(b) **Simulation Details.** All the simulations were performed in GROMACS 4.5.5[30] at constant temperature and volume conditions. We have performed simulations at various solvent viscosities, which were controlled by changing the solvent masses as discussed in Chapter 2.[31-32] While scaling the mass by a factor of λ, the time step has to be scaled by a factor of $\sqrt{\lambda}$ to keep the forces invariant. This scales the diffusion coefficient by a factor of $1/\sqrt{\lambda}$ which in turn changes the viscosity by a factor of $\sqrt{\lambda}$ in accordance with Stokes-

Einstein relationship. Best and coworkers have shown that in this system, the diffusion coefficient shows negligible deviation from the viscosity dependence as expected from the Stokes law and the viscosity scales as square root over the solvent mass.[14] The different masses and time steps used in the simulation are given in Table 2. We changed the barrier curvature and the barrier height by systematically varying n and V_0 of the potential (Eq. 3.1), respectively. Also, we have altered the van der Waals parameter of the LJ particle to change the interaction of the solute particle with the solvent. We also modified the diameter of the solute keeping every other parameter same to see the effect of the size of the particle. Thus, we have varied the different parameters independently and calculated the transmission coefficient at several viscosities to capture their effect on σ.

Table 2. *Different solvent masses and time steps used to change the viscosity in the desired range.* $\boldsymbol{m_0}$ *is standard mass and* $\boldsymbol{\eta_0}$ *is the standard viscosity of water.*

m/m_0	η/η_0	dt(fs)
0.5	0.7	0.7
1	1	1
2	1.4	1.4
4	2	2
6.25	2.5	2
9	3	2

(c) **Reactive flux formalism.** We have calculated transmission coefficients using both MD simulations and numerical evaluation of Grote Hynes Theory (GHT). To numerically estimate the transmission coefficient for each system, 1000 simulations starting from the transition state (defined to be the highest point of the potential, i.e., at $(x = 1)$) were run both forward and backward in time for each solvent viscosity. These trajectories were categorized into reactant or product once they reached the stable states. Trajectories that did not reach either of the stable states (defined to be the region within 10% of the minima) within the simulation time (which varied from 1.2 ps to 6 ps per trajectory depending on the viscosity, barrier

curvature and height) were rejected. Reactive flux formalism[29] defines transmission coefficient as the ratio of the actual rate (where both negative and positive flux contribute to the rate) to the rate predicted by the transition state theory (where only positive flux contributes to rate) as,

$$\kappa = \frac{\sum_{i,+}^{N} \omega_i |v_i| Q_i - \sum_{i,-}^{N} \omega_i |v_i| Q_i}{\sum_{i,\pm}^{N} \omega_i |v_i| Q_i} \qquad \text{Eq. 3.2}$$

Here "+" represents trajectories with initial positive flux, "–" represents trajectories with initial negative flux, v_i is the velocity of the RC, ω_i is the probability to have the initial configuration and Q_i is given by

$$Q_i = \begin{cases} 1 \ if RP \\ 0 \ if RR/PP \\ -1 \ if PR \end{cases} \qquad Eq. 3.3$$

RP denotes a trajectory that starts from the reactant and ends up as product. PR is a trajectory that connects the product to the reactant. RR and PP would thus correspond to recrossed trajectories. We have also verified through committor analysis (see Fig. 2) that the configurations we chose for reactive flux calculations represent the transition state of the system.

(d) **Grote Hynes Theory (GHT) and Kramers Theory (KT).** We have also calculated the transmission coefficient using GHT, which expresses the transmission coefficient (TC) as the ratio of reactive frequency (λ_r) to the barrier frequency (ω_b),

$$\kappa_{GH} = \frac{\lambda_r}{\omega_b} \qquad Eq. 3.4$$

The reactive frequency is the measure of the actual rate of passage across the barrier and can be obtained by solving the following self-consistent equation.

$$\lambda_r^2 - \omega_b^2 + \lambda_r \int_0^\infty e^{-\lambda_r t} \, \zeta_{TS}(t) \, dt = 0 \qquad Eq. 3.5$$

The friction kernel, $\zeta_{TS}(t)$ gauges the coupling of reaction coordinate (RC) to other degrees of freedom and can be obtained from the time correlation function of the forces, $F(t)$, exerted on the RC at the TS; μ_{RC} is the reduced mass of the reaction coordinate, T is the temperature and k_B is the Boltzmann constant.

$$\zeta_{TS}(t) = \frac{\langle F_{RC}(0) F_{RC}(t) \rangle}{\mu_{RC} k_B T} \qquad Eq. 3.6$$

To obtain the friction kernel, the system was frozen at the TS and the forces acting on the RC was calculated following the approach used earlier[33]. Since we already know the form of the potential, we can estimate ω_b from the double derivative of the potential. Thus, with the estimate of barrier frequency and the friction kernel, we calculated the reactive frequency and thereby the transmission coefficient. GHT is more general than Kramers theory. In the limit of zero frequency friction, GHT is shown to produce Kramers result. Therefore, we can express the transmission coefficient from Kramers theory using the following modification of GHT,

$$\lambda_{kr}^2 - \omega_b^2 + \lambda_{kr} \int_0^\infty \zeta_{TS}(t)\, dt = 0 \qquad Eq. 3.7$$

$$\kappa_{kr} = \frac{\lambda_{kr}}{\omega_b} \qquad Eq. 3.8$$

(e) **Mean First Passage Time (MFPT) calculation.** While reactive flux and GHT provides the transmission coefficient, mean first passage time measures the average time that a system takes to cross the barrier. It is a direct method to calculate the rate for a barrier crossing problem. However, for a high barrier process, it requires long simulations. We have used MFPT to compare and validate the rate obtained from the above methods. Also, we used MFPT to estimate the rate in Langevin dynamics simulation.

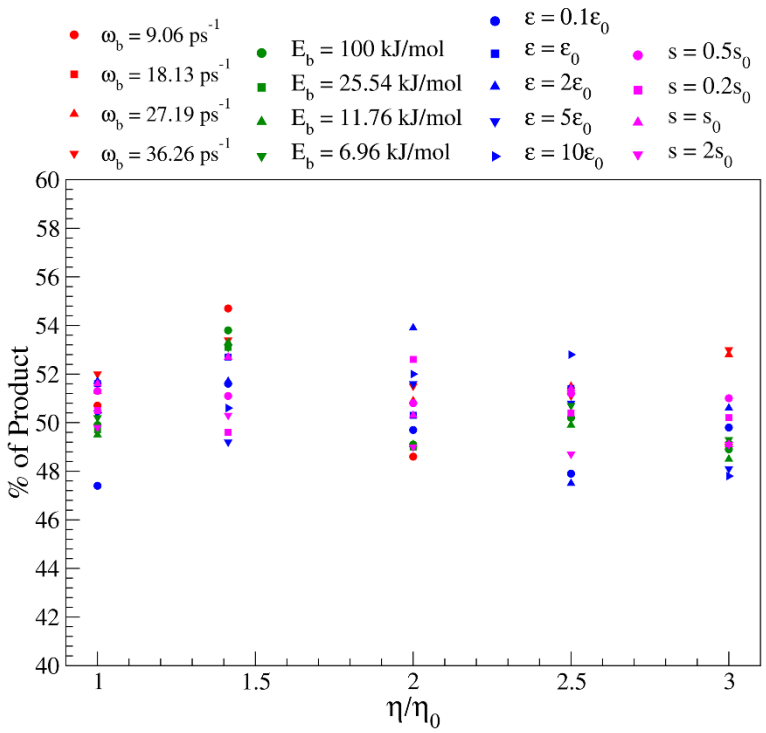

Figure 2: *Here, we show that the variation of solvent viscosity and the different parameters, namely; barrier curvature, barrier height, solute-solvent coupling and solute size do not change the position of the transition state. A trajectory originating from a true transition state has equal probability to end up as reactant or product. For all combinations of the variation in parameter and solvent viscosity, we have plotted the percentage of products formed from a total of 1000 simulations (starting from the transition state) performed for each case, which shows that for all cases, the percentage of product formed lies around 50% with an error bar of 5%.*

(f) Calculation of "*Internal Friction*". The viscosity dependence of the rate was fitted to $\kappa = C/(\eta + \sigma)$ only in the moderate to high friction regime to obtain the estimate of the additional friction term σ. Also, the rate was fitted to check the fractional viscosity dependence using the equation $\kappa = C\,\eta^{-\alpha}$. While $\alpha = 1$ denotes Kramers behavior, $\alpha < 1$ indicates that the rate is lower than Kramers prediction.

3.4 Results

(a) Viscosity dependence of rate - To examine the effect of viscosity on the barrier crossing rate of the particle, we chose an arbitary barrier curvature, ω_b = 13.5 ps^{-1} and height E_b = 6.96 kJ/mol. Since the barrier height is fixed, the rate obtained from the transition state theory would remain unaffected at all viscosities. Therefore, to see the effect of viscosity on the rate,

it suffices to estimate the effect of viscosity on the transmission coefficients. Fig. 3a shows TC obtained from reactive flux (MD-RF) formalism, Grote-Hynes theory, and Kramers theory at various viscosities. All the three different methods show that TC decreases with the increase in viscosity. However, we observe that while the results obtained from MD-RF and GHT match quite well, Kramers theory underestimate the rate significantly. We have also calculated the rate using mean first passage time (MFPT) and shown that $k_{MFPT} \approx \kappa\, k_{TST}$ where the transmission coefficient obtained from each method was multipled by TST rate to estimate the overall rate (see Fig. 3b). This indicates that it is solely the TC that depends on η, thus providing a justification to calculate TC against viscosity to understand the viscosity dependence of the rate. The advantage of using reactive flux method is that we can use any barrier height provided the accurate position of the TS is known, as in the present case. Also, we can compare the result directly with the GHT estimate, as the latter provide the formula for estimating TC.

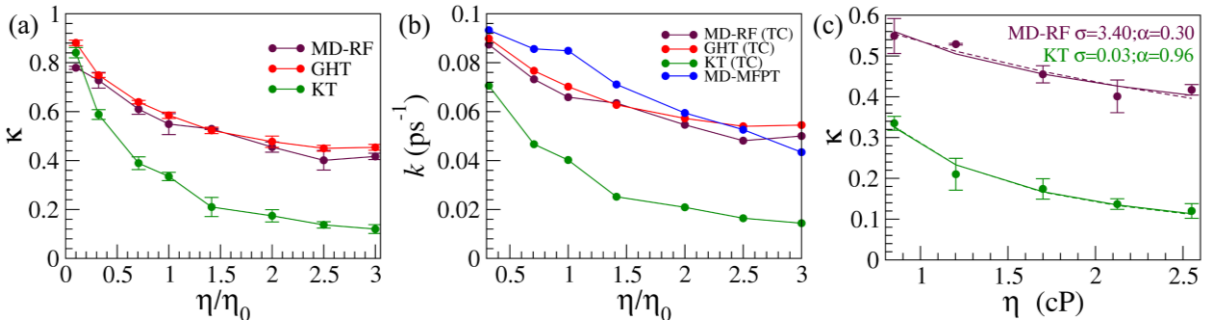

Figure 3: *(a) Transmission coefficients (TC) obtained using reactive flux (MD-RF), Grote-Hynes Theory (GHT), and Kramers theory (KT) at ω_b=13.5 ps^{-1}and E_b=6.96 kJ/mol. (b) The rate obtained from all the four different methods. (c) Fit to the transmission coefficients obtained from MD-RF and KT to power law (solid line) and Ansari's equation (dashed line).*

To investigate the nature of viscosity dependence of rate for this simple system of LJ particle crossing the smooth potential barrier in water bath, we have fitted the transmission coefficients obtained from MD-RF and KT to two equations: the power law fit, $\kappa = c'/\eta^{\alpha}$ and the Ansari's equation, $\kappa = c''/(\eta + \sigma)$, where c' and c'' are proportionality constants

(see Fig. 3c). Note that we fit TC values in the high viscosity regime (normal water viscosity and above) only, where inverse viscosity dependence of Kramers rate is applicable. The fit to the rate obtained from MD-RF yielded a value of α=0.3 and σ=3.4 cP corresponding to the two fits. As is customary in the literature, we can therefore argue that this simple system shows an internal friction of 3.4 cP. However, we know that the present system does not have any internal friction, or any ruggedness. Naturally, we investigated the origin of this non-zero estimate of σ. As anticipated, when we fit the TC values obtained from KT, to both the power law and Ansari's equations, we obtain an almost inverse viscosity dependence of the rate with α nearly equal to unity (0.96) and σ nearly equal to zero (0.03 cP). In fact, Figs. 3a-b show that actual rate is higher than the Kramers prediction. However, it was previously shown that the internal friction, when incorporated in theoretical studies, reduced the rate of the reaction.[34] While the reduction in rate that we observe is supported by $\alpha < 1$ (a higher rate value than KT where $\alpha = 1$), it is puzzling why a positive value of σ in the denominator of Ansari's equation, which will reduce the rate lower than KT, would fit the curve. We will discuss this issue later.

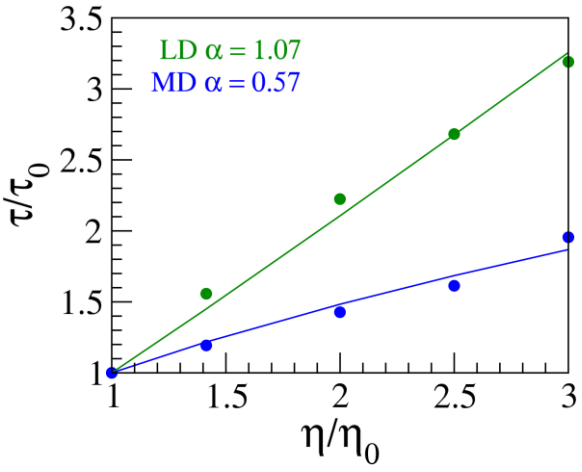

Figure 4: *Mean first passage time (MFPT) calculated from molecular dynamics simulations (MD) with explicit water and Langevin dynamics simulations (LD).*

We will now discuss the origin of non-zero σ for such a simple system as discussed above. First, let us identify the difference between Kramers theory and the simulations that give

positive value of σ. It is obvious that the difference lies in the solvent models used (implicit versus explicit). Kramers solved the barrier crossing problem in Langevin bath which assumes that the friction does not have any memory. However, the agreement with GHT and MD-RF indicates that we need time-dependent friction to accurately estimate the rate in an explicit solvent. To test whether the deviation is indeed caused due to the presence of explicit water, we have calculated MFPT of the same system using MD simulation in explicit water bath and Langevin dynamics at similar friction coefficients. The MFPT obtained from both the systems at multiple viscosities are shown in Fig. 4. The data has been fitted to the equation, $\tau/\tau_0 = \left(\eta/\eta_0\right)^{\alpha}$ to find the deviation from the Kramers behavior (inverse viscosity dependence). We observe that while the results from Langevin dynamics simulations follow Kramers theory (α=1.07), the results from MD simulations where explicit water models were used show a non-linear dependence on the viscosity with α=0.57. Since $\alpha < 1$ corresponds to $\sigma > 0$ in the fit, we can say, following the conventional literature, that the cause of internal friction is explicit water and not the dry system. *However, this goes against the very definition of internal friction which must arise from within the system.* Had the internal friction been present in our system, it would show up in the Langevin simulation as well. Therefore, the above results, the non-zero value of σ or fractional value of α in explicit water, indicate that a fit to Ansari's equation may not provide an estimate of internal friction; rather it indicates only a deviation from Kramers behavior. This deviation is already shown to depend on the frequency-dependent friction first by Bagchi and Oxtoby[12] and later by many others[35-36]. Therefore, it is evident that the origin of the fractional viscosity dependence, i.e., $\alpha < 1$, and thereby the deviation from Kramers theory must arise from the solvent motion that gives rise to the time-dependent friction. It also shows that the fractional viscosity dependence does not necessarily imply the presence of internal friction.

There are only a few studies in the literature that have questioned the interpretation of $\sigma > 0$ or $\alpha < 1$ as internal friction.[36] However, they ascribed the presence of σ to the frequency dependent friction which, as we will show next, is only partially correct. Nevertheless several studies still used viscosity dependence of rate to estimate the internal friction and attributed its origin to various specific interactions. Therefore, we also have investigated how these fit parameters, α and σ, would depend on the variations of some relevant factors in proteins' folding and created four sets of systems which vary in barrier curvature, barrier height, solute-solvent coupling and solute size independently as mentioned in the methods section.

(b) Effect of barrier curvature. Internal friction was found in case of rotational motion of phenylalanine in cytochrome C, unexposed to the solvent.[6] The origin of internal friction was attributed to a change in barrier curvature and a complex dependence of internal friction on viscosity was proposed. Therefore, to investigate the effect of barrier curvature on σ, we have varied the periodicity of the one-dimensional potential to create four different barrier curvatures 9.06 ps^{-1}, 18.13 ps^{-1}, 27.19 ps^{-1}, and 36.26 ps^{-1} corresponding to n = 2, 4, 6 and 8 respectively. The barrier height was fixed at E_b=50 kJ/mol. The choice of the barrier height was arbitrary. We calculated TC at different solvent viscosities for the four different curvatures by performing MD simulations and using reactive flux approach (MD-RF). Figure 5a shows the transmission coefficient obtained from at different curvatures. We fitted the result to both the power law to obtain α (solid line in Fig. 5a) and Ansari's equation to obtain σ (dashed line in Fig. 5a) as was discussed previously. Figure 5e shows the result of the fit in terms of power law exponent α and additional friction σ against the barrier curvature. We see that as the barrier curvature increases, σ increases (and α decreases). For the smallest barrier curvature, 9.06 ps^{-1}, σ=2.4 cP, while for the largest barrier curvature, 36.26 ps^{-1}, σ is 70.73 cP. Similarly α changes from 0.02 to 0.37 from the highest to the lowest barrier curvature.

(c) Effect of barrier height. To investigate the relationship between the barrier height and σ, the barrier height was varied at a constant barrier curvature of 13.5 ps^{-1}. All other parameters were same as the previous system. We have chosen four different heights; 6.96 kJ/mol, 11.76 kJ/mol, 25.54 kJ/mol, and 100 kJ/mol for our study. The TC values obtained at different barrier heights for all viscosities are plotted in Fig. 5b. The fits to power law and Ansari's equation show little change in σ or α with barrier height (Fig. 5f). The variation in σ for almost 14.4 times change in the barrier height is only by 0.18 cP. Similar observation was found for α also, which changes from 0.3 to 0.25 from the lowest to the highest barrier heights. Therefore the results show that σ required to explain the viscosity dependence of rate does not depend on barrier height. We would like to mention at this point that Best and coworkers used this potential to mimic dihedral motion[14] which typically have lower barrier heights (5-10 kJ/mol). We show here that any process, not necessarily the dihedral motion, with curvatures in the range mentioned above may give rise to a positive value of σ irrespective of the barrier height.

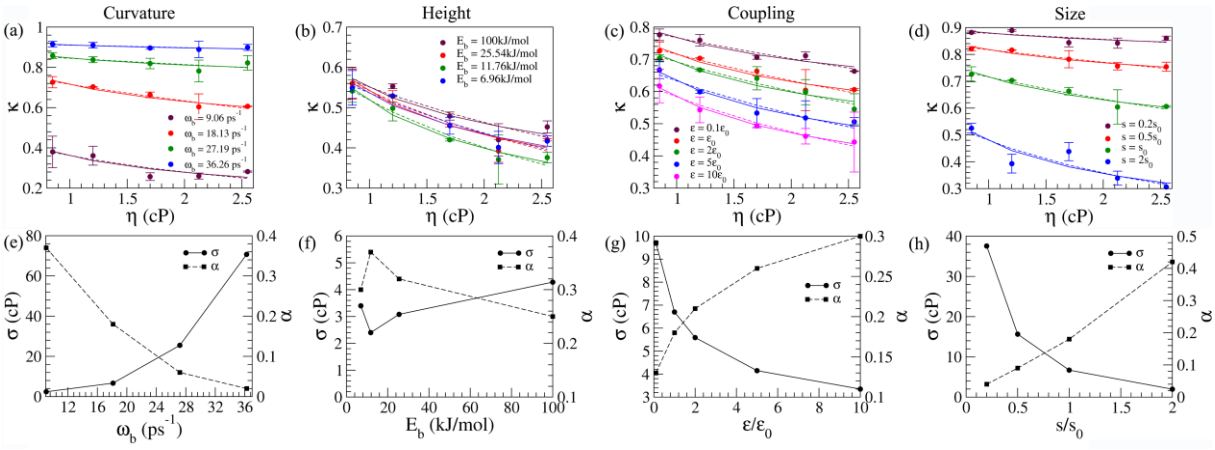

Figure 5: *Variation of transmission coefficient (TC) with viscosity for different (a) barrrier curvatures, (b) barrier heights, (c) solvent couplings, and (d) sizes of the solute particles. Each result is fitted to two equations (the power law- solid line and the Ansari's fit- dashed line) to obtain the corresponding values of additional friction σ and power law exponent α. σ and α are plotted with respect to (e) barrier curvature, (f) barrier heights, (g) solvent coupling, and (h) size of the solute particle.*

(d) Effect of solvent coupling. The interaction between the solvent and the solute was tuned by changing the epsilon parameter in the LJ equation. Note that, unlike barrier curvature and height, this quantity is not an internal parameter. Thus, for a fixed barrier curvature, ω_b=18.13 ps^{-1} and barrier height E_b=50 kJ/mol, we varied the solvent coupling. We observe that when the solvent coupling increases, TC decreases (Fig. 5c), indicating more recrossing. The larger recrossing may be attributed to the increased friction as we will see later. However, when we fit the viscosity dependence of the TC to Ansari's equation, we find that the σ decreases with the increase in solvent coupling (Fig. 5g) from 9.71 cP at the smallest coupling to 3.35 cP at the largest coupling. Similarly, α changes from 0.13 to 0.3 from weakest ($\varepsilon/\varepsilon_0$ =0.1) to strongest coupling ($\varepsilon/\varepsilon_0$ =10). This result also indicates that non-zero σ or $\alpha < 1$ is not a signature of the presence of internal friction as it depends on solvent coupling which is not an internal parameter.

(e) Effect of solute size. We have also varied the size of the solute to probe its role in the value of σ. We have used four different sizes of the solute, $0.2s$, $0.5s$, s and $2s$ where s is taken as the diameter of the carbon atom, keeping all other parameters such as barrier height (50 kJ/mol), barrier curvature (18.13 ps^{-1}), and interactions ($\varepsilon = \varepsilon_0$) constant during this variation. When the size of the solute increases, TC decreases at any given solvent viscosity (Fig. 5d). We also observe that σ decreases with the increase in size of the solute (Fig. 5h). σ varies from 37.56 cP at the smallest size of $0.2s$ to 1.96 cP at ten times the size. This is matched by an increase of α from 0.04 to 0.42.

3.5 Discussion

It is important to mention here that Best and coworkers[14] had used the same model to calculate the viscosity dependence of rate using MFPT for varying curvature. They also found using GHT that an increase in the barrier curvature leads to an increase in the memory

dependence of friction, resulting in a lesser sensitivity towards solvent viscosity with fractional viscosity dependence, $\alpha < 1$. Fig. 5a qualitatively agrees with that of Best and coworkers[14]. However, they attributed the rate behavior to internal friction originating from the ruggedness, caused due to dihedral rotations, in the protein folding landscape.[14] Also, they attributed the memory dependence to the presence of internal friction. However, results from the variations of other parameters presented here indicate otherwise. Since MFPT is computationally expensive for high barriers, we used reactive flux (MD-RF) method to calculate the rate at different viscosity for different barrier heights (Fig. 5b) and show that the fit parameters σ and α remain unchanged (Fig. 5f) across different barrier heights (6.96 kJ/mol to 100 kJ/mol). Thus, it is not just the smaller barriers (or ruggedness, which are small corrugations in the energy profiles[21], as modeled by Best et al.[14]), that gives rise to σ. Moreover, we have shown that the fractional viscosity dependence varies with the solute-solvent coupling and solute size (Fig. 5g-h). If σ truly denotes internal friction, one would not expect it to change with solute-solvent coupling which is strictly a non-internal parameter. Finally, we emphasize that the fractional viscosity dependence is not necessarily a signature of the presence of internal friction. This becomes more evident when we observe that the TC obtained from MD-RF matches with that from GHT for all systems (Figs. 6a-d). The good agreement implies that GHT, which differs from KT in terms of frequency-dependent friction, holds the key to understand the origin of the fit parameters (σ and α) and we don't need to invoke the presence of internal friction to explain the existence of σ.

The primary difference between KT and GHT is that the former includes the entire friction, ζ, for rate calculation whereas GHT uses frequency dependent friction, $\zeta(\lambda_r)$. Therefore, the above quantities may hold the key to the discrepancy from KT and the amount of σ required to fit the rate with Kramers theory. Hence, we calculated the friction kernel, the autocorrelation of the forces acting on the reaction coordinate, at the transition state for all the

cases studied (see Figs. 6e-h) to estimate ζ and $\zeta(\lambda_r)$. We plot $\zeta(\lambda_r)$ and the zero frequency friction ζ, used in KT, for different parameters at normal solvent viscosity in Figs. 6i-l. First, we notice that $\zeta(\lambda_r)$ is always smaller than ζ for all parameters. Therefore, the system encounters lesser friction than that imparted by solvent, especially for a high-frequency (faster) motion across the barrier, for the same static friction. Therefore, it is the lack of full solvent friction rather than the presence of an additional friction that causes the deviation from KT. Second, we observe that the difference between $\zeta(\lambda_r)$ and ζ, except in case of the barrier height, increases with the increase in barrier curvature, solute-solvent coupling, and solute size. We expected the increased deviation with the parameters would correspond to increase in σ. However, although σ increases with the barrier curvature, it decreases with solute-solvent coupling and solute size (Figs. 5e-h). Therefore, deviation from KT or memory friction alone, as previously thought[36], does not correlate with σ.

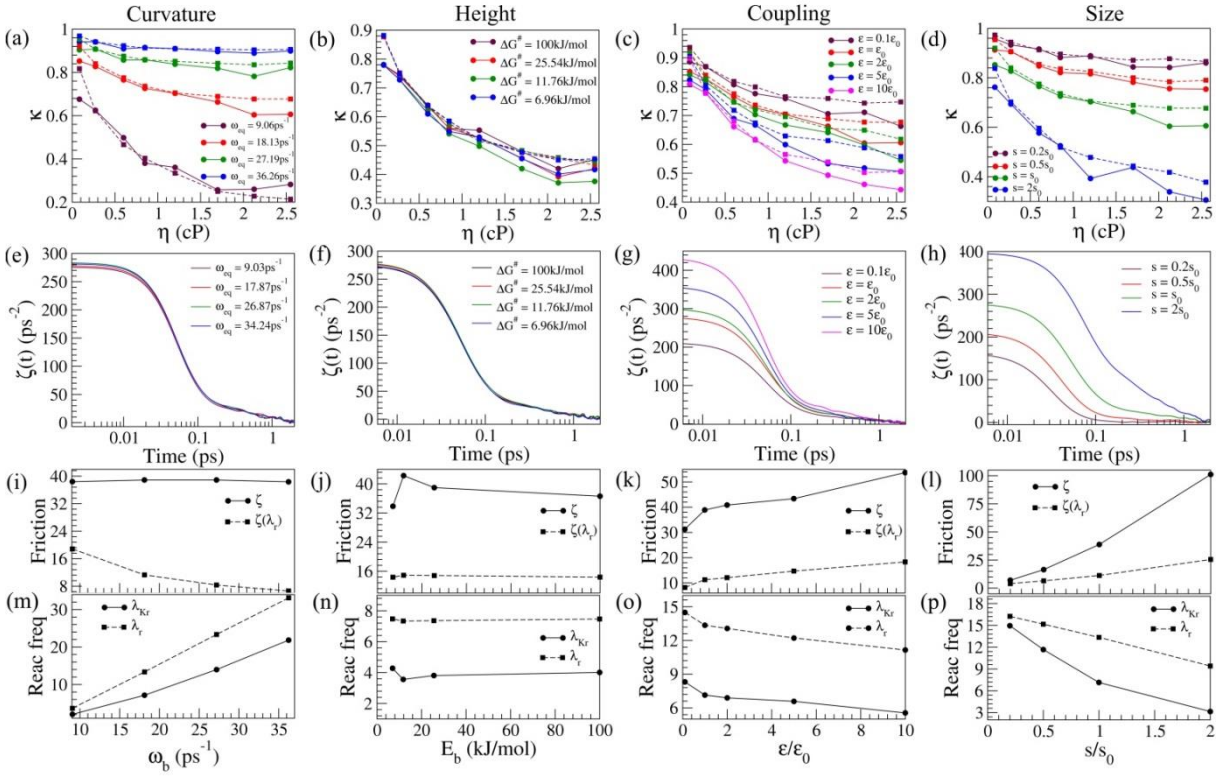

Figure 6: *(a), (b), (c) and (d) shows the comparison of TC values obtained from MD-RF and GHT for different barrier curvatures, barrier heights, solvent couplings and solute sizes respectively. Solid lines are the TC values obtained from the MD using reactive flux*

formalism. Dashed lines are the TC values obtained at different solvent viscosity from GHT. (e), (f), (g) and (h) show the friction kernal obtained at normal water viscosity for the different barrier curvatures, barrier heights, solvent couplings and solute sizes respectively. (i), (j), (k) and (l) compares the frequency dependent friction and the zero frequency friction evaluated at normal solvent viscosity. Similarly, (m), (n), (o) and (p) show the variation of the reactive frequency corresponding to GHT and KT.

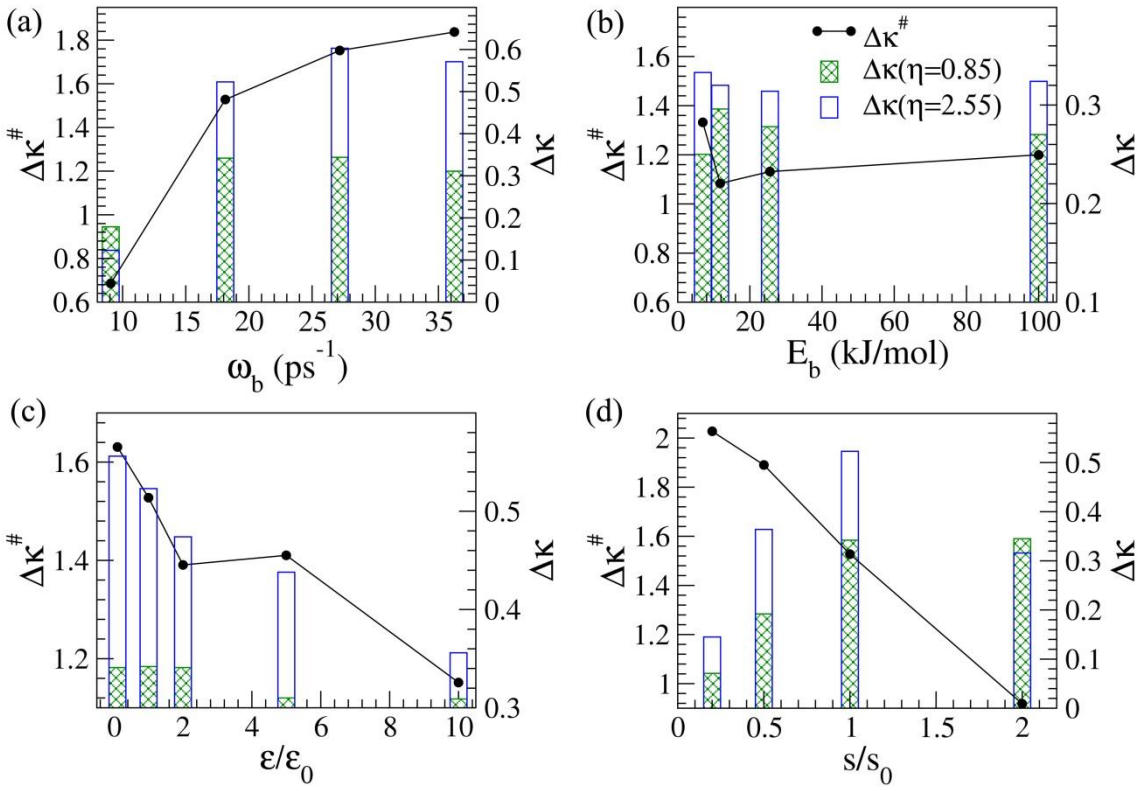

Figure 7: *(a), (b), (c) and (d) show the difference in values of TC obtained from GHT and KT, $\Delta\kappa$, at $\eta = 0.85$ (filled bar) and $\eta = 2.55$ (empty bar) and the line shows their ratios $\Delta\kappa^{\#}$ defined as, $\Delta\kappa^{\#} = \Delta\kappa(\eta = 2.55)/\Delta\kappa(\eta = 0.85)$ for variation in curvature, height, coupling and size respectively. Bars are drawn from the abscissa.*

To find a general trend of σ, we have looked into how the rate changes with viscosity in GHT compared to that in KT. For that, we plotted the difference in TCs obtained from GHT and KT, $\Delta\kappa(\eta) = \kappa_{GH}(\eta) - \kappa_{KR}(\eta)$, at two representative viscosities ($\eta_1 = 2.55$ and $\eta_2 = 0.85$) for all parameter variations (bars in Fig. 7a-d). The results show that $\Delta\kappa$, in either viscosity, does not always correspond to the magnitude of σ. For example, as the solute size increases, $\Delta\kappa$ increases, but σ decreases (Figs. 7d and 5h). Hence, memory dependence of the friction and therefore deviation from KT alone does not ensure a significant value of the σ. However, their ratios, $\Delta\kappa^{\#} = \Delta\kappa(\eta_1)/\Delta\kappa(\eta_2)$, correlate well with the change in σ for all different

parameters. The correlation becomes clear when we plot σ against $\Delta\kappa^{\#}$ for all the systems studied here (Fig. 8). Thus, we argue that when the relative deviation of the rate from KT with viscosity is large, one is expected to observe a large value of σ. The relative deviation will be large when reaction coordinate is insensitive to the change in viscosity as was observed for dihedral rotations.[14, 16, 37] *We emphasize here that it is the insensivity of the reaction coordinate motion to viscosity that gives rise to* σ.

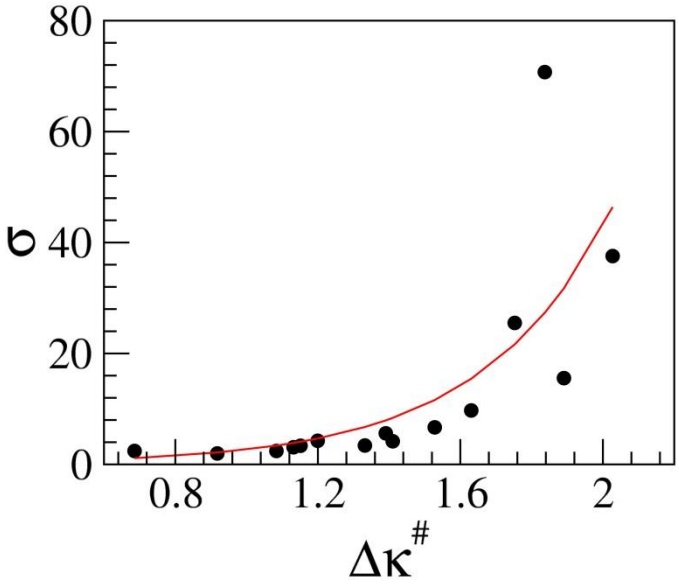

Figure 8: *Variation of* σ *obtained from all parameters studied here with* $\Delta\kappa^{\#}$. *The red line is an exponential fit. Reprinted with permission Copyright (2018) American Chemical Society.*

The changes in σ with the variation in parameters can be understood as follows. If the solvent plays an important role in the dynamics and imparts the full friction, then the sensitivity of the rate to viscosity is similar to that predicted by Kramers, resulting in lower values of σ. In case of barrier curvature variation, solute spends more time at the barrier top when the barrier curvature is lower compared to a higher barrier curvature where the motion across the barrier top is faster. Hence, the solute motion is more affected by the solvent dynamics at a lower barrier curvature compared to that at a higher barrier curvature, thereby resulting in smaller values of σ at lower barrier curvatures. The decrease in σ with increasing solvent coupling results from the fact that at higher solute-solvent interaction, the solvent plays major role in

solute dynamics. Similarly, the number of solvent molecules interacting with the solute is more for bigger solutes, thus making it more sensitive to the changes in the solvent viscosity and resulting in a low value of σ. As height of the barrier does not affect how the solvent dynamics affects the RC motion, σ values are independent of barrier height.

We have come across mainly three equations employed in the literature to confirm the presence of internal friction. These, as discussed before, are the power law fit, Ansari's fit, and the extrapolation of viscosity dependence of rate to zero viscosity. Kramers formulated the rate theory for a particle moving in a one-dimensional potential where the friction felt by the particle is encompassed in the term ζ. Many multi-atom reactions can be described using Kramers theory since the entire reaction can be depicted as a single coordinate motion (typically termed as the reaction coordinate). Then the friction ζ felt by the RC arises from the solvent degrees of freedom as well as other degrees of freedom from the solute. This is what probably prompted Ansari *et al.* to come up with an equation where a new term corresponding to the internal friction was added to the solvent friction to describe the reaction rate. If this extra friction opposed the reaction in addition to the solvent friction, then in a power law fit for a plot of rate versus viscosity, one should obtain α>1. On the contrary, to the best of our knowledge, for all the power law fits in the studies performed so far in the literature, the value of $\alpha \leq 1$. This indicates that the rate observed is always greater than that predicted by the Kramers theory, indicating a lack of full friction. Pradeep and Udgaonkar, in an experimental unfolding study of a small protein barstar by urea, have shown that the protein chain experiences an effective viscosity lower than the bulk solvent viscosity.[27] This warrants for a $\sigma < 0$. However, the fit to Ansari's equation gives $\sigma > 0$ for two reasons: (i) the constant in the numerator becomes unphysically augmented during the fit, (ii) the slope of the equation $k = c/(\eta + \sigma)$ is less than that of $k = c/\eta$, the latter being less sensitive to viscosity changes fits the equation with the lesser slope. Likewise, extrapolation of rate

obtained at high viscosity regime to zero viscosity to calculate the internal friction also suffers from the two following issues. Kramers predicts a different rate behavior for low viscosity regime, where rate is proportional to the friction as confirmed in experimental[38] and simulation studies[39-40]. Secondly, this method assumes that memory effects are absent which are not necessarily satisfied in real systems.[31] Hence, the method of using a non-zero intercept of the rate at zero viscosity to estimate internal friction requires a critical examination. This does not imply that there is no internal friction, i.e., friction offered from the solute degrees of freedom. The frequency-dependent friction may have contributions from both internal and solvent degrees of freedom. However, it does not reflect in the fit parameter σ when the rate is calculated in the presence of solvent since it involves decoupling of the frequency-dependent friction into solvent and solute contributions.

3.6 Conclusion

In this chapter, we have employed a simple model system used earlier by Best *et al.*[14] to arrive at an entirely different conclusion. We have shown in the present study that the lack of the full solvent friction can result in a non-zero value of σ. The solvent coupling and Langevin dynamics studies also proved that solvent affects the value of σ, thereby raising doubts in the interpretation of σ as internal friction. We show that the relative deviation of rate with viscosity from Kramers prediction correlate with σ. We also find that a low barrier curvature, high solvent interaction, and a large solute results in a lower value of σ. In an experimental study to quantify internal friction in unfolded and intrinsically disordered proteins, Soranno *et al*. had found that the magnitude of "*internal friction"* correlates with the compactness of the unfolded protein. They observed that the "*internal friction*" is dominant for compact unfolded state of the protein and decreases for the intrinsically disordered proteins.[17] This correlates with our size-dependent results. We have shown that the deviation

from Kramers law is not necessarily a consequence of ruggedness in the landscape. We believe that deviations from Kramers law can be observed whenever the assumptions used in the theory does not hold for the system under consideration. This may include frequency dependent friction, multidimensional reaction coordinate, anharmonicity in the potential, non-linear solute-solvent interactions, presence of tunneling, etc. Hence, one should consider these possibilities prior to the calculation of internal friction. Breakdown of Kramers theory can give rise to two kinds of situation, one where the deviation from the KT prediction is constant at all viscosities and another case where the deviation is different at different viscosities. The first case will not affect the value of σ whereas the second case would give rise to σ whose amount will be determined by the extent of change in the deviation with viscosity. We propose that when the motion of RC is insensitive to the change in the solvent viscosity, the need for a non-zero σ arises. We believe that if the solvent plays a dominant role in the reaction dynamics then the rate will be sensitive to viscosity changes and one would observe less σ.

Even though we have used a simple model system, we believe that the conclusions that we arrived at are applicable in general for any process with memory dependence in friction. Our results on a simple model system show that non-zero σ can arise as a consequence of memory effects that deviate the rate from Kramers theory differently at different viscosity. Hence, a non-zero value of σ cannot be readily attributed to the presence of internal friction. Therefore, the present study brings out the lacuna in the current practice of evaluating internal friction and its association to various microscopic origins. This necessitates the development of a new approach to estimate internal friction accurately in experimental and simulation studies.

References

1. Kramers, H. A., Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **1940,** *7* (4), 284-304.
2. Spångberg, D.; Rey, R.; Hynes, J. T.; Hermansson, K., Rate and Mechanisms for Water Exchange around Li+(aq) from MD Simulations. *The Journal of Physical Chemistry B* **2003,** *107* (18), 4470-4477.
3. Velsko, S. P.; Fleming, G. R., Photochemical isomerization in solution. Photophysics of diphenyl butadiene. *The Journal of Chemical Physics* **1982,** *76* (7), 3553-3562.
4. Velsko, S. P.; Waldeck, D. H.; Fleming, G. R., Breakdown of Kramers theory description of photochemical isomerization and the possible involvement of frequency dependent friction. *The Journal of Chemical Physics* **1983,** *78* (1), 249-258.
5. Wensley, B. G.; Batey, S.; Bone, F. A. C.; Chan, Z. M.; Tumelty, N. R.; Steward, A.; Kwa, L. G.; Borgia, A.; Clarke, J., Experimental evidence for a frustrated energy landscape in a three-helix-bundle protein family. *Nature* **2010,** *463* (7281), 685-688.
6. Sashi, P.; Ramakrishna, D.; Bhuyan, A. K., Dispersion Forces and the Molecular Origin of Internal Friction in Protein. *Biochemistry* **2016,** *55* (33), 4595-4602.
7. Cellmer, T.; Henry, E. R.; Hofrichter, J.; Eaton, W. A., Measuring internal friction of an ultrafast-folding protein. *Proceedings of the National Academy of Sciences* **2008,** *105* (47), 18320-18325.
8. Qiu, L.; Hagen, S. J., Internal friction in the ultrafast folding of the tryptophan cage. *Chemical Physics* **2005,** *312* (1–3), 327-333.
9. Narayanan, R.; Pelakh, L.; Hagen, S. J., Solvent Friction Changes the Folding Pathway of the Tryptophan Zipper TZ2. *Journal of Molecular Biology* **2009,** *390* (3), 538-546.
10. Northrup, S. H.; Hynes, J. T., The stable states picture of chemical reactions. I. Formulation for rate constants and initial condition effects. *The Journal of Chemical Physics* **1980,** *73* (6), 2700-2714.
11. Grote, R. F.; Hynes, J. T., The stable states picture of chemical reactions. II. Rate constants for condensed and gas phase reaction models. *The Journal of Chemical Physics* **1980,** *73* (6), 2715-2732.
12. Bagchi, B.; Oxtoby, D. W., The effect of frequency dependent friction on isomerization dynamics in solution. *The Journal of Chemical Physics* **1983,** *78* (5), 2735-2741.
13. Ansari, A.; Jones, C. M.; Henry, E. R.; Hofrichter, J.; Eaton, W. A., The role of solvent viscosity in the dynamics of protein conformational changes. *Science* **1992,** *256* (5065), 1796-1798.
14. de Sancho, D.; Sirur, A.; Best, R. B., Molecular origins of internal friction effects on protein-folding rates. *Nat Commun* **2014,** *5*.
15. Zheng, W.; de Sancho, D.; Best, R. B., Modulation of Folding Internal Friction by Local and Global Barrier Heights. *The Journal of Physical Chemistry Letters* **2016,** *7* (6), 1028-1034.
16. Echeverria, I.; Makarov, D. E.; Papoian, G. A., Concerted Dihedral Rotations Give Rise to Internal Friction in Unfolded Proteins. *Journal of the American Chemical Society* **2014,** *136* (24), 8708-8713.
17. Soranno, A.; Buchli, B.; Nettels, D.; Cheng, R. R.; Müller-Späth, S.; Pfeil, S. H.; Hoffmann, A.; Lipman, E. A.; Makarov, D. E.; Schuler, B., Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proceedings of the National Academy of Sciences* **2012,** *109* (44), 17800.

18. Zheng, W.; De Sancho, D.; Hoppe, T.; Best, R. B., Dependence of Internal Friction on Folding Mechanism. *Journal of the American Chemical Society* **2015,** *137* (9), 3283-3290.
19. Schulz, J. C. F.; Schmidt, L.; Best, R. B.; Dzubiella, J.; Netz, R. R., Peptide Chain Dynamics in Light and Heavy Water: Zooming in on Internal Friction. *Journal of the American Chemical Society* **2012,** *134* (14), 6273-6279.
20. Jas, G. S.; Eaton, W. A.; Hofrichter, J., Effect of Viscosity on the Kinetics of α-Helix and β-Hairpin Formation. *The Journal of Physical Chemistry B* **2001,** *105* (1), 261-272.
21. Zwanzig, R., Diffusion in a rough potential. *Proceedings of the National Academy of Sciences of the United States of America* **1988,** *85* (7), 2029-2030.
22. Wensley, B. G.; Kwa, L. G.; Shammas, S. L.; Rogers, J. M.; Clarke, J., Protein Folding: Adding a Nucleus to Guide Helix Docking Reduces Landscape Roughness. *Journal of Molecular Biology* **2012,** *423* (3), 273-283.
23. Kwa, L. G.; Wensley, B. G.; Alexander, C. G.; Browning, S. J.; Lichman, B. R.; Clarke, J., The Folding of a Family of Three-Helix Bundle Proteins: Spectrin R15 Has a Robust Folding Nucleus, Unlike Its Homologous Neighbours. *Journal of Molecular Biology* **2014,** *426* (7), 1600-1610.
24. Chung, H. S.; Piana-Agostinetti, S.; Shaw, D. E.; Eaton, W. A., Structural origin of slow diffusion in protein folding. *Science* **2015,** *349* (6255), 1504-1510.
25. Schulz, J. C. F.; Miettinen, M. S.; Netz, R. R., Unfolding and Folding Internal Friction of β-Hairpins Is Smaller than That of α-Helices. *The Journal of Physical Chemistry B* **2015,** *119* (13), 4565-4574.
26. Yasin, U. M.; Sashi, P.; Bhuyan, A. K., Expansion and Internal Friction in Unfolded Protein Chain. *The Journal of Physical Chemistry B* **2013,** *117* (40), 12059-12064.
27. Pradeep, L.; Udgaonkar, J. B., Diffusional Barrier in the Unfolding of a Small Protein. *Journal of Molecular Biology* **2007,** *366* (3), 1016-1028.
28. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983,** *79* (2), 926-935.
29. Bergsma, J. P.; Reimers, J. R.; Wilson, K. R.; Hynes, J. T., Molecular dynamics of the A+BC reaction in rare gas solution. *The Journal of Chemical Physics* **1986,** *85* (10), 5625-5643.
30. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **2008,** *4* (3), 435-447.
31. Walser, R.; Mark, A. E.; van Gunsteren, W. F., On the validity of Stokes' law at the molecular level. *Chemical Physics Letters* **1999,** *303* (5), 583-586.
32. Walser, R.; Hess, B.; Mark, A. E.; van Gunsteren, W. F., Further investigation on the validity of Stokes–Einstein behaviour at the molecular level. *Chemical Physics Letters* **2001,** *334* (4–6), 337-342.
33. Gertner, B. J.; Wilson, K. R.; Hynes, J. T., Nonequilibrium solvation effects on reaction rates for model SN2 reactions in water. *The Journal of Chemical Physics* **1989,** *90* (7), 3537-3558.
34. Samanta, N.; Chakrabarti, R., End to end loop formation in a single polymer chain with internal friction. *Chemical Physics Letters* **2013,** *582*, 71-77.
35. Hara, K.; Kiyotani, H.; Bulgarevich, D. S., Effect of frequency-dependent friction for the excited-state isomerization dynamics of 2-vinylanthracene in solution. *Chemical Physics Letters* **1995,** *242* (4), 455-459.
36. Chandra Bag, B.; Hu, C.-K.; Suan Li, M., Colored noise, folding rates and departure from Kramers' behavior. *Physical Chemistry Chemical Physics* **2010,** *12* (37), 11753-11762.

37. Daldrop, J. O.; Kappler, J.; Brünig, F. N.; Netz, R. R., Butane dihedral angle dynamics in water is dominated by internal friction. *Proceedings of the National Academy of Sciences* **2018,** *115* (20), 5169.
38. Rondin, L.; Gieseler, J.; Ricci, F.; Quidant, R.; Dellago, C.; Novotny, L., Direct measurement of Kramers turnover with a levitated nanoparticle. *Nature Nanotechnology* **2017,** *12*, 1130.
39. Klimov, D. K.; Thirumalai, D., Viscosity Dependence of the Folding Rates of Proteins. *Physical Review Letters* **1997,** *79* (2), 317-320.
40. Best, R. B.; Hummer, G., Diffusive Model of Protein Folding Dynamics with Kramers Turnover in Rate. *Physical Review Letters* **2006,** *96* (22), 228104.

Chapter 4

Separating memory effects from internal friction in viscosity dependence of rate

4.1 Overview

Deviation of viscosity dependence of the rate from Kramers theory is often attributed to the presence of internal friction in many protein folding studies. We have demonstrated in the previous chapter that the presence of memory dependent friction, due to explicit solvent, can be misinterpreted as internal friction. Here we show that even a simple system such as diatomic particle and butane can exhibit memory dependent friction even in the absence of explicit solvent. Moreover, we show that although both the above can exhibit fractional viscosity dependence, the decoupling is possible by comparing the actual rate and not by its viscosity dependence. This study therefore opens up the need to revisit the interpretation of internal friction is complex systems like proteins.

4.2 Introduction

A reaction is often depicted as motion along a one-dimensional degree of freedom called reaction coordinate (RC). The motion of RC is frequently hindered (sometimes assisted) by the other degrees of freedom which can arise from solvent molecules or from within the system. The latter, friction that originates from within the system is termed as internal friction or dry friction and thus reduces the rate of the reaction.[1]

Internal friction has been typically found in protein folding[2-7], unfolded state dynamics[8-9], chain dynamics of intrinsically disordered proteins[1] and folding transition-path times[10]. It has

been argued that alpha helix proteins have more internal friction than beta hairpin forming peptides.[2-3, 11] This shows that beta hairpin can fold faster and thus is relevant in the context that many neurodegenerative diseases involves the formation of misfolded betasheets.[12] Thus discerning the origin of internal friction can help us to understand why some proteins fold faster while others don't. An interesting study in this regard was carried out by Clarke and co-workers investigating into the folding times of three different α-spectrin domains.[5, 13-14] Even though the domains R15, R16 and R17 are similar in structure and thermodynamic stabilities, R15 folds three orders of magnitude faster than R16 and R17. They ascribed the difference in timescales of folding to landscape roughness surfacing due to internal friction in the core. However, origin of internal friction is a highly debated subject matter. Different studies attribute the origin of internal friction to several factors such as concerted dihedral rotations[8], hydrogen bonding[2, 15], non-native salt bridge formation[10], non-native dispersion forces[16], etc.

In literature, we came across three different methods that are used to calculate internal friction. The first method was introduced by Ansari et al.[17], when they observed that the rate of conformational changes in myoglobin protein did not always vary linearly with the solvent viscosity. They found that the equation ($k = C/(\eta + \sigma)$) could explain the plot obtained. The friction term, σ, which was used in addition to the solvent viscosity, η, was termed internal friction. The second method that is used widely to quantify internal friction is the fractional dependence of the rate on viscosity.[3-4, 7, 15, 18] The reaction time versus viscosity plot is fitted to $\tau/\tau_0 = (\eta/\eta_0)^{\alpha}$ where η_0 is generally taken to be the normal solvent viscosity and τ_0 is the average time taken to complete the reaction at η_0. An α value of unity indicates absence of internal friction and any value of α less than unity is generally attributed to the presence of internal friction. The third method is the extrapolation of reaction time versus viscosity plot

to zero viscosity, $\tau = a\eta + \tau_i$.[1, 4, 8, 15] A non-zero y-intercept, τ_i is commonly accredited to the existence of internal friction in the system.

All the above methods heftily make two assumptions. The first assumption is that Kramers theory[19] is valid in the systems of concern and the second assumption is that the system dynamics falls in the high friction diffusion regime. Only when these two assumptions are valid, one can use deviation from Kramers behavior, $k \propto 1/\eta$, to measure internal friction. However, deviation from Kramers theory can happen even when there is no internal friction. It was shown by Grote and Hynes that if the solvent motions cannot not catch up with the system's motion along the reaction coordinate will give rise to an effective time dependent (memory) friction, which as was shown by Bagchi et al.,[20] can give rise to a fractional viscosity dependence of rate.[21]

Thus an imperative caveat of these three methods mentioned above is that one cannot really demonstrate that the estimated value is in fact surfacing from the internal system or due to memory effects. Nevertheless, these methods have been repeatedly used in the literature for estimation of internal friction without validating the assumptions.

In the previous chapter, we have shown that a single Lennards-Jones particle crossing a potential energy barrier in water, with no scope for experiencing internal friction (since the solute is a single atom) gave rise to fractional viscosity dependence with α less than unity and a non-zero value of σ in the equation $k = C/(\eta + \sigma)$ due to the presence of memory effects from the solvent water molecules.[22] Thus our result showed that memory effects arising from microscopic motions of water can give a false impression of the presence of internal friction.

The objective of the present chapter is therefore to explore a way to extract actual internal friction present in a system. Our previous system did not have a scope of any internal friction.

Therefore in the present study we chose two systems, a diatomic model and butane, that may possess internal friction and yet it is simple enough to investigate in detail. In fact, a recent study by Netz and coworkers claimed that the butane dihedral dynamics is influenced by internal friction and that the internal friction disappears on removing orthogonal degrees of freedom (achieved by constraining three carbon atoms in butane).[23] We are going to show here that it is not trivial to extract the internal friction in such cases. We show in this chapter that both the diatom model and butane exhibit only memory effects.

This study therefore investigates the two physical phenomena (internal friction and memory effects) giving rise to a similar behavior – fractional viscosity dependence of the rate with viscosity due to deviation from Kramers theory. The more complex systems like proteins might exhibit both the effects and therefore, understanding this simple system thoroughly will lead to a better comprehension of the experimental results.

4.3 Method and Design

(a) Diatom model – The model consist of two atoms in the presence of an external double well potential (see Fig. 1a). The particles are free to move in all the three directions while the potential acts along the distance between the particles in the x direction. The potential is the only form of interaction between the two particles. The external potential is of the form $V(x) = \frac{V}{b^4}((x - a/2)^2 - b^2)^2$ where V is the barrier height and $(a/2) \pm b$ are the locations of the two local minimas. For the system considered here $V = 9\ kJ/mol$ and $(a - b) = 0.375$ were used. This is the simplest model with internal interaction leading to a process. In our earlier model in previous chapter, a single particle was used and therefore there was no chance of an internal interaction present.

(b) United butane model – We also considered a simple, yet realistic model system comprising a united atom model of butane, which was solvated with SPC/E[24] water molecules in a cubic box. GROMOS 58 force field[25] was used for butane. All the bonds and angles of the butane were constrained using LINCS[26] algorithm in the simulations. Therefore, this model enables us to monitor the dihedral dynamics.

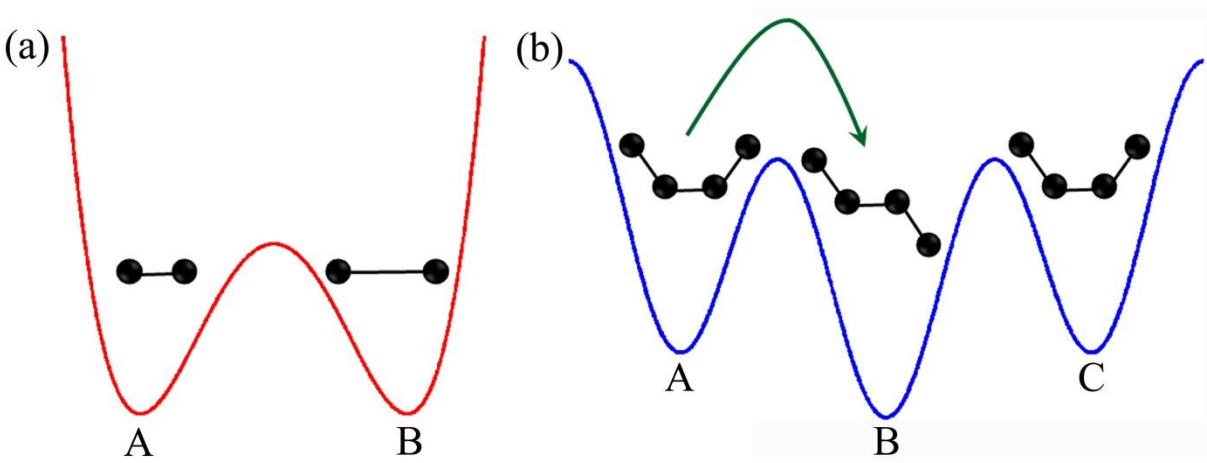

Figure 1: *Schematic figures for (a) diatom model with external potential on the separation along the x direction. (b) the dihedral potential of the united butane model.*

Molecular dynamics simulations and Langevin dynamics simulations were performed using the GROMACS[27] software. In molecular dynamics simulations, the viscosity of the solvent was changed by scaling the solvent masses and time step; thereby not disturbing the equilibrium distribution of properties.[28-29] NVT simulations were done with the temperature fixed at 300K using a velocity rescale thermostat[30] with a coupling constant of 0.1ps. In langevin dynamics simulations, also performed at 300K, the viscosity was changed by changing the friction of the solvent medium. Employing both MD and LD simulations, we have calculated the time taken for distance change in the diatom model along x direction and dihedral transition from cis form to trans form of butane as shown in Fig. 1. The mean first passage time calculated for both the reactions were averaged over 1000 transitions for each viscosity.

4.4 Results and Discussion

(a) **MFPT for barrier crossing in diatom model**. Using molecular dynamics simulations with explicit solvent water molecules to find the rate of barrier crossing in diatom model warrants memory effects due to solvent motions. Presence of memory effects are known to cause deviation from Kramers theory resulting in a non-zero value of σ which can be misinterpreted as the internal friction.[22] Hence, to eliminate memory effects arising from solvent degrees of freedom, we calculate the rate of barrier crossing in diatomic model using Langevin dynamics. Langevin dynamics uses Langevin equation thus replacing the solvent degrees of freedom with a frictional force and a random force with no memory. We carried out 1000 simulations to calculate the mean first passage time (MFPT) at each viscosity using Langevin dynamics and then fit the data to the equation $\tau/\tau_0 = (\eta/\eta_0)^{\alpha}$. We obtain that $\alpha = 0.90$ as shown in Fig. 2a. This value is slightly less than unity. The deviation from Kramers theory, which predicts that $\alpha = 1$, could arise due to memory effects stemming from solute degrees of freedom or due to the presence of internal friction. We then constrained one of the atoms in the diatom model allowing only the other atom to move and then calculated the MFPT using Langevin dynamics. Following the fit to equation $\tau/\tau_0 = (\eta/\eta_0)^{\alpha}$ we obtain $\alpha = 0.97$ which implies that the constrained diatom follows Kramers behavior (see Fig. 2a). Thus, constrained diatom model neither exhibits any sort of memory effects nor possesses internal friction.

Let's now compare the rate between the free and the constrained diatomic models in Langevin bath. If the deviation from Kramers behavior in free diatom model was due to internal friction, the MFPT in case of free diatom at any viscosity should be higher than that of constrained diatom that displays no internal friction since internal friction is known to oppose the motion of RC, as also demonstrated in the paragraph below. However, when we

examine the MFPT for barrier crossing in both free and constrained diatom model, we find that at all viscosities the MFPT for the free diatom is lower than the constrained model (see Fig. 2b). This confirms that it is not the presence of internal friction that causes deviation from Kramers theory in the free diatom model. Further, presence of memory effects are known to give rise to a higher rate (lower MFPT) compared to the Kramers theory.[31-33] Therefore, we can conclude that the memory effects arising from the two independent solute degrees of freedom results in $\alpha = 0.90$ in free diatom model rather than internal friction.

To illustrate that the presence of internal friction indeed decreases the rate (increase the MFPT), we add externally a potential that mimics internal friction following the work of Booji and Wiechen[34]. The form of the potential that mimics internal friction is $F_{IF} = -\xi_{IF} v_{RC}$, where F_{IF} is the force due to the internal friction, ξ_{IF} is the friction coefficient that determines the strength of internal friction, and v_{RC} is the velocity of the reaction coordinate. Since internal friction opposes the motion of RC, the force is directed in the negative direction of motion of the RC. We then calculate the MFPT at different viscosities using Langevin dynamics after adding the internal friction potential to the constrained diatom model. We observe that the time taken to cross the barrier is higher in the case of the system with added internal friction compared to the constrained diatom model which has no internal friction (see Fig. 2b). However, when fit to the equation, $\tau/\tau_0 = (\eta/\eta_0)^{\alpha}$, it yields $\alpha = 0.63$ showing significant deviation from Kramers behavior as can be seen from Fig. 2a. Thus, both the memory effects and internal friction can cause a deviation from Kramers theory resulting in a value of α less than one. However, internal friction always opposes the motion of the RC and thus decreases the rate of the reaction[1, 35] whereas memory effects may increase or decrease[20] the rate in comparison to the Kramers prediction.

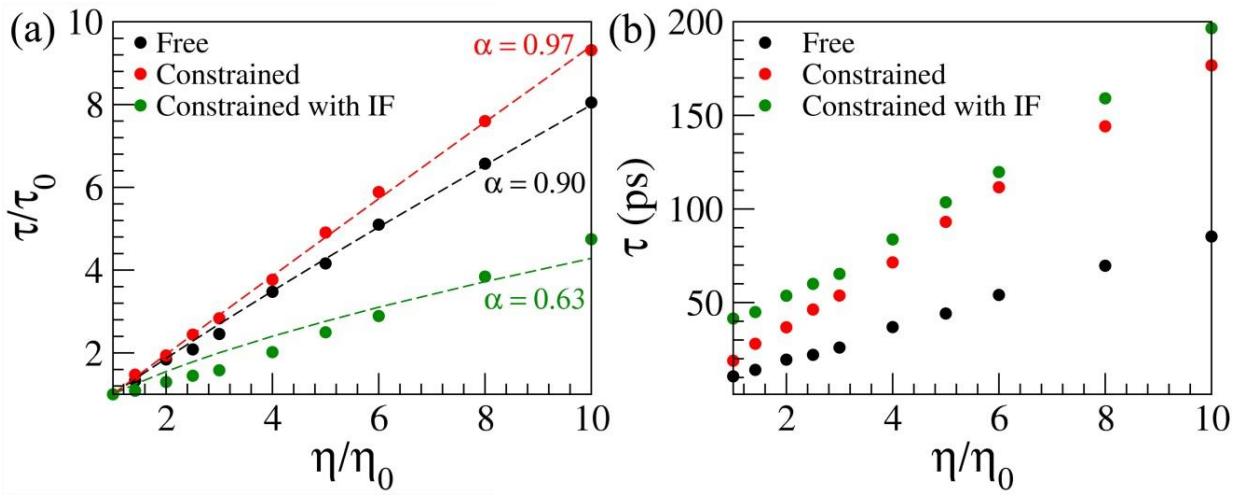

Figure 2: *(a) τ/τ_0 plotted against (η/η_0) where τ_0 is the time taken to cross the barrier at normal solvent viscosity η_0. The dashed lines are fit to the equation $\tau/\tau_0 = (\eta/\eta_0)^{\alpha}$. (b) MFPT at all viscosities for the three different cases: free diatom, constrained diatom and constrained diatom with internal friction.*

We have demonstrated the effect of internal friction and memory effects using simplest possible model systems. Now, we will discuss the effects of the above for a more realistic system. One of the simplest realistic systems where one can investigate the presence of internal friction is the dihedral rotation in butane. Dihedral rotations have also been accredited the origin of internal friction in many experimental and simulation protein folding studies.[3-4, 8, 18] Therefore, here we investigate the dihedral dynamics in butane at different viscosities. We performed several molecular dynamics simulations with explicit solvent water molecules starting from the state *A* (see Fig. 1) to find the MFPT for the dihedral barrier crossing process. We observed that with the increase in viscosity the time taken for the dihedral crossing increased as expected (see Fig. 3a). When we fitted the plot of MFPT versus viscosity to equation, $\tau/\tau_0 = (\eta/\eta_0)^{\alpha}$, we found that $\alpha = 0.25$, which is significantly lesser than unity (see Fig. 3a). As mentioned above for the diatomic model system, solvent memory effects also results in a value less than unity.[22] To verify whether the present scenario arises due to solvent memory effects or internal friction, we calculated dihedral transition time using Langevin dynamics (LD) at different friction values. If internal friction

was the source of fractional values of α, then α value should remain same in both MD and LD simulations since the only difference between the two is in the solvent used. The results are shown in Fig. 3a. The fit of the plot to the same equation yields a value of $\alpha = 0.92$, thus confirming that the majority of the non-linear dependence of butane dihedral crossing on viscosity in MD emerged due to the memory effects arising from the solvent molecules rather than internal friction. However, the persisting small deviation from Kramers theory ($\alpha = 0.92$) in LD of the butane still needs to be explained. This could come from either the internal friction or memory effect within the butane molecule.

Netz and co-workers had recently studied dihedral crossing in united atom model of butane molecule using MD simulations.[23] They found that when three carbon atoms in butane are constrained, the rate of dihedral crossing is ten times slower from that of the free butane at higher viscosities. Since the free energy profiles for both free and constrained butane are same, they argued that the difference in rates must then result from the difference in friction acting on the RC. They claimed that the difference in friction was due to the presence of internal friction in free butane arising from the extra degrees of freedom that was otherwise constrained in the constrained butane. To investigate if internal friction was absent in the dynamics of constrained butane, we performed MD and LD simulations of constrained butane (where three atoms were frozen) to find the dihedral crossing times at different solvent viscosities. When we fit the MFPT against viscosity to the equation, $\tau/\tau_0 = (\eta/\eta_0)^{\alpha}$, we found that $\alpha = 0.80$ for MD simulations and $\alpha = 0.99$ for LD simulations (see Fig. 3b). This indicates that the constrained butane in LD follows Kramers behavior and does not exhibit any internal friction. The deviation observed from Kramers theory in MD simulations of constrained butane can be then attributed to solvent memory effects similar to what we observed for the MD simulation of the free butane system discussed above.

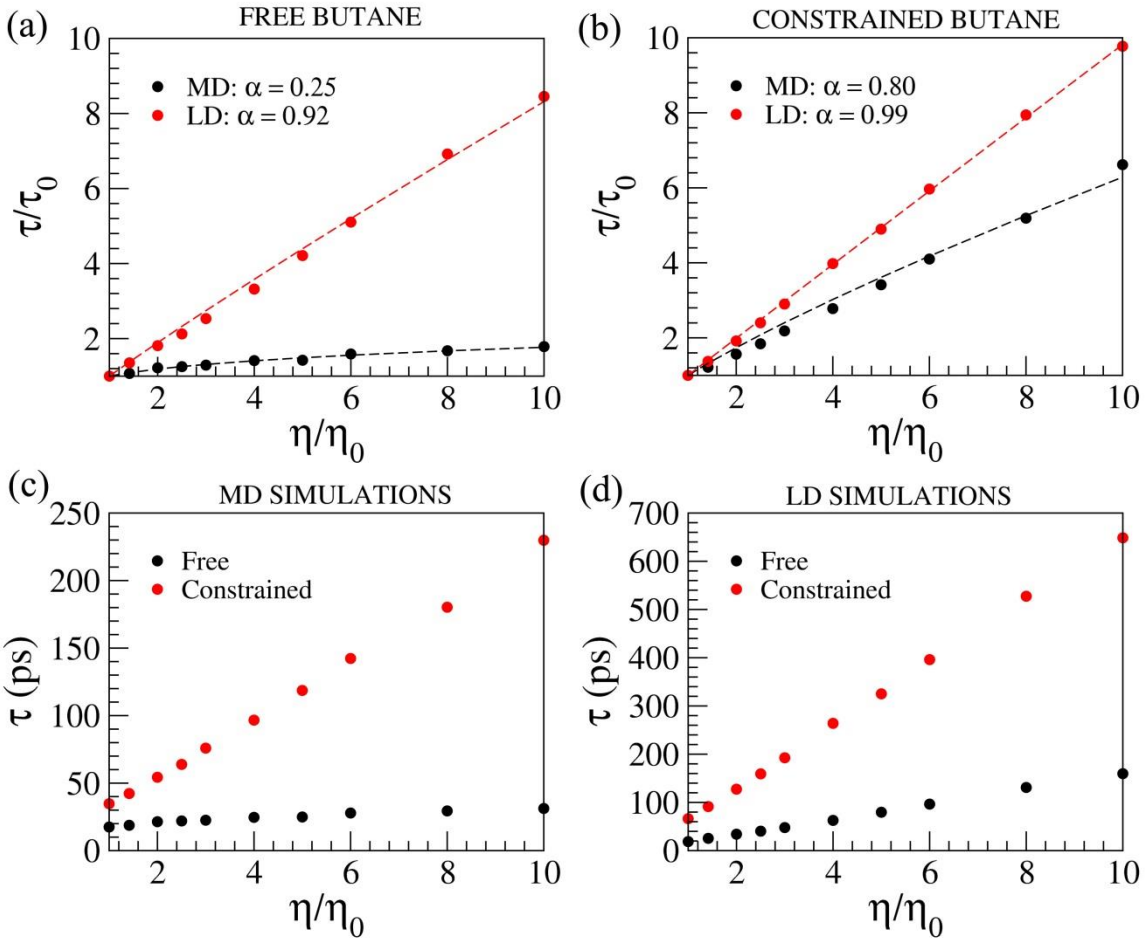

Figure 3: τ/τ_0 *plotted against* (η/η_0) *for MD and LD simulations for (a) free butane (b) constrained butane. Dashed lines are fit to the equation,* $\tau/\tau_0 = (\eta/\eta_0)^{\alpha}$. *MFPT for free and constrained butane at different viscosities as obtained from (c) MD simulations (d) LD simulations.*

Netz and coworkers claimed that the free butane possesses internal friction while the constrained system does not. We can now compare our results in butane system with the diatomic model above and reason that if the deviation from Kramers theory in free butane in LD simulations were due to internal friction as claimed by Netz et al[23], one should observe a higher MFPT for free butane compared to that in the constrained butane at any given viscosity. However, concurrent with the study by Netz et al[23], Fig. 3c-d shows that both in MD and LD simulations the MFPT at each viscosity in free butane system is lower than that of the constrained butane (rate is inverse to the MFPT; so rate for dihedral transition is higher in free butane compared to that of the constrained butane). Thus, the rate difference between

free and constrained butane cannot be ascribed to internal friction arising from the orthogonal degrees of freedom.

While Netz et al[23] maintain that the decrease in rate (increase in MFPT) in constrained butane in their MD simulations is due to the absence of internal friction, we reason that the lowering of rate can arise either from solvent memory effects and/or system memory effects. Even though, the solvent used is same in both free and constrained MD simulations, the solvent friction felt by the reaction coordinate can be very different. In the free butane, motions of all the four atoms may be hindered or assisted by the water molecules whereas in constrained scenario the water molecules affect only the motion of one atom. The possibility of solvent memory effects can be eliminated by performing LD simulations of free and constrained butane, which we have performed. We had obtained $\alpha = 0.99$ for constrained butane and $\alpha = 0.92$ for free butane in LD simulations (see Fig. 3a-b). The difference in α values between free and constrained butane in LD simulations (0.07) is much smaller than that in MD simulations (0.55). This shows that much of the difference in rate dependence on viscosity in free and constrained butane in MD simulations arose due to solvent memory effects (since other elements like system memory effects and internal friction would remain same in both LD and MD simulations). The α value, 0.99, obtained from constrained butane LD simulations shows that there is no internal friction or memory effects in constrained butane. The α value, 0.92, obtained in free butane LD simulations is slightly less than unity and can only ensue from system memory effects or internal friction. Then again, as explained before, since the rate in free butane is higher than that of constrained butane, one cannot attribute this trend to internal friction. Therefore the likely origin of the deviation from Kramers theory for butane dihedral rotation is the memory effects coming from within the butane degrees of freedom.

4.5 Conclusion

The present study therefore shows the interplay between two opposing effects – memory effects due to orthogonal degrees of freedom and internal friction. Both these effects cause deviation from the Kramers theory by having an α value less than unity, which is strictly valid only in case of an one dimensional motion considered in Langevin bath (i.e., constrained diatom and butane in LD). We showed that the majority of memory effects come from the solvents. However, when more than one degrees of freedom are involved (free diatom and butane), memory starts playing a role even when there is no explicit solvent and this memory effect increases the rate with respect to the Kramers prediction. Instead, as we have shown here, when internal friction is present the rate decreases.

Therefore, the present study provides a simple guideline to estimate the presence of internal friction. Typically, an α value less than unity is attributed to the presence of internal friction. However, one should also compare whether the rate is higher or lower than the Kramers prediction. If the rate is higher than Kramers prediction, it would imply the presence of memory effect, otherwise it would generally indicate the presence of internal friction. However if both the effects are concurrently present in the systems, which is probable in complex systems like proteins, the effects of one may dominate the other. This renders the decoupling of the two effects more challenging. Therefore, we reason that new methods need to be developed to estimate internal friction accurately even in the presence of memory effects.

References

1. Soranno, A.; Buchli, B.; Nettels, D.; Cheng, R. R.; Müller-Späth, S.; Pfeil, S. H.; Hoffmann, A.; Lipman, E. A.; Makarov, D. E.; Schuler, B., Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proceedings of the National Academy of Sciences* **2012,** *109* (44), 17800.

2. Schulz, J. C. F.; Miettinen, M. S.; Netz, R. R., Unfolding and Folding Internal Friction of β-Hairpins Is Smaller than That of α-Helices. *The Journal of Physical Chemistry B* **2015,** *119* (13), 4565-4574.
3. Zheng, W.; De Sancho, D.; Hoppe, T.; Best, R. B., Dependence of Internal Friction on Folding Mechanism. *Journal of the American Chemical Society* **2015,** *137* (9), 3283-3290.
4. de Sancho, D.; Sirur, A.; Best, R. B., Molecular origins of internal friction effects on protein-folding rates. *Nat Commun* **2014,** *5*.
5. Wensley, B. G.; Batey, S.; Bone, F. A. C.; Chan, Z. M.; Tumelty, N. R.; Steward, A.; Kwa, L. G.; Borgia, A.; Clarke, J., Experimental evidence for a frustrated energy landscape in a three-helix-bundle protein family. *Nature* **2010,** *463* (7281), 685-688.
6. Cellmer, T.; Henry, E. R.; Hofrichter, J.; Eaton, W. A., Measuring internal friction of an ultrafast-folding protein. *Proceedings of the National Academy of Sciences* **2008,** *105* (47), 18320-18325.
7. Qiu, L.; Hagen, S. J., Internal friction in the ultrafast folding of the tryptophan cage. *Chemical Physics* **2004,** *307* (2), 243-249.
8. Echeverria, I.; Makarov, D. E.; Papoian, G. A., Concerted Dihedral Rotations Give Rise to Internal Friction in Unfolded Proteins. *Journal of the American Chemical Society* **2014,** *136* (24), 8708-8713.
9. Soranno, A.; Holla, A.; Dingfelder, F.; Nettels, D.; Makarov, D. E.; Schuler, B., Integrated view of internal friction in unfolded proteins from single-molecule FRET, contact quenching, theory, and simulations. *Proceedings of the National Academy of Sciences* **2017,** *114* (10), E1833.
10. Chung, H. S.; Piana-Agostinetti, S.; Shaw, D. E.; Eaton, W. A., Structural origin of slow diffusion in protein folding. *Science* **2015,** *349* (6255), 1504-1510.
11. Jas, G. S.; Eaton, W. A.; Hofrichter, J., Effect of Viscosity on the Kinetics of α-Helix and β-Hairpin Formation. *The Journal of Physical Chemistry B* **2001,** *105* (1), 261-272.
12. Ross, C. A.; Poirier, M. A., Protein aggregation and neurodegenerative disease. *Nature Medicine* **2004,** *10*, S10.
13. Wensley, B. G.; Kwa, L. G.; Shammas, S. L.; Rogers, J. M.; Clarke, J., Protein Folding: Adding a Nucleus to Guide Helix Docking Reduces Landscape Roughness. *Journal of Molecular Biology* **2012,** *423* (3), 273-283.
14. Kwa, L. G.; Wensley, B. G.; Alexander, C. G.; Browning, S. J.; Lichman, B. R.; Clarke, J., The Folding of a Family of Three-Helix Bundle Proteins: Spectrin R15 Has a Robust Folding Nucleus, Unlike Its Homologous Neighbours. *Journal of Molecular Biology* **2014,** *426* (7), 1600-1610.
15. Schulz, J. C. F.; Schmidt, L.; Best, R. B.; Dzubiella, J.; Netz, R. R., Peptide Chain Dynamics in Light and Heavy Water: Zooming in on Internal Friction. *Journal of the American Chemical Society* **2012,** *134* (14), 6273-6279.
16. Sashi, P.; Ramakrishna, D.; Bhuyan, A. K., Dispersion Forces and the Molecular Origin of Internal Friction in Protein. *Biochemistry* **2016,** *55* (33), 4595-4602.
17. Ansari, A.; Jones, C. M.; Henry, E. R.; Hofrichter, J.; Eaton, W. A., The role of solvent viscosity in the dynamics of protein conformational changes. *Science* **1992,** *256* (5065), 1796-1798.
18. Zheng, W.; de Sancho, D.; Best, R. B., Modulation of Folding Internal Friction by Local and Global Barrier Heights. *The Journal of Physical Chemistry Letters* **2016,** *7* (6), 1028-1034.
19. Kramers, H. A., Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **1940,** *7* (4), 284-304.

20. Grote, R. F.; Hynes, J. T., The stable states picture of chemical reactions. II. Rate constants for condensed and gas phase reaction models. *The Journal of Chemical Physics* **1980,** *73* (6), 2715-2732.
21. Bagchi, B.; Oxtoby, D. W., The effect of frequency dependent friction on isomerization dynamics in solution. *The Journal of Chemical Physics* **1983,** *78* (5), 2735-2741.
22. Hridya, V. M.; Mukherjee, A., Probing the Viscosity Dependence of Rate: Internal Friction or the Lack of Friction? *The Journal of Physical Chemistry B* **2018,** *122* (39), 9081-9086.
23. Daldrop, J. O.; Kappler, J.; Brünig, F. N.; Netz, R. R., Butane dihedral angle dynamics in water is dominated by internal friction. *Proceedings of the National Academy of Sciences* **2018,** *115* (20), 5169.
24. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P., The missing term in effective pair potentials. *The Journal of Physical Chemistry* **1987,** *91* (24), 6269-6271.
25. Reif, M. M.; Winger, M.; Oostenbrink, C., Testing of the GROMOS Force-Field Parameter Set 54A8: Structural Properties of Electrolyte Solutions, Lipid Bilayers, and Proteins. *Journal of Chemical Theory and Computation* **2013,** *9* (2), 1247-1264.
26. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M., LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **1997,** *18* (12), 1463-1472.
27. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **2008,** *4* (3), 435-447.
28. Walser, R.; Mark, A. E.; van Gunsteren, W. F., On the validity of Stokes' law at the molecular level. *Chemical Physics Letters* **1999,** *303* (5), 583-586.
29. Walser, R.; Hess, B.; Mark, A. E.; van Gunsteren, W. F., Further investigation on the validity of Stokes–Einstein behaviour at the molecular level. *Chemical Physics Letters* **2001,** *334* (4–6), 337-342.
30. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **2007,** *126* (1), 014101.
31. 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†. *The Journal of Physical Chemistry B* **2008,** *112* (2), 529-534.
32. 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. *Journal of the American Chemical Society* **2008,** *130* (23), 7477-7488.
33. Tuñón, I.; Laage, D.; Hynes, J. T., Are there dynamical effects in enzyme catalysis? Some thoughts concerning the enzymatic chemical step. *Archives of Biochemistry and Biophysics* **2015,** *582*, 42-55.
34. Booij, H. C.; van Wiechen, P. H., Effect of Internal Viscosity on the Deformation of a Linear Macromolecule in a Sheared Solution. *The Journal of Chemical Physics* **1970,** *52* (10), 5056-5068.
35. Samanta, N.; Chakrabarti, R., End to end loop formation in a single polymer chain with internal friction. *Chemical Physics Letters* **2013,** *582*, 71-77.

Chapter 5

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

5.1 Overview

We have investigated internal friction in two model systems and in butane in the previous chapters. Next, we intent to estimate internal friction in a complex process – intercalation of drug into DNA. Since intercalation is a millisecond timescale process, calculation of rate is extremely expensive even at a single viscosity whilst estimation of internal friction involves rate calculation at multiple viscosities. Therefore, as learnt from previous chapters, we have adopted the method of calculating transmission coefficients, which along with the TST rate, would provide the rate of the process and thereby the internal friction. Hence, as the first step towards this goal, in this chapter 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 κ=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. Our 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. This work was done in collaboration with Prof. James T. Hynes from Ecole Normale Supérieure-PSL Research University, Paris.

5.2 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. A recent work 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 κ, 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

$$m\ddot{x}(t) = -\frac{\partial G[(x,t)]}{\partial x(t)} - \int_0^t m\,\xi(\tau)\,\dot{x}(t-\tau)\,d\tau + \delta F_R(t), \qquad \text{Eq. 5.1}$$

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.

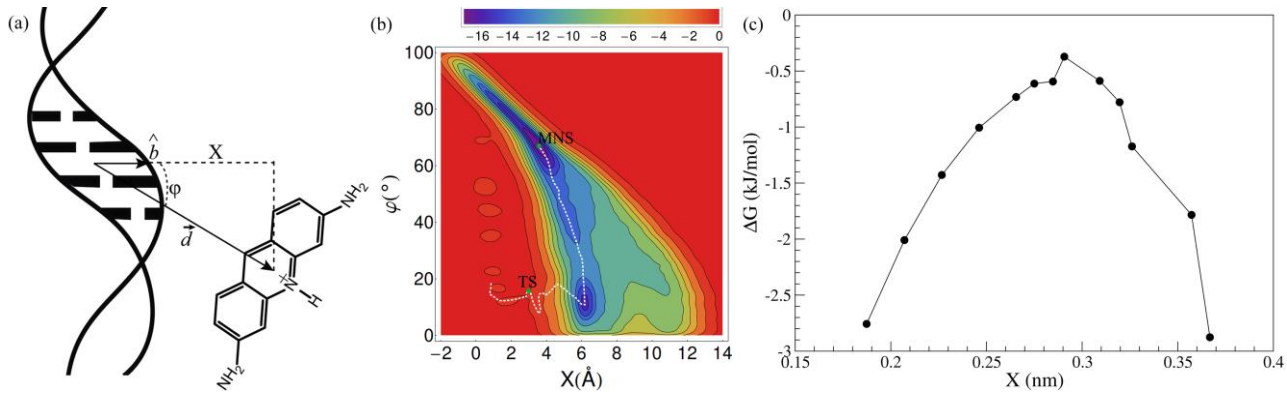

Figure 1. *(a) Schematic representation of the reaction coordinates X and φ. $\vec{d}$ is the vector from the center of mass of the intercalating base pairs to the center of mass of the proflavine. $\hat{b}$ is the unit vector from the center of mass of the same base pairs to the center of mass of the two sugars. $X = \hat{b} \cdot \vec{d}$ and $\varphi = cos^{-1}(\hat{b} \cdot \vec{d}/|\vec{d}|)$. (b) The two-dimensional free energy surface of the intercalation process from minor groove against X and φ.*[7-8] *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. Reprinted with permission Copyright (2019) American Chemical Society*

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 $\kappa_{GH} = \frac{\omega_r}{\omega_{eq}}$. 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 a 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 extensive 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.

5.3 Methods

(a) Simulation details. The simulated system consists of a twelve base pair DNA d(GCGCTCGAGCGC)2 and a protonated proflavine, 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 intercalation, 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 $X \leq 1$ Å) or reactant (away from DNA with $X \geq 6$ Å) 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 (300^{th} ps) of each trajectory.

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. 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) trajectories. 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

$$\kappa = \frac{\sum_{i,+}^{N} p_i |v_i| Q_i}{\sum_{i,+}^{N} p_i |v_i|}, \qquad \text{Eq. 5.2}$$

where "+" represents trajectories with an initial positive flux, p_i is the probability to have the i^{th} 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

$$Q_i = \begin{matrix} 1 & \text{if RP} \\ 0 & \text{if RR/PP} \\ -1 & \text{if PR} \end{matrix} \qquad \text{Eq. 5.3}$$

(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]

$$\kappa_{GH} = \frac{\omega_r}{\omega_{eq}} \qquad \text{Eq. 5.4}$$

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

$$\omega_r^2 - \omega_{eq}^2 + \omega_r \int_0^\infty e^{-\omega_r t}\, \xi_{TS}(t) dt = 0 \qquad \text{Eq. 5.5}$$

$$\xi(t)_{TS} = \frac{\langle F_{RC}(0) F_{RC}(t) \rangle}{\mu_{RC} k_B T} \qquad \text{Eq. 5.6}$$

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

$$\Delta \mathrm{PM}F(X) = -\frac{1}{2}K_{eq}(X - X^{\#})^2 \qquad \text{Eq. 5.7}$$

Here $\Delta \mathrm{PM}F(X)$ 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

$$\omega_{eq} = \frac{1}{2\pi c}\sqrt{\frac{K_{eq}}{\mu_{RC}}}. \qquad \text{Eq. 5.8}$$

Finally, the average square of the RC velocity, v_{RC} at the TS provides the reduced mass via the equipartition theorem, $\frac{1}{2}\,\mu_{RC}\,\langle v_{RC,TS}^2\rangle = \frac{1}{2}\,k_B T$.

5.4 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 metadynamics-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. But the non-equilibrium character of metadynamics need not provide the sufficient TS region sampling demanded by the committor analysis. Nonetheless, we can infer from the 93% product formation result that this CN1 point in fact lies 1.18 kcal/mol towards the product. Thus, despite the extreme difficulty -- in a system with more than 30,000 atoms -- to precisely locate a point providing equal probability for

generation of reactant and product, the prior metadynamics simulation can aid us to narrow our TS search.

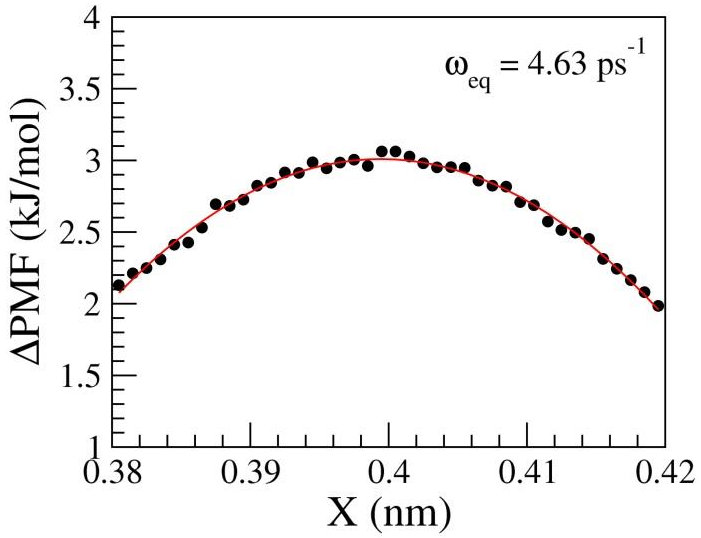

Figure 2. *The parabolic fit to the potential of mean force, plotted against the reaction coordinate X, near the transition state from umbrella sampling simulations to obtain the equilibrium barrier frequency via Eqs. 7 and 8.*

Accordingly, we recalculated the free energy profile along X in the neighborhood of the metadynamics TS. Umbrella sampling simulations[45] 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 candidates 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. Figure 3 shows the locations of some of the representative TS candidates in the free energy surface (FES).

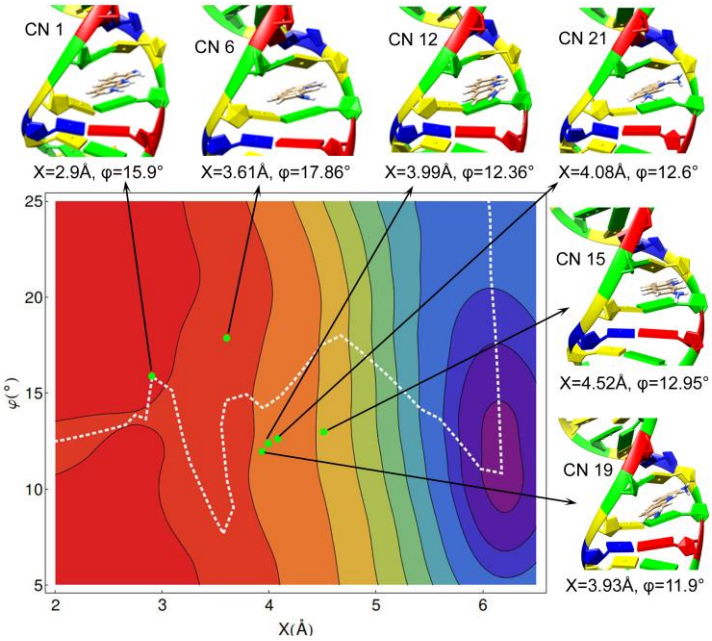

Figure 3. *The free energy surface (FES) in Figure 1b is zoomed in near the transition state. The dotted line represents the minimum free energy path. The green circles in the FES represent some of the transition state candidates chosen for committor probability calculation. The reaction coordinate values are indicated below each structure. Reprinted with permission Copyright (2019) American Chemical Society*

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.

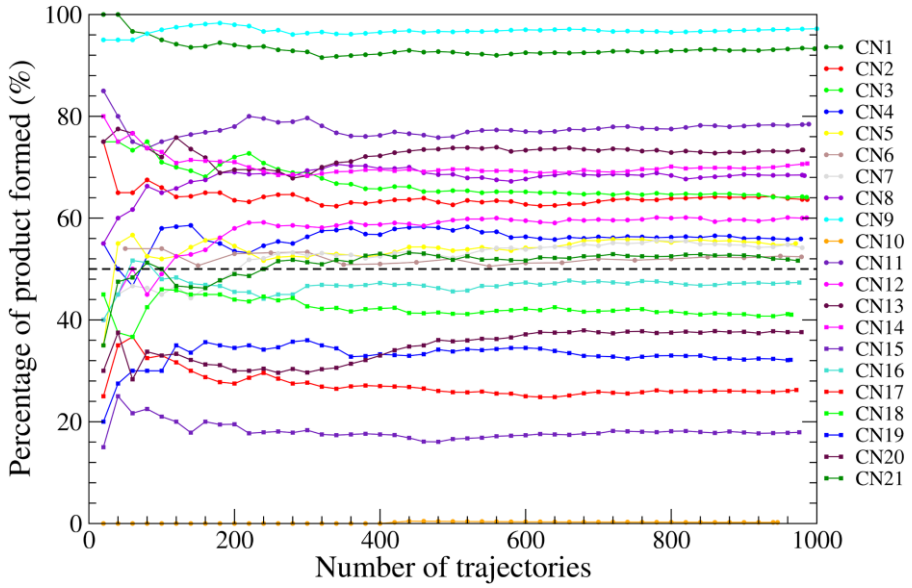

Figure 4. *Percentage of Stable product formation is plotted against the number of trajectories. The CN # corresponds to the different initial configurations. The black dotted line at 50% is to guide the eye. Reprinted with permission Copyright (2019) American Chemical Society*

(b) MD-simulated reactive flux and transmission coefficient. Once CN21 was identified as the TS using committor analysis just described, we sampled the configurations around the TS by freezing the intercalating base pairs and proflavine. These sampled states of CN21 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 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.

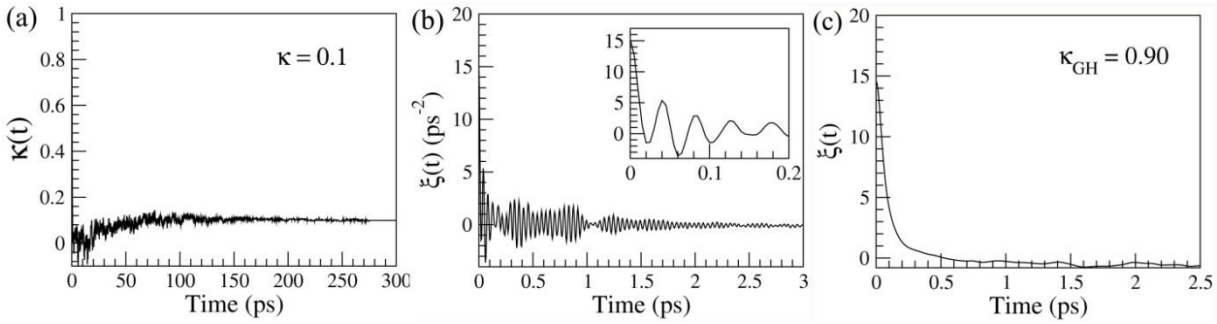

Figure 5. *(a) The time-dependent transmission coefficient $\kappa(t)$ obtained from MD simulations using the reactive flux formalism. $\kappa(t)$ plateaus to the value κ = 0.1. (b) The friction kernel for the RC fixed at the transition state, with the inset showing the initial decay. (c) The friction kernel obtained for the model system used.*

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; 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). We have used a model system to assess the friction kernel calculation 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}. The friction kernel (see Fig. 5c) is characterized by a weak friction similar to the proflavine intercalation result in Fig. 5b (but without an oscillatory behavior); calculation of the associated transmission coefficients in the GHT and KT perspective yields $\kappa_{GH} = 0.90$ and $\kappa_{Kr} = 0.88$.

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.

5.5 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 are now discussed. 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 ~ k_BT 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. 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, but in the present case, they are almost identical. This can be seen from 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 negligible frictional recrossing is the same in both theories.

Next we discuss the marked discrepancy of the predictions of both the theories 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[46]). While these assumptions evidently hold in quite a wide range of reaction systems[12-13, 28, 47-48], they are very clearly violated here. A number of trajectories recross the TS after wandering rather far from the TS, as now recounted.

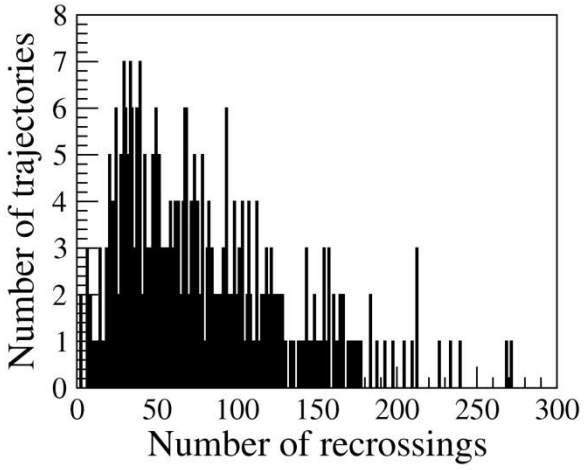

Figure 6. *The histogram of proflavine intercalation process trajectories characterized by a given number of TS recrossings. Reprinted with permission Copyright (2019) American Chemical Society*

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. 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. [46, 49-50]

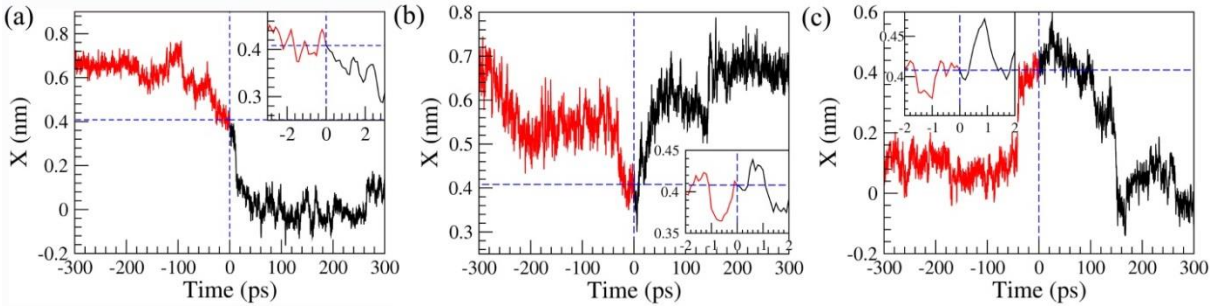

Figure 7. *Three representative trajectory results following trajectories that crossed the TS from the Stable reactant R to the Stable product P side at t=0. The vertical dashed blue line denotes time t=0 and the horizontal dashed blue line denotes the reaction coordinate X's value at t=0. Black solid 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 larger amplitude 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. Reprinted with permission Copyright (2019) American Chemical Society*

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

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.

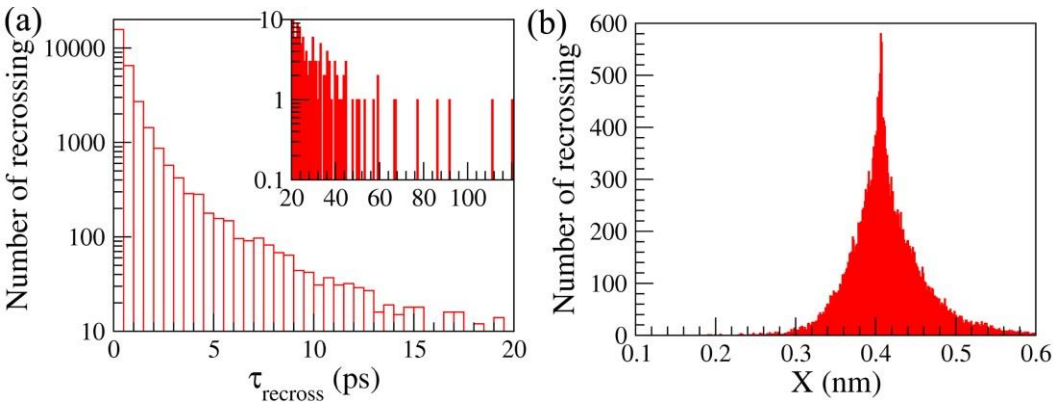

Figure 8. *(a) A histogram of the time taken between consecutive recrossings up to 20 ps. The inset shows the corresponding results for the times larger than 20 ps. (b) The histogram of the largest magnitude value of X reached away from TS by a trajectory before recrossing. Reprinted with permission Copyright (2019) American Chemical Society*

(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 all the different trajectory types, especially the successful trajectories RP, and have examined the behavior of several quantities in addition to the selected reaction coordinate X. 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, 51] 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 all the 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 RP trajectory results (magenta graphs in Fig. 9) 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 (P). 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,

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

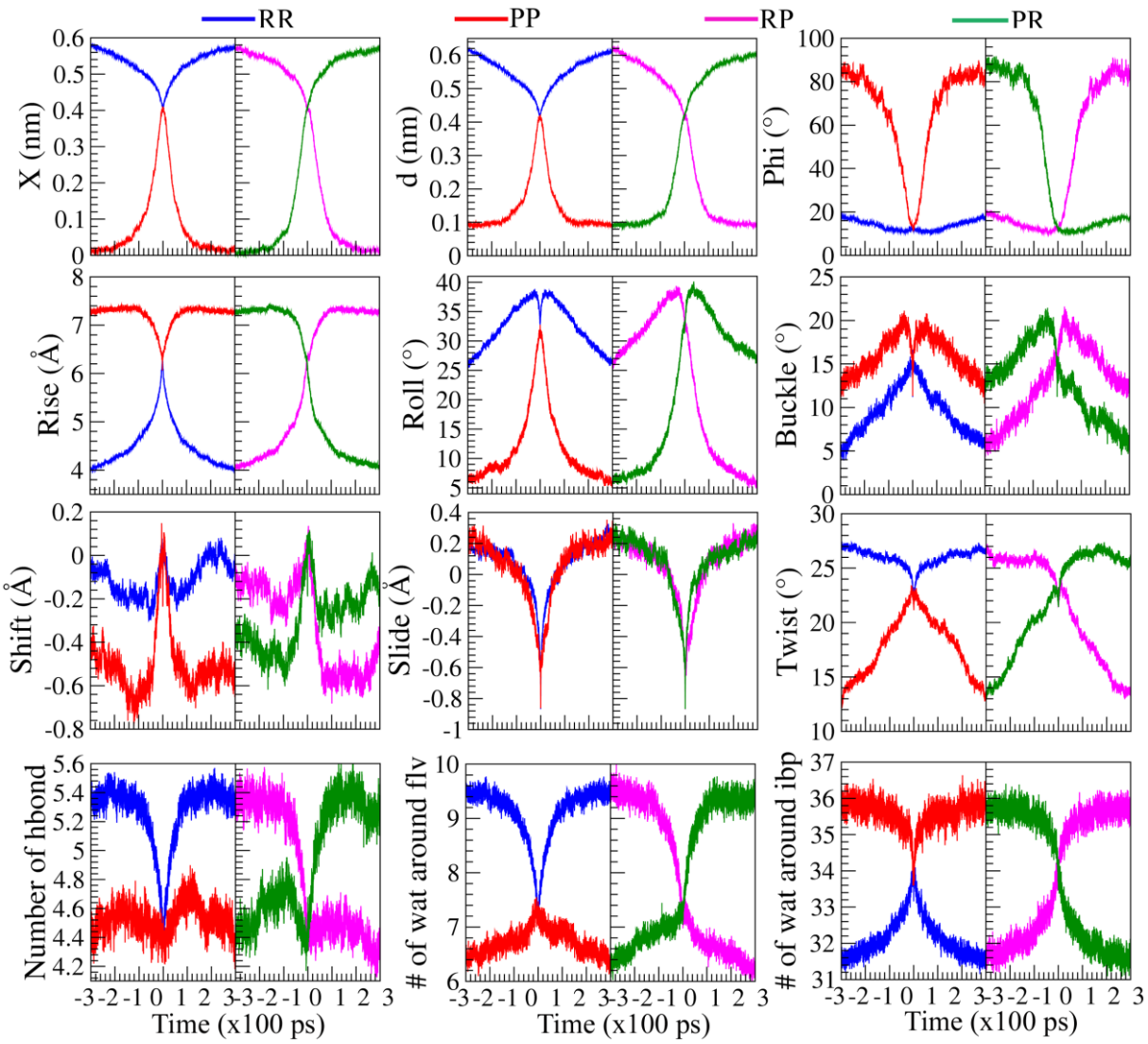

Figure 9. *The average successful RP reaction trajectory values, versus time before and after the TS, of the coordinates X, d, φ; the IBP variables Rise, Roll, Buckle, Shift, Slide, Twist; and the water-related variables the number of hydrogen bonds between water and drug, the number of waters around 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 behavior is striking: it 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. 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[52-53]. 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 molecules

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.[54-56] 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.

5.6 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 of these variables in an improved reaction coordinate description seems worthy of investigation. However, the overall mechanism corroborates 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.

References

1. Lerman, L. S., Structural considerations in the interaction of DNA and acridines. *Journal of Molecular Biology* **1961,** *3* (1), 18-IN14.
2. Chaires, J. B., A thermodynamic signature for drug–DNA binding mode. *Archives of Biochemistry and Biophysics* **2006,** *453* (1), 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* (17), 3933-3940.
4. Brana, M. F.; Cacho, M.; Gradillas, A.; De Pascual-Teresa, B.; Ramos, A., Intercalators as anticancer drugs. *Current Pharmaceutical Design* **2001,** *7* (17), 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 Research* **2003,** *31* (14), 4138-4146.
6. Li, H. J.; Crothers, D. M., Relaxation studies of the proflavine-DNA complex: The kinetics of an intercalation reaction. *Journal of Molecular Biology* **1969,** *39* (3), 461-477.
7. Sasikala, W. D.; Mukherjee, A., Molecular Mechanism of Direct Proflavine–DNA Intercalation: Evidence for Drug-Induced Minimum Base-Stacking Penalty Pathway. *The Journal of Physical Chemistry B* **2012,** *116* (40), 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. *Physical Chemistry Chemical Physics* **2013,** *15* (17), 6446-6455.
9. Eyring, H., The Activated Complex in Chemical Reactions. *The Journal of Chemical Physics* **1935,** *3* (2), 107-115.
10. Roca, M.; Pascual-Ahuir, J.-L.; Tuñón, I., Reversibility and Diffusion in Mandelythiamin Decarboxylation. Searching Dynamical Effects in Decarboxylation Reactions. *Journal of the American Chemical Society* **2012,** *134* (25), 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. *Journal of the American Chemical Society* **2008,** *130* (23), 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. *Journal of the American Chemical Society* **2006,** *128* (18), 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. *Journal of the American Chemical Society* **2005,** *127* (30), 10648-10655.
15. Mullen, R. G.; Shea, J.-E.; Peters, B., Transmission Coefficients, Committors, and Solvent Coordinates in Ion-Pair Dissociation. *Journal of Chemical Theory and Computation* **2014,** *10* (2), 659-667.
16. Kramers, H. A., Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **1940,** *7* (4), 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. *The Journal of Chemical Physics* **1980,** *73* (6), 2715-2732.
18. Grote, R. F.; Hynes, J. T., Saddle point model for atom transfer reactions in solution. *The Journal of Chemical Physics* **1981,** *75* (5), 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. *The Journal of Chemical Physics* **1980,** *73* (6), 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. *The Journal of Physical Chemistry B* **2010,** *114* (42), 13593-13600.
21. Pecina, O.; Schmickler, W., On the dynamics of electrochemical ion-transfer reactions. *Journal of Electroanalytical Chemistry* **1998,** *450* (2), 303-311.

22. Gertner, B. J.; Wilson, K. R.; Hynes, J. T., Nonequilibrium solvation effects on reaction rates for model SN2 reactions in water. *The Journal of Chemical Physics* **1989,** *90* (7), 3537-3558.
23. Keirstead, W. P.; Wilson, K. R.; Hynes, J. T., Molecular dynamics of a model SN1 reaction in water. *The Journal of Chemical Physics* **1991,** *95* (7), 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. *The Journal of Physical Chemistry A* **1997,** *101* (15), 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. *The Journal of Physical Chemistry* **1995,** *99* (33), 12647-12654.
26. Smith, B. B.; Hynes, J. T., Electronic friction and electron transfer rates at metallic electrodes. *The Journal of Chemical Physics* **1993,** *99* (9), 6517-6530.
27. Staib, A.; Borgis, D.; Hynes, J. T., Proton transfer in hydrogen-bonded acid–base complexes in polar solvents. *The Journal of Chemical Physics* **1995,** *102* (6), 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†. *The Journal of Physical Chemistry B* **2008,** *112* (2), 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. *The Journal of Chemical Physics* **1983,** *79* (2), 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? *Journal of Computational Chemistry* **2000,** *21* (12), 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. *Biophysical Journal* **2007,** *92* (11), 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. *Journal of Computational Chemistry* **2004,** *25* (9), 1157-1174.
33. Nosé, S., A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics* **1984,** *52* (2), 255-268.
34. Hoover, W. G., Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* **1985,** *31* (3), 1695-1697.
35. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993,** *98* (12), 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. *Journal of Chemical Theory and Computation* **2008,** *4* (3), 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. *Journal of the American Chemical Society* **2008,** *130* (30), 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. *Journal of the American Chemical Society* **2012,** *134* (20), 8588-8596.
39. Onsager, L., Initial Recombination of Ions. *Physical Review* **1938,** *54* (8), 554-557.
40. Király, P.; Kiss, D. J.; Tóth, G., Committor of elementary reactions on multistate systems. *The Journal of Chemical Physics* **2018,** *148* (13), 134107.

41. Du, R.; Pande, V. S.; Grosberg, A. Y.; Tanaka, T.; Shakhnovich, E. S., On the transition coordinate for protein folding. *The Journal of Chemical Physics* **1998,** *108* (1), 334-350.
42. Martys, N. S.; Mountain, R. D., Velocity Verlet algorithm for dissipative-particle-dynamics-based models of suspensions. *Physical Review E* **1999,** *59* (3), 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. *The Journal of Chemical Physics* **1986,** *85* (10), 5625-5643.
44. Peters, B., Chapter 17 - Grote-Hynes theory. In *Reaction Rate Theory and Rare Events Simulations*, Peters, B., Ed. Elsevier: Amsterdam, 2017; pp 451-471.
45. Torrie, G. M.; Valleau, J. P., Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **1977,** *23* (2), 187-199.
46. Spångberg, D.; Rey, R.; Hynes, J. T.; Hermansson, K., Rate and Mechanisms for Water Exchange around Li+(aq) from MD Simulations. *The Journal of Physical Chemistry B* **2003,** *107* (18), 4470-4477.
47. Tolokh, I. S.; White, G. W. N.; Goldman, S.; Gray, C. G., Prediction of ion channel transport from Grote—Hynes and Kramers theories. *Molecular Physics* **2002,** *100* (14), 2351-2359.
48. Roux, B.; Karplus, M., Ion transport in a gramicidin-like channel: dynamics and mobility. *The Journal of Physical Chemistry* **1991,** *95* (12), 4856-4868.
49. Rey, R.; Hynes, J. T., Hydration Shell Exchange Kinetics: An MD Study for Na+(aq). *The Journal of Physical Chemistry* **1996,** *100* (14), 5611-5615.
50. Rey, R.; Hynes, J. T., Hydration shell exchange dynamics for in water. *Journal of Physics: Condensed Matter* **1996,** *8* (47), 9411-9416.
51. Dickerson, R. E., Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Research* **1989,** *17* (5), 1797-1803.
52. Günther, K.; Mertig, M.; Seidel, R., Mechanical and structural properties of YOYO-1 complexed DNA. *Nucleic Acids Research* **2010,** *38* (19), 6526-6532.
53. Celedon, A.; Wirtz, D.; Sun, S., Torsional Mechanics of DNA Are Regulated by Small-Molecule Intercalation. *The Journal of Physical Chemistry B* **2010,** *114* (50), 16929-16935.
54. Rey, R.; Ingrosso, F.; Elsaesser, T.; Hynes, J. T., Pathways for H2O Bend Vibrational Relaxation in Liquid Water. *The Journal of Physical Chemistry A* **2009,** *113* (31), 8949-8962.
55. 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. *Journal of the American Chemical Society* **1991,** *113* (1), 74-87.
56. Whitnell, R. M.; Wilson, K. R.; Hynes, J. T., Vibrational relaxation of a dipolar molecule in water. *The Journal of Chemical Physics* **1992,** *96* (7), 5354-5369.

Conclusion

6. 1 Summary

This thesis investigates the phenomenon of dynamical recrossing at the transition state which stems from friction arising due to the environment. Transmission coefficient – an estimate of recrossing in the system -- was used to investigate the origin of highly-debated internal friction. We calculated recrossing and the rate of crossing a barrier in model systems at different viscosities to investigate the internal friction in terms of the equations currently used in the literature. We found that even for the simplest system of a particle crossing a one-dimensional barrier, these equations gave rise to the so-called internal friction. Further investigations revealed that the presence of memory effects was falsely attributed as the internal friction. This thesis work thus draws attention to the flaws of current methods applied to estimate internal friction. Through this thesis, we advocate for the need to develop new methods to assess the presence of internal friction. Although we showed that the current methods cannot unequivocally capture internal friction, we don't deny the presence of it in a real system. We showed that both the memory effects and internal friction can give rise to similar trend in the viscosity dependence of the rate. In an effort to see if we can decouple the two, we chose two simple models; two atoms interacting via a symmetric one-dimensional double well potential and a united butane atom model. However, we find that neither of the systems exhibits internal friction. Nevertheless, systems with higher degrees of freedom like proteins can display both internal friction and memory effects. In this direction, we studied the process of intercalation of an anticancer agent proflavine into DNA through the minor-groove side. Since intercalation is a milli-second long reaction, we calculated the transmission coefficient instead of rate to estimate the internal friction. In the process, we realized that there is significant recrossing in the system and dynamical effects are complex in nature. We find that theories like Kramers theory and Grote-Hynes theory are inadequate

to explain recrossing in the system, thus demanding more detailed theoretical approaches for their elucidation. Internal friction for this system still remains to be determined.

6.2 Future directions

This thesis brings out the lacunae in estimating internal friction by currents methods when memory effects are present. Hence, the first and foremost future objective is to formulate a novel method that can accurately measure internal friction in the presence of memory effects. This is a challenging quest since internal friction, the friction arising from the solute degrees of freedom, can also exhibit memory effects. Nonetheless, this is an essential path worth pursuing to unearth the molecular origins of internal friction. Once such a method is devised, it can be used to calculate internal friction in complex systems like DNA-drug intercalation, protein folding etc.

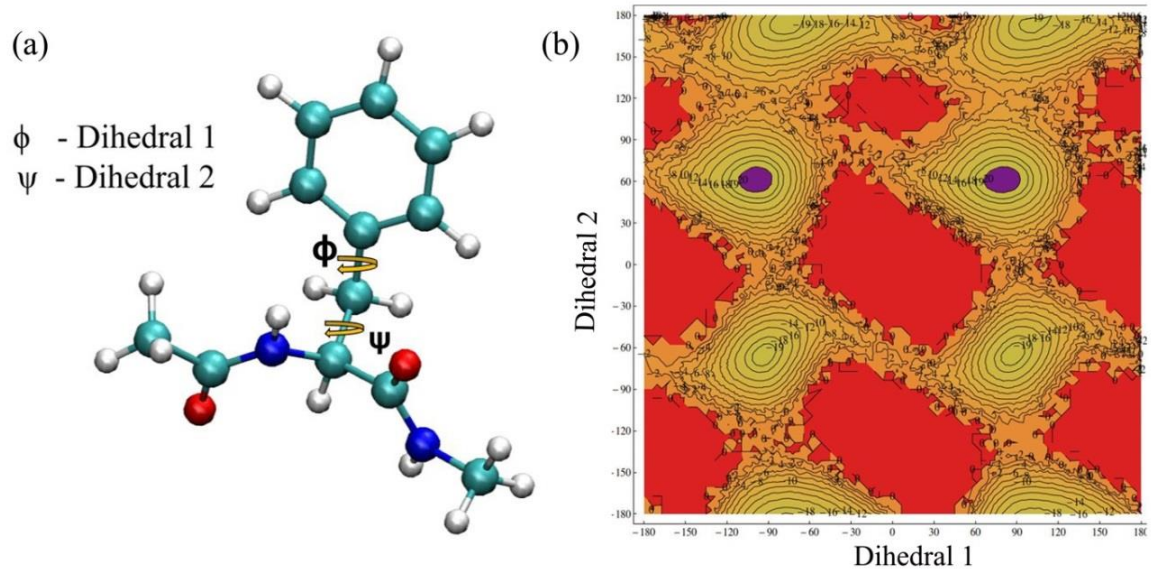

Figure 1: *(a) The two dihedral angles in phenylalanine (b) The free energy surface for both the dihedral angles obtained via metadynamics simulations*

From our studies, we deem that not only memory effects, but other factors eliciting failure of Kramers prediction can lead to a deviation from the inverse dependence of rate on viscosity. This can again lead to misinterpretation of internal friction since current methods used to estimate internal friction inherently assume the validity of Kramers theory. One of the

unapparent factors that may cause breakdown of Kramers prediction is the multi-pathway reactions. For example, the ring flip in phenylalanine is caused by the rotation of dihedral angle ϕ (see Fig. 1a). In addition to the direct change in ϕ, the dihedral rotation in ϕ can also occur via changes in dihedral angle Ψ as is evident from the free energy surface for both the dihedral angles shown in Fig. 1b. Thus, the dihedral rotation of ϕ can occur through multiple pathways. However, measuring the rate of dihedral rotation in experiment would provide an overall rate which includes an weighted average of all the different pathways. Then the overall rate may not be an inverse function of viscosity and might give a false positive for internal friction. The ring flip in phenylalanine is a simple system through which this hypothesis can be verified.

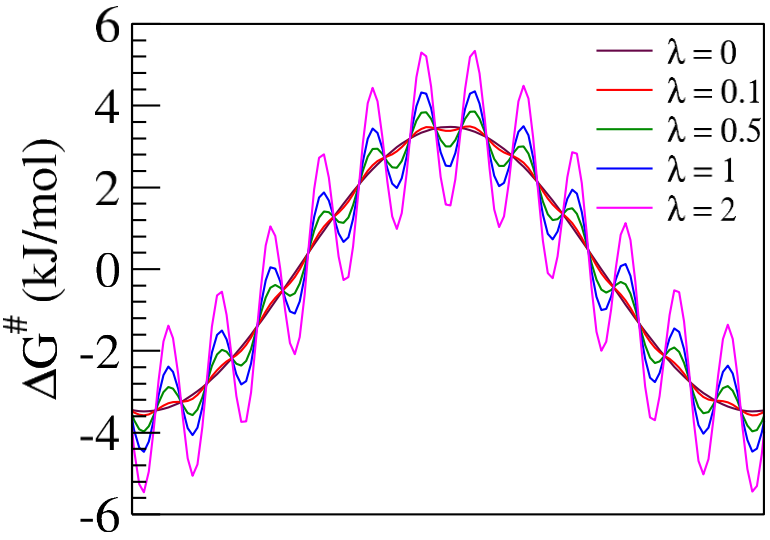

Figure 2: *Potential energy surfaces with different amount of ruggedness. The form of the potential is* $V = V_0 Cos(2\pi nx/L) + \lambda sin(300x)$ *where the first term is the potential used in Chapter 3 and* λ *is the measure of ruggedness in the potential energy surface.*

A third future goal is to investigate the relation of ruggedness to internal friction. Zwanzig had shown in 1988 that the diffusion in rugged potential is slower than that in a smooth potential and the effective diffusion exponentially depends on the height of the ruggedness.[1] The slow folding of some of the proteins had been attributed to the presence of ruggedness in the free energy landscape and has been related to the presence of internal friction.[2] Using Langevin dynamics (thus eliminating the memory effects), one can calculate the rate of a

Lennard-Jones particle crossing a rugged barrier. Monitoring the rate behavior at different viscosities and different amount of ruggedness in the barrier (see Fig. 2) can reveal the exact nature of relationship between ruggedness and internal friction.

References

1. Zwanzig, R., Diffusion in a rough potential. *Proceedings of the National Academy of Sciences of the United States of America* **1988,** *85* (7), 2029-2030.
2. Wensley, B. G.; Batey, S.; Bone, F. A. C.; Chan, Z. M.; Tumelty, N. R.; Steward, A.; Kwa, L. G.; Borgia, A.; Clarke, J., Experimental evidence for a frustrated energy landscape in a three-helix-bundle protein family. *Nature* **2010,** *463* (7281), 685-688.

Addendum

MD Study of Bivalent Nucleic Acid Ligands for Recognition of Repeated Expansions in DNA and RNA

A.1 Introduction

Many neurodegenerative diseases like Huntington's disease, muscular dystrophy etc. are associated with trinucleotide repeat expansions in the genes.[1-2] Huntington's disease is an autosomal dominant disorder that affects muscular coordination and leads to behavioral changes, cognitive decline, and dementia.[3] It is caused by an expansion of CAG repeats in the first exon of the huntingtin (htt) gene, from a normal range of 6-29 to a pathogenic range of 40-180.[4] The length of the CAG repeats is inversely related to the age of onset.[5] Similarly, myotonic dystrophy, the most common form of muscular dystrophy is caused by CTG-repeat expansion in the 3′-untranslated region (3′-UTR) of the dystrophia myotonica protein kinase (DMPK) gene, from a normal range of 5–35 repeats to a pathogenic range of 80 to >2500.[6] Currently there are no effective therapeutic methods to counter these repeat formations. Designing modified nucleobases with good binding affinity that will open up these repeats and bind with them is a plausible approach to curtail these repeat expansions. However, due to the lack of binding energy, modified nucleobases cannot invade stable double DNA helices. Nielsen and coworkers[7-8] have shown that PNA, a nucleic acid mimic comprising a pseudopeptic backbone (Fig.1A) can invade B-DNA where the strand invasion can occur through Watson-Crick or Hoogsten base-pairing depending on the sequence context. Yet, with the original backbone design, PNA recognition is restricted to mostly homopurine and homopyrimidine targets. Mixed-sequence PNA generally does not have sufficient binding energy to invade B-DNA. More recently it was shown that PNA, when preorganized into a

right-handed helical motif by installing an (*R*)-stereogenic center at the γ-backbone (Fig.1B), can invade B-DNA without sequence restriction.[9] In spite of this recent advance, the challenge in being able to accomplish such a task at elevated (physiologically relevant) ionic strengths remains. At higher ionic strengths, the ions like Mg^{2+} make DNA double helix more stable. Hence, more energy is needed for invasion of the DNA strand. The required binding energy could be attained by replacing natural nucleobases with synthetic analogues having improved hydrogen-bonding and base-stacking interactions. However, such an approach could compromise recognition-specificity, due to the propensity of ultra-high affinity probes to bind DNA (or RNA) (Fig.1C) with closely related sequences, in addition to the intended targets and also due to the tendency of PNA strands to self-hybridize. To overcome these challenges, the group of researchers under Prof. Danith Ly in Carnegie Mellon University developed a novel class of bifacial nucleic acid recognition elements for targeting canonical DNA base pairs of DNA (Fig.1D); the dual binding mode improves the binding energy and the sequence specificity. We study the feasibility of these bifacial interactions using computer simulations.

A.2 Methods

(a) **Design of the nucleotide**: To augment the binding energy of γPNA, while at the same time improving its recognition specificity, the group at Carnegie Mellon envisioned the application of Janus bases (or JBs) as recognition elements due to their potential for forming hydrogen-bonding interactions with nucleobases in both strands of the DNA double helix (Fig.1D). Their interactions with double helical DNA were expected to be more favorable than that of natural nucleobases (Fig.1E) due to a significant increase in the number of hydrogen-bonds and in the degree of base-stacking interactions of the resulting triplex as compared to that of a duplex (Fig.1F). An improvement in base-stacking interactions was anticipated due to the expanded aromatic ring-systems of JBs and as a result of the formation

of base-triads. For every two hydrogen-bonds in A-T or T-A pair, and for every three in C-G or G-C that are broken, five new ones are formed upon the invasion of DNA by JBγPNA. Moreover, the binding will be more sequence-specific because a single base-mismatch that would normally occur on one face of a natural nucleobase would occur on both faces of a JB. The concept of bifacial nucleic acid recognition is not new. It was conceived by Lehn[10] more than two decades ago and, subsequently, expounded upon by others in the development of Janus-wedges. Despite concerted efforts from several research groups, only a small set of Janus-wedges has been developed, and they vary considerably in shapes and sizes, such that they cannot be effectively combined in a modular format for recognition of a non-homogenous nucleic acid sequence. In contrast, the JBs under development were strategically designed and optimized so that they contain the appropriate shapes, sizes, chemical functionalities, and tautomers for proper recognition of the respective DNA base-pairs (Fig.1F).

(b) **Modeling of the triplex:** To assess the feasibility of the method, the stability of the DNA-JBγPNA-DNA triplex needs to be determined. Further, one needs to investigate whether JBγPNA-JBγPNA is more favored over DNA-JBγPNA-DNA due to the possibility of self-hybridization between γPNAs. Hence, molecular dynamics simulations were performed to examine the stability of these complexes. A dodecameric γPNA containing a mixture of all four JBs, H-Lys-EBFDBEFDFDFB-NH_2, was chosen as a model system (Fig.2A, *i*). We built the structure of CEG, GFC, ABT and TDA triads using chimera[11] and optimized them using HF/6-31G* basis set in Gaussian[12]. Using these optimized triads, the helical structure of the DNA-JBγPNA-DNA was constructed using the NAB module of Ambertools[13] (Fig.2A, *ii*) and that of JBγPNA-JBγPNA was adopted from an existing NMR structure[14] (Fig.2B, *i* and *ii*). The PNA backbone from X-ray crystal structure, PDB-ID 3PA0[15] (note that the MP-side chain was replaced with methyl group), was grafted onto the

DNA-JBγPNA-DNA helix and then energy minimized to obtain the initial structure used for simulations.

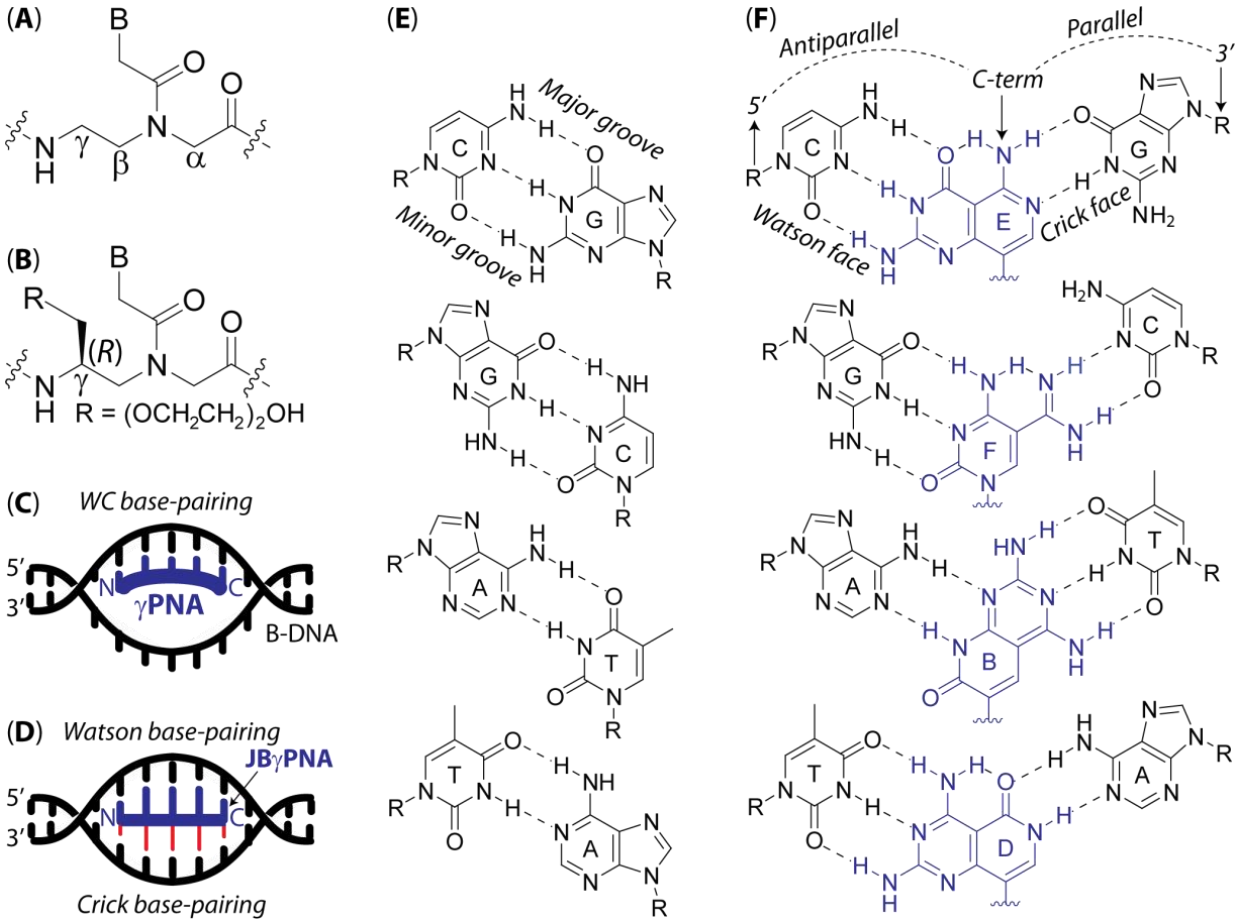

Figure 1: *(**A**) PNA, (**B**) γPNA, (**C**) the binding mode of γPNA containing natural nucleobases, and (**D**) that containing JBs. (**E**) Hydrogen-bonding interactions of natural base-pairs, and (**F**) that of JB-triads. Figure reprinted with permission from natureresearch Copyright (2018).*

(c) **Molecular dynamics simulations:** MD simulations were performed for DNA-JBγPNA-DNA (W-P-C) and JBγPNA-JBγPNA (P-P) complexes. Each complex was solvated with TIP3P[16] water molecules in a cubic box, and ions were added to maintain the physiological concentration. The systems were energy minimized using the steepest descent method[17] and then heated to 300 K under a harmonic restraint of 25 kcal/mol/Å on all the heavy atoms. In a series of six short simulations, the restraint was gradually released. The final unrestrained simulation was performed for 500 ns for each system. Simulations were done in NPT conditions, where the Nose-Hoover thermostat[18-19] was used to maintain the temperature at

300K with a coupling constant of 0.4 ps and the Parinello-Rahman barostat[20] with a coupling constant of 1.0 ps was used to maintain the pressure at 1 bar. A time step of 2 fs was used, and the electrostatic interactions were treated using particle mesh Ewald method[21] with a cutoff at 10 Å. All the simulations were performed using GROMACS-5.1[22-23]. The AMBER parmbsc1[24] force field was used in simulations for both the DNA and PNA. The charges for the non-standard PNA bases and the backbones were derived using RED[25].

A.3 Results

The result of MD simulations showed that the W-P portion of the triplex was stable for 500 ns (Fig.1A, iii and *iv*) – the entire duration of the simulation, while the P-C segment displayed a significant structural distortion (Fig.2A, *iii* and *v*). This is reflected in the number of hydrogen-bonds and in the inter-strand interaction energy (Fig.3). The weaker interaction of P-C, as compared to that of W-P, could be attributed to the number of hydrogen-bonds being fewer and to the fact that hybridization occurs in a less favorable parallel orientation in case of the former.[26] In contrast, the structure of JBγPNA-JBγPNA unraveled upon restraint release, probably as a result of steric repulsion in the backbone (Fig.2B). Self-hybridization was an impending concern in the design of JBγPNA due to the complementary nature of the two faces of JBs, although it was considered less likely with γPNA as a backbone than with PNA due to the former adopting helical chirality. Overall, the MD simulations showed that not only can JBγPNA hybridize to both strands of the DNA double helix, but it is unable to hybridize to each other—an important requirement for a successful design of a bifacial nucleic acid system.

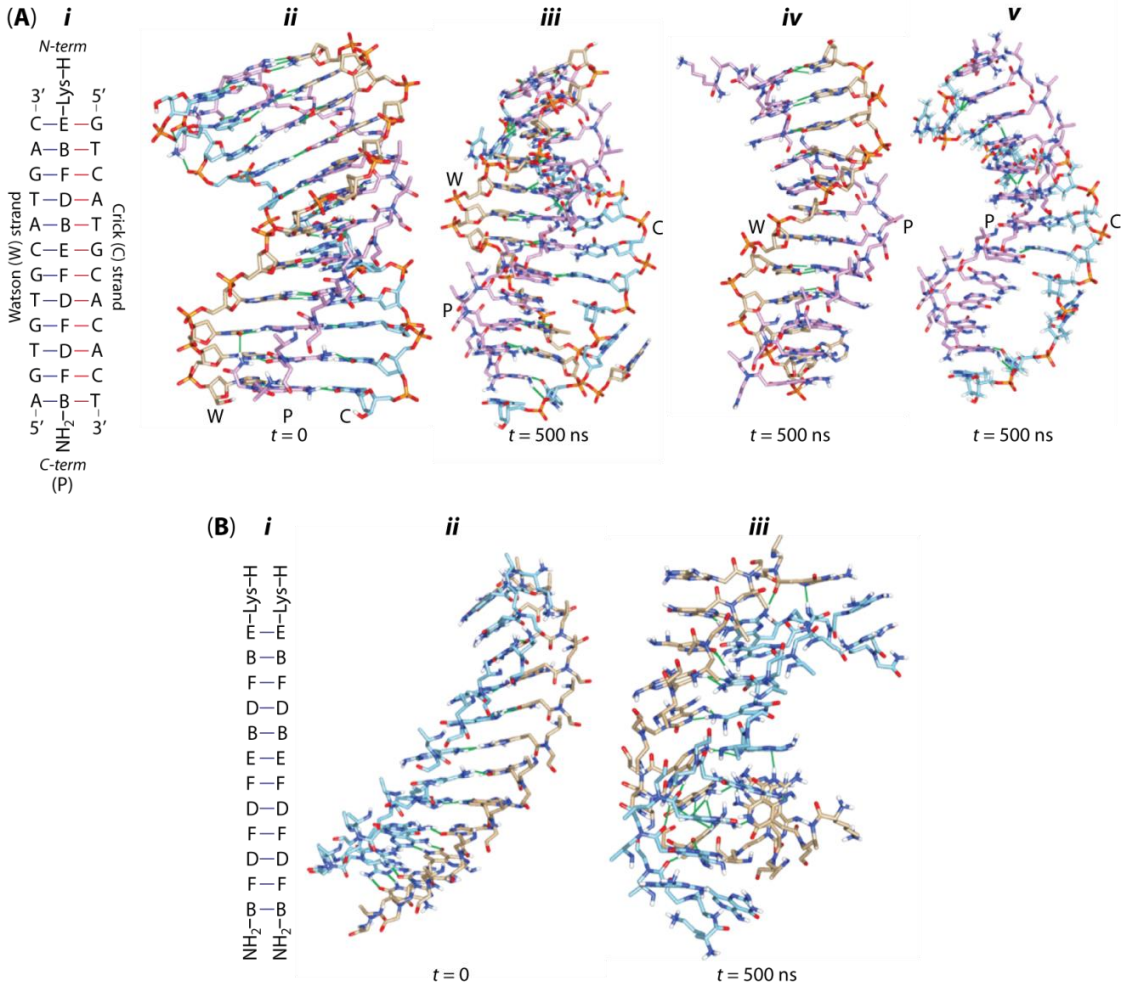

Figure 2: *(**A**) Simulated structure of a DNA-JBγPNA-DNA triplex. (i) Sequence of the triplex, (ii) initial triplex structure, and (iii) simulated triplex structure. (iv) & (v) Simulated structure of the triplex with the respective C and W strands removed for clarity. (**B**) Simulated structure of a self-hybridized JBγPNA-JBγPNA duplex. Figure reprinted with permission from natureresearch Copyright (2018).*

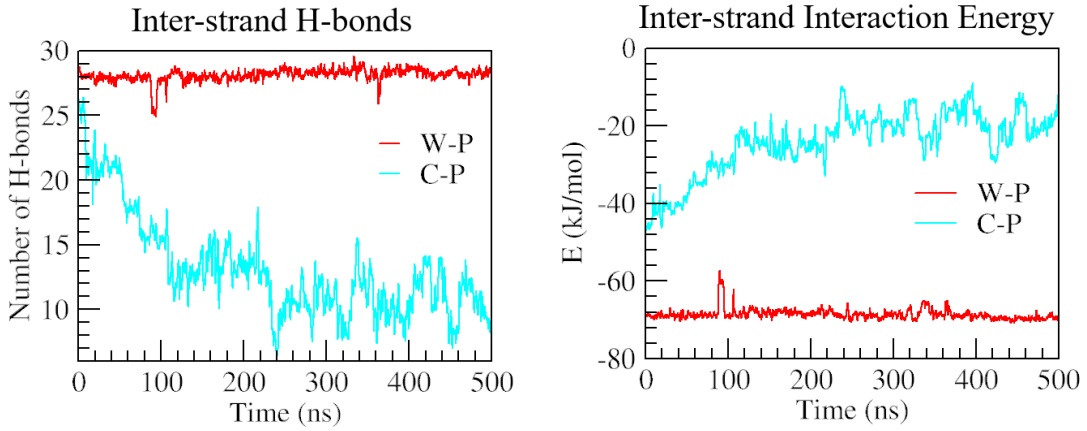

Figure 3: *(**Left**) H-bonding and (**right**) inter-strand energy of W-P and C-P per base-pair (excluding terminal base-pair). Figure reprinted with permission from natureresearch Copyright (2018).*

We also tested the development of a new class of nucleic acid ligands, comprising Janus bases and MPγPNA backbone that are capable of binding rCAG-repeats prevalent in Huntington's disease. We carried out MD simulations of LG1 ligand (comprising of three JBs E,I and F; see Fig. 4B) bound to an RNA duplex containing four rCAG-repeats. The complex remained stable throughout the simulation, with the four LG1 ligands fitting snuggly in between the two RNA strands (Fig. 4A). The number of H-bonds, five for each of the CEG and GFC triads and four for AIA (Fig. 4B), remained intact throughout the simulation (Fig. 2C). Attempts to simulate binding with fewer than four LG1 ligands, however, were unsuccessful, as the complexes unraveled upon restraint release (Fig. 4D-E). This finding suggests that ligand binding occurs in a cooperative manner with preference for an expanded rCAG-hairpin structure, presumably due to intermolecular base-stacking.

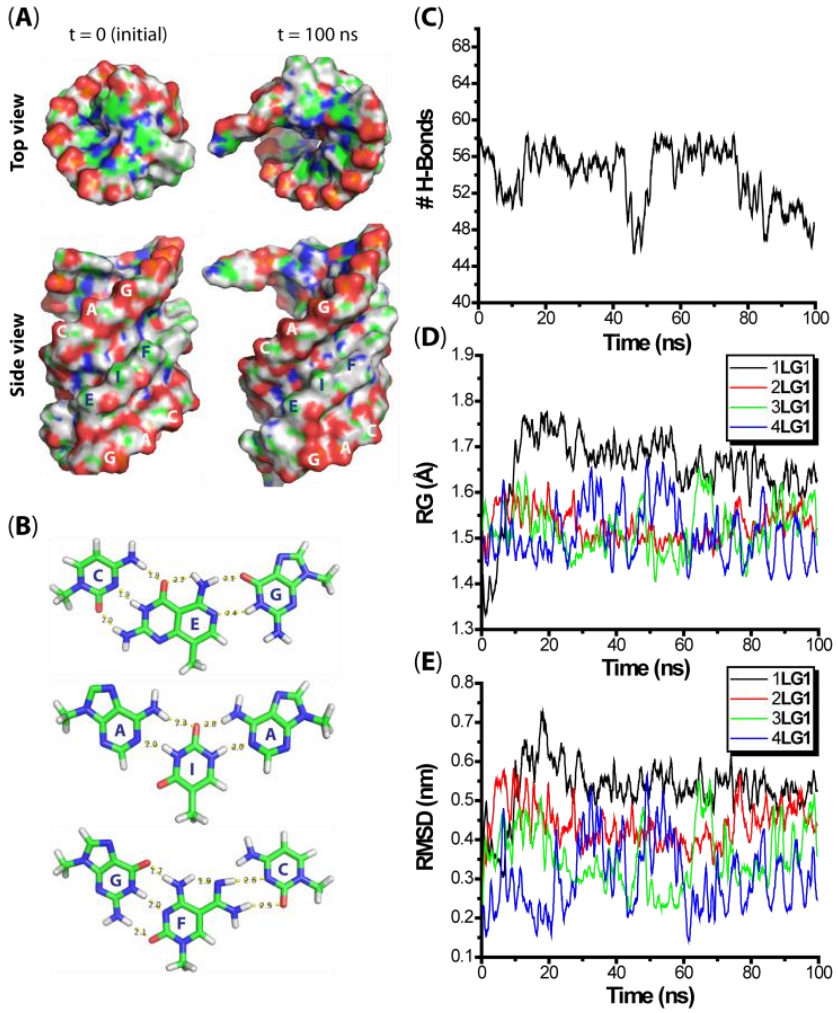

Figure 4: *The results of MD simulations of LG1 ligand binding to an RNA duplex containing the sequence r(CAG)$_4$/r(CAG)$_4$. LG1: NH$_2$-EIF-H. (**A**) Surface representation of the bound complex with four separate LG1 ligands and an RNA-duplex containing four CAG-repeats before (t=0) and after (t=100 ns) the simulation. (**B**) H-bonding interaction of the CEG, AIA, and GFC triads after the simulation. (**C**) Number of H-bonds, (**D**) radius of gyration of PNA-RNA complex, and (**E**) root-mean square deviation (RMSD) of the PNA-RNA complex with respect to the initial structure. Figure reprinted with permission from American Chemical Society Copyright (2018).*

A.4 Conclusion

In summary, we have shown that γPNA containing a selected set of JBs could be developed and can hybridize to strands of the DNA and RNA double helix without undergoing self-hybridization. JBγPNA is able to invade a highly stable double helical DNA at a physiologically relevant ionic strength, whereas a homologue containing natural nucleobases is not. Due to their tight binding, significantly shorter probes could potentially be used to target the secondary and tertiary structures of RNA in an attempt to elucidate their physiological functions. We have also demonstrated the applicability of the methods as a means to treat different neurodegenerative diseases. We illustrated that a relatively short nucleic ligand, three units in length, comprising J-bases and γPNA backbone, is capable of binding rCAG-repeats in a sequence-specific and selective manner. Thus this work, done in collaboration with Prof. Danith Ly from Carnegie Mellon University, has an important implication for the development of millamolecular nucleic acid ligands for targeting repeated expansions in DNA and RNA.

References

1. Orr, H. T.; Zoghbi, H. Y., Trinucleotide Repeat Disorders. *Annual Review of Neuroscience* **2007,** *30* (1), 575-621.
2. La Spada, A. R.; Taylor, J. P., Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nature Reviews Genetics* **2010,** *11*, 247.
3. Cattaneo, E.; Zuccato, C.; Tartari, M., Normal huntingtin function: an alternative approach to Huntington's disease. *Nature Reviews Neuroscience* **2005,** *6* (12), 919-930.
4. MacDonald, M. E.; Ambrose, C. M.; Duyao, M. P.; Myers, R. H.; Lin, C.; Srinidhi, L.; Barnes, G.; Taylor, S. A.; James, M.; Groot, N.; MacFarlane, H.; Jenkins, B.; Anderson, M. A.; Wexler, N. S.; Gusella, J. F.; Bates, G. P.; Baxendale, S.; Hummerich, H.; Kirby, S.; North, M.; Youngman, S.; Mott, R.; Zehetner, G.; Sedlacek, Z.; Poustka, A.; Frischauf, A.-

M.; Lehrach, H.; Buckler, A. J.; Church, D.; Doucette-Stamm, L.; O'Donovan, M. C.; Riba-Ramirez, L.; Shah, M.; Stanton, V. P.; Strobel, S. A.; Draths, K. M.; Wales, J. L.; Dervan, P.; Housman, D. E.; Altherr, M.; Shiang, R.; Thompson, L.; Fielder, T.; Wasmuth, J. J.; Tagle, D.; Valdes, J.; Elmer, L.; Allard, M.; Castilla, L.; Swaroop, M.; Blanchard, K.; Collins, F. S.; Snell, R.; Holloway, T.; Gillespie, K.; Datson, N.; Shaw, D.; Harper, P. S., A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **1993,** *72* (6), 971-983.

5. Duyao, M.; Ambrose, C.; Myers, R.; Novelletto, A.; Persichetti, F.; Frontali, M.; Folstein, S.; Ross, C.; Franz, M.; Abbott, M.; Gray, J.; Conneally, P.; Young, A.; Penney, J.; Hollingsworth, Z.; Shoulson, I.; Lazzarini, A.; Falek, A.; Koroshetz, W.; Sax, D.; Bird, E.; Vonsattel, J.; Bonilla, E.; Alvir, J.; Bickham Conde, J.; Cha, J. H.; Dure, L.; Gomez, F.; Ramos, M.; Sanchez-Ramos, J.; Snodgrass, S.; de Young, M.; Wexler, N.; Moscowitz, C.; Penchaszadeh, G.; MacFarlane, H.; Anderson, M.; Jenkins, B.; Srinidhi, J.; Barnes, G.; Gusella, J.; MacDonald, M., Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nature Genetics* **1993,** *4* (4), 387-392.

6. Brook, J. D.; McCurrach, M. E.; Harley, H. G.; Buckler, A. J.; Church, D.; Aburatani, H.; Hunter, K.; Stanton, V. P.; Thirion, J.-P.; Hudson, T.; Sohn, R.; Zemelman, B.; Snell, R. G.; Rundle, S. A.; Crow, S.; Davies, J.; Shelbourne, P.; Buxton, J.; Jones, C.; Juvonen, V.; Johnson, K.; Harper, P. S.; Shaw, D. J.; Housman, D. E., Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. *Cell* **1992,** *68* (4), 799-808.

7. Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O., Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* **1991,** *254* (5037), 1497.

8. Nielsen, P. E., Peptide Nucleic Acid. A Molecule with Two Identities. *Accounts of Chemical Research* **1999,** *32* (7), 624-630.

9. Bahal, R.; Sahu, B.; Rapireddy, S.; Lee, C.-M.; Ly, D. H., Sequence-Unrestricted, Watson–Crick Recognition of Double Helical B-DNA by (R)-MiniPEG-γPNAs. *ChemBioChem* **2012,** *13* (1), 56-60.

10. Branda, N.; Kurz, G.; Lehn, J.-M., JANUS WEDGES: a new approach towards nucleobase-pair recognition. *Chemical Communications* **1996,** (21), 2443-2444.

11. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **2004,** *25* (13), 1605-1612.

12. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.

13. Macke, T. J.; Case, D. A., Modeling Unusual Nucleic Acid Structures. In *Molecular Modeling of Nucleic Acids*, American Chemical Society: 1997; Vol. 682, pp 379-393.

14. He, W.; Crawford, M. J.; Rapireddy, S.; Madrid, M.; Gil, R. R.; Ly, D. H.; Achim, C., The structure of a γ-modified peptide nucleic acid duplex. *Molecular BioSystems* **2010,** *6* (9), 1619-1629.
15. Yeh, J. I.; Shivachev, B.; Rapireddy, S.; Crawford, M. J.; Gil, R. R.; Du, S.; Madrid, M.; Ly, D. H., Crystal Structure of Chiral γPNA with Complementary DNA Strand: Insights into the Stability and Specificity of Recognition and Conformational Preorganization. *Journal of the American Chemical Society* **2010,** *132* (31), 10717-10727.
16. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983,** *79* (2), 926-935.
17. Perrin, C. L., Numerical Recipes in Fortran 90: The Art of Scientific Computing, Second Edition, Volume 2 (3 CD-ROMs and Manual) By William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. Cambridge University Press: New York, 1996. *Journal of the American Chemical Society* **1997,** *119* (37), 8748-8748.
18. Nosé, S., A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics* **1984,** *52* (2), 255-268.
19. Hoover, W. G., Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* **1985,** *31* (3), 1695-1697.
20. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **1981,** *52* (12), 7182-7190.
21. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993,** *98* (12), 10089-10092.
22. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **2008,** *4* (3), 435-447.
23. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19-25.
24. Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpí, J. L.; González, C.; Vendruscolo, M.; Laughton, C. A.; Harris, S. A.; Case, D. A.; Orozco, M., Parmbsc1: a refined force field for DNA simulations. *Nature Methods* **2015,** *13*, 55.
25. Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P., The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. *Physical Chemistry Chemical Physics* **2010,** *12* (28), 7821-7839.
26. Uhlmann, E.; Peyman, A.; Breipohl, G.; Will, D. W., PNA: Synthetic Polyamide Nucleic Acids with Unusual Binding Properties. *Angewandte Chemie International Edition* **1998,** *37* (20), 2796-2823.

Title: The role of solvent viscosity in the dynamics of protein conformational changes

Author: A Ansari,CM Jones,ER Henry,J Hofrichter,WA Eaton

Publication: Science

Publisher: The American Association for the Advancement of Science

Date: Jun 26, 1992

Order Completed

Thank you for your order.

This Agreement between Ms. Hridya V M ("You") and The American Association for the Advancement of Science ("The American Association for the Advancement of Science") consists of your license details and the terms and conditions provided by The American Association for the Advancement of Science and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

printable details

License Number	4657660540897
License date	Aug 28, 2019
Licensed Content Publisher	The American Association for the Advancement of Science
Licensed Content Publication	Science
Licensed Content Title	The role of solvent viscosity in the dynamics of protein conformational changes
Licensed Content Author	A Ansari,CM Jones,ER Henry,J Hofrichter,WA Eaton
Licensed Content Date	Jun 26, 1992
Licensed Content Volume	256
Licensed Content Issue	5065
Volume number	256
Issue number	5065
Type of Use	Thesis / Dissertation
Requestor type	Scientist/individual at a research institution
Format	Print and electronic
Portion	Figure
Number of figures/tables	1
Order reference number	
Title of your thesis / dissertation	Dynamical recrossing, Internal friction and memory effects: Investigating model systems and DNA-drug intercalation process
Expected completion date	Dec 2019
Estimated size(pages)	150
Requestor Location	Ms. Hridya V M IISER-Pune Pashan

SPRINGER NATURE

Title:	Shape selective bifacial recognition of double helical DNA
Author:	Shivaji A. Thadke et al
Publication:	Communications Chemistry
Publisher:	Springer Nature
Date:	Nov 7, 2018

Creative Commons

Title: Probing the Viscosity Dependence of Rate: Internal Friction or the Lack of Friction?

Author: V. M. Hridya, Arnab Mukherjee

Publication: The Journal of Physical Chemistry B

Publisher: American Chemical Society

Date: Oct 1, 2018

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

BACK CLOSE WINDOW

Title: Design of Bivalent Nucleic Acid Ligands for Recognition of RNA-Repeated Expansion Associated with Huntington's Disease

Author: Shivaji A. Thadke, J. Dinithi R. Perera, V. M. Hridya, et al

Publication: Biochemistry

Publisher: American Chemical Society

Date: Apr 1, 2018

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

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

Author: V. M. Hridya, James T. Hynes, Arnab Mukherjee

Publication: The Journal of Physical Chemistry B

Publisher: American Chemical Society

Date: Nov 1, 2019

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

BACK CLOSE WINDOW